brain sciences

Advances in the Neurocognition of Music and Language

Edited by
Daniela Sammler and Stefan Elmer

Printed Edition of the Special Issue Published in *Brain Sciences*

www.mdpi.com/journal/brainsci

MDPI

Advances in the Neurocognition of Music and Language

Advances in the Neurocognition of Music and Language

Editors

Daniela Sammler
Stefan Elmer

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors
Daniela Sammler
Max Planck Institute for Human Cognitive and Brain Sciences
Germany

Stefan Elmer
University of Zurich
Switzerland

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Brain Sciences* (ISSN 2076-3425) (available at: https://www.mdpi.com/journal/brainsci/special_issues/Neurocognition_Music_and_Language).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Article Number*, Page Range.

ISBN 978-3-03943-126-7 (Hbk)
ISBN 978-3-03943-127-4 (PDF)

Contents

About the Editors

Daniela Sammler, PD Dr., is the leader of the research group Neural Bases of Intonation in Speech and Music at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. She combines behavioral, neuroimaging, and brain stimulation techniques in healthy and brain damaged adults and professional musicians to elucidate the mechanisms of music/language perception and production and their neural overlap. Sammler received her Ph.D. (2008) and P.D. (2018) from the University of Leipzig and was a recipient of the Otto Hahn Award of the Max Planck Society.

Stefan Elmer, Dr., received his Ph.D. (2010) from the University of Zurich, where since 2012 he has been responsible for the Auditory Research Group Zurich at the Department of Neuropsychology. He has extensive experience in using brain imaging and neurophysiology techniques, and in the last few years, he has contributed to a better understanding of the electrophysiological, functional, and anatomical markers of music and language expertise, transfer effects, language learning, as well as plasticity in the auditory system in general.

brain sciences

MDPI

Editorial

Advances in the Neurocognition of Music and Language

Daniela Sammler [1,*] and Stefan Elmer [2,*]

[1] Otto Hahn Group Neural Bases of Intonation in Speech and Music, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany
[2] Auditory Research Group Zurich (ARGZ), Division Neuropsychology, Institute of Psychology, University of Zurich, 8050 Zurich, Switzerland
* Correspondence: sammler@cbs.mpg.de (D.S.); s.elmer@psychologie.uzh.ch (S.E.)

Received: 27 July 2020; Accepted: 30 July 2020; Published: 2 August 2020

Abstract: Neurocomparative music and language research has seen major advances over the past two decades. The goal of this Special Issue "Advances in the Neurocognition of Music and Language" was to showcase the multiple neural analogies between musical and linguistic information processing, their entwined organization in human perception and cognition and to infer the applicability of the combined knowledge in pedagogy and therapy. Here, we summarize the main insights provided by the contributions and integrate them into current frameworks of rhythm processing, neuronal entrainment, predictive coding and cognitive control.

Keywords: music; language; brain; rhythm; prosody; musical training; dyslexia; reading; oscillations; statistical learning; cognitive control

---

The scholarly fascination for the relationships between music and language (M&L) is as old as antiquity. To this day, continuous methodological progress and, in part, radical conceptual shifts paved the way for new directions of research. In the 1990s, technological revolutions in neuroimaging revealed partial neural overlap between the two domains [1], despite dissociable clinical deficits in M&L [2]. Together with known benefits of music for speech and language functions [3] this nurtured the idea that—once we understand what holds M&L together at their biological core—music interventions could constitute a bridge to prevent, alleviate, or even reverse speech and language disorders [4,5].

This Special Issue took stock of recent advances in the neurocognition of M&L to examine the current status of this vein of research. Sixteen research papers and reviews from 48 experts in linguistics, musicology, cognitive neuroscience, biological psychology and educational sciences demonstrate that research has been active on all fronts. As we will see, the studies follow two burgeoning trends in M&L research: First, they focused on common auditory processing of temporal regularities [6–9] that are thought to promote higher-level linguistic functions [8,10–14], possibly via mechanisms of neuronal entrainment [15]. Second, they explored top-down modulations of common auditory processes [16–18] by domain-general cognitive [19,20] and motor functions in both perception and production [21]. These topics were addressed using a broad toolkit of well-designed behavioral and computational approaches combined with functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS) or electroencephalography (EEG) in different cohorts of participants.

The starting point for most of the included studies was that speech and music have similar acoustic [9,18] and structural features [6–8,13,15–17,19]. As argued in the review article of Reybrouck and Podlipniak [9], some of these sound features and their common preconceptual affective meanings may even reflect joint evolutionary roots of M&L that still prevail today, for example, in musical expressivity and speech prosody. Notably, a feature that was particularly central to half of all

contributions is the *temporal* structure of M&L, i.e., the patterning of strong and weak syllables or beats that make up rhythm, meter and prosodic stress [6–8,10–13,15].

The rhythmic patterning of both speech and music has been proposed to draw on domain-general abilities which are required to perceive and process temporal features of sound [22,23]. Accordingly, three studies present data in line with common rhythm processing resources in M&L. First, Lagrois et al. [6] found that individuals with beat finding deficits in music—so called "beat-deaf" individuals—also show deficits in synchronizing their taps with speech rhythm, and more generally, in regular tapping without external rhythms. The authors argue that this pattern of deficits may arise from a basic deficiency in timekeeping mechanisms that affect rhythm perception across domains. Second, Boll-Avetisyan et al. [7] used multiple regression analyses and found that musical rhythm perception abilities predicted rhythmic grouping preferences in speech in adults with and without dyslexia. Similarly, in an EEG study, Fotidzis et al. [8] found that musical rhythmic skills predicted children's neural sensitivity to mismatches between the speech rhythm of a written word and an auditory rhythm. Interestingly, both studies further report connections between rhythm perception in music and reading skills. Hence, these findings not only speak for a common cross-domain basis of rhythmic processes in M&L but also suggest that deficient or enhanced rhythmic abilities may have an impact on higher-level language functions.

Potential downstream effects of general rhythmic processing skills on higher-order linguistic abilities are currently being extensively investigated, particularly in the context of first language acquisition (for a recent review, see [24]). Accordingly, several studies in this Special Issue probe whether the acoustic properties of speech rhythm can serve as scaffolding for the acquisition of stable phonological representations [12], for the segmentation of words from continuous speech and the construction of lexical representations [13], for the recognition of syntactic units in sentences [10] and for reading [7,8,11]. For example, Richards and Goswami [10] explain that prosody, particularly the hierarchical structuring of stressed and unstressed syllables, provides reliable cues to the syntactic structure of speech [25] and can hence facilitate learning of syntactic language organization [26]. Early perturbations at this rhythm-syntax interface may, in turn, hinder normal language acquisition, such as in developmental language disorders (DLD). The authors found that children with DLD indeed had difficulties in noticing conflicting alignments between prosodic and syntactic boundaries in rhythmic children's stories, and that these deficits coincided with enhanced perceptual thresholds for acoustic cues to prosodic stress. With these data at hand, Richards and Goswami support the assertion that basic processing of rhythmic-prosodic cues may be a key foundation onto which higher aspects of language are scaffolded during development.

In a similar vein, rhythmic-prosodic sensitivity has been proposed as fundamental stepping stone into literacy [27–29] as well as implicit driver for skilled reading [30]. Breen et al. [11] and Fotidzis et al. [8] present converging EEG evidence for implicit rhythmic processing in silent reading of words in literate adults and children. In particular, they both found a robust fronto-central negativity in response to stress patterns in written words that mismatched the rhythm of silently read limericks [11] or auditory click trains [8]. These results suggest that rhythmic context—no matter whether implicit in written text or explicit in sound—can induce expectations of prosodic word stress that facilitate visual word recognition and reading speed.

Current neurophysiological models assume that speech and music processing as well as the catalytic role of rhythm in language development are based on the synchronization of internal neuronal oscillations with temporally regular stimuli [27,31–33]. The review article by Myers et al. [15] summarizes the current state of knowledge about neuronal entrainment to the speech envelope reflecting quasi-regular amplitude fluctuations over time. This neural tracking occurs simultaneously at multiple time scales corresponding to the rates of phonemes, syllables and phrases [34,35]. In this context, Myers and colleagues argue that the slowest rate—corresponding to prosodic stress and rhythmic pacing in the delta range (~2Hz)—constitutes a particularly strong source of neuronal entrainment which is crucial for normal language development. Correspondingly, atypical entrainment

to rhythmic prosodic cues due to deficits in fine-grained auditory perception may constitute a risk for the development of speech and language disorders such as DLD and developmental dyslexia (DD) [24,36].

If rhythmic processing disabilities are indeed the basis of speech and language disorders, then useful avenues for prevention and intervention could lie in (i) increasing the regularity of stimuli, or (ii) strengthening individual rhythmic abilities with the aim at improving neuronal entrainment [37–39]. Several studies in this Special Issue deal directly or indirectly with these ideas, either by exploring processing benefits of rhythmically highly regular stimuli such as songs [13,14] or poems [10,11], or by discussing potential protective or curative effects of music-based rhythm training on language skills [7,8,10,12,15,16]. Even though the results are promising, they also raise a number of questions. For example, using EEG Snijders et al. [13] found that 10-month-old infants were able to segment words in natural children's songs. However, they did equally well in infant-directed speech. Similarly, Rossi et al. [14] found no differences between speech and songs in a combined EEG-NIRS study on semantic processing in healthy adults. Taken together, these data suggest that the presentation of verbal material as song may not be sufficient to enhance vocabulary learning or language comprehension in healthy individuals (but see [40]). The longitudinal study of Frey et al. [12] zoomed in on training effects. Using EEG, the authors demonstrate that 6 months of music but not painting training positively influenced the pre-attentive processing of voice onset time in speech in children with DD. However, no effects were found in behavioral measures of phonological processing or reading ability. This raises the questions of how much training is required and which aspects the training should include to translate to behavior, both inside and outside the laboratory setting. Clearly, the identification of optimal interventions is a joint mission for future research that goes hand in hand with the development of solid conceptual [41,42] and neurophysiological frameworks [27] to identify the key variables underlying the amelioration of speech and language processing through rhythm and music [43–46].

The studies of this Special Issue introduced so far primarily focused on links between M&L that are bottom-up driven by shared acoustic features between the two domains. The remaining articles took a different approach and examined domain-general top-down modulations of M&L from both the perspectives of perception and production. Four articles illustrate the continuous interaction between bottom-up and top-down processes. In line with significant trends in predictive coding [47,48], Daikoku [16] reviews the conceptual, computational, experimental and neural similarities of statistical learning in M&L acquisition and perception with links to rehabilitation. Bidirectional interactions between perceptual (bottom-up) and predictive (top-down) processes are a core feature in the framework of statistical learning. Experimental evidence for the top-down adjustment of M&L perception is provided by the behavioral modelling study of Silva et al. [17] who found that listeners placed break patterns in ambiguous speech-song stimuli differently depending on whether they believed they were listening to speech prosody or contemporary music. Similarly, the fMRI study of Tsai and Li [18] found that the strength with which an ambiguous stimulus was perceived as song rather than speech depended not only on the acoustics of the stimulus itself, but also on the sound category of the preceding stimulus. Finally, Mathias et al. [21] show with EEG that pianists gradually anticipated the sounds of their actions during music production, similar to mechanisms of auditory feedback control during speech production [49,50]. Taken together, these studies suggest that the listening context, one's own motor plans as well as statistical and domain-specific expectations may influence the top-down anticipation and perception of acoustic features in speech and music.

Finally, the last two articles focus on the relevance of domain-general cognitive functions for M&L interactions. Lee et al. [19] argues that well-known syntax interference effects between M&L [51,52] may emerge from shared domain-general attentional resources. Accordingly, they show that the top-down allocation of attention similarly modulated EEG markers of syntax processing in M&L, particularly at late processing stages associated with cognitive reanalysis and integration. Otherwise, Christiner and Reiterer [20] found that links between musical aptitude and phonetic language abilities

in pre-school children (i.e., imitation of foreign speech) were mediated by domain-general working memory resources. While none of these studies denies auditory-perceptual connections between M&L, they remind us that what we have seen so far is perhaps only the tip of the iceberg, with more complex entwinements still to be discovered.

To sum up, this Special Issue indicates that questions have shifted from mapping to mechanisms. Initial descriptions of M&L analogies have turned into a determined search for explanations of M&L links in human neurophysiology, general perceptual principles and cognitive computations. Accordingly, the obvious next questions are of a mechanistic nature: Can musical training enhance the neuronal entrainment to speech (and vice versa)? How exactly does entrainment promote higher-order linguistic functions? How can working memory and attention be included in the equation? These are only a few questions, but we are confident that the joint efforts of this multidisciplinary field of research will be rewarded by a better understanding of the M&L interface and the necessary tools to optimize interventions for music- and language-related dysfunctions.

Author Contributions: D.S. and S.E. edited this Special Issue, D.S. wrote the original draft of the editorial, S.E. and D.S. revised the editorial. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: New York, NY, USA, 2008.
2. Peretz, I.; Coltheart, M. Modularity of Music Processing. *Nat. Neurosci.* **2003**, *6*, 688–691. [CrossRef] [PubMed]
3. Sparks, R.; Helm, N.; Albert, M. Aphasia Rehabilitation Resulting from Melodic Intonation Therapy. *Cortex* **1974**, *10*, 303–316. [CrossRef]
4. Patel, A.D. Why Would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front. Psychol.* **2011**, *2*, 142. [CrossRef] [PubMed]
5. Schön, D.; Tillmann, B. Short- and Long-Term Rhythmic Interventions: Perspectives for Language Rehabilitation. *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 32–39. [CrossRef]
6. Lagrois, M.-E.; Palmer, C.; Peretz, I. Poor Synchronization to Musical Beat Generalizes to Speech. *Brain Sci.* **2019**, *9*, 157. [CrossRef] [PubMed]
7. Boll-Avetisyan, N.; Bhatara, A.; Höhle, B. Processing of Rhythm in Speech and Music in Adult Dyslexia. *Brain Sci.* **2020**, *10*, 261. [CrossRef]
8. Fotidzis, T.S.; Moon, H.; Steele, J.R.; Magne, C.L. Cross-Modal Priming Effect of Rhythm on Visual Word Recognition and Its Relationships to Music Aptitude and Reading Achievement. *Brain Sci.* **2018**, *8*, 210. [CrossRef]
9. Reybrouck, M.; Podlipniak, P. Preconceptual Spectral and Temporal Cues as Source of Meaning in Speech and Music. *Brain Sci.* **2019**, *9*, 53. [CrossRef]
10. Richards, S.; Goswami, U. Impaired Recognition of Metrical and Syntactic Boundaries in Children with Developmental Language Disorders. *Brain Sci.* **2019**, *9*, 33. [CrossRef]
11. Breen, M.; Fitzroy, A.B.; Oraa Ali, M. Event-Related Potential Evidence of Implicit Metric Structure during Silent Reading. *Brain Sci.* **2019**, *9*, 192. [CrossRef]
12. Frey, A.; François, C.; Chobert, J.; Velay, J.-L.; Habib, M.; Besson, M. Music Training Positively Influences the Preattentive Perception of Voice Onset Time in Children with Dyslexia: A Longitudinal Study. *Brain Sci.* **2019**, *9*, 91. [CrossRef] [PubMed]
13. Snijders, T.M.; Benders, T.; Fikkert, P. Infants Segment Words from Songs—An EEG Study. *Brain Sci.* **2020**, *10*, 39. [CrossRef] [PubMed]
14. Rossi, S.; Gugler, M.F.; Rungger, M.; Galvan, O.; Zorowka, P.G.; Seebacher, J. How the Brain Understands Spoken and Sung Sentences. *Brain Sci.* **2020**, *10*, 36. [CrossRef] [PubMed]
15. Myers, B.R.; Lense, M.D.; Gordon, R.L. Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception. *Brain Sci.* **2019**, *9*, 70. [CrossRef]

16. Daikoku, T. Neurophysiological Markers of Statistical Learning in Music and Language: Hierarchy, Entropy, and Uncertainty. *Brain Sci.* **2018**, *8*, 114. [CrossRef]
17. Silva, S.; Dias, C.; Castro, S.L. Domain-Specific Expectations in Music Segmentation. *Brain Sci.* **2019**, *9*, 169. [CrossRef]
18. Tsai, C.G.; Li, C.W. Is It Speech or Song? Effect of Melody Priming on Pitch Perception of Modified Mandarin Speech. *Brain Sci.* **2019**, *9*, 286. [CrossRef]
19. Lee, D.J.; Jung, H.; Loui, P. Attention Modulates Electrophysiological Responses to Simultaneous Music and Language Syntax Processing. *Brain Sci.* **2019**, *9*, 305. [CrossRef]
20. Christiner, M.; Reiterer, S.M. Early Influence of Musical Abilities and Working Memory on Speech Imitation Abilities: Study with Pre-School Children. *Brain Sci.* **2018**, *8*, 169. [CrossRef]
21. Mathias, B.; Gehring, W.J.; Palmer, C. Electrical Brain Responses Reveal Sequential Constraints on Planning during Music Performance. *Brain Sci.* **2019**, *9*, 25. [CrossRef]
22. Kotz, S.A.; Ravignani, A.; Fitch, W.T. The Evolution of Rhythm Processing. *Trends Cogn. Sci.* **2018**, *22*, 896–910. [CrossRef] [PubMed]
23. Jones, M.R. *Time Will Tell: A Theory of Dynamic Attending*; Oxford University Press: New York, NY, USA, 2019.
24. Ladányi, E.; Persici, V.; Fiveash, A.; Tillmann, B.; Gordon, R.L. Is Atypical Rhythm a Risk Factor for Developmental Speech and Language Disorders? *Wiley Interdiscip. Rev. Cogn. Sci.* **2020**, e1528. [CrossRef] [PubMed]
25. Selkirk, E. *Phonology and Syntax: The Relation between Sound and Structure*; MIT Press: Cambridge, MA, USA, 1984.
26. Cumming, R.; Wilson, A.; Goswami, U. Basic Auditory Processing and Sensitivity to Prosodic Structure in Children with Specific Language Impairments: A New Look at a Perceptual Hypothesis. *Front. Psychol.* **2015**, *6*, 972. [CrossRef] [PubMed]
27. Goswami, U. A Neural Oscillations Perspective on Phonological Development and Phonological Processing in Developmental Dyslexia. *Lang. Linguist. Compass* **2019**, *13*, e12328. [CrossRef]
28. Tierney, A.; Kraus, N. Music Training for the Development of Reading Skills. *Prog. Brain Res.* **2013**, *207*, 209–241.
29. Wade-Woolley, L.; Heggi, L. The Contributions of Prosodic and Phonological Awareness to Reading: A Review. In *Linguistic Rhythm and Literacy*; Thomson, J., Jarmulowicz, L., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2016; pp. 3–24.
30. Breen, M. Empirical Investigations of Implicit Prosody. In *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor*; Frazier, L., Gibson, E., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 177–192.
31. Poeppel, D.; Assaneo, M.F. Speech Rhythms and Their Neural Foundations. *Nat. Rev. Neurosci.* **2020**, *21*, 322–334. [CrossRef]
32. Lakatos, P.; Gross, J.; Thut, G. A New Unifying Account of the Roles of Neuronal Entrainment. *Curr. Biol.* **2019**, *29*, R890–R905. [CrossRef]
33. Large, E.W.; Herrera, J.A.; Velasco, M.J. Neural Networks for Beat Perception in Musical Rhythm. *Front. Syst. Neurosci.* **2015**, *9*, 159. [CrossRef]
34. Gross, J.; Hoogenboom, N.; Thut, G.; Schyns, P.; Panzeri, S.; Belin, P.; Garrod, S. Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biol.* **2013**, *11*, e1001752. [CrossRef]
35. Giraud, A.L.; Poeppel, D. Cortical Oscillations and Speech Processing: Emerging Computational Principles and Operations. *Nat. Neurosci.* **2012**, *15*, 511–517. [CrossRef]
36. Goswami, U. A Temporal Sampling Framework for Developmental Dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
37. Vanden Bosch der Nederlanden, C.M.; Joanisse, M.F.; Grahn, J.A. Music as a Scaffold for Listening to Speech: Better Neural Phase-Locking to Song than Speech. *Neuroimage* **2020**, *214*, 116767. [CrossRef] [PubMed]
38. Harding, E.E.; Sammler, D.; Henry, M.J.; Large, E.; Kotz, S.A. Cortical Tracking of Rhythm in Music and Speech. *Neuroimage* **2019**, *185*, 96–101. [CrossRef] [PubMed]
39. Doelling, K.B.; Poeppel, D. Cortical Entrainment to Music and Its Modulation by Expertise. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6233–E6242. [CrossRef]

40. François, C.; Teixidó, M.; Takerkart, S.; Agut, T.; Bosch, L.; Rodriguez-Fornells, A. Enhanced Neonatal Brain Responses to Sung Streams Predict Vocabulary Outcomes by Age 18 Months. *Sci. Rep.* **2017**, *7*, 12451. [CrossRef]
41. Tierney, A.; Kraus, N. Auditory-Motor Entrainment and Phonological Skills: Precise Auditory Timing Hypothesis (PATH). *Front. Hum. Neurosci.* **2014**, *8*, 949. [CrossRef]
42. Ozernov-Palchik, O.; Patel, A.D. Musical Rhythm and Reading Development: Does Beat Processing Matter? *Ann. N. Y. Acad. Sci.* **2018**, *1423*, 166–175. [CrossRef]
43. Besson, M.; Chobert, J.; Marie, C. Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Front. Psychol.* **2011**, *2*, 94. [CrossRef]
44. Virtala, P.; Partanen, E. Can Very Early Music Interventions Promote At-Risk Infants' Development? *Ann. N. Y. Acad. Sci.* **2018**, *1423*, 92–101. [CrossRef]
45. Elmer, S.; Dittinger, E.; Besson, M. One Step Beyond: Musical Expertise and Word Learning. In *the Oxford Handbook of Voice Perception*; Frühholz, S., Belin, P., Eds.; Oxford University Press: Oxford, UK, 2019; pp. 209–234.
46. Torppa, R.; Huotilainen, M. Why and How Music Can Be Used to Rehabilitate and Develop Speech and Language Skills in Hearing-Impaired Children. *Hear. Res.* **2019**, *380*, 108–122. [CrossRef]
47. Koelsch, S.; Vuust, P.; Friston, K. Predictive Processes and the Peculiar Case of Music. *Trends Cogn. Sci.* **2019**, *23*, 63–77. [CrossRef] [PubMed]
48. Erickson, L.C.; Thiessen, E.D. Statistical Learning of Language: Theory, Validity, and Predictions of a Statistical Learning Account of Language Acquisition. *Dev. Rev.* **2015**, *37*, 66–108. [CrossRef]
49. Palmer, C.; Pfordresher, P.Q. Incremental Planning in Sequence Production. *Psychol. Rev.* **2003**, *110*, 683–712. [CrossRef] [PubMed]
50. Hickok, G. Computational Neuroanatomy of Speech Production. *Nat. Rev. Neurosci.* **2012**, *13*, 135–145. [CrossRef]
51. Koelsch, S.; Gunter, T.C.; Wittfoth, M.; Sammler, D. Interaction between Syntax Processing in Language and in Music: An ERP Study. *J. Cogn. Neurosci.* **2005**, *17*, 1565–1577. [CrossRef]
52. Slevc, L.R.; Rosenberg, J.C.; Patel, A.D. Making Psycholinguistics Musical: Self-Paced Reading Time Evidence for Shared Processing of Linguistic and Musical Syntax. *Psychon. Bull. Rev.* **2009**, *16*, 374–381. [CrossRef]

brain sciences

MDPI

Review

Preconceptual Spectral and Temporal Cues as a Source of Meaning in Speech and Music

Mark Reybrouck [1,2,*] and Piotr Podlipniak [3]

[1] Musicology Research Group, KU Leuven–University of Leuven, 3000 Leuven, Belgium
[2] IPEM–Department of Musicology, Ghent University, 9000 Ghent, Belgium
[3] Institute of Musicology, Adam Mickiewicz University in Poznań, ul. Umultowska 89D, 61-614 Poznań, Poland; podlip@amu.edu.pl
* Correspondence: Mark.Reybrouck@kuleuven.be; Tel.: +32-478-603-479

Received: 23 January 2019; Accepted: 26 February 2019; Published: 1 March 2019

Abstract: This paper explores the importance of preconceptual meaning in speech and music, stressing the role of affective vocalizations as a common ancestral instrument in communicative interactions. Speech and music are sensory rich stimuli, both at the level of production and perception, which involve different body channels, mainly the face and the voice. However, this bimodal approach has been challenged as being too restrictive. A broader conception argues for an action-oriented embodied approach that stresses the reciprocity between multisensory processing and articulatory-motor routines. There is, however, a distinction between language and music, with the latter being largely unable to function referentially. Contrary to the centrifugal tendency of language to direct the attention of the receiver away from the text or speech proper, music is centripetal in directing the listener's attention to the auditory material itself. Sound, therefore, can be considered as the meeting point between speech and music and the question can be raised as to the shared components between the interpretation of sound in the domain of speech and music. In order to answer these questions, this paper elaborates on the following topics: (i) The relationship between speech and music with a special focus on early vocalizations in humans and non-human primates; (ii) the transition from sound to meaning in speech and music; (iii) the role of emotion and affect in early sound processing; (iv) vocalizations and nonverbal affect burst in communicative sound comprehension; and (v) the acoustic features of affective sound with a special emphasis on temporal and spectrographic cues as parts of speech prosody and musical expressiveness.

Keywords: preconceptual meaning; affective vocalizations; action-oriented embodied approach; affect burst; speech prosody; musical expressiveness

1. Introduction

The problem of meaning extraction in speech and music has received a lot of concern in different fields, such as infant-directed speech and singing, the origins of music perception and cognition, and the primary use of acoustic cues in emotion-driven and affect-laden preverbal communication. This kind of research saw its heyday in the 1990s with major contributions in the field of early music perception [1,2] and preference. Many efforts have been directed towards the study of *motherese* and *infant-directed speech* and *singing* [3–7], the acoustic basis of young children's preference for such kinds of vocal communication [8,9], the musical elements in early affective communication between newborns and caregivers [10–12], and the role of prosodic features in preverbal and early musical communication [13–15]. Most of this research has stressed the extreme sensitivity of young infants for acoustic features of speech and music [16] as well as the existence of early musical predispositions [17–19].

Many of these studies—and also some subsequent ones—have emphasized certain commonalities between language and music [19–21], most of them being related to the *prosodic* and *paralinguistic features* of language, which can be considered as being musical to some extent. Besides, there are lots of empirical results which indicate that both music and language can prime the meaning of a word and that music meaning is represented in a very similar fashion to language meaning in the human brain [22,23]. This observation suggests that propositional semantics that is specific solely to language can be based on broader meaning categories. which are less precise, but not language specific.

Recent developments have provided additional evidence from the domains of *comparative* [24,25] and *evolutionary musicology* [26–33], which deal with the evolutionary origins of music by adopting a comparative approach to vocal communication in animals and an evolutionary psychological approach to the emergence of music in the hominin line [34]. These approaches make it possible to tease apart those processes that appear to be innate from those that develop with maturation or acculturation [35]. The animal research provides insights in the role of acoustic cues in nonverbal and preverbal communication [36,37], which are related to affective speech and which can be considered emotional antecedents of music and language [15]. Some of these findings seem to corroborate, to some extent, Darwin's hypothesis on musical protolanguage, which stated that speech and music originated from a common precursor that developed from "the imitation and modification of various natural sounds, the voices of other animals, and man's own instinctive cries" [38] (p. 3). Such a primitive system would have been especially useful in the expression of emotion and music, as we know it nowadays, and should be a behavioral remnant of this early system of communication (see also [39–41]). It is a hypothesis which has been elaborated and restated by modern researchers under the umbrella of the musical protolanguage hypothesis [20,24,40,42,43].

2. Meaning Before Language and Music

Meaning can be considered as the result of the interpretation of stimuli by the nervous system. Such interpretation is often described in terms of internal mental representations that animals have of the things, events, and situations in their environment, and it is evolutionary older than their corresponding expressions in language [44] (pp. 4–6). From a phylogenetic perspective, there are three major kinds of meaning which have evolved over time: Meaning as a way to orient oneself in the environment [45], emotions and emotional communication as integral decision mechanisms [46] and as motivational states which are meaningful for survival by providing also a primordial way of interpretation of the external world [47], and referential meaning as the outcome of the appearance of a conceptual mind [48]. Meaning, moreover, can be considered as a basis for communication, using sound as the main source of information as is the case in primate and human vocalizations [49,50]. The latter, however, should not be identified solely with speech and singing, which can be contrasted clearly with nonverbal utterances, such as laughter, crying, and mourning. As such, there is a hierarchy in the kind of meaning that is conveyed by sound: There is, first, a distinction between the digital (speech and music) and analog (prosody, laughter, etc.) usage of sound [51]; there is, second, a hierarchical distinction between preconceptual and conceptual meaning with a first level of simple spectral and temporal cues as a source of reflexes and a second level of expressive dynamics and emotional communication [51,52]; there is, third, a level of meaning that is conveyed by means of syntax, such as, e.g., tonality in music and grammatical correctness in speech [53]; and finally, there is a fourth level of propositional semantics and associative meaning [48], such as language's lexicons and, most probably, chimpanzees' referential grunts [54].

Speech is closely related to vocal production and can be studied from a broader stance, including the articulatory, linguistic, and information-conveying point of view. The *articulatory approach* describes lexical units in terms of gestures that are characterizations of discrete, physical events that unfold during the speech production process [55]. They can be considered basic units of articulatory action, allowing phonology to be described as a set of relations among "phonological events" [56]. These basic units of articulatory routines are discrete gestures, which emerge pre-linguistically as early and gross

versions of their adult use [55], calling forth the *linguistic level* of speech processing with articulatory routines that gradually develop into higher-level phonological units that can be contrasted with each other [57]. Linguistic meaning, however, is discrete-digital rather than analog-continuous. It relies on propositional knowledge without direct coupling with the speech signals—as sounding and thus sensory phenomena—and combines referential meaning with particular sound patterns that function as vehicles to convey symbolic meaning. Such a "vehicle mode" of meaning involves referential meaning, which is a representational mode of conveying information, as against the "acoustic mode", which refers merely to the local modulations in sound that are involved in expressive communication of meanings [58].

Speech, as opposed to language as a system, is articulated in real time. As such, it is a sensory rich stimulus. It provides information across multiple modalities, combining both the auditory and visual modalities, as exemplified most typically in the facial expression of audio-visual emotional speech. The latter, together with prosody, cannot be reduced to the control of voice qualities alone, but is closely related to the integration of sensory modalities—with facial and vocal expressions reinforcing each other [13]—and even with the movements of body [59]. Much research on emotional speech (see e.g., [60]), however, has been oriented rather narrowly to facial expressions since it has been hypothesized over a long period of time that judges are more accurate in inferring distinct emotions from facial expressions than from vocal ones. Acoustic cues, on the other hand, have been considered merely as additional features to facial expression, marking only levels of physiological arousal that are less distinctive than those expressed by the face. This conclusion, however, has proved to be erroneous since previous studies have studied only a limited number of acoustic cues, and the arousal differences within emotion families have also been largely neglected [61]. This has been shown in recent studies that used a comprehensive path model of vocal emotion communication, encompassing encoding, transmission, and decoding processes [62,63] to empirically model data sets on emotional expression and recognition from two different cultures and languages. Results of their extended Brunswikian "lens model" [64]—lens equations, hierarchical regression, and multivariate path analysis—, all reflect the strong evidence from past work on the role of arousal in affective communication that vocal sounds primarily convey the arousal state of the sender. It was stated that the "voice is the privileged modality for the expression and communication of arousal and activation, whereas the face is vastly superior with respect to valence" [62] (p. 24).

Additional evidence comes from studies of infants' reactions to parental communicative signals, which have stressed their outstanding discriminative abilities for timing patterns, pitch, loudness, harmonic interval, and voice quality [65]. It seems, moreover, that newborns are very sensitive also to facial expressions, vocalizations, and hand movements, which they can largely imitate to some extent. Such a kind of *communicative musicality*, as it has been coined [11,66], shows children's awareness of human communicative signals. It is a faculty which is comprehensive, multimodal, and coherent at birth and in the first months after birth [67]. It stresses the conflation of perceptual and motor aspects in speech recognition and vocal expression, bringing together audio-visual, visual-motor, and audio-motor integration.

Music is related to this preverbal communicative expressivity. It precedes or bypasses verbal communication by stressing the sensory richness of the stimuli. As such, it is directed primarily to itself with meaning being self-referential rather than referring to something else. Contrary to the centrifugal tendency of linguistic meaning, where the attention is directed away from the text proper (centrifugal) to grasp the meaning of what is referred to, music has a centripetal tendency in directing the listener's attention to the auditory material of the sounding music itself [68,69]. As such, there seems to be a major distinction between language and music, though there are also some commonalities, which stress a number of shared components. This applies in particular to vocal music and its communicative possibilities.

Music, seen from an evolutionarily point of view, is one of the most ancient forms of human communication, with the human voice being probably the most ancestral instrument in human

music [70]. It can even be questioned in this regard whether music and speech are different, and if so, to what extent [71]. There are two ways to address this question, either by intraspecies or interspecies comparison. An example of the former is the study of para-musical elements in language and para-lingual elements in music [72], such as the use of lexical tone in tone languages and prosody (para-musical) or the use of Leitmotive in music (para-lingual) [20]. Also, the languages based on musical tone systems, such as drum and whistle languages, can be studied in this context [73]. The interspecies comparison, on the other hand, is still more challenging and embraces a very extensive body of research. It has been hypothesized, for example, that singing could have evolved from loud calls by nonhuman primates, such as the Old-World monkeys and apes, which have been considered to be possible precursors of human singing and music. Gibbons, in particular, use vocalizations that elicit emotional responses from human listeners by using acoustic characteristics, such as loudness, acceleration of note rhythm, a final slow-down in rhythm, sounds consisting of alternated exhalation and inhalation, higher pitch frequencies in the central section of the call, pure tone of notes, and frequent accompaniment with piloerection and locomotor displays [36]. All these elements, however, are also used to a different degree in speech.

As such, there is an ability of communication by means of sounds that touches on an evolutionarily old layer of sound communication, which is older than singing and speech. This level is involved in the development of functional sensitivity to a specific class of sounds in ancestral vertebrates both as an aid in identifying and localizing predators and for capturing prey [74]. It is exemplified most typically in the use of alarm calls, which can be considered as a class of punctuate sounds that tend to be short, with sharp and abrupt signal onset, dramatic frequency and amplitude fluctuations, and a chaotic broadband spectral content. There is also a broad class of vocalizations that has been labeled "squeaks, shrieks, and screams" and which have direct impact on animal perception [75]. Their specific designs make them stand out against background noise so as to make them easy to localize. Moreover, they may provoke immediate orienting reactions by other animals in the direction of the calls, in combination with reflexive movements that prepare for flight [76]. Such generalized startle responses are induced also in very young infants, even in the absence of significant previous experience. They are, in fact, reducible to the operation of low-level brainstem and subcortical processes, which are associated with sound localization, orienting, and autonomic responding [77,78]. These vocalizations, however, can be exemplary of an intentional, communicative use of sounds which differ functionally from simple auditory sensations, which are prelinguistic default labels of sound sources [79], such as sensation of loudness and low pitch, as a tag of a big animal.

Such vocalizations by animals are not gratuitous. They are used frequently by youngsters as an opportunity to influence the behavior of older and larger individuals by engaging their attention, arousal, and concomitant behavior, sometimes in a very compelling way [80]. It can be questioned, however, whether primates have a theory of mind or act intentionally to influence others. A tentative answer can be found in comparable research in humans into to the neurocognitive mechanisms (auditory prosodic activations) that allow listeners to read the intentions of speakers from vocal prosodic patterns, and which illustrates their anchoring at the interface between auditory and social cognition, involving the cooperation of distributed auditory prosodic, sociocognitive, and cingulo-opercular brain areas [81].

These attention-capturing sounds in animals are often characterized by loud protracted bouts of harsh and variable vocalizations, which include rapidly varying combinations of loud, noisy screams and piercing high-frequency tonal cries, with dramatic amplitude and frequency modulations, which together are able to increase the arousal state of the mother, including human ones [74,82]. It has been shown, moreover, that screaming is one of the most relevant communication signals in humans for survival. By using a recently developed, neurally informed characterization of sounds (modulation power spectrum) see [83,84], it has been demonstrated that human screams cluster within a rather restricted portion of the acoustic space between about 30 and 150 Hz, which corresponds to the

perceptual attribute of roughness. This acoustic roughness has been found also to engage subcortical structures, which are critical to the rapid appraisal of danger [85].

The vocal repertoire of most primate species, however, is not limited to these attention-capturing sounds. There is also an additional class of sounds, which are referred to as "sonants and gruffs" and which may be considered as structural opposites of these arousal-increasing sounds [74]. Instead of being unpatterned and chaotic, they are tonal and harmonically rich, with a more diffuse regularly patterned broadband spectral structure. Rather than having direct impact on listener's arousal and affect, they seem to induce a less inherent affective force. Their richly structured spectra, moreover, make them even suited for revealing clear cues to the caller's identity since their individual idiosyncrasies impart individually distinctive voice cues that are associated either with the dynamic action of the vocal folds or with the resonance properties of the vocal tract cavities [86,87]. Chimpanzees, likewise, are able to intentionally use grunts as referential calls and to learn new calls from other individuals [54], which represents most probably an early stage of the evolution of lexical meaning (but see [88]). However, although the monkeys' vocal tract is ready to generate speech sounds [89], language and music seem to necessitate more elaborate neural processing mechanisms and vocal control [46].

3. Affective Sounds and Vocalizations

Speech—at least in its most primitive appearance—and music seem to share a common affective substrate. Studying emotional communication by means of speech and music, therefore, can benefit from a thorough investigation of their underlying mechanisms. One field of research that has been particularly fruitful in this regard has been the study of auditory affective processing that was conducted in the context of *speech prosody* [13]. It has been argued, in fact, that two separate neuroanatomic channels with different phylogenetic histories participate in human acoustic communication to support either nonverbal affective vocalization or articulate speech [90,91]. This *dual-pathway model* of human acoustic communication clearly distinguishes the propositional and emotional contents of spoken language, which rely on channels that are seated in separate brain networks that create different data structures, which are known as analogue versus digital (see below). Both channels, however, must coordinate to some extent, but the functional mechanisms and neuroanatomic pathways underlying their intertwined integration are still not totally clear [92].

Affective prosody, further, is opposed to the discrete coding of speech, which is used in the case of phonemes, words, and those aspects of music that consist of pitches and durations. Its expressive dynamics can be modelled more effectively by continuous variables, as is the case with emotional gestures that are shared not only by all humans, but also by a broader group of animals, including many taxa of mammals and even other vertebrates [51]. The same dynamics of affective prosody—as an evolutionarily old form of communication—are to be found, in fact, in the prosody of human language and in the vocal expressions of different mammalian species, which could mean that its use in human acoustic communication has deep phylogenetic roots that are present in the vocal communication systems of nonhuman animals as well. Consistent structures, in fact, can be seen in acoustic signals that communicate affective states, such as high-pitched, tonal sounds in expressions of submission and fear, and low, loud, broadband sounds in expressions of threats and aggression. Animal signals may thus have direct effects on listeners. They may not simply provide information about the caller, but may effectively manage or manipulate the behavior of listeners [93] (see also [76]). This *prehuman origin hypothesis* of affective prosody locates its grounding in innate mechanisms, which have a prehuman basis and which are used to discriminate between different emotions, both qualitatively (anger, fear, joy, sadness, boredom, etc.) and quantitatively (affect intensity) [52]. It has been shown, moreover, that there exists a functional dissociation between brain regions that process the quality of acoustically conveyed emotions (orbitofrontal cortex) and those that process the intensity of that emotion (amygdala) [94]. Current research has also revealed a high degree of acoustic flexibility in

attention-attracting sounds in nonhuman mammalian species, which points in the direction of more complex acoustic signaling and processing mechanisms [95].

As such, it can be argued that the study of the faculties of language and music can benefit from a comparative approach that includes communication and cognition in humans and nonhuman animals alike [46]. The capacity to learn language, in fact, requires multiple, separable mechanisms, which include the ability to produce, perceive, and learn complex signals as well as to interpret and control them. Some of them seem to have figured already in the common ancestors of both humans and animals, some others evolved later. Relying on comparative data from living animals, therefore, may be definitively helpful to address these issues. Acoustic signaling in humans, in this view, may have roots in the vocal production, auditory perception, and cognitive processing capabilities of nonhuman mammals, and the study of affective prosody, as a shared component of human speech, music, and nonverbal acoustic communication, in particular, may shed some light on the evolutionary roots of human speech and music as well as the evolution of meaning itself. It is important, in this regard, to consider also the role of *iconicity*—the similarity between some aspects of sound to some aspects of meaning—in linking the sound to meaning in language. It should be noted, in fact, that affective prosody is considered a paralinguistic property, which accompanies the semantic meaning arising from the symbolic system of human language. The question of how meaning emerges from symbolic signs, therefore, cannot be fully understood by focusing only on prosodical features of language, which work in parallel to the semantic processing. Here, an iconic relationship between sound and the meaning of words that has traditionally been considered as only a marginal property of language (e.g., onomatopoeia, and to some extent also phonaesthemes, i.e., a phoneme or group of phonemes, which has recognizable semantic associations as the result of appearing in a number of words with similar meanings, such as, e.g., the English onset /sn-/ in snarl, snout, sniff, snuffle), has been assumed to serve as an interface for accomplishing the need to map linguistic form to human experience as a vital part of meaning making. Iconicity, thus, has been shown to play an important role for both phylogenetic language evolution (e.g., [96]) and ontogenetic language development (e.g., [97]). This holds in particular for the correspondences between the sound and meaning of words in the affective domain, termed *affective iconicity* [98], which have been supported by recent empirical results indicating that the specific sound profile of a word can be attributed to a specific affective state, which, in turn, can contribute to the perception of the affective meaning of that word, such as, e.g., whether it designates something positive/negative or arousing/calming [99]. Importantly, the affectivity in the sound of words in a language has been shown to be processed in similar brain regions that are involved in processing other types of affective sounds, such as emotional vocalization and affective prosody [100,101]. In addition, such affective potential in the sound of words is even capable of interacting with higher cognitive processes, such as affective evaluation of the words' meaning [102]. All this suggests that consciously experienced meaning is inferred from a number of cues that reflects a hierarchy of sound processing.

It is possible, further, to conceive of this hierarchy in the processing of sounds, reflecting the evolutionary history of human sound communication from early mammals, showing an extension of the perceivable spectrum of sound frequency related to the evolution of the mammalian ear [103], to primates. Non-human primates and early hominins, for example, are an especially interesting group in which to consider the potential affective influence of vocalizations on listeners. Because of their large brains and their phylogenetic proximity to humans, traditional research has focused mostly on "higher-level" cognitive processes that organize communication in higher primates. Yet, they still can rely on the neurophysiological substrates for affective influence, which are still very broadly conserved. It is likely, therefore, that affective influence is an important part of the vocal signals of non-human primates [74]. As such, it is possible to conceive of hierarchical levels of affective signaling, starting from loud calls and vocalizations of early hominids, over prelinguistic affective processing of sound by neonates to infant-directed speech, affective speech, and even music. The step via onomatopoeia and iconicity, finally, could be added as a last step from affective to referential signaling.

The loud calls of *early hominins* are exemplified most typically in a broad class of vocalizations with acoustic features that have direct impact on animal perception, as mentioned already above: Sharp signal onsets, dramatic frequency and amplitude fluctuations, and chaotic spectral structures [104]. *Neonates* are another interesting group for the study of prelinguistic affective processing of sound. They have been shown to possess complex endowments for perceiving and stimulating parental communicative signals by discriminating timing patterns, pitch, loudness, harmonic interval, and voice quality [65]. They also seem to react to the human voice and display imitations of facial expressions, vocalizations, and hand movements, showing an awareness of human signals that is already comprehensive, multimodal, and coherent at birth [67]. As a result, people, all over the world, have capitalized on this sensitivity by developing *infant-directed speech* or *motherese* (see below), which is obviously more simplified than adult speech, and which involves exaggerated prosodic features, such as wider excursions of voice pitch, more variable amplitude, tempo, and delivery, and more varied patterns of word stress [74]. All these features have been the subject of research on auditory affective processing, which has been conducted mainly in the context of speech prosody, which has been coined also the "third element of language" [105]. Vocal emotion perception in speech, further, has been studied by using test materials consisting of speech, spoken with various emotional tones by actors, and nonverbal interjections or *affect bursts*, such as laughter or screams of fear [106] (see for an overview). These vocal expressions, which usually accompany intense emotional feelings, along with the corresponding facial expressions, are closely related to *animal affect vocalizations* [107], which can be defined as short, emotional non-speech expressions, which comprise both clear non-speech sounds (e.g., laughter) and interjections with a phonemic structure (e.g., 'Wow'), but which exclude verbal interjections that can occur as a different part of speech (like 'Heaven', 'No', etc.)" [108].

These nonverbal affect bursts have proven to be useful for the study of meaning. They provide an interesting class of affective sounds, which have been collected in validated sets of auditory stimuli—such as the Montreal Affective Voices (MAV) [106] and the "Musical Emotional Burst (MEB) for musical equivalents [109]. Using nonverbal sounds, moreover, presents several advantages over verbal ones: The stimuli do not contain semantic information, there are no linguistic barriers, the expression of emotion is more primitive and closer to the affect expressions of animals or human babies, and they are more similar to the Ekman faces [110] used in the visual modality than emotional speech. As such, they avoid possible interactions between affective and semantic content, they can be used for the study of cross-cultural differences, and they allow better comparisons across modalities, as well as studies of cross-modal emotional integration [106].

Affect bursts, however, are limited in their semantic content, but are able to communicate by sound [51,111]. Being evolutionarily older than singing and speech, they have been considered as their precursors to some extent. Singing is one of the interesting ways of sound expression, which goes beyond the transmission of semantic information. It can be questioned, however, whether every kind of music—as an evolved and cultural product—exploits such pre-existing perceptual sensitivities, which were originally evolved thanks to a variety of auditory functions, such as navigating sonic environments and communication by means of singing. Cultural evolution, in this regard, has led to increasingly complex and cumulative musical developments through processes of sensory exploitation [112].

4. Calls, Vocalizations, and Human Music: Affectively-Based Sound–Meaning Relationships

Music has inductive power. It can move listeners emotionally and physically by means of the information-processing mechanisms it engages. The majority of these mechanisms, however, did not evolve as music-specific traits. Some of them are related to the processing of sound that is recognized as being similar to voices, objects that are approaching, or the sounds of animals. As such, this processing seems to involve cognitive processes of attraction and cultural transmission mechanisms that have cumulatively and adaptively shaped an enormous variety of signals for social relationships [112]. Music, in this view, is an inherently social phenomenon, and the same holds true for loud calls of

nonhuman primates, especially those of the Old-World monkeys, which, most likely, were the substrate from which singing could evolve [36].

This brings us to the question of the origins of language and music and their mutual relationship. It has been hypothesized, e.g., that language seems to be more related to logic and the human mind, whereas music should be grounded in emotion and the human body [113] (see for an overview). This dichotomous approach has been questioned, however, in the sense that language and music could evolve from common roots, a common musical protolanguage [24,42]. Especially, the *loud calls* in modern apes and music in modern humans seem to be derived from such a common ancestral form. The calls are believed to serve a variety of functions, such as territorial advertisement, inter-group intimidation and spacing, announcing the precise locality of specific individuals, food sources, or danger, and strengthening intra-group cohesion. The most likely function of early hominin music, on the other hand, was to display and reinforce the unity of a social group toward other groups [36]. This is obvious in vocalizing and gesturing together in time, where the ability to act musically underlies and supports human companionship. It seems likely, moreover, that the elements of communicative musicality are necessary for joint human expressiveness to arise and that they underlie all human communication [11,66].

As such, it seems that a major ancestral function of calls, protolanguage, and music may be related to several kinds of signaling, attention capturing, affective influence, and group cohesion rather than conveying propositional knowledge that is related to higher level cognitive processes that are involved in the communication of contemporary humans. This brings us to the role of *affective semantics*, as the domain that studies semantic constructs that are grounded in the perceptual-affective impacts of sound structure [74]. Empirical grounding for that kind of signaling has been provided by a typical class of primate vocalizations, which are known as *referential emotive vocalizations* [58] and separation calls [114]. There are, in fact, a number of important affective effects of sounds and vocalizations, such as, e.g., attention capturing mechanisms, which are used also in speech directed to young infants with the function to focus and maintain attention and to modulate arousal by using dramatic frequency variations. As such, there is a whole domain of acoustic signals which goes beyond the lexico-semantic level of communication and which is shared between humans and non-human animals. There are, as such, acoustic attributes of aroused vocalizations which are shared across many mammalian species and which humans can use also to infer emotional content. Humans, as a rule, use multiple acoustic parameters to infer relative arousal in vocalizations, but they mainly rely on the fundamental frequency and spectral centre of gravity to identify higher arousal vocalizations across animal species, thus suggesting the existence of fundamental mechanisms of vocal expressions that are shared among vertebrates, and which could represent a homologous signaling system [115].

Such core affective effects of vocal signals may be functional. Yet they do not undercut the role of cognition and the possibility of more complex communicative processes and outcomes, such as speech communication in people. The latter can be seen as a refinement of phylogenetically older vocal production and perception abilities that are shared with non-human animals [91]. These abilities may scaffold, in part, an increasing communicative complexity, which means that at least some of the semantic complexity of human language might capitalize on affectively-based sound–meaning relationships. It is probable, therefore, that evolutionarily older ways of interpreting acoustical cues can be involved in the construction of more complex meaning. Such preprepared or early acquired sound–sense relationships represent a form of intrinsic or original meaning that provides a natural foundation from which increasingly complex semantic systems may be constructed, both developmentally and evolutionarily. This foundation can explain the universal tendency first observed by Köhler [116] (pp. 224-225) to associate pseudowords, such as *takete* or *kiki*, with spiky shapes whereas *malumba* or *bouba* are associated with round shapes [117]. It has been shown, moreover, that the communicative importance of the affective influence of vocal signals does not disappear when brains get larger and their potential for cognitive, evaluative control of behavior increases. It is likely,

therefore, that complex communicative processes exploit and build on the phylogenetically-ancient and widespread affective effects of vocal signals [74] (p. 183).

5. Sound Communication, Emotion, and Affective Speech

Sounds can have a considerable affective effect on listeners and this holds true also for non-human animals that use many of their vocal signals precisely to exert these effects. There is, as such, a relationship between the acoustic structure in animal signals and the communicative purposes they purport [74,112]. This is obvious in vocalizations of non-human primates, which bear the mark of design for direct effects on the listener's affect and behavior, as exemplified most typically in alarm vocalizations that are produced during encounters with predators [91]. These alarm calls tend to be short, broadband calls, with an abrupt-onset, standing out against background noise, thus being easy to localize. As such, they display acoustic features for capturing and manipulating the attention and arousal in listeners. They have been studied already in the 1970s in the context of agonistic vocalizations that are involved in confrontations or competitions with others. Among their most important features is a low fundamental frequency (F_0) and a tendency towards aperiodicity, with a possible explanation that low, broadband sounds with a wide frequency range are often tied to body size and hostile intent. Such sounds, presumably, can induce fear in the receivers. High pitched sounds with tone-like high F_0, on the contrary, are related to appeasement and are often produced to reduce fear in listeners [118,119]. This illustrates again how sound is often more important than semantic meaning in animals' signals.

Similar findings have been reported also for humans. Prohibitive utterances across cultures, for example, contain similar acoustic features, such as a fast rising amplitude, lowered pitch, and small repertoires [112]. A more elaborated field of research, however, is the study of *motherese* or *infant-directed speech* [65]. Mothers, as a rule, speak in short bursts and talk in an inviting sing-song manner with the baby occasionally answering back. Young infants, moreover, stimulate their caregivers to a kind of musical or poetic speech, which can move into wordless song with imitative, rhythmic, and repetitive nonsense sounds. Such baby–mother interactions imply communicative interactions, which have also been called "communicative musicality" [11]. They suggest an awareness of human signals which is present at birth, with newborns reacting to the human voice and imitating facial expressions, vocalizations, and hand movements. It means that young infants possess complex endowments for perceiving and stimulating parental communicative signals by discriminating timing patterns, pitch, loudness, harmonic interval, and voice quality [65]. Effective communication, in this view, must be held by means other than lexical meaning, grammar, and syntax, with mothers and babies being highly "attuned" to the vocal and physical gestures of the mother. Both seem to explore pitch-space in a methodical manner over short and long intervals of time [11]. This has been reported extensively by the Papoušeks [6,19], who both have stressed the importance of early childhood musical behaviors as forms of play to nurture children's exploratory competence. They have studied intensively infant–caregiver interactions and focused on the musicality of these interactions, stressing the indivisibility of music and movement. It has been found, in fact, that music and movement share a dynamic structure that supports universal expressions of emotion as exemplified in particular in infants' predispositions for perceptual correspondences between music and movement. This ability, further, seems to be possible by the existence of prototypical emotion-specific dynamic contours, but also by isomorphic structural relationships between music and movement [120].

They found out that the parent's multimodal stimulation is, so to say, tailored to the infant's early competence for perceiving information through different senses and that "regular synchronization of vocal and kinaesthetic patterns provides the infant with multimodal sensory information including tactile, kinaesthetic and visual information." [6] (p. 100). Similar findings have been reported by Trevarthen [121], who has centered on the temporal characteristics of the infant–caregiver interaction. The rhythmicity of this interaction can be described as the capacity of the infant to follow and respond to temporal regularities in vocalization and movement, and to initiate temporally regular sets of

vocalizations and movements. What he proposes is a conceptual framework to explore the expression and development of communication or intersubjectivity through empirical observations and analyses of infant–caregiver interaction. It enables the sharing of patterned time with others and facilitates harmonizing the affective state and interaction [27].

As such, there seems to be an evolutionarily old layer of sound communication that exists in speech, but that arouses emotion in singing as well. This happens in a hierarchic order with the evolutionarily older elements being most basic and effective, and those which are acquired in processes of socialization being most subtle and conventional. Primitive affective vocalizations, therefore, are considered as more authentic and more truly felt information than conventional and ritual information [10,122], and a great deal of music is also designed specifically to give rise to these affective effects [74].

6. Sound/Speech Understanding and the Gestural Approach

Language and music can be considered as sound-signal using communication systems. There is, however, a distinction with respect to their respective semantics, which can be either lexico-semantic or action-oriented. In language, as well as in music, the vocal or acoustic characteristics may help to convey an impression, but it has been shown that the position of the eyebrows and the facial expression as a whole, may have the same function [119]. Many facial gestures, in fact, are part of a multi-modal array of signals, and facial expressions may even influence the acoustic cues of the expression by vocal tract deformation [13].

This brings us to the question of bimodality and audiovisual integration of emotional expressions [123]. Even in visible emotion, for example, the auditory modality can carry strong information, which is not only related to the consequences of the facial gestures [13]. In this context, it is important to remind the musicality of infant–caregiver interactions with synchronous stimulation that provides continuous multimodal sensory information (see above). This multimodal stimulation, further, entails processes of affective and behavioral resonance in the sense that the neurophysiological organization of behavior depends on a reciprocal influence between systems that guides both the production, perception, interpretation, and response to the behavior of others, somewhat reminiscent of the discovery of mirror and canonical neuron systems in primate brains [124]. This means that seeing an object or an action performed by someone else can activate the same neurons as when one is performing this action oneself. However, the multimodal stimulation can be even stronger. It has been shown, for example, that if acoustic speech is the main medium for phonetic decoding, some integration with the visual modus cannot be avoided [125]. As such, there is a lot of interest in the role of the co-occurrence of sight and sound, with a special focus on research on emotion effects on voice and speech [61].

Multimodal stimulation entails interactions between individuals, which is obvious in the ability to vocalize and gesture together—as in synchronous chorusing and gesturing—both in humans and nonhuman primates [126]. The ability to act musically and to move sympathetically with each other, accordingly, seems to be the vehicle for carrying emotions from one to someone else. It underlies human companionship in the sense that elements of communicative musicality are necessary for joint human expressiveness to arise [11].

Speech, as a later evolutionarily development, pays tribute to this interactive, gestural approach. It is a basic claim of articulatory phonology, which states that articulatory gestures and gestural organization can be used to capture both categorical and gradient information [55]. They can be described as events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators. Gestures, in this view, are dynamic articulatory structures, which consist of the formation and release of constrictions in the vocal tract. As such, they can be described in terms of task-dynamics, which have been used to model different kinds of coordinated multi-articulator actions, such as reaching and speaking. It means also that the same gestural structures

may simultaneously characterize phonological properties of the utterance (contrastive units and syntagmatic organization) and physical properties.

7. Sound Comprehension in Speech and Music: Spectral and Temporal Cues

Articulatory gestures are situated at the productive level of vocal communication. There is, however, also the receptive level, which is related to the recognition of acoustic parameters, such as, for example, spectral cues when we discriminate pitch in music [127] and intonation patterns in speech [128]. Sound comprehension, in this view, should be related to the recognition of the acoustic profiles of vocal expression, as exemplified most typically in emotional expression. It has been stated erroneously that the voice might only reflect arousal. Recent research, using a larger number of parameters, has shown that spectro-temporal parameters play a major role in differentiating qualitative differences between emotions [129]. This is obvious, for example, in the vocal repertoire of most primate species with a clear distinction between squeaks, shrieks, and screams, with direct impact on the listener's arousal and affect, and sonants and gruffs, with structured spectra that provide an excellent medium for revealing clear cues to the identity of the caller (see above). These cues, which are highly idiosyncratic, impart distinctive voice cues in the acoustic features of these calls, which are associated with the patterns of dynamic action of the vocal folds or with the resonance properties of the vocal tract cavities [74,87]. Human infants, accordingly show an impressive acoustic sensitivity, which allow them to discriminate timing patterns, pitch, loudness, harmonic interval, and voice quality [11], with many perceptual biases being in place before articulated speech evolved [112]. Importantly, although all these features depend on acoustic parameters, they are in fact auditory phenomena [79]. It means that the discrimination of vocal cues is the interpretation of sound stimuli by the nervous system influenced by genetic (both species specific and shared with other taxa) and environmental (including cultural) factors.

Music as well as speech can be considered as dynamic stimuli with sounds changing continuously across the time of presentation. This means that new sensory information is added serially during sound presentation, with physiological systems that respond to simple changes in the physical stimulus being continuously active. Sounds, moreover, are dynamic and often require an accrual of information over time to be interpreted [130]. The effects of speech and music, therefore, are related in important ways to the information-processing mechanisms they engage. As a result, humans interpret speech and music sounds not only as expressive information, but also as coherent sound structures, which convey the whole pack of information. Even at this level, however, both speech and music structures are auditory phenomena which rely to a different degree on acoustical cues. In the case of phonemes recognition [131] and timbre discrimination in music [132], the most important cues are spectro-temporal. Spectral cues, in contrast, are crucial in the discrimination of intonation patterns in speech and pitch class structure in music [127].

The main difference between speech and music in this regard consists in the role of particular acoustic cues played in the transmission of meaning. While spectro-temporal cues are crucial for the recognition of words, they seem to be less important as far as the music structure is concerned. It means that spectro-temporal cues evolved in humans as a main source of transmitting lexical meaning. In contrast, spectral cues are important for discrete pitch class discrimination in music—one of the main elements of musical structure—which is deprived of lexical meaning. Nonetheless, spectral cues can contribute to the lexical meaning in tone languages where the relative change of pitch influences the interpretation of the word meaning [133]. Even in tone languages, however, lexical meaning is conveyed mainly by the means of spectro-temporal cues. Similarly, temporal cues can be used as an additional source of information which influences lexical meaning in "quantity languages", which are sensitive to the duration of the segments for the assignment of their meaning [134,135]. It has been shown also that spectral and temporal cues contribute to the signaling of the word meaning in non-tonal languages as well [136], with the extent to which these cues are important for the transmission of lexical meaning being dependent on the particular language.

8. Conclusions and Perspectives

In this paper, we described the role of preconceptual spectral and temporal cues in sound communication and in the emergence of meaning in speech and music, stressing the role of affective vocalizations as a common ancestral instrument in communicative interactions. In an attempt to search for shared components between speech and music, we have stressed their commonalities by defining speech and music as sensory rich stimuli. Their experience, moreover, involves different body channels, such as the face and the voice, but this bimodal approach has proven to be too restrictive. It has been argued, therefore, that an action-oriented approach is more likely to describe the reciprocity between multisensory processing and articulatory-motor routines as phonological primitives. As such, a distinction should be made between language and speech, with the latter being more centripetal in directing the attention of the listener to the sounding material itself, whereas language is mainly centrifugal in directing the attention away from the text to function referentially. There are, however, commonalities as well and the shared component between speech and music is not meaning, but sound. Therefore, to describe quite systematically the transition from sound to meaning in speech and music, one must stress the role of emotion and affect in early sound processing, the role of vocalizations and nonverbal affect burst in communicative sound comprehension, and the acoustic features of affective sound with a special emphasis on temporal and spectrographic cues as parts of speech prosody and musical expressiveness.

One of the major findings in this regard was a kind of hierarchy in the type of meaning that is conveyed, with a distinction between analog and digital usage of the sound. Especially, the role of affective prosody seems to be important here. As a typical example of analog processing, it goes beyond a mere discrete coding of speech and music, stressing the wider possibilities of sound-signal communications systems rather than relying merely on semantic content and propositional knowledge. As such, there seems to be a major ancestral function of affect burst, calls, protolanguage, and music which are related to several kinds of signaling, attention capturing, affective influence, and group cohesion. They hold a place in a developmental continuum at the phylogenetic and ontogenetic level.

The view presented thus suggests that meaning in language and music is a complex phenomenon which is composed of hierarchically organized features, which are mostly related to the interpretation of acoustical cues by the nervous system. The bulk of this interpretation, moreover, is processed at an unconscious level. More studies are needed, however, to better understand the role of spectral and temporal cues as sources of information in the complex process of human communication. Inter-species and inter-cultural comparative studies are especially promising in this respect, but equally important are developmental investigations, which together with genetic research can elucidate the interconnection between the environmental and hereditary information in the process of the development of human vocal communication.

Author Contributions: The first draft of this article was written by M.R. The final version was prepared jointly by M.R and P.P.

Funding: This research received no external funding.

Acknowledgments: We thank the anonymous reviewers. Their critical remarks were very helpful in updating our summary of the current available research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Trehub, S.E.; Hannon, E.E. Infant Music Perception: Domain-General or Domain-Specific Mechanisms? *Cognition* **2006**, *100*, 73–99. [CrossRef] [PubMed]
2. Zentner, M.R.; Kagan, J. Perception of Music by Infants. *Nature* **1996**, *383*, 29. [CrossRef] [PubMed]
3. Cooper, R.P.; Aslin, R.N. Preference for Infant-Directed Speech in the First Month after Birth. *Child. Dev.* **1990**, *61*, 1584–1595. [CrossRef] [PubMed]

4. Fernald, A.; Kuhl, P. Acoustic Determinants of Infant Preference for Motherese Speech. *Infant Behav. Dev.* **1987**, *10*, 279–293. [CrossRef]
5. Masataka, N. Preference for Infant-Directed Singing in 2-Day-Old Hearing Infants of Deaf Parents. *Dev. Psychol.* **1999**, *35*, 1001–1005. [CrossRef] [PubMed]
6. Papoušek, M. Intuitive Parenting: A Hidden Source of Musical Stimulation in Infancy. In *Musical Beginnings Origins and Development of Musical Competence*; Deliège, I., Sloboda, J., Eds.; Oxford University Press: Oxford, NY, USA, 1996; pp. 88–112.
7. Werker, J.F.; McLeod, P.J. Infant Preference for Both Male and Female Infant-Directed Talk: A Developmental Study of Attentional and Affective Responsiveness. *Can. J. Psychol. Can. Psychol.* **1989**, *43*, 230–246. [CrossRef]
8. Trainor, L.J. Infant Preferences for Infant-Directed versus Noninfant-Directed Playsongs and Lullabies. *Infant Behav. Dev.* **1996**, *19*, 83–92. [CrossRef]
9. Trainor, L.J.; Clark, E.D.; Huntley, A.; Adams, B.A. The Acoustic Basis of Preferences for Infant-Directed Singing. *Infant Behav. Dev.* **1997**, *20*, 383–396. [CrossRef]
10. Gorzelańczyk, E.J.; Podlipniak, P. Human Singing as a Form of Bio-Communication. *Bio-Algorithms Med. Syst.* **2011**, *7*, 79–83.
11. Malloch, S.N. Mothers and Infants and Communicative Musicality. *Music. Sci.* **1999**, *3*, 29–57. [CrossRef]
12. Wermke, K.; Mende, W. Musical Elements in Human Infants' Cries: In the Beginning Is the Melody. *Music. Sci.* **2009**, *13*, 151–175. [CrossRef]
13. Aubergé, V.; Cathiard, M. Can We Hear the Prosody of Smile? *Speech Commun.* **2003**, *40*, 87–97. [CrossRef]
14. Fernald, A.; Mazzie, C. Prosody and Focus in Speech to Infants and Adults. *Dev. Psychol.* **1991**, *27*, 209–221. [CrossRef]
15. Panksepp, J.; Bernatzky, G. Emotional Sounds and the Brain: The Neuro-Affective Foundations of Musical Appreciation. *Behav. Process.* **2002**, *60*, 133–155. [CrossRef]
16. Fassbender, C. Infants' Auditory Sensitivity towards Acoustic Parameters of Speech and Music. In *Musical Beginnings Origins and Development of Musical Competence*; Deliège, I., Sloboda, J., Eds.; Oxford University Press: Oxford, NY, USA, 1996; pp. 56–87. [CrossRef]
17. Trehub, S.E. Musical Predispositions in Infancy: An Update. In *The Cognitive Neuroscience of Music*; Peretz, I., Zatorre, R.J., Eds.; Oxford University Press: Oxford, NY, USA, 2003; pp. 2–20. [CrossRef]
18. Trehub, S.E.; Schellenberg, E.; Glenn Hill, D.S. The Origins of Music Perception and Cognition: A Developmental Perspective. In *The Origins of Music Perception and Cognition: A Developmental Perspective*; Deliège, I., Sloboda, J.A., Eds.; Psychology Press/Erlbaum (UK) Taylor & Francis: Hove, UK, 1997; pp. 103–128.
19. Papoušek, H. Musicality in Infancy Research: Biological and Cultural Origins of Early Musicality. In *Musical Beginnings Origins and Development of Musical Competence*; Deliège, I., Sloboda, J.A., Eds.; Oxford University Press: Oxford, NY, USA, 1996; pp. 37–55. [CrossRef]
20. Brown, S. The "Musilanguage" Model of Musical Evolution. In *The Origins of Music*; Wallin, N.L., Merker, B., Brown, S., Eds.; The MIT Press: Cambridge, MA, USA, 2000; pp. 271–300.
21. Fenk-Oczlon, G.; Fenk, A. Some Parallels between Language and Music from a Cognitive and Evolutionary Perspective. *Music. Sci.* **2009**, *13*, 201–226. [CrossRef]
22. Koelsch, S.; Kasper, E.; Sammler, D.; Schulze, K.; Gunter, T.; Friederici, A.D. Music, Language and Meaning: Brain Signatures of Semantic Processing. *Nat. Neurosci.* **2004**, *7*, 302–307. [CrossRef] [PubMed]
23. Steinbeis, N.; Koelsch, S. Comparing the Processing of Music and Language Meaning Using EEG and FMRI Provides Evidence for Similar and Distinct Neural Representations. *PLoS ONE* **2008**, *3*, e2226. [CrossRef]
24. Fitch, W.T. The Biology and Evolution of Music: A Comparative Perspective. *Cognition* **2006**, *100*, 173–215. [CrossRef] [PubMed]
25. Hauser, M.D.; McDermott, J. The Evolution of the Music Faculty: A Comparative Perspective. *Nat. Neurosci.* **2003**, *6*, 663–668. [CrossRef] [PubMed]
26. Cross, I. Music, Cognition, Culture, and Evolution. *Ann. N. Y. Acad. Sci.* **2001**, *930*, 28–42. [CrossRef] [PubMed]
27. Cross, I. Music, Mind and Evolution. *Psychol. Music* **2001**, *29*, 95–102. [CrossRef]
28. Cross, I. The Evolutionary Nature of Musical Meaning. *Music. Sci.* **2009**, *13*, 179–200. [CrossRef]

29. Huron, D. Is Music an Evolutionary Adaptation? In *The Cognitive Neuroscience of Music*; Peretz, I., Zatorre, R.J., Eds.; Oxford University Press: Oxford, NY, USA, 2003; pp. 57–75. [CrossRef]
30. McDermott, J.; Hauser, M. The Origins of Music: Innateness, Uniqueness, and Evolution. *Music Percept.* **2005**, *23*, 29–59. [CrossRef]
31. Reybrouck, M. From Sound to Music: An Evolutionary Approach to Musical Semantics. *Biosemiotics* **2013**, *6*, 585–606. [CrossRef]
32. Tolbert, E. Music and Meaning: An Evolutionary Story. *Psychol. Music* **2001**, *29*, 84–94. [CrossRef]
33. Wallin, N.L. *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music*; Pendragon Press: Stuyvesant, NY, USA, 1991.
34. Brown, S.; Merker, B.; Wallin, N.L. An Introduction to Evolutionary Musicology. In *The Origins of Music*; Wallin, N.L., Merker, B., Brown, S., Eds.; The MIT Press: Cambridge, MA, USA, 2000; pp. 3–24. [CrossRef]
35. Drake, C.; Bertrand, D. The Quest for Universals in Temporal Processing in Music. In *The Cognitive Neuroscience of Music*; Peretz, I., Zatorre, R.J., Eds.; Oxford University Press: Oxford, NY, USA, 2003; pp. 21–31. [CrossRef]
36. Geissmann, T. Gibbon Songs and Human Music from an Evolutionary Perspective. In *The Origins of Music*; Wallin, N.L., Merker, B., Brown, S., Eds.; The MIT Press: Cambridge, MA, USA, 2000; pp. 103–123.
37. Huron, D.B. The Other Semiotic Legacy of Charles Sanders Peirce: Ethology and Music-Related Emotion. In *Music, Analysis, Experience. New Perspectives in Musical Semiotics*; Meader, C., Reybrouck, M., Eds.; Leuven University Press: Leuven/Louvain, Belgium, 2015; pp. 185–208.
38. Darwin, C. *The Descent of Man, and Selection in Relation to Sex*, 1st ed.; John Murray: London, UK, 1871.
39. Ma, W.; Thompson, W.F. Human Emotions Track Changes in the Acoustic Environment. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14563–14568. [CrossRef] [PubMed]
40. Fitch, W.T. On the Biology and Evolution of Music. *Music Percept.* **2006**, *24*, 85–88. [CrossRef]
41. Fitch, W.T. Musical Protolanguage: Darwin's Theory of Language Evolution Revisited. In *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*; Bolhuis, J.J., Everaert, M., Eds.; The MIT Press: Cambridge, MA, USA, 2013; pp. 489–503.
42. Mithen, S.J. *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*; Harvard University Press: Cambridge, UK, 2006.
43. Thompson, W.F.; Marin, M.M.; Stewart, L. Reduced Sensitivity to Emotional Prosody in Congenital Amusia Rekindles the Musical Protolanguage Hypothesis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19027–19032. [CrossRef] [PubMed]
44. Hurford, J.R. *The Origins of Meaning: Language in the Light of Evolution*; Oxford University Press: Oxford, UK, 2007.
45. Hauser, M.D. *The Evolution of Communication*; The MIT Press: Cambridge, MA, USA, 1996.
46. Fitch, W.T.; Zuberbühler, K. Primate Precursors to Human Language: Beyond Discontinuity. In *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*; Altenmüller, E., Schmidt, S., Zimmermann, E., Eds.; Oxford University Press: Oxford, UK, 2013; Volume 16, pp. 27–48.
47. Panksepp, J. Affective Consciousness: Core Emotional Feelings in Animals and Humans. *Conscious. Cogn.* **2005**, *14*, 30–80. [CrossRef] [PubMed]
48. Bickerton, D. *Adam's Tongue: How Humans Made Language, How Language Made Humans*; Hill and Wang: New York, NY, USA, 2010.
49. Roederer, J.G. On the Concept of Information and Its Role in Nature. *Entropy* **2003**, *5*, 3–33. [CrossRef]
50. Hauser, M.D.; Konishi, M. *The Design of Animal Communication*; The MIT Press: Cambridge, MA, USA, 1999.
51. Merker, B. Is There a Biology of Music? And Why Does It Matter? In *Proceedings of the 5th Triennial ESCOM Conference*; Kopiez, R., Lehmann, A.C., Wolther, I., Wolf, C., Eds.; Hanover University of Music and Drama: Hanover, Germany, 2003; pp. 402–405.
52. Zimmermann, E.; Leliveld, L.; Schehka, S. Toward the Evolutionary Roots of Affective Prosody in Human Acoustic Communication: A Comparative Approach to Mammalian Voices. In *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*; Altenmüller, E., Schmidt, S., Zimmermann, E., Eds.; Oxford University Press: Oxford, NY, USA, 2013; pp. 116–132.
53. Merker, B. Music: The Missing Humboldt System. *Music. Sci.* **2002**, *6*, 3–21. [CrossRef]

54. Watson, S.K.; Townsend, S.W.; Schel, A.M.; Wilke, C.; Wallace, E.K.; Cheng, L.; West, V.; Slocombe, K.E. Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Curr. Biol.* **2015**, *25*, 495–499. [CrossRef] [PubMed]
55. Browman, C.P.; Goldstein, L. Articulatory Phonology: An Overview. *Phonetica* **1992**, *49*, 155–180. [CrossRef] [PubMed]
56. Bird, S.; Klein, E. Phonological Events. *J. Linguist.* **1990**, *26*, 33–56. [CrossRef]
57. Studdert-Kennedy, M. The Phoneme as a Perceptuomotor Structure. In *Cognitive Science Series Language Perception and Production: Relationships Between Listening, Speaking, Reading and Writing*; Allport, A., MacKay, D., Prinz, W., Scheerer, E., Eds.; Academic Press: London, UK, 1987; pp. 67–84.
58. Frayer, D.W.; Nicolay, C. Fossil Evidence for the Origin of Speech Sounds. In *The Origin of Music*; Wallin, N.L., Merker, B., Brown, S., Eds.; MIT Press: Cambridge, MA, USA, 2000; pp. 271–300.
59. Clynes, M. *Sentics: The Touch of Emotions*; Anchor Press: Garden City, NY, USA, 1977.
60. Arias, P.; Belin, P.; Aucouturier, J.-J. Auditory Smiles Trigger Unconscious Facial Imitation. *Curr. Biol. CB* **2018**, *28*, R782–R783. [CrossRef] [PubMed]
61. Scherer, K.R. Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Commun.* **2003**, *40*, 227–256. [CrossRef]
62. Bänziger, T.; Hosoya, G.; Scherer, K.R. Path Models of Vocal Emotion Communication. *PLoS ONE* **2015**, *10*, e0136675. [CrossRef] [PubMed]
63. Scherer, K.R.; Clark-Polner, E.; Mortillaro, M. In the Eye of the Beholder? Universality and Cultural Specificity in the Expression and Perception of Emotion. *Int. J. Psychol.* **2011**, *46*, 401–435. [CrossRef] [PubMed]
64. Scherer, K.R. Personality Inference from Voice Quality: The Loud Voice of Extroversion. *Eur. J. Soc. Psychol.* **1978**, *8*, 467–487. [CrossRef]
65. Trehub, S.E.; Unyk, A.M.; Trainor, L.J. Maternal Singing in Cross-Cultural Perspective. *Infant Behav. Dev.* **1993**, *16*, 285–295. [CrossRef]
66. Malloch, S.; Trevarthen, C. The Human Nature of Music. *Front. Psychol.* **2018**, *9*, 1680. [CrossRef] [PubMed]
67. Nadel, J.; Butterworth, G. *Imitation in Infancy*; Nadel, J., Butterworth, G., Eds.; Cambridge University Press: Cambridge, NY, USA, 1999.
68. Kyndrup, M. Mediality and Literature: Literature versus Literature. In *Why Study Literature*; Nielsen, H.S., Kraglund, R., Eds.; Aarhus University Press: Aarhus, Denmark, 2011; pp. 85–96.
69. Wierød, L.M.L. Where to Draw the Line? In *Music, Analysis, Experience*; Maeder, C., Reybrouck, M., Eds.; New Perspectives in Musical Semiotics; Leuven University Press: Leuven/Louvain, Belgium, 2015; pp. 135–148.
70. Ewens, G. *Die Klänge Afrikas: Zeitgenössische Musik von Kairo Bis Kapstadt*; Marino-Verlag: München, Germany, 1995.
71. Brandt, A.; Gebrian, M.; Slevc, L.R. Music and Early Language Acquisition. *Front. Psychol.* **2012**, *3*, 1–17. [CrossRef] [PubMed]
72. Rubin, D.C. *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads, and Counting-Out Rhymes*; Oxford University Press: Oxford, NY, USA, 1995.
73. Meyer, J. Typology and Acoustic Strategies of Whistled Languages: Phonetic Comparison and Perceptual Cues of Whistled Vowels. *J. Int. Phon. Assoc.* **2008**, *38*, 69–94. [CrossRef]
74. Rendall, D.; Owren, M.J. Vocalizations as Tools for Influencing the Affect and Behavior of Others. *Handb. Behav. Neurosci.* **2010**, *19*, 177–185. [CrossRef]
75. Gil-da-Costa, R.; Braun, A.; Lopes, M.; Hauser, M.D.; Carson, R.E.; Herscovitch, P.; Martin, A. Toward an Evolutionary Perspective on Conceptual Representation: Species-Specific Calls Activate Visual and Affective Processing Systems in the Macaque. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17516–17521. [CrossRef] [PubMed]
76. Owings, D.H.; Morton, E.S. *Animal Vocal Communication: A New Approach*; Cambridge University Press: Cambridge, UK, 1998.
77. Herzog, M.; Hopf, S. Behavioral Responses to Species-Specific Warning Calls in Infant Squirrel Monkeys Reared in Social Isolation. *Am. J. Primatol.* **1984**, *7*, 99–106. [CrossRef]
78. Seyfarth, R.M.; Cheney, D.L.; Marler, P. Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science* **1980**, *210*, 801–803. [CrossRef] [PubMed]
79. Huron, D.B. *Voice Leading: The Science Behind a Musical Art*; The MIT Press: Cambridge, MA, USA, 2016.

80. Owren, M.J.; Dieter, J.A.; Seyfarth, R.M.; Cheney, D.L. Vocalizations of Rhesus (Macaca Mulatta) and Japanese (M. Fuscata) Macaques Cross-Fostered between Species Show Evidence of Only Limited Modification. *Dev. Psychobiol.* **1993**, *26*, 389–406. [CrossRef] [PubMed]
81. Hellbernd, N.; Sammler, D. Neural Bases of Social Communicative Intentions in Speech. *Soc. Cogn. Affect. Neurosci.* **2018**, *13*, 604–615. [CrossRef] [PubMed]
82. Owings, D.H.; Zeifman, D. Human Infant Crying as an Animal Communication System: Insights from an Assessment/Management Approach. In *Evolution of Communication Systems: A Comparative Approach*; Oller, D.K., Griebel, U., Eds.; MIT Press: Cambridge, MA, USA, 2004; pp. 151–170.
83. Chi, T.; Gao, Y.; Guyton, M.C.; Ru, P.; Shamma, S. Spectro-Temporal Modulation Transfer Functions and Speech Intelligibility. *J. Acoust. Soc. Am.* **1999**, *106*, 2719–2732. [CrossRef] [PubMed]
84. Theunissen, F.E.; Elie, J.E. Neural Processing of Natural Sounds. *Nat. Rev. Neurosci.* **2014**, *15*, 355–366. [CrossRef] [PubMed]
85. Arnal, L.H.; Flinker, A.; Kleinschmidt, A.; Giraud, A.-L.; Poeppel, D. Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Curr. Biol. CB* **2015**, *25*, 2051–2056. [CrossRef] [PubMed]
86. Rendall, D.; Owren, M.J.; Rodman, P.S. The Role of Vocal Tract Filtering in Identity Cueing in Rhesus Monkey (Macaca Mulatta) Vocalizations. *J. Acoust. Soc. Am.* **1998**, *103*, 602–614. [CrossRef] [PubMed]
87. Ghazanfar, A.A.; Turesson, H.K.; Maier, J.X.; Van Dinther, R.; Patterson, R.D.; Logothetis, N.K. Vocal-Tract Resonances as Indexical Cues in Rhesus Monkeys. *Curr. Biol.* **2007**, *17*, 425–430. [CrossRef] [PubMed]
88. Fischer, J.; Price, T. Meaning, Intention, and Inference in Primate Vocal Communication. *Neurosci. Biobehav. Rev.* **2017**, *82*, 22–31. [CrossRef] [PubMed]
89. Fitch, W.T.; De Boer, B.; Mathur, N.; Ghazanfar, A.A. Monkey Vocal Tracts Are Speech-Ready. *Sci. Adv.* **2016**, *2*, e1600723. [CrossRef] [PubMed]
90. Ackermann, H.; Hage, S.R.; Ziegler, W. Brain Mechanisms of Acoustic Communication in Humans and Nonhuman Primates: An Evolutionary Perspective. *Behav. Brain Sci.* **2014**, *37*, 529–546. [CrossRef] [PubMed]
91. Owren, M.J.; Amoss, R.T.; Rendall, D. Two Organizing Principles of Vocal Production: Implications for Nonhuman and Human Primates. *Am. J. Primatol.* **2011**, *73*, 530–544. [CrossRef] [PubMed]
92. Scherer, K.R.; Johnstone, T.; Klasmeyer, G. Vocal Expression of Emotion. In *Handbook of Affective Sciences*; Series in affective science; Oxford University Press: New York, NY, USA, 2003; pp. 433–456.
93. Snowdon, C.T.; Teie, D. Affective Responses in Tamarins Elicited by Species-Specific Music. *Biol. Lett.* **2010**, *6*, 30–32. [CrossRef] [PubMed]
94. Lewis, P.A.; Critchley, H.D.; Rotshtein, P.; Dolan, R.J. Neural Correlates of Processing Valence and Arousal in Affective Words. *Cereb. cortex* **2007**, *17*, 742–748. [CrossRef] [PubMed]
95. Altenmüller, E.; Schmidt, S.; Zimmermann, E. *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*; Altenmüller, E., Schmidt, S., Zimmermann, E., Eds.; Oxford University Press: Oxford, UK, 2013.
96. Roberts, G.; Lewandowski, J.; Galantucci, B. How Communication Changes When We Cannot Mime the World: Experimental Evidence for the Effect of Iconicity on Combinatoriality. *Cognition* **2015**, *141*, 52–66. [CrossRef] [PubMed]
97. Monaghan, P.; Shillcock, R.C.; Christiansen, M.H.; Kirby, S. How Arbitrary Is Language? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2014**, *369*, 20130299. [CrossRef] [PubMed]
98. Aryani, A. Affective Iconicity in Language and Poetry, Freie Universität Berlin. Ph.D. Dissertation, 2018. Available online: https://refubium.fu-berlin.de/handle/fub188/22744 (accessed on 1 March 2019).
99. Aryani, A.; Conrad, M.; Schmidtke, D.; Jacobs, A. Why "piss" Is Ruder than "Pee"? The Role of Sound in Affective Meaning Making. *PLoS ONE* **2018**, *13*, e0198430. [CrossRef] [PubMed]
100. Aryani, A.; Hsu, C.-T.; Jacobs, A.M. The Sound of Words Evokes Affective Brain Responses. *Brain Sci.* **2018**, *8*, 94. [CrossRef] [PubMed]
101. Ullrich, S.; Kotz, S.A.; Schmidtke, D.S.; Aryani, A.; Conrad, M. Phonological Iconicity Electrifies: An ERP Study on Affective Sound-to-Meaning Correspondences in German. *Front. Psychol.* **2016**, *7*, 1200. [CrossRef] [PubMed]
102. Aryani, A.; Jacobs, A.M. Affective Congruence between Sound and Meaning of Words Facilitates Semantic Decision. *Behav. Sci.* **2018**, *8*, 56. [CrossRef] [PubMed]
103. Heffner, H.E.; Heffner, R.S. The Evolution of Mammalian Sound Localization. *Acoust. Today* **2016**, *12*, 20–27.

104. Rendall, D.; Notman, H.; Owren, M.J. Asymmetries in the Individual Distinctiveness and Maternal Recognition of Infant Contact Calls and Distress Screams in Baboons. *J. Acoust. Soc. Am.* **2009**, *125*, 1792–1805. [CrossRef] [PubMed]
105. Monrad-Krohn, G.H. The Third Element of Speech: Prosody and Its Disorders. In *Problems in Dynamic Neurology*; Halpern, L., Ed.; Hebrew University Press: Jerusalem, Israel, 1963; pp. 101–117.
106. Belin, P.; Fillion-Bilodeau, S.; Gosselin, F. The Montreal Affective Voices: A Validated Set of Nonverbal Affect Bursts for Research on Auditory Affective Processing. *Behav. Res. Methods* **2008**, *40*, 531–539. [CrossRef] [PubMed]
107. Scherer, K.R. Expression of Emotion in Voice and Music. *J. Voice* **1995**, *9*, 235–248. [CrossRef]
108. Schröder, M. Experimental Study of Affect Bursts. *Speech Commun.* **2003**, *40*, 99–116. [CrossRef]
109. Paquette, S.; Peretz, I.; Belin, P. The "Musical Emotional Bursts": A Validated Set of Musical Affect Bursts to Investigate Auditory Affective Processing. *Front. Psychol.* **2013**, *4*, 509. [CrossRef] [PubMed]
110. Ekman, P.; Friesen, W.V.; Hager, J.C.; A Human Face (Firm). *Facial Action Coding System*; A Human Face: Salt Lake City, UT, USA, 2002.
111. Hauser, M.D. The Sound and the Fury: Primate Vocalizations as Reflections of Emotion and Thought. In *The Origins of Music*; Wallin, N.L., Brown, S., Merker, B., Eds.; The MIT Press: Cambridge, MA, USA, 2000; pp. 77–102.
112. Bryant, G.A. Animal Signals and Emotion in Music: Coordinating Affect across Groups. *Front. Psychol.* **2013**, *4*, 1–13. [CrossRef] [PubMed]
113. Seifert, U.; Verschure, P.F.M.J.; Arbib, M.A.; Cohen, A.J.; Fogassi, L.; Fritz, T.; Kuperberg, G.; Manzolli, J.; Rickard, N. Semantics of Internal and External Worlds. In *Language, Music, and the Brain*; Arbib, M.A., Ed.; The MIT Press: Cambridge, MA, USA, 2013; pp. 203–230.
114. Newman, J.D. Neural Circuits Underlying Crying and Cry Responding in Mammals. *Behav. Brain Res.* **2007**, *182*, 155–165. [CrossRef] [PubMed]
115. Filippi, P.; Congdon, J.V.; Hoang, J.; Bowling, D.L.; Reber, S.A.; Pašukonis, A.; Hoeschele, M.; Ocklenburg, S.; De Boer, B.; Sturdy, C.B.; et al. Humans Recognize Emotional Arousal in Vocalizations across All Classes of Terrestrial Vertebrates: Evidence for Acoustic Universals. *Proc. R. Soc. B Biol. Sci.* **2017**, *284*, 20170990. [CrossRef] [PubMed]
116. Köhler, W. *Gestalt Psychology; An Introduction to New Concepts in Modern Psychology*, Rev. ed.; Liveright: Oxford, UK, 1947.
117. Fort, M.; Martin, A.; Peperkamp, S. Consonants Are More Important than Vowels in the Bouba-Kiki Effect. *Lang. Speech* **2015**, *58*, 247–266. [CrossRef] [PubMed]
118. Morton, E.S. On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *Am. Nat.* **1977**, *111*, 855–869. [CrossRef]
119. Ohala, J. Signaling with the Eyebrows–Commentary on Huron, Dahl, and Johnson. *Empir. Musicol. Rev.* **2009**, *4*, 101–102. [CrossRef]
120. Sievers, B.; Polansky, L.; Casey, M.; Wheatley, T. Music and Movement Share a Dynamic Structure That Supports Universal Expressions of Emotion. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 70–75. [CrossRef] [PubMed]
121. Trevarthen, C. Musicality and the Intrinsic Motive Pulse: Evidence from Human Psychobiology and Infant Communication. *Music. Sci.* **1999**, *3*, 155–215. [CrossRef]
122. Johnstone, T.; Scherer, K.R. Vocal Communication of Emotion. In *Handbook of Emotions*; Lewis, M., Haviland-Jones, J.M., Eds.; The Guilford Press: New York, NY, USA, 2000.
123. Massaro, D.W. Multimodal Emotion Perception: Analogous to Speech Processes. In Proceedings of the ISCA Workshop on Speech and Emotion, Newcastle, Northern Ireland, UK, 5–7 September 2000; pp. 114–121.
124. Gallese, V.; Fadiga, L.; Fogassi, L.; Rizzolatti, G. Action Recognition in the Premotor Cortex. *Brain* **1996**, *119*, 593–609. [CrossRef] [PubMed]
125. McGurk, H.; MacDonald, J. Hearing Lips and Seeing Voices. *Nature* **1976**, *264*, 746–748. [CrossRef] [PubMed]
126. Merker, B. Synchronous Chorusing and Human Origins. In *The Origins of Music*; Wallin, N.L., Merker, B., Brown, S., Eds.; The MIT Press: Cambridge, MA, USA, 2000; pp. 315–327.
127. Stainsby, T.; Cross, I. The Perception of Pitch. In *The Oxford Handbook of Music Psychology*; Hallam, S., Cross, I., Thaut, M., Eds.; Oxford University Press: Oxford, NY, USA, 2008; Volume 1, pp. 47–58.

128. Peng, S.-C.; Chatterjee, M.; Lu, N. Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants. *Trends Amplif.* **2012**, *16*, 67–82. [CrossRef] [PubMed]
129. Banse, R.; Scherer, K.R. Acoustic Profiles in Vocal Emotion Expression. *J. Personal. Soc. Psychol.* **1996**, 614–636. [CrossRef]
130. Bradley, M.M.; Lang, P.J. Affective Reactions to Acoustic Stimuli. *Psychophysiology* **2000**, *37*, 204–215. [CrossRef] [PubMed]
131. Xu, L.; Thompson, C.S.; Pfingst, B.E. Relative Contributions of Spectral and Temporal Cues for Phoneme Recognition. *J. Acoust. Soc. Am.* **2005**, *117*, 3255–3267. [CrossRef] [PubMed]
132. McAdams, S.; Giordano, B.L. The Perception of Musical Timbre. In *The Oxford Handbook of Music Psychology*; Hallam, S., Cross, I., Thaut, M., Eds.; Oxford University Press: Oxford, NY, USA, 2008; Volume 1, pp. 72–80. [CrossRef]
133. Fu, Q.-J.; Zeng, F.-G.; Shannon, R.V.; Soli, S.D. Importance of Tonal Envelope Cues in Chinese Speech Recognition. *J. Acoust. Soc. Am.* **1998**, *104*, 505–510. [CrossRef] [PubMed]
134. Lehiste, I. The Function of Quantity in Finnish and Estonian. *Language* **1965**, *41*, 447–456. [CrossRef]
135. Suomi, K. Temporal Conspiracies for a Tonal End: Segmental Durations and Accentual F0 Movement in a Quantity Language. *J. Phon.* **2005**, *33*, 291–309. [CrossRef]
136. Järvikivi, J.; Vainio, M.; Aalto, D. Real-Time Correlates of Phonological Quantity Reveal Unity of Tonal and Non-Tonal Languages. *PLoS ONE* **2010**, *5*, e12603. [CrossRef] [PubMed]

brain sciences

MDPI

Article

Poor Synchronization to Musical Beat Generalizes to Speech†

Marie-Élaine Lagrois [1,2,*], Caroline Palmer [1,3] and Isabelle Peretz [1,2]

1 International Laboratory for Brain, Music, and Sound Research, Montreal, QC H3C 3J7, Canada
2 Department of Psychology, University of Montreal, Montreal, QC H3C 3J7, Canada
3 Department of Psychology, McGill University, Montreal, QC H3A 1B1, Canada
* Correspondence: marie-elaine.lagrois@umontreal.ca; Tel.: +1-514-343-5840
† This paper is part of Marie-Élaine Lagrois's PhD thesis—Étude de la modularité de la synchronisation à la pulsation musicale: synchronisation sensorimotrice dans l'amusie congénitale, from the Department of Psychology, University of Montreal, delivered in August 2018.

Received: 17 June 2019; Accepted: 1 July 2019; Published: 4 July 2019

Abstract: The rhythmic nature of speech may recruit entrainment mechanisms in a manner similar to music. In the current study, we tested the hypothesis that individuals who display a severe deficit in synchronizing their taps to a musical beat (called beat-deaf here) would also experience difficulties entraining to speech. The beat-deaf participants and their matched controls were required to align taps with the perceived regularity in the rhythm of naturally spoken, regularly spoken, and sung sentences. The results showed that beat-deaf individuals synchronized their taps less accurately than the control group across conditions. In addition, participants from both groups exhibited more inter-tap variability to natural speech than to regularly spoken and sung sentences. The findings support the idea that acoustic periodicity is a major factor in domain-general entrainment to both music and speech. Therefore, a beat-finding deficit may affect periodic auditory rhythms in general, not just those for music.

Keywords: beat deafness; music; speech; entrainment; sensorimotor synchronization; beat-finding impairment; brain oscillations

1. Introduction

Music is quite unique in the way it compels us to engage in rhythmic behaviors. Most people will spontaneously nod their heads, tap their feet, or clap their hands when listening to music. In early infancy, children already show spontaneous movements to music [1]. This coupling between movements and music is achieved through entrainment. Entrainment can be broadly defined as the tendency of behavioral and brain responses to synchronize with external rhythmic signals [2,3]. Currently, the predominant models of entrainment are based on the dynamic attending theory (DAT) [2,4–6]. According to this theory, alignment between internal neural oscillators and external rhythms enables listeners to anticipate recurring acoustic events in the signal, allowing for maximum attentional energy to occur at the onset of these events, thus facilitating a response to these events [2]. Multiple internal oscillators that are hierarchically organized in terms of their natural frequency or period are likely involved in this process. Interaction of these oscillators would permit the extraction of regularities in complex rhythms that are periodic or quasi-periodic in nature, such as music [7–9]. Of note, entrainment to rhythms, as modeled by oscillators, would apply not only to music but also to speech [10–18].

The periodicities contained in musical rhythms typically induce the perception of a beat, that is, the sensation of a regular pulsation, on which timed behaviors are built [19]. Simple movements in response to beat perception, like taps, are usually produced within a few tens of milliseconds of the

beat onset, indicating the precision of the temporal predictions made about the timing of upcoming beats [20–22]. Listeners can extract the beat from various complex rhythms, without the need for a one-to-one correspondence between acoustic events and beat occurrences [23–27] and across a large range of tempi (~94–174 beats per minute) [20,28–31]. Beat extraction is also robust to moderate tempo fluctuations [8,32,33]. Beat induction from music has in fact been proposed as one of the fundamental and universal traits of music [34,35].

Musical meter, which corresponds to the hierarchical organization of beats, where some beats are perceived as stronger than others, leads to higher-order periodicities of strong and weak beats (for example, a march versus a waltz). Similarly, speech has a hierarchically organized temporal structure, with phonemes, syllables, and prosodic cues, each occurring at different time scales [16,36–38]. As in music, metrical hierarchy in speech may rely on the occurrence of stressed or accented acoustic events, typically associated with syllables [11,17,39–41]. Stress patterns in speech vary and depend on different acoustic cues according to language. The meter of "stress-timed" languages, such as English, is usually clearer than the meter of "syllable-timed" languages like French [14,42]. However, regardless of the language studied, temporal intervals between stressed syllables are not as regular in speech as in music [41,43–46].

Despite this variability in the regularity of stress or beat in spoken language, individuals seem to be able to entrain to speech. Initial evidence in this regard is the finding that the timing of speech can be synchronized with a metronome [11]. Speakers can not only adapt their speech rate to match another speaker [47,48], but they also entrain to each other's syllables rate in conversational turn taking [18,49]. In a prior study using a similar experimental design to the present study [14], French and English monolingual speakers and French–English bilingual speakers were invited to tap their finger along with the beat they perceived in French and English sentences spoken with natural prosody. The variability of intervocalic intervals (IVIs) in these sentences predicted the participants' inter-tap variability, suggesting that the participants were able to entrain to the speech stimuli.

While there is evidence of entrainment to speech, a puzzling difference exists between the absence of synchronous ("choral") speech and the widespread and exquisite synchronization observed in music. To address this issue, Cummins [50,51] proposed that synchronous speech should be possible because (1) speakers of the same language have mastered the association between motor actions and speech sounds of their language, and (2) they share knowledge of speech timing. He supports his claim by showing that speakers can synchronize while reading an unfamiliar text without prior practice, which the author considered an indication of aperiodic synchronization [10,52–54]. According to this perspective, entrainment to speech and music would reflect a fundamental propensity of humans to time their actions with the rhythm of an external event.

Entrainment to speech and music has rarely been compared behaviorally, with few previous studies in this regard. In one of these [55], the influence of music and speech on entrainment was assessed through interference. The main task was to synchronize finger taps to a metronome while hearing highly isochronous computer-generated music or regularly spoken poems. When the metronome tones and the musical beats or stressed syllables were perfectly aligned, higher variability in the asynchronies between taps and metronome was found with the speech distractor compared to the musical one. When misaligned, both music and speech led to synchronization interference by increasing the asynchrony between taps and metronome onsets, and music induced the largest asynchrony. In a second experiment in this study, the stimuli were better matched: songs, either sung with lyrics, sung with a single syllable, or spoken with a regular pace, were presented. In this case, misaligned stimuli had identical detrimental effects on the variability of tapping to the metronome, whether spoken or sung. Therefore, when isochrony is equalized between music and speech, entrainment appears to be very similar.

However, natural speech is typically not isochronous. In a second study comparing music and speech [56], using the same paradigm as the current study, native French and English speakers tapped along with French and English sentences in three conditions: naturally spoken, regularly spoken, and

sung with a simple melody. The inter-tap intervals (ITIs) were more variable in the naturally spoken sentences than in the other conditions. The taps were also more closely aligned to the beat (the nearest implied metronome click to which the singer synchronized her renditions of the stimuli) for sung than for regularly spoken sentences. These results show an overall effect of regularity on entrainment, with music being more suitable to elicit entrainment than regular speech.

Here, we tested the same materials as those used by Lidji and collaborators [56] with individuals who have a documented deficit in tracking the beat in music. This disorder is characterized by an inability to synchronize whole-body movements, clapping, or tapping to the beat of music [57–61], to amplitude-modulated noise derived from music [60], and to metronome-like rhythms [62,63]. This beat-finding deficit occurs in the absence of intellectual disability or acquired brain damage. Study of this "beat-deaf" population provides an opportunity to test the domain specificity of entrainment mechanisms. If the beat-finding disorder initially diagnosed with music also disrupts entrainment to speech, then the association will provide evidence for the domain-general nature of entrainment mechanisms to auditory rhythms.

Beat-deaf individuals and matched control participants who did not exhibit a beat processing disorder were asked to tap to spoken and sung sentences. If entrainment abilities are domain-general, then beat-deaf participants should show deficits to adapt their tapping period to the intervocalic period between syllables to all versions of sentences, compared to the control group. The control group was expected to replicate the findings of [56] showing largest inter-tap interval variability to natural speech, next largest to regularly spoken sentences, and smallest inter-tap variability to sung sentences, and with more accurate synchronization to the intervocalic period between syllables of sung sentences than regularly spoken sentences. Alternatively, if entrainment is domain-specific, beat-deaf participants' tapping should be most impaired for sung sentences and unimpaired (meaning similar to the control group) for speech.

2. Materials and Methods

2.1. Participants

Thirteen beat-deaf French-speaking adults (10 females) and 13 French-speaking matched control participants (11 females) took part in the study. The groups were matched for age, education, and years of music and dance training (detailed in Table 1). One beat-deaf participant was completing an undergraduate degree in contemporary dance at the time of testing. Accordingly, a trained contemporary dancer was also included in the control group. All participants were non-musicians and had no history of neurological, cognitive, hearing, or motor disorders. In addition, all had normal verbal auditory working memory and non-verbal reasoning abilities, as assessed by the Digit Span and Matrix Reasoning subtests of the WAIS-III (Wechsler Adult Intelligence Scale) [64], with no differences between groups on these measures (p-values > 0.34; Table 1). Participants provided written consent to take part in the study and received monetary compensation for their participation. All procedures were approved by the Research Ethics Council for the Faculty of Arts and Sciences at the University of Montreal (CERAS-2014-15-102-D).

Procedure Prior to Inclusion of Participants in the Study

Participants in the beat-deaf group had taken part in previous studies in our lab [63,65] and were identified as being unable to synchronize simple movements to the beat of music. Control participants had either taken part in previous studies in the lab or were recruited via online advertisements directed toward Montreal's general population or through on-campus advertisements at the University of Montreal.

Table 1. Characteristics of the beat-deaf and control groups.

Variables	Beat-deaf (SD) $n = 13$	Control (SD) $n = 13$
Age (years)	37.4 (17.6)	38.7 (17.8)
Education (years)	18.2 (2.2)	17.6 (3.2)
Musical Training (years)	1.0 (2.3)	1.1 (2.1)
Dance Training (years)	1.3 (3.0)	1.8 (3.1)
WAIS-III Digit Span (ss)	10.0 (3.0)	11.0 (3.0)
WAIS-III Matrix Reasoning (ss) [a]	13.0 (3.0)	14.0 (1.0)

ss—standard score. [a], Scores from 12 beat-deaf and 10 control participants. Some participants did not complete the Matrix Reasoning test because they were Ph.D. students in a clinical neuropsychology program and were too familiar with the test.

Inclusion in the current study was based on performance on the Montreal Beat Alignment Test (M-BAT) [66]. In a beat production task, participants were asked to align taps to the beat of 10 song excerpts from various musical genres. Tempo varied across the excerpts from 82 beats per minute (bpm) to 170 bpm. Each song was presented twice, for a total of 20 trials. Control participants successfully matched the period of their taps to the songs' beat in at least 85% of the trials (M = 96.9%, SD = 5.2%); successful period matching was determined through evaluation of *p*-values on the Rayleigh *z* test of periodicity, with values smaller than 0.05 considered successful. In the beat-deaf group, the average percentage of trials with successful tempo matching was 39.2% (range of mean values: 10–65%, SD = 18.3%). As shown in Figure 1, there was no overlap between the groups' performance on this task, confirming that the participants in the beat-deaf group showed a deficit in synchronizing their taps to the beat of music.

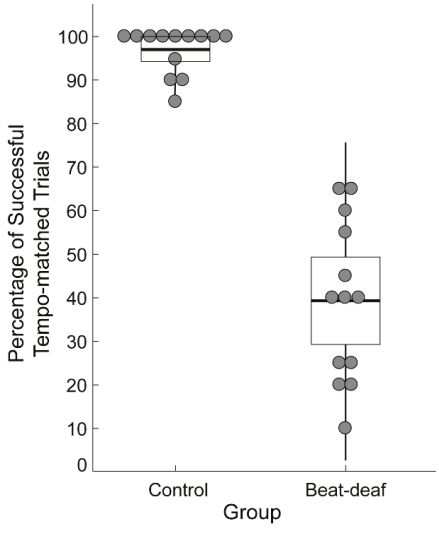


Figure 1. Performance of participants in the control and beat-deaf groups in the beat production task of the Montreal Beat Alignment Test (M-BAT). Each dot represents a participant. Boxes correspond to a 95% confidence interval from the mean based on the standard error of the mean (SEM). The black horizontal line within each box indicates the group mean. The vertical lines represent two standard deviations from the mean.

Prior to their participation in the current study, participants completed the online test of amusia to screen for the presence of a musical pitch perception impairment [67]. The online test is composed of three tests: Scale, Off-beat, and Off-key. The Scale test requires the comparison of 30 pairs of melodies

that differ by an out-of-key note in half of the trials. The Off-beat and Off-key tests consist of the detection of either an out-of-time or an out-of-key note, respectively. A score lying 2-SD below the mean of a large population on both the Scale and Off-key tests indicates the likely presence of pitch deafness (also called congenital amusia) [67,68]. Based on the data from Peretz and Vuvan [67], a cut-off score of 22 out of 30 was used for the Scale test and 16 out of 24 for the Off-key test. Table 2 indicates the individual scores of beat-deaf participants on the online test. Half of the beat-deaf group scored at or below the cut-off on both the Scale and Off-key tests. As these cases of beat-deaf participants could also be considered pitch-deaf, the influence of musical pitch perception will be taken into account in the analysis and interpretation of the results. All control participants had scores above the 2-SD cut-offs.

Table 2. Individual scores of the beat-deaf participants and the group average of their matched controls in the online test of amusia.

	Group															
	Beat-Deaf													**Control (n = 13)**		
Participant	B1	B2	B3	B4	B5	B6	B7 †	B8 †	B9 †	B10 †	B11 †	B12 †	B13 †	M	M	SD
Online test																
Scale (22/30)	23	24	23	23	23	24	21 †	21 †	20 †	22 †	19 †	18 †	22 †	21.8	27.7	2.2
Off-key (16/24)	20	14 †	19	14 †	16 †	14 †	13 †	16 †	15 †	9 †	13 †	14 †	13 †	14.6	19.8	2.2
Off-beat (17/24)	23	21	19	17 †	20	16 †	15 †	17 †	18	17 †	18	18	19	18.3	19.8	1.4

Scores in parentheses represent the cut-off scores taken from Peretz and Vuvan [67]. Participants with co-occurring pitch deafness are marked with †.

2.2. Stimulus Materials

The 12 French sentences used in this experiment were taken from Lidji et al. [56]. Each sentence contained 13 monosyllabic words and was recorded in three conditions as depicted in Figure 2. The recordings were made by a native Québec French/English female speaker in her twenties who had singing training. Recordings were made with a Neumann TLM 103 microphone in a sound-attenuated studio. In the naturally spoken condition, the speaker was asked to speak with a natural prosody (generating a non-periodic pattern of stressed syllables). In the regularly spoken condition, sentences were recorded by the speaker to align every other syllable with the beat of a metronome set to 120 bpm, heard over headphones. In the sung condition, the sentences were sung by the same speaker, again with every other syllable aligned to a metronome at 120 bpm, heard over headphones. Each sung sentence was set to a simple melody, with each syllable aligned with one note of the melody. Twelve unique melodies composed in the Western tonal style in binary meter, in major or minor modes, were taken from Lidji et al. [56]. These melodies were novel to all participants. Although each sentence was paired with two different melodies, participants only heard one melody version of each sung sentence, counterbalanced across participants.

Additional trials for all three conditions (naturally spoken, regularly spoken, sung) were then created from the same utterances at a slower rate (80% of original stimulus rate, i.e., around a tempo of 96 bpm) using the digital audio production software Reaper (v4.611, 2014; time stretch mode 2.28 SOLOIST: speech, Cockos Inc., New York, United States). This ensured that the beat-deaf participants adapted their taps to the rate of each stimulus and could comply with the task requirements. All the stimuli were edited to have a 400 ms silent period before the beginning of the sentence and a 1000 ms silent period at the end of the sentence. Stimuli amplitudes were also equalized in root mean square (RMS) intensity. Preliminary analyses indicated that all participants from both groups adapted the rate of their taps from the original stimulus rate to the slower stimulus rate, with a Group × Material (naturally spoken, regularly spoken, sung) × Tempo (original, slow) ANOVA on mean inter-tap interval (ITI) showing a main effect of Tempo, $F(1,24) = 383.6$, $p < 0.001$, $\eta^2 = 0.94$, with no significant Group × Tempo interaction, $F(1,24) = 0.0004$, $p = 0.98$ or Group × Material × Tempo interaction, $F(2,48) = 1.91$, $p = 0.17$ (mean ITI results are detailed in Table 3). Therefore, the data obtained for the slower stimuli are not reported here for simplicity.

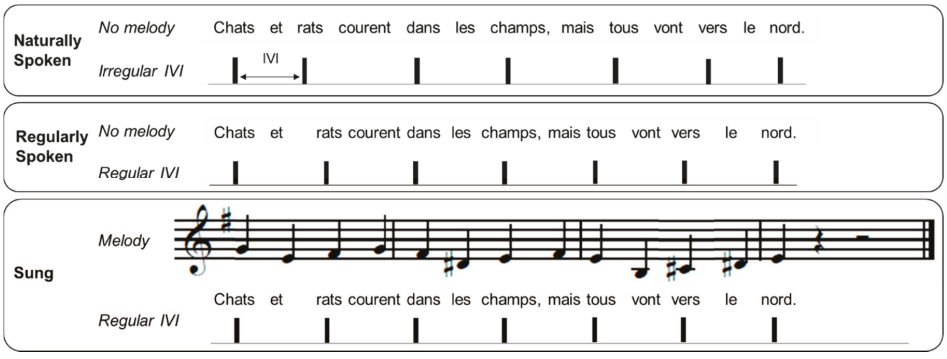


Figure 2. Example of a sentence in the naturally spoken, regularly spoken, and sung conditions. IVI refers to the intervocalic interval between stressed syllables.

Table 3. Mean inter-tap interval (ITI) in ms for each group according to material type and tempo.

	Original Tempo		Slowed Tempo	
Condition	**Control**	**Beat-Deaf**	**Control**	**Beat-Deaf**
Naturally Spoken Sentences (SE)	480.00 (7.70)	489.80 (12.80)	591.77 (8.90)	585.58 (11.70)
Regularly Spoken Sentences (SE)	497.43 (6.52)	522.15 (12.83)	615.20 (4.97)	632.00 (17.69)
Sung Sentences (SE)	488.70 (4.28)	517.08 (13.39)	611.24 (4.12)	664.25 (17.30)

For the comparison of mean ITI between groups and conditions, the mean ITIs were scaled to the ITI corresponding to tapping once every two words (or stressed syllables).

Table 4 describes the features of the rhythmic structure of the stimuli in each condition. Phoneme boundaries were marked by hand using Praat [69], and were classified as vowels or consonants based on criteria defined by Ramus, Nespor, and Mehler [70]. Note that the analyses reported below include the stimuli at the original tempo only. Once the segmentation was completed, a MATLAB script was used to export the onset, offset, and duration of vocalic (a vowel or a cluster of vowels) and consonantal (a consonant or a cluster of consonants) intervals. The Normalized Pairwise Variability Index for Vocalic Intervals (V-nPVI), an indication of duration variability between successive vowels [71], was used to measure the rhythmic characteristics of the stimuli. A higher V-nPVI indicates greater differences in duration between consecutive vocalic intervals. Comparison of sentences in the naturally spoken, regularly spoken, and sung conditions showed a significant difference between conditions, $F(2,22) = 21.6$, $p < 0.001$, $\eta^2 = 0.66$. The V-nPVI was higher in the naturally and regularly spoken conditions than in the sung condition (Table 4). The coefficient of variation (CV, calculated as SD/mean) of IVIs (vowel onset to onset) is another indication of rhythmic variability [14]. A small CV for IVIs indicates similar time intervals between vowel onsets across the sentence. Here the CV was measured between every other syllable's vowel onset, corresponding to stressed syllables (see IVI in Figure 2). Once again, a significant difference between conditions was observed, $F(2,22) = 64.6$, $p < 0.001$, $\eta^2 = 0.85$. Naturally spoken sentences had the largest timing variations between vowel onsets (M = 0.21), followed by regularly spoken sentences (M = 0.08), while sung sentences showed the smallest variability (M = 0.05). To ensure that the female performer was comparably accurate in timing the sentences with the metronome in the regularly spoken and sung conditions, the relative asynchrony between each vowel onset and the closest metronome pulsation was measured. In this context, a negative mean asynchrony indicates that the vowel onset preceded the metronome tone onset, while a positive asynchrony means that the vowel onset followed the metronome tone (Table 4). There was no significant difference between conditions, indicating similar timing with the metronome in the regularly spoken and sung conditions, $t(11) = 1.146$, $p = 0.28$.

Table 4. Stimuli characteristics related to rhythm.

Variable	Naturally Spoken Sentences (SE)		Regularly Spoken Sentences (SE)		Sung Sentences (SE)	
Mean IVI (ms)	458.00	(10.00) *	503.00	(3.00)	501.00	(1.000)
V-nPVI	49.40	(2.50) *	42.30	(2.00) *	31.10	(1.800)
CV(IVI)	0.21	(0.02) *	0.08	(0.01) *	0.05	(0.004) *
Beat Asynchrony from Vowel Onset (ms)	-		14.00	(11.00)	−2.00	(5.000)

Values indicate means; standard errors appear in parentheses. IVI—intervocalic interval (in ms); V-nPVI—normalized Pairwise Variability Index for Vocalic Intervals; CV—coefficient of variation (SD IVI/Mean IVI between stressed syllables); Beat asynchrony corresponds to the average of signed values from subtracting metronome tone onset from the closest spoken/sung vowel onset, in milliseconds. *, indicate significant differences.

2.3. Design and Procedure

Participants performed three tasks. First, they performed a spontaneous tapping task to assess their spontaneous tapping rate (mean and variance) in the absence of a pacing stimulus. They were asked to tap as regularly as possible for 30 seconds, as if they were a metronome or the "tick-tock" of a clock (as in [30]). Participants were asked to tap with the index finger of their dominant hand. Next, participants performed the tapping task with the spoken/sung sentences, as described below. Then the participants repeated the spontaneous tapping task to determine whether their spontaneous rate had changed, and finally, they tapped at a fixed rate with a metronome set to 120 bpm (inter-beat interval of 500 ms) and 96 bpm (inter-beat interval of 625 ms), chosen to match the tempi of the spoken/sung stimuli used in the experiment. The experiment had a total duration of approximately 60 minutes.

In the spoken/sung tapping blocks, each participant was presented with 12 each of naturally spoken sentences, regularly spoken sentences, and sung sentences at the original rate (120 bpm), and six sentences in each condition at the slower rate (96 bpm). These stimuli were mixed and divided into three blocks of 18 trials each. Two pseudo-random orders were created such that not more than two sentences from the same condition occurred consecutively and that the same sentence was never repeated. On each trial, participants first listened to the stimulus; then, for two additional presentations of the same stimulus, they were asked to tap along to the beat that they perceived in the stimulus (as in [56]). The action to perform (listen or tap) was prompted by instructions displayed on a computer screen. Participants pressed a key to start the next trial. Prior to commencing the task, a demonstration video was presented to participants, which showed an individual finger tapping on the sensor with one example stimulus from each condition. In the demonstration, a different sentence was used for each condition, and each was presented at a different rate (84 bpm or 108 bpm) than the ones used in the experiment. The sung sentence example was also presented with a different melody than any heard by participants in the task. After the demonstration, participants completed a practice trial for each type of sentence.

For the metronome task, there were two trials at each metronome tempo (120 bpm and 96 bpm), and the presentation order of the two metronome tempi was counterbalanced across participants. Each metronome stimulus contained sixty 50 ms 440 Hz sine tones. Each metronome trial began with seven tones at the specific tempo, during which participants were instructed to listen and prepare to tap with the metronome. A practice trial was also first performed with a metronome set to 108 bpm. As mentioned previously, since all participants could adapt their tapping rate to the stimuli at both 120 bpm and 96 bpm, only the results of tapping to the metronome at 120 bpm (rate of the original speech stimuli) are reported here.

The experiment took place in a large sound-attenuated studio. The tasks were programed with MAX/MSP (https://cycling74.com). Taps were recorded on a square force-sensitive resistor (3.81 cm, Interlink FSR 406) connected to an Arduino UNO (R3; arduino.cc) running the Tap Arduino script (fsr_silence_cont.ino; [72,73]) and transmitting timing information to a PC (HP ProDesk 600 G1, Windows 7) via the serial USB port. The stimuli were delivered at a comfortable volume through closed

headphones (DT 770 PRO, Beyerdynamic, Heilbronn, Germany) controlled by an audio interface (RME Fireface 800). No auditory feedback was provided for participants' tapping.

2.4. Data Analyses

2.4.1. Tapping Data Preprocessing

In the spontaneous tapping task, the first five taps produced were discarded and the following 30 ITIs were used, in line with McAuley et al.'s procedure [30]. If participants produced fewer than 30 taps, the data included all taps produced (the smallest number of taps produced was 16 in this task). Due to recording problems, taps were missing from one beat-deaf participant's first spontaneous tapping trial.

Recorded taps were first pre-processed to remove ITIs smaller than 100 ms in the spontaneous tapping task, and ITIs smaller than 150 ms in the spoken/sung tapping task and the metronome task. In the three tasks, taps were also considered outliers and were removed if they were more than 50% smaller or larger than the median ITI produced by each participant (median ITI ± (median ITI × 0.5)). Pre-processing of tapping data was based on the procedure described by [74]. Accordingly, the 100 ms criterion was used at first for the spoken/sung task but the number of outliers mean ITIs was high in both groups of participants. A 150 ms criterion was chosen instead considering that it remained smaller than two standard deviations from the average time interval between consecutive syllables across stimuli (M = 245 ms, SD = 47 ms, M − 2SD = 152 ms), thus allowing the removal of more artefact taps while still limiting the risk of removing intended taps. As a result, 1.6% of the taps were removed (range: 0.0–6.4%) in the spontaneous tapping task. In the spoken/sung tapping task, 0.85% of taps per trial were removed (range: 0–36.4% taps/trial). In the metronome task, 5.27% of taps were removed on average (range: 3.4–8.1%), leaving between 54 and 76 taps per trial, of which the first 50 taps produced by each participant were used for analysis.

2.4.2. Analysis of Tapping Data

The mean ITI was calculated for all tapping tasks. In the spoken/sung tapping task, since each participant tapped twice on each utterance in succession, the mean ITIs per stimulus were averaged across the two presentations. However, in 0.16% of the trials, participants did not tap at the same hierarchical level in the two presentations of the stimulus. For example, they tapped on every syllable in the first presentation, and every other syllable in the second presentation. These trials were not included in the calculations of CV, to avoid averaging together taps with differing mean ITIs. Nevertheless, at least 11 of the 12 trials at 120 bpm for each participant in each condition were included in the analyses. In the metronome task, data were also averaged across the two trials with the metronome at 120 bpm.

In the spoken/sung tapping task, inter-tap variability (CV *SD* ITI/mean ITI) was computed for each condition. As Table 4 indicates, the CVs of taps to naturally spoken sentences should be larger than the CVs to regular stimuli. To assess this, we examined how produced ITIs matched the stimulus IVIs (as done by [75–77]). ITI deviation was calculated by averaging the absolute difference between each ITI and the corresponding IVI of the stimulus. To control for differences in IVI for each stimulus, the ITI deviation was normalized to the mean IVI of that stimulus and converted to a percentage of deviation (% ITI deviation) with formula (1) below, where x is the current interval and n the number of ITI produced:

$$\% \text{ ITI deviation} = (\Sigma|\text{ITIx} - \text{IVIx}|/n)/\text{mean IVI} \times 100 \quad (1)$$

This measure of period deviation gives an indication of how participants' taps matched the rhythmic structure of the stimuli, whether regular or not.

Period-matching between spoken/sung sentences and taps was further assessed for the stimuli that contained regular beat periods (i.e., regularly spoken, sung, and metronome stimuli) with circular statistics using the Circular Statistics Toolbox for MATLAB [78]. With this technique, taps are transposed as angles on a circle from 0° to 360°, where a full circle corresponds to the period of the IVI of the

stimulus. The position of each tap on the circle is used to compute a mean resultant vector. The length of the mean resultant vector (vector length, VL) indicates how clustered the data points are around the circle. Values of VL range from 0 to 1; the larger the value, the more the points on the circle are clustered together, indicating that the time interval between taps matches the IVI of the stimulus more consistently. For statistical analyses, since the data were skewed in the control group for the spoken/sung task (skewness: −0.635, SE: 0.144) and in the metronome tapping task for participants of both groups (skewness: −1.728, SE: 0.427), we used a logit transform of VL (logVL = $-1 \times \log(1 - \text{VL})$), as is typically done with synchronization data (e.g., [57,58,60,61,74]). However, for simplicity, untransformed VL is reported when considering group means and individual data. The Rayleigh z test of periodicity was employed to assess whether a participant's taps period-matched the IVI of each stimulus consistently [79]. A significant Rayleigh z test (p-value < 0.05) demonstrates successful period matching. An advantage of the Rayleigh test is that it considers the number of taps available in determining if there is a significant direction in the data or not [78]. Using linear statistics, the accuracy of synchronization was further measured using the mean relative asynchrony between taps and beats' onset time in milliseconds. Note that this measure only included trials for which participants could successfully match the inter-beat interval of the stimuli, as assessed by the Rayleigh test, since the asynchrony would otherwise be meaningless.

The period used to perform the Rayleigh test was adjusted to fit the hierarchical level at which participants tapped on each trial. Since the stimuli had a tempo of 120 bpm (where one beat = two syllables), this meant that if a participant tapped to every word, the period used was 250 ms, if a participant tapped every two words, then 500 ms, and every four words, 1000 ms. This approach was chosen, as suggested by recent studies using circular statistics to assess synchronization to stimuli with multiple metric level (or subdivisions of the beat period), in order to avoid bimodal distributions or to underestimate tapping consistency [61,80,81]. Given this adaptation, in the spoken/sung tapping task, we first looked at the closest hierarchical level at which participants tapped. This was approximated based on the tapping level that fitted best the majority of ITIs within a trial (i.e., the modal tapping level).

2.4.3. Correlation between Pitch Perception and Tapping to Spoken/Sung Sentences

In order to assess the contribution of musical pitch perception to synchronization with the spoken and sung sentences, the scores from the online test of amusia were correlated with measures of tapping variability (CV) and period-matching (% ITI deviation) from the spoken/sung tapping task.

2.5. Statistical Analyses

Statistical analyses were performed in SPSS (IBM SPSS Statistics, Armonk, United States, version 24, 2016). A mixed repeated-measures ANOVA with Group as the between-subjects factor was used whenever the two groups were compared on a dependent variable with more than one condition. Because of the small group sample size, a statistical approach based on sensitivity analysis was applied, ensuring that significant effects were reliable when assumptions regarding residuals' normality distribution and homogeneity of variance were violated [82]. When these assumptions were violated, the approach employed was as follows: (1) inspect residuals to identify outliers (identified using Q—Q plot and box plot), (2) re-run the mixed-design ANOVA without the outliers and assess the consistency of the previous significant results, and (3) confirm the results with a non-parametric test of the significant comparisons [82]. If the effect was robust to this procedure, the original ANOVA was reported. Bonferroni correction was used for post-hoc comparisons. Other group comparisons were performed with Welch's test, which corrects for unequal variance. Paired t-tests were utilized for within-group comparisons on a repeated measure with only two conditions. Effect sizes are reported for all comparisons with p-values smaller than 0.50 [83]. To indicate the estimated effect sizes, partial eta-squared values are reported for repeated-measures ANOVA, and Hedge's g was computed for the other comparisons.

3. Results

3.1. Spontaneous Tapping

The mean ITI of the spontaneous tapping task ranged from 365 to 1109 ms in control participants and from 348 to 1443 ms in the beat-deaf group (Table 5). There was no significant group difference in the mean ITIs, $F(1,23) = 1.2$, $p = 0.27$, $\eta^2 = 0.05$, and no significant effect of Time, $F(1,23) = 0.48$, $p = 0.49$, $\eta^2 = 0.02$, and no interaction, $F(1,23) = 0.19$, $p = 0.66$, indicating that spontaneous tapping was performed similarly before and after the spoken/sung tapping task. In contrast, a main effect of Group emerged in the CV for spontaneous tapping, $F(1,23) = 18.2$, $p < 0.001$, $\eta^2 = 0.44$, with no effect of Time, $F(1,23) = 0.30$, $p = 0.59$, and no interaction, $F(1,23) = 0.19$, $p = 0.67$. The CV for spontaneous tapping was higher in the beat-deaf group than in the control group (Table 5). As observed by Tranchant and Peretz [63], the beat-deaf individuals showed more inter-tap variability than control participants when trying to tap regularly without a pacing stimulus.

Table 5. Mean inter-tap interval (ITI) and coefficient of variation (CV) of spontaneous tapping.

Group	Spontaneous Tapping-Pre		Spontaneous Tapping-Post	
	Mean ITI [a]	CV	Mean ITI [a]	CV
Control (SE)	603 (56)	0.06 (0.003)	565 (57)	0.06 (0.004)
Beat-deaf (SE)	680[b] (47)	0.08[b] (0.01)	659 (80)	0.08 (0.01)

Numerical values represent group means, with the standard error of the mean in parentheses. [a], Values are in milliseconds. [b], $n = 12$; otherwise, $n = 13$.

3.2. Tapping to Speech and Song

As expected, participants' inter-tap variability (CV) for the naturally spoken sentences was higher than the CV in the other two conditions. Figure 3a depicts the mean CV of the stimulus IVIs and Figure 3b depicts the mean CV for tapping in each condition. The CV for participants' taps was larger for the naturally spoken sentences (M = 0.13) than for the regularly spoken (M = 0.10) and sung (M = 0.10) sentences, $F(1.5,35.4) = 15.2$, $p < 0.001$, $\eta^2 = 0.39$. The groups did not differ significantly, $F(1,24) = 2.8$, $p = 0.10$, $\eta^2 = 0.11$, and there was no interaction with material type, $F(1.5,35.4) = 0.58$, $p = 0.52$. One control participant had a larger tapping CV than the rest of the group for natural speech. Three beat-deaf participants also had larger CVs across conditions. However, removing the outliers did not change the results of the analysis. Thus, the inter-tap variability only discriminated natural speech from the regularly paced stimuli for both groups.

Deviation in period matching between ITIs and IVIs of stimuli indicated that control participants exhibited better performance than beat-deaf participants, whether the stimuli were regular or not. Control participants showed a smaller percentage of deviation between the inter-tap period produced and the corresponding stimulus IVI across stimulus conditions (% ITI deviation; Figure 4), with a main effect of Group, $F(1,24) = 8.2$, $p = 0.008$, $\eta^2 = 0.26$, a main effect of Material, $F(1.4,32.5) = 95.9$, $p < 0.001$, $\eta^2 = 0.80$, and no interaction, $F(1.4,32.5) = 0.19$, $p = 0.74$. Post-hoc comparisons showed a significant difference between all conditions: the % ITI deviation was the largest for naturally spoken sentences (20.9% and 27.2% for the control and beat-deaf group, respectively), followed by regular speech (11% and 18.2%) and sung sentences (8.5%, and 15.7%; see Figure 4). These results held even when outliers were removed.

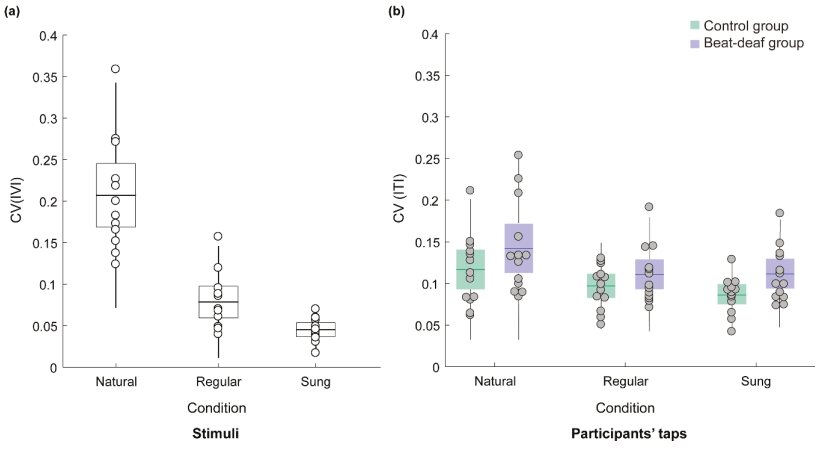


Figure 3. (**a**) Coefficient of variation (CV) of the intervocalic interval (IVI) between stressed syllables of the stimuli. Each dot represents a sentence. (**b**) Mean CV of the inter-tap interval (ITI) produced by the beat-deaf and control group as a function of sentence type. Each dot represents a participant. Boxes correspond to a 95% confidence interval from the mean based on the standard error of the mean (SEM). The darker horizontal line within each box indicates the group mean, while the vertical lines represent two standard deviations from the mean.

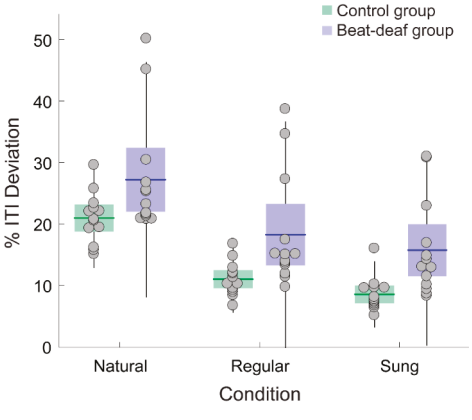


Figure 4. Mean percentage of deviation between the inter-tap intervals (ITIs) produced by each participant and the IVIs of the sentences. Each dot represents a participant. Boxes corresponds to a 95% confidence interval from the mean based on standard error mean (SEM). The black horizontal line within each box indicates the group mean. The vertical lines represent two standard deviations from the mean.

In order to measure synchronization more precisely, we first examined the hierarchical level at which participants tapped. A chi-squared analysis of the number of participants who tapped at each hierarchical level (1, 2, or 4 words) by Condition and Group indicated a main effect of Group, $\chi^2(2,78) = 7.4$, $p = 0.024$. In both groups, participants tapped preferentially every two words (see Figure 5), although control participants were more systematic in this choice than beat-deaf participants. Both groups were consistent in the hierarchical level chosen for tapping across conditions. The hierarchical level at which a participant tapped determined the period used in the following analysis of synchronization to the regular stimuli.

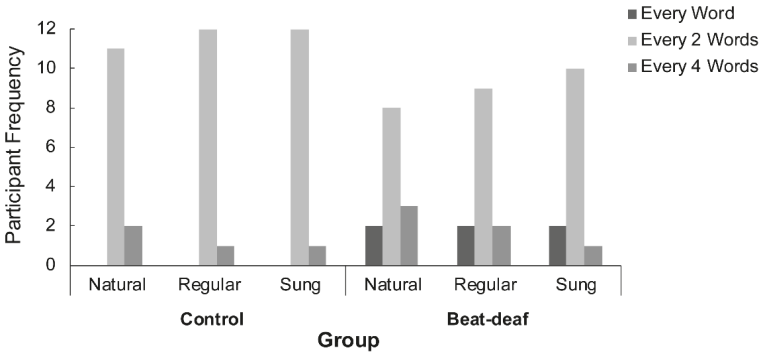


Figure 5. Number of participants in each group who tapped at every word, every two words, or every four words, according to each sentence condition (natural, regular, or sung).

The average percentage of trials with successful period matching (using Rayleigh's z test) for the control group was 91.7% (range: 58–100%) for regularly spoken sentences and 90.4% (range: 50–100%) for sung ones. In the beat-deaf group, the mean percentage of successful period-matched trials was much lower, with 30.4% (range: 0–75%) and 23.8% (range: 0–66.7%) for regularly spoken and sung sentences, respectively. The percentage of trials with successful period matching did not differ between the regular and sung conditions, $t(25) = 1.297$, $p = 0.21$, $g = 0.10$.

We next examined if synchronization was more consistent and accurate for sung than for regularly spoken sentences. These analyses were conducted on trials for which participants were able to synchronize successfully with the beat (i.e., Rayleigh p-value < 0.05). Because most beat-deaf participants failed to synchronize with the stimuli, the analyses are limited to the control group. The analyses of the log transform of the mean vector length (logVL) revealed that the control group's tapping was as constant with regularly spoken sentences ($M = 1.79$, range: 1.18 to 3.09) as with sung ones ($M = 1.85$, range: 1.10 to 3.45), $t(12) = -0.755$, $p = 0.46$, $g = 0.09$. Accuracy of synchronization was assessed with the mean relative asynchrony between taps and beats in milliseconds. Control participants anticipated the beat onsets of sung sentences significantly earlier (M = −14 ms, range: −51 to 19 ms) than the beat onsets of regularly spoken sentences (M = 1 ms, range: −30 to 34 ms), $t(12) = 3.802$, $p = 0.003$, $g = 0.74$. This result suggests that beat onsets were better anticipated in sung sentences than in regularly spoken ones, corroborating results found by Lidji and collaborators [56]. Of note, the two beat-deaf participants (B2 and B4) who could successfully period-match the stimuli on more than 50% percent of trials showed similar consistency (logVL range: 1.08 to 1.53) and accuracy (mean asynchrony range: −3 ms to 22 ms) of synchronization to control participants.

3.3. Tapping to Metronome

All participants could successfully match their taps to the period of the metronome, as assessed by the Rayleigh z test, except for one beat-deaf participant (B10) who tapped too fast compared to the 120 bpm tempo (mean ITI = 409 ms for a metronome inter-onset interval of 500 ms). Thus, this participant and a matched control were removed from subsequent analyses in this task. As in previous analyses, control participants had smaller inter-tap variability than beat-deaf participants. This was confirmed by a group comparison with Welch's test on the CV, $t(14.0) = 11.698$, $p = 0.004$, $g = 1.35$ (control: M = 0.06, SE = 0.003; beat-deaf: M = 0.09, SE = 0.01). Period-matching consistency, using the logVL, also showed a significant group difference, $t(22.0) = 9.314$, $p = 0.006$, $g = 1.20$. The difference between groups was not significant, however, for the mean relative asynchrony between taps and metronome tones, $t(20.4) = 0.066$, $p = 0.80$ (control: M = −56 ms, range: −120 ms to 0 ms; beat-deaf: M = −53 ms, range: −104 ms to −11 ms).

3.4. Contribution of Musical Pitch Perception to Entrainment to Utterances

To assess the impact of musical pitch perception on tapping performance, we correlated the scores from the online test of amusia with tapping variability (CV) and period matching (%ITI deviation) for all conditions and participant groups (Table 6). The correlations between CV and musical pitch-related tests did not reach significance, while the % of ITI deviation did for two of the three stimulus conditions when considering participants from both groups. The significant correlation between the Scale test and the % ITI deviation was driven mostly by the beat-deaf group ($r_{(8)} = -0.61$) rather than the control group ($r_{(11)} = -0.05$). There was also a significant correlation between the Off-key test and % ITI deviation. None of the correlations reached significance with the Off-beat test. However, tapping variability (CV) to sentences and to music (M-BAT) were highly correlated in control but not beat-deaf participants.

Table 6. Spearman correlations between tapping and music perception.

Variable	Scale Test			Off-Key Test			Off-Beat Test			CV M-BAT Production Test		
CV—natural sentences	−0.20 [a]	−0.10 [c]	−0.15 [d]	−0.22 [a]	−0.06 [c]	−0.53 [d]	−0.12 [a]	0.12 [c]	−0.25 [d]	0.47 [a]	0.78 [c]	0.14 [d]
CV—regular sentences	−0.24 [a]	−0.42 [c]	−0.27 [d]	−0.29 [a]	−0.41 [c]	−0.48 [d]	−0.03 [a]	0.03 [c]	0.08 [d]	0.32 [a]	0.79 [c]	−0.19 [d]
CV—sung sentences	−0.26 [a]	−0.15 [c]	−0.11 [d]	−0.35 [a]	−0.16 [c]	−0.37 [d]	−0.24 [a]	−0.05 [c]	−0.24 [d]	0.31 [a]	0.43 [c]	−0.11 [d]
%ITI deviation—natural	−0.30 [b]	−0.01 [c]	0.08 [e]	−0.34 [b]	−0.39 [c]	−0.01 [e]	0.11 [b]	0.30 [c]	0.17 [e]	-	-	-
%ITI deviation—regular	−0.32 [b]	0.30 [c]	−0.18 [e]	−0.55 [b]	−0.35 [c]	−0.12 [e]	−0.19 [b]	0.05 [c]	0.17 [e]	-	-	-
%ITI deviation—sung	−0.54 [b]	−0.05 [c]	−0.61 [e]	−0.43 [b]	−0.21 [c]	0.43 [e]	−0.21 [b]	0.26 [c]	0.12 [e]	-	-	-

CV—coefficient of variation; ITI—inter-tap interval. Outliers from the beat-deaf group were removed, with [a] n = 24, [b] n = 23, [c] n = 13, [d] n = 11, [e] n = 10. Columns in white indicate correlations with participants of both groups, light blue with control participants only, and darker blue beat-deaf participants only. Significant correlations after correcting for multiple comparisons are marked in orange ($p \leq 0.015$).

These results raise the possibility that beat-deaf individuals with an additional deficit in pitch perception have a more severe impairment in finding the beat. If we compare beat-deaf participants with and without a co-occurring musical pitch deficit, the difference between groups does not reach significance on period-matching consistency of tapping (mean logVL) in the M-BAT beat production test, $t(6.2) = 1.874$, $p = 0.11$, $g = 1.0$. Thus, musical pitch perception seems to have little impact on synchronization to both musical (see [65]) and verbal stimuli.

4. Discussion

This study investigated the specialization of beat-based entrainment to music and to speech. We show that a deficit with beat finding initially uncovered with music can similarly affect entrainment to speech. The beat-deaf group, in the current study identified on the basis of abnormal tapping to various pre-existing songs, also show more variable tapping to sentences, whether naturally spoken or spoken to a (silent) metronome, as compared to matched control participants. These results could argue for the domain generality of beat-based entrainment mechanisms to both music and speech. However, even tapping to a metronome or tapping at their own pace is more irregular in beat-deaf individuals than in typical non-musicians. Thus, the results point to the presence of a basic deficiency in timekeeping mechanisms that are relevant to entrainment to both music and speech and might not be specific to either domain.

Such a general deficiency in timekeeping mechanisms does not appear related to an anomalous speed of tapping. The spontaneous tapping tempo of the beat-deaf participants is not different from the tempo of neurotypical controls. What differs is the regularity of their tapping. This anomalous variability in spontaneous tapping has not been reported previously [57–60,62]. Two beat-deaf cases previously reported from the same lab [62], not included in the present sample, had higher inter-tap variability in unpaced tapping; however, the difference was not statistically significant in comparison to a control group. However, our study includes one of the largest samples of individuals with a beat-finding disorder so far (compared to 10 poor synchronizers in [60] for example), which might explain some discrepancies with previous studies. Only recently (in our lab, [63]) has an anomalously high variability in spontaneous regular tapping in beat-deaf individuals been observed irrespective of the tapping tempo.

A similar lack of precision was noted among beat-deaf participants compared to matched controls when tapping to a metronome, which is in line with Palmer et al. [62]. These similar findings suggest that temporal coordination (both in the presence of auditory feedback from a metronome and in its absence during spontaneous tapping) is impaired in beat-deaf individuals. These individuals also display more difficulty with adapting their tapping to temporally changing signals, such as phase and period perturbations in a metronome sequence [62]. Sowiński and Dalla Bella [60] also reported that poor beat synchronizers had more difficulty with correcting their synchronization errors when tapping to a metronome beat, as reflected in lag -1 analyses. Therefore, a deficient error correction mechanism in beat-deaf individuals may explain the generalized deficit for tapping with and without an external rhythm. This error correction mechanism may in turn result from a lack of precision in internal timekeeping mechanism, sometimes called "intrinsic rhythmicity" [63].

However, the deficit in intrinsic rhythmicity in beat-deaf individuals is subtle. The beat-deaf participants appear sensitive to the acoustic regularity of both music and speech, albeit not as precisely as the control participants. All participants tapped more consistently to regularly spoken and sung sentences than to naturally spoken ones. All showed reduced tapping variability for the regular stimuli, with little difference between regularly spoken and sung sentences, while normal control participants also showed greater anticipation of beat onsets in the sung condition. The latter result suggests that entrainment was easier for music than speech, even when speech is artificially made regular. However, the results may simply reflect acoustic regularity, which was higher in the sung versions than in the spoken versions: the sung sentences had lower intervocalic variability and V-nPVI than regular speech, which may facilitate the prediction of beat occurrences, and, therefore, entrainment. These results corroborate previous studies proposing that acoustic regularity is the main factor supporting entrainment across domains [56].

Another factor that may account for better anticipation of beats in the sung condition is the presence of pitch variations. There is evidence that pitch can influence meter perception and entrainment in music [84–93]. The possible contribution of musical pitch in tapping to sung sentences is supported by the correlations between perception of musical pitch and period-matching performance (as measured by ITI deviation) in tapping to the sung sentences. However, the correlation was similar for the regularly spoken sentences where pitch contributes little to the acoustic structure. Moreover, the beat-deaf participants who also had a musical-pitch deficit, corresponding to about half the group, did not perform significantly poorer than those who displayed normal musical pitch processing. Altogether, the results suggest that pitch-related aspects of musical structure are not significant factors in entrainment [55,56,94,95].

Thus, a key question remains: What is the faulty mechanism that best explains the deficit exhibited by beat-deaf individuals? One useful model to conceptualize the imprecision in regular tapping that seems to characterize beat deafness, while maintaining sensitivity to external rhythm, is to posit broader tuning of self-sustained neural oscillations in the beat-impaired brain. An idea that is currently gaining increasing strength is that auditory-motor synchronization capitalizes on the tempi of the naturally occurring oscillatory brain dynamics, such that moments of heightened excitability (corresponding to particular oscillatory phases) become aligned to the timing of relevant external events (for a recent review, see [96]). In the beat-impaired brain, the alignment of the internal neural oscillations to the external auditory beats would take place, as shown by their sensitivity to acoustic regularities, but it would not be sufficiently well calibrated to allow precise entrainment.

This account of beat deafness accords well with what is known about oscillatory brain responses to speech and music rhythms [12,97–104]. These oscillatory responses match the period of relevant linguistic units, such as phoneme onsets, syllable onsets, and prosodic cues, in the beta/gamma, theta, and delta rhythms, respectively [16,38,105,106]. Oscillatory responses can also entrain to musical beat, and this oscillatory response may be modulated by the perceived beat structure [81,107–111]. Oscillatory responses may even occur in the absence of an acoustic event on every beat, and not just in response to the frequencies present in the signal envelope, indicating the contribution of oscillatory

responses to beat perception [23,25,110]. Ding and Simon [98] propose that a common entrainment mechanism for speech and music could occur in the delta band (1–4 Hz). If so, we predict that oscillations in the delta band would not be as sharply aligned with the acoustic regularities present in both music and speech in the beat-deaf brain as in a normal brain. This prediction is currently under study in our laboratory.

One major implication of the present study is that the rhythmic disorder identified with music extends to speech. This is the first time that such an association across domains is reported. In contrast, there are frequent reports of reverse associations between speech disorders and impaired musical rhythm [112–116]. Speech-related skills, such as phonological awareness and reading, are associated to variability in synchronization with a metronome beat [117–119]. Stutterers are also less consistent than control participants in synchronizing taps to a musical beat [120,121]. However, in none of these prior studies [112,116,118] was a deficit noted in spontaneous tapping, hence in intrinsic rhythmicity. Thus, it remains to be seen if their deficit with speech rhythm is related to a poor calibration of intrinsic rhythmicity as indicated here.

The design used in this study, which presented stimuli in the native language of the participants, creates some limitations in generalization of these findings across languages. It is possible, for example, that stress- and syllable-timed languages might elicit different patterns of entrainment [14]. French is usually considered a less "rhythmic" language than English [70,71]. One's native language has also been shown to influence perception of speech rhythm [14,56,122,123]. For example, Lidji et al. [14] found that tapping was more variable to French sentences than English sentences, and that English speakers tapped more regularly to sentences of both languages. However, using the same protocol as the one used here, Lidji et al. [56] found that tapping was more variable to English than French stimuli, irrespective of participants' native language. Thus, it is presently unclear whether participants' native language influence tapping to speech. This should be explored in future studies.

The use of a tapping task has also limited ecological value for entrainment to speech. A shadowing task, for example, where the natural tendency of speakers to entrain to another speaker's speech rate is measured, could be an interesting paradigm to investigate further entrainment to speech [18,47–49] in beat-deaf individuals. The use of behavioral paradigms (tapping tasks), in the absence of neural measurements (such as electroencephalography), leaves open the question of the order in which timing mechanisms contribute to entrainment in speech and music. For example, it is possible that entrainment with music, which typically establishes a highly regular rhythm, is processed at a faster (earlier) timescale than language, which requires syntactic and semantic processing, known to elicit different timescales in language comprehension tasks [124,125]. These questions offer interesting avenues for future directions in comparison of rhythmic entrainment across speech and music.

5. Conclusions

In summary, our results indicate that beat deafness is not specific to music, but extends to any auditory rhythm, whether a metronome, speech or song. Furthermore, as proposed in previous studies [55,56], regularity or isochrony of the stimulus period seems to be the core feature through which entrainment is possible.

Author Contributions: Conceptualization, M.-É.L., C.P. and I.P.; data curation, M.-É.L.; formal analysis, M.-É.L.; funding acquisition, I.P.; investigation, M.-É.L.; methodology, M.-É.L., C.P. and I.P.; project administration, M.-É.L. and I.P.; resources, I.P.; software, M.-É.L. and C.P.; supervision, I.P.; validation, M.-É.L., C.P. and I.P.; visualization, M.-É.L.; writing—original draft, M.-É.L., C.P. and I.P.; writing—review and editing, M.-É.L., C.P. and I.P.

Funding: This research was funded by Natural Sciences and Engineering Research Council of Canada grant number 2014-04-068, the Canada Research Chairs program.

Acknowledgments: We would like to thank Pauline Tranchant for help with recruitment and insightful comments on data analysis, and Mailis Rodrigues for help with programming the task. We also thank Dawn Merrett for help with editing.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

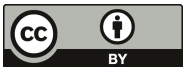
1. Zentner, M.; Eerola, T. Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5768–5773. [CrossRef] [PubMed]
2. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
3. Phillips-Silver, J.; Keller, P.E. Searching for Roots of Entrainment and Joint Action in Early Musical Interactions. *Front. Hum. Neurosci.* **2012**, *6*, 26. [CrossRef] [PubMed]
4. Jones, M.R. Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychol. Rev.* **1976**, *83*, 323–355. [CrossRef] [PubMed]
5. Jones, M.R. Dynamic pattern structure in music: Recent theory and research. *Percept. Psychophys.* **1987**, *41*, 621–634. [CrossRef] [PubMed]
6. Jones, M.R.; Boltz, M. Dynamic attending and responses to time. *Psychol. Rev.* **1989**, *96*, 459–491. [CrossRef] [PubMed]
7. Large, E.W. Resonating to musical rhythm: Theory and experiment. In *The Psychology of Time*; Grondin, S., Ed.; Emerald Group Publishing: Bingley, UK, 2008; pp. 189–232.
8. Large, E.W.; Palmer, C. Perceiving temporal regularity in music. *Cogn. Sci.* **2002**, *26*, 1–37. [CrossRef]
9. Large, E.W.; Snyder, J.S. Pulse and Meter as Neural Resonance. *Ann. New York Acad. Sci.* **2009**, *1169*, 46–57. [CrossRef]
10. Cummins, F. Rhythm as entrainment: The case of synchronous speech. *J. Phon.* **2009**, *37*, 16–28. [CrossRef]
11. Cummins, F.; Port, R. Rhythmic constraints on stress timing in English. *J. Phon.* **1998**, *26*, 145–171. [CrossRef]
12. Giraud, A.-L.; Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **2012**, *15*, 511–517. [CrossRef] [PubMed]
13. Goswami, U. Entraining the Brain: Applications to Language Research and Links to Musical Entrainment. *Empir. Music. Rev.* **2012**, *7*, 57–63. [CrossRef]
14. Lidji, P.; Palmer, C.; Peretz, I.; Morningstar, M. Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychon. Bull. Rev.* **2011**, *18*, 1035–1041. [CrossRef] [PubMed]
15. O'Dell, M.; Nieminen, T. Coupled oscillator model of speech rhythm. In Proceedings of the XIVth International Congress of Phonetic Sciences, Berkeley, CA, USA, 7 August 1999; Volume 2, pp. 1075–1078.
16. Peelle, J.E.; Davis, M.H.; Davis, M. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* **2012**, *3*, 320. [CrossRef] [PubMed]
17. Port, R.F. Meter and speech. *J. Phon.* **2003**, *31*, 599–611. [CrossRef]
18. Wilson, M.; Wilson, T.P. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **2005**, *12*, 957–968. [CrossRef]
19. Lerdahl, F.; Jackendoff, R. An Overview of Hierarchical Structure in Music. *Music Percept. Interdiscip. J.* **1983**, *1*, 229–252. [CrossRef]
20. Repp, B.H. Sensorimotor synchronization: A review of the tapping literature. *Psychon. Bull. Rev.* **2005**, *12*, 969–992. [CrossRef]
21. Repp, B.H.; Su, Y.-H. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychon. Bull. Rev.* **2013**, *20*, 403–452. [CrossRef]
22. Van Der Steen, M.C.; Keller, P.E. The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Front. Hum. Neurosci.* **2013**, *7*, 253. [CrossRef]
23. Chapin, H.L.; Zanto, T.; Jantzen, K.J.; Kelso, S.J.A.; Steinberg, F.; Large, E.W. Neural Responses to Complex Auditory Rhythms: The Role of Attending. *Front. Psychol.* **2010**, *1*, 224. [CrossRef] [PubMed]
24. Drake, C.; Jones, M.R.; Baruch, C. The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition* **2000**, *77*, 251–288. [CrossRef]
25. Large, E.W.; Herrera, J.A.; Velasco, M.J. Neural Networks for Beat Perception in Musical Rhythm. *Front. Syst. Neurosci.* **2015**, *9*, 583. [CrossRef] [PubMed]

26. Palmer, C.; Krumhansl, C.L. Mental representations for musical meter. *J. Exp. Psychol. Hum. Percept. Perform.* **1990**, *16*, 728–741. [CrossRef] [PubMed]
27. Repp, B.H.; Iversen, J.R.; Patel, A.D. Tracking an Imposed Beat within a Metrical Grid. *Music Percept. Interdiscip. J.* **2008**, *26*, 1–18. [CrossRef]
28. London, J. Cognitive Constraints on Metric Systems: Some Observations and Hypotheses. *Music Percept. Interdiscip. J.* **2002**, *19*, 529–550. [CrossRef]
29. McAuley, J.D. Tempo and Rhythm. In *Music Perception*; Riess Jones, M., Fay, R.R., Popper, A.N., Eds.; Springer: New York, NY, USA, 2010; pp. 165–199.
30. McAuley, J.D.; Jones, M.R.; Holub, S.; Johnston, H.M.; Miller, N.S. The time of our lives: Life span development of timing and event tracking. *J. Exp. Psychol. Gen.* **2006**, *135*, 348–367. [CrossRef] [PubMed]
31. Repp, B.H. Rate Limits in Sensorimotor Synchronization With Auditory and Visual Sequences: The Synchronization Threshold and the Benefits and Costs of Interval Subdivision. *J. Mot. Behav.* **2003**, *35*, 355–370. [CrossRef] [PubMed]
32. Drake, C.; Penel, A.; Bigand, E. Tapping in Time with Mechanically and Expressively Performed Music. *Music Percept. Interdiscip. J.* **2000**, *18*, 1–23. [CrossRef]
33. Palmer, C. Music performance. *Annu. Rev. Psychol.* **1997**, *48*, 115–138. [CrossRef]
34. Honing, H. Without it no music: Beat induction as a fundamental musical trait. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 85–91. [CrossRef] [PubMed]
35. Iversen, J.R. In the beginning was the beat. In *The Cambridge Companion to Percussion*; Hartenberger, R., Ed.; Cambridge University Press: Cambridge, UK, 2016; pp. 281–295.
36. Brown, S.; Pfordresher, P.Q.; Chow, I. A musical model of speech rhythm. *Psychomusicol. Music Mind Brain* **2017**, *27*, 95–112. [CrossRef]
37. Leong, V.; Stone, M.A.; Turner, R.E.; Goswami, U. A role for amplitude modulation phase relationships in speech rhythm perception. *J. Acoust. Soc. Am.* **2014**, *136*, 366–381. [CrossRef] [PubMed]
38. Meyer, L. The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *Eur. J. Neurosci.* **2018**, *48*, 2609–2621. [CrossRef] [PubMed]
39. Kotz, S.A.; Schwartze, M. Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends Cogn. Sci.* **2010**, *14*, 392–399. [CrossRef] [PubMed]
40. Selkirk, E.O. *Phonology and Syntax: The Relationship between Sound and Structure*; MIT Press: Cambridge, MA, USA, 1986; p. 494.
41. Turk, A.; Shattuck-Hufnagel, S. What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Lab. Phonol. J. Assoc. Lab. Phonol.* **2013**, *4*, 93–118. [CrossRef]
42. Liberman, M.; Prince, A. On Stress and Linguistic Rhythm. *Linguist. Inq.* **1977**, *8*, 249–336.
43. Dauer, R.M. Stress-timing and syllable-timing reanalyzed. *J. Phon.* **1983**, *11*, 51–62.
44. Jadoul, Y.; Ravignani, A.; Thompson, B.; Filippi, P.; De Boer, B. Seeking Temporal Predictability in Speech: Comparing Statistical Approaches on 18 World Languages. *Front. Hum. Neurosci.* **2016**, *10*, 351. [CrossRef]
45. Nolan, F.; Jeon, H.S. Speech rhythm: A metaphor? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2014**, *369*, 20130396. [CrossRef]
46. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: New York, NY, USA, 2008; p. 513.
47. Borrie, S.A.; Liss, J.M. Rhythm as a Coordinating Device: Entrainment with Disordered Speech. *J. Speech Lang. Hear. Res.* **2014**, *57*, 815–824. [CrossRef] [PubMed]
48. Jungers, M.K.; Palmer, C.; Speer, S.R. Time after time: The coordinating influence of tempo in music and speech. *Cogn. Process.* **2002**, *1*, 21–35.
49. Schultz, B.G.; O'Brien, I.; Phillips, N.; McFarland, D.H.; Titone, D.; Palmer, C. Speech rates converge in scripted turn-taking conversations. *Appl. Psycholinguist.* **2015**, *37*, 1201–1220. [CrossRef]
50. Cummins, F. Periodic and Aperiodic Synchronization in Skilled Action. *Front. Hum. Neurosci.* **2011**, *5*, 170. [CrossRef] [PubMed]
51. Cummins, F. Joint speech: The missing link between speech and music? *Percepta* **2013**, *1*, 17.
52. Cummins, F. On synchronous speech. *Acoust. Res. Lett. Online* **2002**, *3*, 7–11. [CrossRef]
53. Cummins, F. Entraining speech with speech and metronomes. *Cadernos de Estudos Lingüísticos* **2002**, *43*, 55–70. [CrossRef]
54. Cummins, F.; Li, C.; Wang, B. Coupling among speakers during synchronous speaking in English and Mandarin. *J. Phon.* **2013**, *41*, 432–441. [CrossRef]

55. Bella, S.D.; Białuńska, A.; Sowinski, J. Why Movement Is Captured by Music, but Less by Speech: Role of Temporal Regularity. *PLOS ONE* **2013**, *8*, e71945. [CrossRef]
56. Lidji, P.; Palmer, C.; Peretz, I.; Morningstar, M. Entrainment to speech and song. In Proceedings of the International Symposium on Performance Science, Utrecht, The Netherlands, 24 August 2011; pp. 123–128.
57. Bégel, V.; Benoit, C.-E.; Correa, A.; Cutanda, D.; Kotz, S.A.; Bella, S.D. "Lost in time" but still moving to the beat. *Neuropsychology* **2017**, *94*, 129–138. [CrossRef]
58. Bella, S.D.; Sowinski, J. Uncovering Beat Deafness: Detecting Rhythm Disorders with Synchronized Finger Tapping and Perceptual Timing Tasks. *J. Vis. Exp.* **2015**, *97*, 51761. [CrossRef] [PubMed]
59. Phillips-Silver, J.; Toiviainen, P.; Gosselin, N.; Piché, O.; Nozaradan, S.; Palmer, C.; Peretz, I. Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychology* **2011**, *49*, 961–969. [CrossRef] [PubMed]
60. Sowinski, J.; Bella, S.D. Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychology* **2013**, *51*, 1952–1963. [CrossRef] [PubMed]
61. Tranchant, P.; Vuvan, D.T.; Peretz, I. Keeping the Beat: A Large Sample Study of Bouncing and Clapping to Music. *PLOS ONE* **2016**, *11*, e0160178. [CrossRef] [PubMed]
62. Palmer, C.; Lidji, P.; Peretz, I. Losing the beat: deficits in temporal coordination. *Philos. Trans. R. Soc. B Boil. Sci.* **2014**, *369*, 20130405. [CrossRef] [PubMed]
63. Tranchant, P.; Peretz, I. Faulty Internal Rhythm in the Beat-based Form of Congenital Amusia. Unpublished work. (in preperation)
64. Wechsler, D.; Coalson, D.L.; Raiford, S.E. *WAIS-III: Wechsler Adult Intelligence Scale*; Psychological Corporation: San Antonio, TX, USA, 1997.
65. Lagrois, M.-É.; Peretz, I. The co-occurrence of pitch and rhythm disorders in congenital amusia. *Cortex* **2019**, *113*, 229–238. [CrossRef] [PubMed]
66. Tranchant, P.; Lagrois, M.-É.; Bellemare Pépin, A.; Schultz, B.G.; Peretz, I. Beat alignment test of the motor origin of musical entrainment deficits. *Neuropsychologia* **2019**, submitted.
67. Peretz, I.; Vuvan, D.T. Prevalence of congenital amusia. *Eur. J. Hum. Genet.* **2017**, *25*, 625–630. [CrossRef] [PubMed]
68. Vuvan, D.T.; Paquette, S.; Mignault Goulet, G.; Royal, I.; Felezeu, M.; Peretz, I. The Montreal Protocol for Identification of Amusia. *Behav. Res. Methods* **2018**, *50*, 662–672. [CrossRef] [PubMed]
69. Boersma, P.; Weenink, D. Praat: Doing phonetics by computer [Computer program], Version 6.0.2. 2017. Available online: http://www.praat.org/ (accessed on 17 January 2017).
70. Ramus, F.; Nespor, M.; Mehler, J. Correlates of linguistic rhythm in the speech signal. *Cognition* **1999**, *73*, 265–292. [CrossRef]
71. Grabe, E.; Low, E.L. Durational variability in speech and the rhythm class hypothesis. In *Laboratory Phonology 7*; Gussenhoven, C., Warner, N., Eds.; De Gruyter Mouton: Berlin, Germany, 2002; pp. 515–546.
72. Schultz, B.G.; van Vugt, F.T. Tap Arduino: An Arduino microcontroller for low-latency auditory feedback in sensorimotor synchronization experiments. *Behav. Res. Methods* **2016**, *48*, 1591–1607. [CrossRef] [PubMed]
73. Van Vugt, F.T.; Schultz, B.G. Taparduino v1.01. Zenodo 16178. 2015. Available online: https://doi.org/10.5281/zenodo.16178 (accessed on 20 March 2015).
74. Dalla Bella, S.; Farrugia, N.; Benoit, C.-E.; Begel, V.; Verga, L.; Harding, E.; Kotz, S.A. BAASTA: Battery for the assessment of auditory sensorimotor and timing abilities. *Behav. Res. Methods* **2017**, *49*, 1128–1145. [CrossRef] [PubMed]
75. Chen, J.L.; Penhune, V.B.; Zatorre, R.J. Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *J. Cogn. Neurosci.* **2008**, *20*, 226–239. [CrossRef] [PubMed]
76. Giovannelli, F.; Innocenti, I.; Rossi, S.; Borgheresi, A.; Ragazzoni, A.; Zaccara, G.; Viggiano, M.P.; Cincotta, M. Role of the dorsal premotor cortex in rhythmic auditory-motor entrainment: A perturbational approach by rTMS. *Cereb. Cortex* **2014**, *24*, 1009–1016. [CrossRef] [PubMed]
77. Leow, L.-A.; Parrott, T.; Grahn, J.A. Individual Differences in Beat Perception Affect Gait Responses to Low- and High-Groove Music. *Front. Hum. Neurosci.* **2014**, *8*, 811. [CrossRef] [PubMed]
78. Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. *J. Stat. Softw.* **2009**, *31*, 1–21. [CrossRef]
79. Wilkie, D. Rayleigh Test for Randomness of Circular Data. *J. R. Stat. Soc. Ser. C Applied Stat.* **1983**, *32*, 311. [CrossRef]

80. Dalla Bella, S.; Sonja, K. Method and apparatus for the synchronization of data sequences including filtering. Google Patents US20180199859A1, 2018.
81. Nozaradan, S.; Peretz, I.; Keller, P.E. Individual Differences in Rhythmic Cortical Entrainment Correlate with Predictive Behavior in Sensorimotor Synchronization. *Sci. Rep.* **2016**, *6*, 20612. [CrossRef]
82. Thabane, L.; Mbuagbaw, L.; Zhang, S.; Samaan, Z.; Marcucci, M.; Ye, C.; Thabane, M.; Giangregorio, L.; Dennis, B.; Kosa, D.; et al. A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Med. Res. Methodol.* **2013**, *13*, 92. [CrossRef]
83. Kover, S.T.; Atwood, A.K. Establishing Equivalence: Methodological Progress in Group-Matching Design and Analysis. *Am. J. Intellect. Dev. Disabil.* **2013**, *118*, 3–15. [CrossRef]
84. Ammirante, P.; Thompson, W.F.; Russo, F.A. Ideomotor effects of pitch on continuation tapping. *Q. J. Exp. Psychol.* **2011**, *64*, 381–393. [CrossRef] [PubMed]
85. Boasson, A.D.; Granot, R. Melodic direction's effect on tapping. In Proceedings of the 12th International Conference on Music Perception and Cognition, Thessaloniki, Greece, 23–28 July 2012.
86. Ellis, R.J.; Jones, M.R. The role of accent salience and joint accent structure in meter perception. *J. Exp. Psychol. Hum. Percept. Perform.* **2009**, *35*, 264–280. [CrossRef] [PubMed]
87. Hannon, E.E.; Snyder, J.S.; Eerola, T.; Krumhansl, C.L. The Role of Melodic and Temporal Cues in Perceiving Musical Meter. *J. Exp. Psychol. Hum. Percept. Perform.* **2004**, *30*, 956–974. [CrossRef] [PubMed]
88. Jones, M.R.; Pfordresher, P.Q. Tracking musical patterns using joint accent structure. *Can. J. Exper. Psychol.* **1997**, *51*, 271–291. [CrossRef]
89. McKinney, M.F.; Moelants, D. Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels? *Music. Percept. Interdiscip. J.* **2006**, *24*, 155–166. [CrossRef]
90. Pfordresher, P.Q. The Role of Melodic and Rhythmic Accents in Musical Structure. *Music. Percept. Interdiscip. J.* **2003**, *20*, 431–464. [CrossRef]
91. Prince, J.B. The integration of stimulus dimensions in the perception of music. *Q. J. Exp. Psychol.* **2011**, *64*, 2125–2152. [CrossRef]
92. Prince, J.B. Pitch structure, but not selective attention, affects accent weightings in metrical grouping. *J. Exp. Psychol. Hum. Percept. Perform.* **2014**, *40*, 2073–2090. [CrossRef]
93. Prince, J.B.; Pfordresher, P.Q. The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychol.* **2012**, *141*, 184–198. [CrossRef]
94. Palmer, C.; Krumhansl, C.L. Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Percept. Psychophys.* **1987**, *41*, 505–518. [CrossRef]
95. Snyder, J.; Krumhansl, C.L. Tapping to Ragtime: Cues to Pulse Finding. *Music Percept. Interdiscip. J.* **2001**, *18*, 455–489. [CrossRef]
96. Nobre, A.C.; van Ede, F. Anticipated moments: Temporal structure in attention. *Nat. Rev. Neurosci.* **2018**, *19*, 34–48. [CrossRef] [PubMed]
97. Di Liberto, G.M.; O'Sullivan, J.A.; Lalor, E.C. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Curr. Boil.* **2015**, *25*, 2457–2465. [CrossRef] [PubMed]
98. Ding, N.; Simon, J.Z. Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* **2014**, *8*, 311. [CrossRef] [PubMed]
99. Ghitza, O. Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2011**, *2*, 130. [CrossRef] [PubMed]
100. Ghitza, O. The theta-syllable: A unit of speech information defined by cortical function. *Front. Psychol.* **2013**, *4*, 138. [CrossRef]
101. Gross, J.; Hoogenboom, N.; Thut, G.; Schyns, P.; Panzeri, S.; Belin, P.; Garrod, S. Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Boil.* **2013**, *11*, e1001752. [CrossRef]
102. Kayser, S.J.; Ince, R.A.; Gross, J.; Kayser, C. Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. *J. Neurosci.* **2015**, *35*, 14691–14701. [CrossRef]
103. Peelle, J.E.; Gross, J.; Davis, M.H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **2013**, *23*, 1378–1387. [CrossRef]
104. Zhang, W.; Ding, N. Time-domain analysis of neural tracking of hierarchical linguistic structures. *NeuroImage* **2017**, *146*, 333–340. [CrossRef]
105. Doelling, K.B.; Poeppel, D. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6233–E6242. [CrossRef] [PubMed]

106. Mai, G.; Minett, J.W.; Wang, W.S.-Y. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage* **2016**, *133*, 516–528. [CrossRef] [PubMed]
107. Nozaradan, S. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. B Boil. Sci.* **2014**, *369*, 20130393. [CrossRef] [PubMed]
108. Nozaradan, S.; Peretz, I.; Missal, M.; Mouraux, A. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* **2011**, *31*, 10234–10240. [CrossRef] [PubMed]
109. Stupacher, J.; Wood, G.; Witte, M. Neural Entrainment to Polyrhythms: A Comparison of Musicians and Non-musicians. *Front. Mol. Neurosci.* **2017**, *11*, 208. [CrossRef] [PubMed]
110. Tal, I.; Large, E.W.; Rabinovitch, E.; Wei, Y.; Schroeder, C.E.; Poeppel, D.; Golumbic, E.Z. Neural Entrainment to the Beat: The "Missing-Pulse" Phenomenon. *J. Neurosci.* **2017**, *37*, 6331–6341. [CrossRef] [PubMed]
111. Tierney, A.; Kraus, N. Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* **2014**, *8*, 949. [CrossRef] [PubMed]
112. Corriveau, K.H.; Goswami, U. Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex* **2009**, *45*, 119–130. [CrossRef]
113. Cumming, R.; Wilson, A.; Leong, V.; Colling, L.J.; Goswami, U. Awareness of Rhythm Patterns in Speech and Music in Children with Specific Language Impairments. *Front. Hum. Neurosci.* **2015**, *9*, 200. [CrossRef]
114. Flaugnacco, E.; Lopez, L.; Terribili, C.; Zoia, S.; Buda, S.; Tilli, S.; Monasta, L.; Montico, M.; Sila, A.; Ronfani, L.; et al. Rhythm perception and production predict reading abilities in developmental dyslexia. *Front. Hum. Neurosci.* **2014**, *8*. [CrossRef]
115. Thomson, J.M.; Fryer, B.; Maltby, J.; Goswami, U.; Thomson, J.; Fryer, B. Auditory and motor rhythm awareness in adults with dyslexia. *J. Res. Read.* **2006**, *29*, 334–348. [CrossRef]
116. Thomson, J.M.; Goswami, U.; Thomson, J. Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *J. Physiol.* **2008**, *102*, 120–129. [CrossRef] [PubMed]
117. Bonacina, S.; Krizman, J.; White-Schwoch, T.; Kraus, N. Clapping in time parallels literacy and calls upon overlapping neural mechanisms in early readers. *Ann. New York Acad. Sci.* **2018**, *1423*, 338–348. [CrossRef] [PubMed]
118. Tierney, A.T.; Kraus, N. The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain Lang.* **2013**, *124*, 225–231. [CrossRef] [PubMed]
119. Carr, K.W.; Fitzroy, A.B.; Tierney, A.; White-Schwoch, T.; Kraus, N. Incorporation of feedback during beat synchronization is an index of neural maturation and reading skills. *Brain Lang.* **2017**, *164*, 43–52. [CrossRef] [PubMed]
120. Falk, S.; Müller, T.; Bella, S.D. Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Front. Psychol.* **2015**, *6*, 847. [CrossRef] [PubMed]
121. Gracco, V.L.; Van De Vorst, R. Atypical non-verbal sensorimotor synchronization in adults who stutter may be modulated by auditory feedback. *J. Fluen. Disord.* **2017**, *53*, 14–25.
122. Cutler, A. Listening to a second language through the ears of a first. *Interpreting. Int. J. Res. Pr. Interpreting* **2000**, *5*, 1–23. [CrossRef]
123. Iversen, J.R.; Patel, A.D.; Ohgushi, K. Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* **2008**, *124*, 2263–2271. [CrossRef]
124. Friederici, A.D.; Hahne, A.; Mecklinger, A. Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *J. Exp. Psychol. Learn. Mem. Cogn.* **1996**, *22*, 1219–1248. [CrossRef]
125. Kutas, M.; Hillyard, S. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* **1980**, *207*, 203–205. [CrossRef] [PubMed]



brain sciences

MDPI

Article

Processing of Rhythm in Speech and Music in Adult Dyslexia

Natalie Boll-Avetisyan [1,*], Anjali Bhatara [2] and Barbara Höhle [1]

[1] SFB1287, Research Focus Cognitive Sciences, Faculty of Human Sciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; hoehle@uni-potsdam.de

[2] CNRS, (Integrative Neuroscience and Cognition Center, UMR 8002), Université de Paris, 45 rue des Saints-Pères, 75270 Paris, France; bhatara@gmail.com

* Correspondence: nboll@uni-potsdam.de; Tel.: +49-331-977-2374

Received: 23 September 2019; Accepted: 29 April 2020; Published: 30 April 2020

Abstract: Recent studies have suggested that musical rhythm perception ability can affect the phonological system. The most prevalent causal account for developmental dyslexia is the phonological deficit hypothesis. As rhythm is a subpart of phonology, we hypothesized that reading deficits in dyslexia are associated with rhythm processing in speech and in music. In a rhythmic grouping task, adults with diagnosed dyslexia and age-matched controls listened to speech streams with syllables alternating in intensity, duration, or neither, and indicated whether they perceived a strong-weak or weak-strong rhythm pattern. Additionally, their reading and musical rhythm abilities were measured. Results showed that adults with dyslexia had lower musical rhythm abilities than adults without dyslexia. Moreover, lower musical rhythm ability was associated with lower reading ability in dyslexia. However, speech grouping by adults with dyslexia was not impaired when musical rhythm perception ability was controlled: like adults without dyslexia, they showed consistent preferences. However, rhythmic grouping was predicted by musical rhythm perception ability, irrespective of dyslexia. The results suggest associations among musical rhythm perception ability, speech rhythm perception, and reading ability. This highlights the importance of considering individual variability to better understand dyslexia and raises the possibility that musical rhythm perception ability is a key to phonological and reading acquisition.

Keywords: developmental dyslexia; Iambic/Trochaic Law; rhythmic grouping; musicality; speech perception; rhythm perception

1. Introduction

Developmental dyslexia (henceforth, dyslexia) affects the acquisition of reading and writing skills despite adequate cognitive and motoric abilities and appropriate access to education. Beyond literacy, dyslexia is also characterized by deficits in spoken language processing, particularly in processing phonological information. For this reason, researchers have proposed that deficits in the processing of phonological information may be the bridge connecting the deficits in spoken and written language, e.g., [1–4]. One prominent theory of dyslexia proposes that phonological processing difficulties are a consequence of impaired auditory processing abilities, in particular when processing rhythm information in speech and music [5]. The present paper aims to connect these hypotheses by investigating the processing of one specific type of phonological information, namely, rhythm information in speech, and its potential associations with literacy and the ability to perceive musical rhythms in dyslexia.

1.1. The Phonological Deficit Hypothesis

The original phonological deficit hypothesis proposes that a deficit in phonological skills underlies dyslexia, as evidenced by difficulties with tasks that tap into phoneme awareness, letter-sound knowledge, verbal short-term memory, and rapid automatized naming. For a recent review see [6]. Research on this hypothesis has primarily concentrated on deficits regarding segmental (i.e., phoneme) information and has, for example, established that children with dyslexia do not seem to perceive phonemes in the same way as children without dyslexia. Specifically, they have a reduced sensitivity to phonemically relevant distinctions (e.g., when discriminating /p/ from /b/) and an enhanced sensitivity to allophonic variants (e.g., when discriminating different realizations of /b/) compared to listeners without dyslexia, who show clear effects of categorical perception of consonants (for a meta-analysis see [7]). As categorical perception is assumed to result from effects of the native language phonological system to speech perception, e.g., [8], weak categorical perception may indicate that the language-specific phoneme categories are not sufficiently well-established. In the case of dyslexia, less well-defined phoneme categories may create difficulties in the phoneme-grapheme mappings that are relevant for the acquisition and/or processing of written language.

1.2. Rhythm Perception Deficits in Dyslexia

More recent developments in dyslexia research have shown that the phonological deficits in dyslexia are not restricted to processing segmental information, but also affect the processing of suprasegmental (i.e., prosodic) information, and, in particular, the processing of rhythm. Rhythm is established by the regular occurrence of an element or a pattern in time. Rhythm is an important feature of languages' prosody, and can be characterized by, for example, an alternation of more prominent (i.e., strong) syllables with less prominent (i.e., weak) ones. Languages differ in their rhythmic structure as the organization of speech in alternations of strong and weak syllables is determined by language-specific "metrical stress" rules [9–11]. For example, in English and German, the basic rhythmic unit has a strong-weak (i.e., trochaic) pattern, but in other languages such as Hebrew, the basic rhythmic unit is weak-strong (i.e., iambic) [9]. Compared to groups without dyslexia, groups of individuals with dyslexia show lower performance in tasks that require perceptual sensitivity to and/or knowledge of stress rules. For example, this is the case in discrimination tasks with words or phrases pronounced with correct or incorrect stress patterns, e.g., [12–14]–an effect that is even present in young children with a familial risk for dyslexia [15]. In addition, these abilities have been found to correlate with reading skills [16–18].

Goswami [5,19] has proposed that a fundamental deficit in the processing of rhythmic information is associated with dyslexia. This account focuses on the periodic modulations of amplitude (amplitude envelope) that are crucial to establish speech rhythm with amplitude peaks being aligned with the strong (stressed) syllables of a speech sequence and Goswami assumes that the processing of this amplitude envelope is impeded in dyslexia. These difficulties may result from atypical basic auditory processing: numerous studies have found that individuals with dyslexia show low performance in the perception of rise time (i.e., the velocity of the amplitude increase) and that the perception of rise time is related to the discrimination of word stress patterns (for a review see [20]. Research on the neural basis of this impairment suggests that in dyslexia, neural oscillations are not synchronized with auditory rhythms in the same way as in populations without dyslexia [21–24]. Independent of whether the basis of the impairment is perceptual or neural, according to Goswami, the problem in the processing of rhythm hinders the segmentation of speech into syllables and also the perception of subsyllabic units like rhymes and single phonemes, the latter case explaining the segmental phonological problems in dyslexia. Although at this point any causal interpretations of associations in neural rhythmic entrainment and dyslexia have to be taken with care, it is relevant to note that Goswami's theory has the potential to account for a broader range of deficits that have been observed in dyslexia. Low performance in the perception of speech rhythm seems to extend to non-linguistic domains such as beat perception in music [25–27], and even to motor synchronization

abilities such as rhythmic tapping [28–30], which suggests a domain-general rhythm processing deficit in dyslexia.

With its focus on rhythm processing, Goswami and colleagues' work offers a substantial approach to the potential mechanisms underlying the performance of individuals with dyslexia in different domains. In our study, we intend to broaden the view on rhythm perception in dyslexia by looking at duration and intensity as acoustic cues of speech rhythm perception. Acoustically, strong and weak syllables can be distinguished on the basis of specific cues such as intensity, duration, and pitch, with strong syllables often being louder, longer, and higher than weak ones [31,32]. Interestingly, these different cues have different effects on rhythmic grouping and segmentation: while alternations in syllables' duration lead to the perception of weak-strong patterns, alternations in intensity and pitch lead to the perception of strong-weak patterns [9,33–38]; for more details see 1.3. The main goal of this paper is to investigate rhythmic grouping according to this bias in individuals with dyslexia.

Of course, not only speech is rhythmically structured. Rhythm is a domain-general phenomenon. Similar organizational rhythmic principles with regular alternations of strong and weak elements are also found in music [9,39], where the same acoustic cues (intensity, duration, and pitch) are relevant for conveying rhythm, and the same tendency to use these cues differently at the beginning or the end of a unit is often exhibited [39–41]. If rhythm perception in speech and music relies on shared perceptual mechanisms or shared rhythm representations, then it should be the case that individuals with better music abilities should also show enhanced language abilities [42–44]. In line with this, [45] reported that, within a group of adults with dyslexia, musicians outperformed non-musicians on several auditory measures, including rise time, frequency, intensity, and timing perception, even reaching the same levels of performance as musicians without dyslexia. However, the advantage that musicians with dyslexia experienced in the auditory perception tasks did not extend to their literacy and phonological awareness. Accordingly, other researchers doubt that dyslexia relates to poor rhythm perception, e.g., [46]. A second goal of the present paper, therefore, is to further examine whether rhythm processing deficits in speech and music are linked with reading deficits in dyslexia.

1.3. Biases on Auditory Rhythmic Grouping

In this paper, we will investigate for the first time how biases on auditory rhythm perception affect speech rhythm perception by adults with dyslexia. For this, we focus on rhythmic grouping of speech following the Iambic/Trochaic Law (ITL) [9]. According to the ITL, rhythmic perception is guided by universal biases. These biases have the effect that sequences of sounds varying in intensity tend to be perceived as trochees (strong-weak), whereas sound sequences varying in duration tend to be perceived as iambs (weak-strong; e.g., [33–36]). These biases have been attested for speakers of various languages, including English [33,34,36,47,48], German, French [38,49], Spanish [48], and Italian [37]. Since rhythmic grouping preferences are asymmetrical between the perceived acoustic cues, these biases cannot simply be accounted for by a tracking of acoustic cues to prominence in the signal. Importantly, asymmetries in rhythmic grouping are mirrored in the rhythm structures in language and music where final prominence is usually marked by a long syllable or note, and initial prominence by a loud syllable or beat, which supports the assumption of the ITL as universal [19]. More recent research, however, indicates that rhythmic grouping preferences are subject to individual variation and depend to some degree on aspects such as individuals' language background [38,47–50] and their musical abilities [51,52].

1.3.1. Effects of Language Background on Rhythmic Grouping

Language background's effects on perception may relate to differences in the function of stress between the languages: [38] hypothesized that when perceiving speech, the ability to draw on abstract phonological representations of lexical stress would facilitate German speakers' rhythm processing. This is because German uses lexical stress contrastively (e.g., /'te,nor/ 'common sense' vs. /te'nor/ 'singer'), while French does not. In [38]'s rhythmic grouping experiment, German and French

listeners listened to syllable streams, in which syllables alternated in intensity (loud-soft-loud-soft …), pitch (high-low-high-low …), duration (long-short-long-short …), or neither (flat control condition). Participants were asked to indicate via button presses whether they perceived strong-weak or weak-strong groupings. The result was that both groups perceived iambs and trochees as predicted by the ITL, but the German listeners were more consistent and had clearer rhythmic grouping preferences than the French listeners. Moreover, German but not French listeners experienced the illusion of hearing strong-weak groupings when listening to the control sequences that did not contain any acoustic cues to rhythm. Ref. [38] argue that this effect is likely to also be driven by the presence of abstract phonological representations of stress in German: As words in German are pre-dominantly trochaic, German listeners might apply a default grouping to sound sequences based on their linguistic experience. Since French has no lexical stress, there may be no reason for a default grouping based on their experience.

1.3.2. Effects of Musical Background on Rhythmic Grouping

Musical ***experience***as defined by the number of acquired musical instruments, the duration of musical training, and the earliest age of acquiring a musical instrument has been found to influence rhythmic grouping. However, this seems to be modulated by the individuals' language background and has, to this point, only been found to affect native speakers of French and not native speakers of German [38,49,51,52]. French speakers who are musically experienced have clearer preferences for grouping acoustically complex non-speech sounds [49] as well as for grouping speech, though only if they are also proficient speakers of German [51]. While general musical experience never predicted monolingual German speakers' grouping of speech, their ability to perceive musical rhythm as measured by a standardized musical ability test (the Musical Ear Test, henceforward MET [53]) did (though their ability to perceive melodies did not) [52]. Musical abilities can be, but not necessarily, correlated with musical experience [46]. Instead, they may relate to more general auditory perception abilities, which vary widely among individuals [29]. In addition, the abilities to perceive and discriminate musical rhythms do not always correlate with musical melody perception abilities [54]. Together, these results suggest a specific connection between language and music via rhythmical properties.

1.4. Hypotheses and Predictions

Individuals with dyslexia have repeatedly exhibited relatively weak stress and rhythm processing abilities, even in domains other than language (e.g., in tapping and music perception). This suggests that their rhythmic grouping preferences will also be weak, especially since rhythmic grouping depends on native language phonological knowledge. Given the findings that musical ability also influences rhythmic speech grouping, the present study set out to investigate the relations among speech rhythm processing, musical rhythm perception ability, reading ability, and dyslexia in German listeners. We aimed at investigating the following research questions:

(1) Do adults with dyslexia have less consistent rhythmic speech grouping preferences than adults without dyslexia?
(2) Do adults with dyslexia show lower musical rhythm perception abilities than adults without dyslexia?
(3) Does musical rhythm perception ability predict rhythmic speech grouping in dyslexia?
(4) Does musical rhythm perception ability predict reading ability in dyslexia?

We hypothesized the following:

(1) Based on the hypothesis that adults with dyslexia have difficulties in processing rhythm, we expect them to show weak grouping preferences. Hence, they should show less asymmetrical grouping preferences when hearing sequences varying in intensity or duration than adults without dyslexia. Further, if this rhythmic deficit hinders the establishment of phonological representations for metrical structure, adults with dyslexia should not show grouping preferences when hearing rhythmically invariant sequences.

(2) We assumed that musical rhythm perception ability would be lower in individuals with dyslexia than in individuals without.
(3) If rhythm perception in music and speech share cognitive underpinnings, we expect that higher musical rhythm perception ability would be associated with more consistent preferences in rhythmic grouping of speech in adults with dyslexia.
(4) If reading difficulties are linked with general underlying difficulties with rhythm processing, then musical rhythm perception ability should predict reading ability in dyslexia.

To investigate these hypotheses, we conducted a rhythmic grouping experiment with adults with and without dyslexia and measured their musical rhythm ability by means of the MET [53], and their reading ability by means of the Salzburger Lese- und Rechtschreibtest SLRT-II [55]. In order to avoid pre-selecting or grouping participants based on their musicality and cognitive abilities, we applied regression modeling for data analysis, with musical rhythm ability, musical experience, and cognitive abilities as covariates.

2. Materials and Methods

2.1. Participants

Participants were 23 monolingually-raised adult native speakers of German with dyslexia (nine women, 14 men, mean age = 24 years, age range: 17–35 years) and 23 (12 women, 11 men) age-matched controls. An additional participant with dyslexia was raised bilingually, and, hence, excluded together with the age-matched control. Participants gave informed consent before taking part.

The inclusion criterion for participants with dyslexia was that they showed us their formal testimonial of their developmental dyslexia diagnosis. In Germany, there are no nation-wide standards for dyslexia diagnosis. To verify the diagnosis provided by the participants, we compared how the groups with and without dyslexia fared at a reading test (i.e., the SLRT-II; [55], see below). Results of a linear regression indicated significantly lower nonword reading ability for the group with dyslexia compared to the group without ($\beta = 41$, SE = 6.06, t = 6.77, $p < 0.001$). This result allowed us to conclude that the testimonial of the dyslexia diagnosis did justify the division of the participants into two groups (with vs. without dyslexia). Hence, we used group (rather than reading ability scores) as a factor to test assumptions regarding dyslexia. Other than these, there were no further constraints on recruitment. Participants of both groups were recruited in the cities of Berlin and Potsdam by means of distribution of flyers and online advertisements on social media, to make sure that our sample would not only consist of university students. For a detailed summary of the groups' background information and the groups' average performance in the tasks described in Section 2.2, see Table 1.

The sample size is justified, as effect sizes of prior rhythmic grouping studies were high: for example in [38] for comparisons between French and German listeners in the intensity condition Cohen's d = 1.4 (large) and Cohen's d = 1.1 (large) in the duration condition; for comparisons between conditions (duration vs. intensity, and duration vs. control) within native speakers of German Cohen's d = 4.4 (large). Moreover, we tested our design using the PANGEA software (https://jakewestfall.shinyapps.io/pangea/, see [56]), which revealed a high power (0.91) for a study design including a four-way interaction with 23 participants per group with the alpha level set at 0.05, and an assumed medium effect size of 0.45. This effect size of 0.45 is conservative given the large effect sizes found in prior studies, however, since prior studies on rhythmic grouping have, as yet, not included adults with dyslexia, power calculations have to be taken with caution.

Table 1. Summary of the results of all questions from the questionnaire as well as all musical and cognitive tests for both the group of adults with versus without dyslexia.

General Participant Information (in N)	With Dyslexia (N = 23)	Without Dyslexia (N = 23)
Age (mean, range)	23.781 (7–35)	23.95 (18–35)
Gender	9 women, 14 men	12 women, 11 men
Handedness	19 right, 1 left, 3 both	22 right, 1 left
Native language = German	23	23
Mother with native language other than German	1	2
Father with native language other than German	3	0
Vision problems (short- or far-sighted, usually compensated by glasses)	10	9
Hearing problems	Auditory perception disorder (1), Otitis media with effusion in childhood (1), Un-defined (1)	0
Language problems	Stuttering (1), Specific Language Impairment (1)	0
Learning problems	Attention Deficit Hyperactivity Disorder (2)	0
Highest Education (in N)		
Without degree	1	1
Hauptschule	0	0
Realschule	7	2
Fachhochschulreife	2	0
Hochschulreife (Abitur)	8	13
Berufsausbildung	8	1
Hochschulabschluss	0	7
Promotion	0	0
Other	1	0
Foreign Language Experience (Max. Number of Learned Foreign Languages, in N)		
One	8	-
Two	9	11
Three	5	7
Four	1	3
Five	0	2

Table 1. *Cont.*

General Participant Information (in N)	With Dyslexia (N = 23)	Without Dyslexia (N = 23)
General Musical Experience (in N)	19 yes, 4 no	18 yes, 5 no
Specific Musical Experience (Average, Range in Bracket)		
Number of instruments (or musical activities such as choir, dance)	2.05 (1–6)	2.94 (1–6)
Age of first musical instrument or activity acquisition	10.72 (4–24)	8.35 (4–20)
Years of practicing a musical instrument or activity	6.11 (1–16)	13.06 (1–30)
Hours spent singing per week	3.74 (0–35)	2.18 (0–10)
Hours spent dancing per week	0.72 (0–6)	1.09 (0–7)
Hours spent with instrument play per week (excl. participants without musical experience)	4.26 (0–40)	2.53 (0–20)
Hours spent listening to music per week	13.04 (0–80)	13.86 (0–40)
Musical Abilities (Self-estimated, Likert Scale 0 (No Ability)–10 (Perfect) (Average, Range in Bracket)		
Musical instrument (excl. participants without musical experience)	3.6 (0–9)	5.35 (0–9)
Dancing	1.9 (0–8)	3.82 (0–9)
Singing	2.59 (0–7)	4.55 (0–9)
Preferred Music Styles (in N)		
Classical music	8	11
Pop	13	15
Rock	16	14
Hiphop	14	7
Jazz	7	7
Popular folk (Schlager)	2	1
Reggae	11	3
Techno	8	7
Heavy Metal	4	4
World music	5	3
Country	2	4
Other	1 (dubstep)	1 (child music)

Table 1. *Cont.*

General Participant Information (in N)	With Dyslexia (N = 23)	Without Dyslexia (N = 23)
Dyslexia Therapy (in N)		
Received dyslexia therapy	20 of 23	
Therapy included music therapy	3 of 23	
Dyslexia Characteristics (in N)		
Reading and writing difficulties	9 of 23	
Reading difficulties alone	3 of 23	
Writing difficulties alone	11 of 23	
Performance in Musicality Tests (in % Correct Responses, Range)		
Musical Ear Test: rhythm test	62% (44–79)	74% (52–92)
Performance in Cognitive Tasks (in N Correct Responses, Range)		
Salzburger Lese-Rechtschreib-Test: word reading	85.13 (18–119)	127.82 (92–156)
Salzburger Lese-Rechtschreib-Test: pseudoword reading	44.13 (14–81)	85.82 (24–124)
Wechsler Adult Intelligence Scale, Verbal Comprehension: similarities	23.13 (14–33)	26.77 (14–34)
Wechsler Adult Intelligence Scale, Working Memory: digit span	24 (15–38)	28.86 (22–37)
Wechsler Adult Intelligence Scale, Processing Speed: symbol search	37.04 (20–56)	43.18 (25–65)
Wechsler Adult Intelligence Scale, Processing Speed: coding	64.83 (34–96)	81.14 (63–115)

2.2. Task Battery

2.2.1. Rhythmic Grouping Preferences

In order to assess rhythmic speech grouping preferences, we used the stimuli and procedure from [37], Experiment 1. The stimuli were 90 speech-like streams that consisted of different simple syllables in which one consonant was always followed by one vowel (e.g., / . . . zulebolilozimube . . . /). The streams were text-to-speech synthesized with a German pronunciation and flat F0. There were three conditions: An intensity condition in which every second syllable was louder than the preceding one, a duration condition in which every second syllable was longer than the preceding one, and a control condition, in which all syllables were of equal intensity and duration. The task was to listen to each of the nonsense speech streams and to indicate by button press whether this pattern consisted of strong-weak or weak-strong disyllables. The proportion of strong-weak responses in the three conditions (intensity/duration/control) served as a dependent variable (Section 3.1/Section 3.3); for details, see [38].

2.2.2. Musical Rhythm Perception Ability

Receptive musical rhythm abilities were assessed using the Musical Ear Test [53]. Participants heard 52 pairs of rhythmic sequences, which are containing 4–11 wood block beats, and had to decide whether the two sequences were the same or different. The obtained proportion of correct responses was used as a dependent measure to evaluate whether the group with dyslexia showed lower performance than the group without dyslexia (Section 3.2). Furthermore, this measure was used as an independent variable to understand its role as a predictor of rhythmic grouping (Section 3.3) and reading ability (Section 3.4).

2.2.3. Questionnaire

An interview based on a questionnaire was used to collect information on the participants' musical and language background, and, if applicable, their dyslexia status and therapy experience (for details and a summary of the results, see Table 1). Questions were read out by the experimenter, who also filled out the questionnaire based on the responses. Following [49,51,52], a predictor of musical experience was extracted using the answers to questions regarding the number of acquired instruments, the age of acquiring the first instrument, and the duration of years of musical practice. In the following analyses, it was tested whether this predictor was correlated with musical rhythm ability (Section 3.2), rhythmic grouping (Section 3.3), and reading ability (Section 3.4).

2.2.4. Reading Ability

Participants completed the reading fluency test of the Salzburger Lese- und Rechtschreibtests (SLRT-II) [55], a standardized test for the diagnosis of dyslexia. They were asked to read aloud lists of words and nonwords within a time limit (one minute per list). It allows for a separate diagnosis of deficits in automatic word recognition versus synthetic sound-based reading. The latter is predicted to be particularly weak in individuals with dyslexia. Note that we did not use this test to diagnose any of the participants with dyslexia. The purpose of this test was to verify that the groups defined on presence vs. absence of formal diagnosis of dyslexia truly differed in reading ability (see Section 2.1), and to test whether musical rhythm ability predicted reading ability (Section 3.4).

2.2.5. Cognitive Ability

Many studies suggest that individual variability in cognitive abilities such as general verbal comprehension, short-term memory, and processing speed can influence performance in psycholinguistic experiments (for a systematic review, see [57]). In order to verify that potential differences between the groups with or without dyslexia in the experimental task are not due to differences in such general

cognitive abilities, participants completed four subsets from the Wechsler Adult Intelligence Scale WAIS-IV (a version adapted for German, [58]), a standardized tool for determining the intelligence quotient. To test verbal comprehension (specifically, verbal reasoning and semantic knowledge), participants performed the subtest Similarities, in which participants heard 18 pairs of words (e.g., piano & drum, or friend & enemy), for which they had to describe which attributes they share. Next, to test short-term memory, they performed subtests that measured their digit span. Specifically, they listened to sequences of orally presented numbers, and, in three subsequent sub-tests, they were required to repeat them in as heard, backward, or in sequential (ascending) order.

For measuring their processing speed, we selected two subtests: Symbol search and Coding. In Symbol Search, participants were required to search for two target symbols in a row of different symbols, and to indicate whether the target symbols were present or not. In Coding, nine different numbers (1–9) are assigned a different symbol. In the task, participants are presented with a list of numbers and are required to draw the corresponding symbol next to each of the numbers. (Participants additionally completed a nonword repetition task for adults [59], which was based on the Mottier test, a standardized test for German-speaking children [60]. Because of redundancy with the digit span tests (Section 2.2.5), which also test verbal memory, we did not include the data of the nonword repetition test in the analyses.) A composite score of the results of all subtests served as a covariate in analyses of rhythmic grouping (Section 3.1/Section 3.3), musical rhythm perception ability (Section 3.2) and reading ability (Section 3.4).

2.3. Data Processing and Analyses

For the analyses, we included data from both the groups of adults with dyslexia (N = 23) and without dyslexia (N = 23). The analysis (Section 3) consisted of four parts.

First, to address hypothesis (1), we tested whether rhythmic speech grouping preferences by the two groups (with vs. without dyslexia) differed from chance in the three acoustic conditions (intensity, control, and duration) by means of generalized linear mixed-effects (Section 3.1).

Second, to address hypothesis (2), a linear regression analysis with the MET scores as a dependent variable was performed in order to determine whether group differences existed, while controlling for general cognitive ability, and musical experience (Section 3.2).

Third, to address hypothesis (1) and (3), we tested whether rhythmic grouping preferences differed between groups, and whether it depended on individuals' musical rhythm perception ability. For this, we performed a generalized linear mixed-effects model analysis. In a stepwise fashion, we incrementally increased the models' complexity to understand the effects of the factors group (which we predicted to have an effect) and musical rhythm perception ability (which we predicted to have an effect) on the three conditions (intensity, duration, control), while, ultimately, controlling for general cognitive ability and musical experience. Our method was to compare mixed-effects models that either included or excluded predictors to find the combination of predictors that accounted for most variance in the data, following the recommended procedures [61–63] (Section 3.3).

Fourth, to address hypothesis (4), we assessed the association of musical rhythm perception ability and reading ability in both the group with and the group without dyslexia, while again controlling for musical experience and cognitive abilities. For this, we performed a linear regression analysis with nonword reading ability (i.e., SLRT nonword reading scores) as the dependent variable, and group, musical rhythm perception ability, cognitive ability and musical experience in the fixed part (Section 3.4).

For the control variable "cognitive ability," a composite score was generated that combined the averaged WAIS-IV subtest scores. For the control variable "musical experience", we generated a composite score on the basis of three questions from the questionnaire representing the participants' years of musical training, their age of beginning musical training, and the number of learned musical instruments/activities. Both composite scores were created by means of Principal Component analysis (see Appendix A, Table A1) to avoid collinearity. Collinearity occurs when a number of independent

variables are correlated, which poses a problem to regression analyses. Principal component regression is a commonly used method to reduce collinearity, as it eliminates the dimensions that are causing the collinearity problem [64] (p. 446). Following the classical procedures, we included the first principal components (PCs) as independent factors in our subsequent regression analyses. The first PC reflecting cognitive ability accounted for 58% of the variance contained in the data of the 4 WAIS-IV subtests, which were represented by this PC to a comparable degree (see Appendix A for details). The first PC reflecting musical experience accounted for 82% of the variance of the 3 questions that were equally represented by this variable (see Appendix B, Table A2).

All analyses were performed in R [65] using the package lme4 [66]; graphs were generated using the package ggplot2 ([67]). For plotting modeled data, the package effects [68] was used to extract the model estimates and respective *SE*s.

3. Results

3.1. Rhythmic Grouping Preferences

Tests against chance (see Appendix C, Table A3) revealed that in both the intensity and control condition, trochaic (strong-weak) responses were above chance for both the group with dyslexia (intensity: $p < 0.001$; control: $p = 0.03$) and the group without dyslexia (both p's < 0.001). In the duration condition, both groups gave more iambic (weak-strong) responses than expected by chance (both p's < 0.001, see Figure 1).

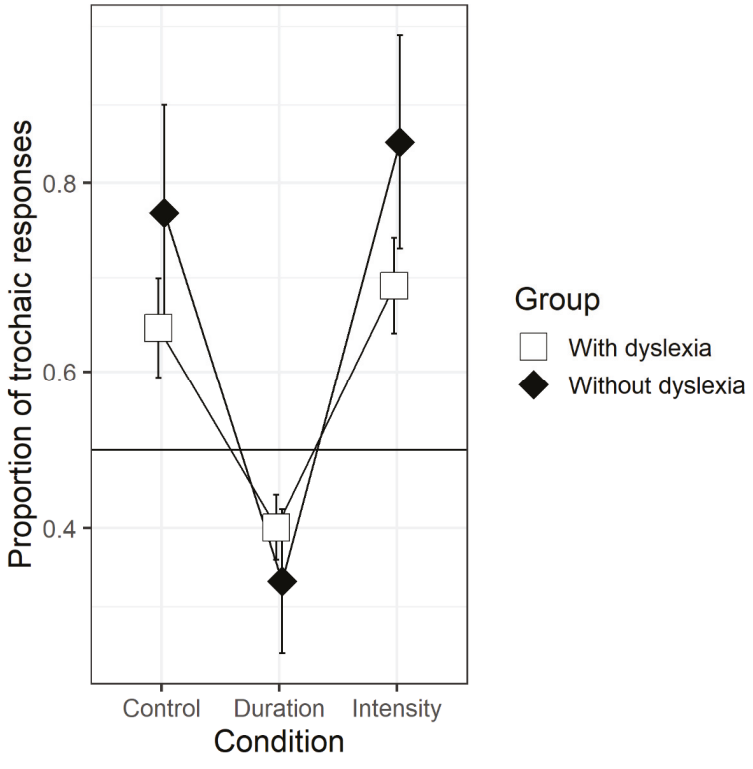


Figure 1. Proportions of trochaic responses (back-transformed, y-axis adjusted to the logit space) in the three acoustic conditions for both groups. The graph reflects the estimates of a simple logit linear mixed-effects model (responses ~ condition * group + (condition + 1||participants) + (1|items) . . .).

3.2. Musical Rhythm Ability

We compared how the two groups (with vs. without dyslexia) fared at the MET for rhythm, while controlling for general cognitive ability and musical experience. Control participants' average rhythm MET scores were higher (73.49% correct, SD = 9.96) than those by participants with dyslexia (61.87% correct, SD = 9.92) with large effect size (Cohen's d = 1.17). Results of a linear regression confirmed that differences between groups were significant ($p = 0.03$). Moreover, rhythm MET scores were significantly predicted by cognitive ability ($p < 0.01$), but not by musical experience ($p = 0.28$, for the full results, see Appendix D, Table A4). Groups did not, however, differ with regards to their musical experience ($\beta = 0.77$, SD = 0.46, t = 1.68, $p = 0.1$). This suggests that dyslexia is associated with reduced musical rhythm perception ability that is independent of musical experience.

3.3. Predictors of Rhythmic Grouping Preferences

Next, we tested whether rhythmic grouping preferences differed in strength between groups and explored the role of individual differences in musical rhythm perception ability, cognitive ability, and musical experience. For this, we report the main results of all models that entered our stepwise regression analysis. To measure the consistency of rhythmic grouping preferences (i.e., how consistent participants were in grouping duration variation as iambs and intensity as trochees), we entered contrasts between conditions into our models. In all models (see Appendix E, Tables A5–A9), significant effects were obtained in the Duration-Intensity contrast and in the Control-Duration contrast (both p's < 0.001), indicating that in both the intensity and control condition, more trochaic responses were given than in the duration condition.

Model 1, serving as a basis, included only the interaction of Condition and Group in the fixed part (Formula: Response ~ Condition/(Group) + (1 + Duration-Intensity + Control-Duration || participant) + (1 | item) to test the hypothesis that adults with versus without dyslexia differ in their rhythmic speech grouping preferences. Model results (fully reported in Table A5 and depicted in Figure 2) show significant group differences in both the intensity ($p = 0.003$) and the control condition ($p = 0.02$), with more trochaic responses by adults without dyslexia than by adults with dyslexia. In the duration condition, no group differences were found.

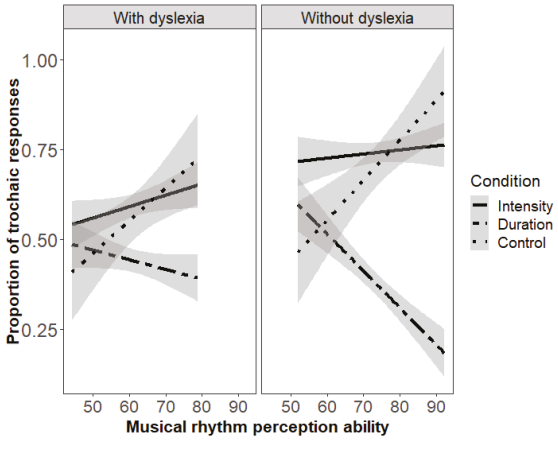


Figure 2. Linear regression lines illustrating the effects of musical rhythm perception ability (Musical Ear Test scores) on rhythmic grouping in the three acoustic conditions separated by group (left panel: with dyslexia, right panel: without dyslexia).

Model 2 included only the interaction of Condition and Musical rhythm perception ability in the fixed part, excluding group (Formula: Response ~ Condition/(Musical rhythm perception ability) +

(1 + Duration-Intensity + Control-Duration || participant) + (1 | item) to evaluate the general contribution of musical rhythm perception ability on rhythmic speech grouping preferences. Model results (see Table A6) show significant effects of musical rhythm perception ability on all conditions: higher musical rhythm perception ability was associated with more trochaic groupings in the intensity (Intensity*Musical rhythm perception ability: $p = 0.04$) and control condition (Control*Musical rhythm perception ability: $p < 0.001$), and more iambic groupings in the duration condition (Duration*Musical rhythm perception ability, $p < 0.001$, see Figure 2). Model comparisons revealed that Model 2 was a better fit than Model 1 ($\chi^2 = 39.53$, $p < 0.001$).

Model 3 included the three-way interaction of Condition, Group and Musical rhythm perception ability in the fixed part (Formula: Response ~ Condition/(Group* Musical rhythm perception ability) + (1 + Duration-Intensity + Control-Duration || participant) + (1 | item) to understand whether group and musical rhythm perception ability predict speech grouping preferences independently. Results (see Table A7 and Figure 2) suggest that this is not the case. Interactions of Group with the Intensity and Control conditions that were present in Model 1 no longer reached significance in Model 3, and the interaction of Duration*Group did not reach significance either. However, the interactions Duration*Musical rhythm perception ability ($p < 0.001$) and Control*Musical rhythm perception ability ($p < 0.001$) that were present in Model 2 remained highly significant in Model 3. This suggests that group differences in the Control condition as attested in Model 1 are due to differences in musical rhythm perception ability between the groups. Moreover, the results suggest that variance in the Duration and Control condition is better captured by differences among individuals' musical rhythm perception ability than by dyslexia status. There were no three-way interactions of any of the conditions with group and musical rhythm perception ability. Model comparisons revealed that Model 3 was a better fit than Model 2 ($\chi^2 = 13.62$, $p < 0.001$).

Two further models tested the potential effects of two control variables: cognitive ability (Model 4, reported in Table A8) and musical experience (Model 5, reported in Table A9). These models revealed the same effects that were also present in Model 3. However, because neither of these control variables significantly influenced participants' grouping preferences in any of the conditions, nor did an inclusion of these factors improve the model fit, we do not discuss these models further (detail and model outputs are provided in Tables A8 and A9).

To summarize, Model 3 (Table A7), which included interactions of condition with group and musical rhythm perception ability, accounted best for the data, which revealed effects of musical rhythm perception ability on the control and duration (but not the intensity) condition, but no significant effects of the group factor.

3.4. Predictors of Nonword Reading Ability

Results of a linear regression analysis (see Appendix F, Table A10 for details) revealed neither effects of cognitive ability nor of musical experience on reading ability (no main effect, no interaction). There was, however, a significant main effect of group, indicating that—as expected—the group without dyslexia had higher reading ability than the group with dyslexia ($p < 0.001$). Moreover, there was a marginal interaction of musical rhythm perception ability and group ($p = 0.056$, Cohen's $f^2 = 0.10$ (medium)). To understand the interaction, we tested the effect of musical rhythm perception ability on reading ability per group. Results were that musical rhythm perception ability positively predicted reading ability by individuals with dyslexia ($\beta = 97.77$, SE = 35.76, t = 2.73, $p = 0.01$, Cohen's $f^2 = 0.36$ (large)) but not by individuals without dyslexia ($\beta = -11.693$, SE = 48.25, t = −0.24, $p = 0.81$; see Figure 3).

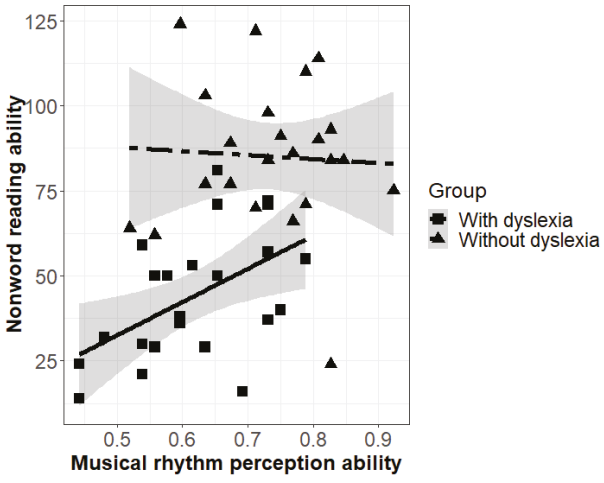


Figure 3. Linear regression lines reflecting the association between nonword reading ability (Salzburger Lese- und Rechtschreibtests (SLRT) scores) and musical rhythm perception ability (Musical Ear Test (MET) scores) split by group, shades indicate confidence intervals, rectangles (with dyslexia) and triangles (without dyslexia) the individuals' averages.

4. Discussion

The present study is based on the theory that there is a deficit in rhythm processing in dyslexia [1–4], which we studied by exploring the modulating effects of musical rhythm perception ability on rhythmic speech grouping by adults with and without dyslexia. Populations with dyslexia have previously been demonstrated to have difficulties with processing stress and rhythm information [12–14], which suggests that the phonological deficit affects not only segmental but also suprasegmental aspects of speech. Hence, we investigated whether adults with dyslexia have reduced abilities in rhythmic grouping of speech, an ability that has previously been found to depend on native language phonological knowledge [38,47,48]. Rhythm, however, is not only part of phonology, but is also an integral aspect of other auditory domains, such as music. Previous studies have established that there are links between individuals' musical abilities and their rhythm processing [52]. Hence, we hypothesized to find links between rhythm processing in speech and music and reading ability in dyslexia.

Specifically, our research was intended to provide answers to the following questions, and delivered the following central findings:

(1) Do adults with dyslexia have less consistent rhythmic speech grouping preferences than adults without dyslexia? Our results indicate that this is not the case, as adults with dyslexia show clear rhythmic speech grouping preferences (Section 3.1), and rhythmic grouping preferences are not different between adults with versus without dyslexia (Section 3.3).

(2) Do adults with dyslexia have lower musical rhythm perception abilities than adults without dyslexia? Our results (Section 3.2) suggest that this is the case.

(3) Does musical rhythm perception ability predict rhythmic speech grouping in dyslexia? Our results suggest that this is the case: Musical rhythm ability predicted rhythmic grouping preferences (Section 3.3).

(4) Does musical rhythm perception ability predict reading ability in dyslexia? Our results suggest that this is the case: We found that musical rhythm perception ability predicts reading ability in dyslexia (Section 3.4).

The results are discussed below.

4.1. Dyslexia, Rhythmic Grouping Preferences, and Musical Rhythm Ability

First, regarding the link between dyslexia and rhythmic speech grouping, results revealed significant preferences for groupings as predicted by the ITL in all conditions (iambic in the duration condition, trochaic in both the control and intensity condition), by native speakers of German with and without dyslexia. This result was unexpected. Our original hypothesis was that rhythmic grouping preferences would be weakened in dyslexia. This hypothesis was motivated by results from prior studies showing that individuals with dyslexia have weakened stress perception abilities, e.g., [12–14]. In prior studies, we found that native speakers of French had weakened rhythmic grouping preferences compared to native speakers of German—a result argued to relate to differences in the phonological systems of German and French (due to the lack of contrastive lexical stress in the French language). Since French speakers have, moreover, repeatedly been found to have weakened stress perception abilities, e.g., [69,70], and the same is true for individuals with dyslexia, e.g., [12–14], we drew a parallel. Unexpectedly, results attested that German speakers with dyslexia show consistent grouping preferences at the group level, just like German speakers without dyslexia. This result replicates our previous findings with native speakers of German without dyslexia and extends them to native speakers of German with dyslexia.

It is important to further explore why listeners with dyslexia generally show the same pattern of responses as those without in the rhythmic speech grouping task: At first glance, this conflicts with the assumption that individuals with dyslexia predominantly have a deficit in rhythm processing. However, the present results can be better understood by considering the results of the model comparisons that addressed the fourth research question about the association of dyslexia, rhythmic speech perception, and musical rhythm perception ability. Model 1 (Table A5, the baseline model that excluded the musical rhythm perception ability factor), suggested a detrimental impact of dyslexia on rhythmic speech perception in both the intensity and control condition. This is in line with prior studies that suggested links between speech rhythm perception and dyslexia, e.g., [12–14].

Importantly, these effects disappeared when, in Model 3 (Table A7, as well as Models 4 and 5 (Tables A8 and A9) that additionally controlled for cognitive ability and musical experience), musical rhythm perception ability was added as a predictor. This suggests that differences in individuals' musical rhythm perception ability better capture the variance in the data than the individuals' dyslexia status (i.e., there are individuals with dyslexia who have high musical rhythm perception ability with consistent grouping preferences, and individuals without dyslexia with low musical rhythm perception ability with inconsistent grouping preferences). However, even though this suggests that rhythmic speech perception is independent of dyslexia and only modulated by musical rhythm perception, adults with dyslexia had overall lower musical rhythm perception ability than adults without dyslexia, which implies an indirect effect of group on rhythmic speech perception.

Notably, as predicted, musical rhythm perception ability accounted for variance in the rhythmic grouping of duration-varied and rhythmically invariant control speech sequences, but, in contrast with our predictions, not of intensity-varied speech sequences. These findings suggest that the relation between musical rhythm perception ability and speech rhythm processing cannot be simply explained by a general deficit in perceiving the acoustic information that is relevant for perceiving rhythm (otherwise the processing of intensity-varied speech sequences should also be related to musical rhythm perception ability). The results of the control condition, in which acoustic cues to rhythm were absent, suggest that the relation between musical rhythm perception ability and speech rhythm processing is established via phonological knowledge. In previous studies [38], it was proposed that German listeners might perceive trochees even in the absence of acoustic cues to rhythm, because German has trochaic metrical stress and their abstract knowledge about this phonological property of their language affects German listener's perception (as commonly seen also in sound perception which is also affected by native language phonemic categories). Ref. [52] also observed that musical rhythm perception ability was associated with this default grouping procedure, and it was speculated whether individual differences in basic auditory perception abilities might lead to differences in how listeners

establish phonological knowledge. The fact that the present study replicates the effect of the connection between musical rhythm perception ability and default trochaic groupings with adults with dyslexia is interesting, as it offers additional support for the interpretation that that default rhythm perception procedures are subject to individual variation and are associated with more general auditory rhythm perception abilities.

In order to explain why musical rhythm perception ability, contrary to our predictions, did not affect grouping of intensity-varied sequences by adults with dyslexia, we consider previous studies on the ITL. Infants have been found to use pitch and intensity cues for trochaic groupings (pitch: [37,71], intensity: [72]) more readily than duration cues for iambic groupings [37,71,72]. Based on these findings it has been proposed that the use of duration cues for grouping is acquired, while the use of other rhythmic cues for trochaic groupings is innate (more evidence for this proposal comes from studies with rats [73,74] but c.f. [75,76] for evidence that the use of duration for iambic groupings is also innate). The present finding that intensity-based grouping is unmodulated by musical rhythm perception ability is consistent with the assumption of an innate preference for trochaic groupings when intensity alternations are perceived. Speculatively, this might suggest that in dyslexia, perception of innately biased speech processing routines is unimpaired. It would be interesting if future studies followed up on this.

4.2. Dyslexia and Musical Rhythm Ability

Regarding the associations among dyslexia, reading ability and musical rhythm perception ability, results were as predicted: First, adults with versus without dyslexia differed in their musical rhythm perception ability: as a group, adults with dyslexia showed a lower performance in the MET rhythm subtest than the control group. This result is in line with previous findings indicating deficits in musical rhythm processing abilities in dyslexia [25–27]. However, it must be noted that performance in the MET is associated with short-term memory performance [58]. This is in line with literature that has found short-term memory ability to be enhanced in musical people [77–79]. However, it is also well-known that groups with dyslexia have less efficient short-term memory abilities than groups without dyslexia (see Table 1, it is also true for the present sample: WAIS, Working Memory: digit span: with dyslexia: 24 (15–38); without dyslexia: 28.86 (22–37)), and it is debated whether this reduced short-term memory efficiency is the basis of the impairment or is an effect of the phonological deficit [3,80]. In fact, we found that musical rhythm ability was predicted by cognitive ability, a composite variable that included the participants' digit span scores. In order to better understand if musical rhythm ability is lower in dyslexia independently of related cognitive abilities, future studies should aim to control for this confounding factor.

4.3. Dyslexia, Nonword Reading Ability, and Musical Rhythm Ability

We tested whether reading ability was predicted by group and by musical rhythm perception ability. As expected, the groups differed in their reading ability. Interestingly, we furthermore found a marginally significant interaction ($p = 0.056$) of group and musical rhythm perception ability to account for reading ability. We explored this interaction (although results based on this insignificant interaction have to be taken with care), and found that, in particular, reading ability of adults with dyslexia was predicted by musical rhythm perception ability: the lower their performance in the MET rhythm subtest, the lower was their score in the nonword reading test. This finding is consistent with prior studies, which found links between musical rhythm perception ability and reading ability (e.g., [81] and references therein). Both these results support theories of links between general rhythm processing abilities and dyslexia and, accordingly, with reading ability.

Note that the lack of a relation between musical rhythm perception ability and reading ability in adults without dyslexia does not justify the conclusion that this association is exclusive to dyslexia. Potentially, a link between musical rhythm perception ability and reading ability could also be found if a reading test were used with adults without dyslexia that elicits greater variability in this groups' reading ability than the SLRT, a test particularly designed for identifying dyslexia in

adulthood. Moreover, the relation between musical rhythm perception ability and reading ability may be non-linear, with ceiling effects of musical rhythm perception ability at a certain level of high reading ability that adults without dyslexia typically reach. It will be interesting to address these questions in future research.

5. Conclusions

In sum, the main findings of the present study are the following: First, rhythmic grouping of speech is not predicted by dyslexia status, but by musical rhythm ability. That is, the present study does not provide direct evidence for the theory that there is a specific speech rhythm processing deficit in dyslexia. However, the fact that we found the group of adults with dyslexia to show lower musical rhythm perception ability than the group of adults without dyslexia, and that musical rhythm ability predicted speech rhythm grouping indicates that there is a link between rhythm processing in music and speech and dyslexia. Second, musical rhythmic skills predict reading in dyslexia. The results suggest clear links between dyslexia (i.e., reading ability), musical rhythm perception ability, and speech rhythm processing, not only when rhythmic cues are available but also when a lack of cues triggers knowledge-driven default processing routines. All in all, the results point to individual differences in the group of adults with dyslexia that are explained by their musical rhythm perception ability.

The present findings cannot inform about causal relationships between musical rhythm perception ability and dyslexia. However, they raise the possibility that rhythm perception ability is a key to phonological and reading acquisition. The present results are in line with two assumptions about the underlying reasons for these links. The first assumption is that deficits connected with dyslexia can be compensated by rhythm perception ability. The second assumption is that the deficits connected with dyslexia are a consequence of lower rhythm perception ability. That is, potentially, individuals with lower rhythm perception ability have a higher risk for developing phonological and reading deficits.

Future studies should address the question of how musical rhythm perception ability, speech rhythm perception, and reading are causally connected. To explore the first assumption, studies should assess the potential of rhythmic interventions in dyslexia therapy, and therewith, follow a line of research that has already been initiated, e.g., [82,83]. Ideally, future research should explore pre-/post-test paradigms to explore whether musical rhythm perception ability can be enhanced by training. This can then be extended to other types of rhythmic behavior, such as motor synchronization (e.g., tapping) with rhythmic beats, to pave the way for targeted rhythm-based therapeutic approaches. To explore the second assumption, future research should conduct longitudinal studies with very young infants with a familial risk for dyslexia (for a similar suggestion, see [84]), to pave the way for our understanding of whether the ability to perceive rhythm in music (and other sensory domains) is a reliable early marker of developmental dyslexia.

Author Contributions: Conceptualization, N.B.-A., A.B. and B.H.; methodology, N.B.-A., A.B. and B.H.; analysis, N.B.-A.; investigation, N.B.-A.; data curation, N.B.-A.; writing—original draft preparation, N.B.-A. and B.H.; writing—review and editing, N.B.-A., A.B. and B.H.; visualization, N.B.-A.; project administration, N.B.-A.; funding acquisition, B.H. All authors have read and agree to the published version of the manuscript.

Funding: This research was funded by two Agence Nationale de la Recherche-Deutsche Forschungsgemeinschaft grants (# 09-FASHS-018 and HO 1960/14–1 to Barbara Höhle and Thierry Nazzi, and HO 1960/15–1 and ANR-13-FRAL-0010 to Ranka Bijeljac-Babic and Barbara Höhle), and partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–project number 317633480–SFB 1287, Project C03.

Acknowledgments: We thank Sophie Gruhn, Franz Hildebrandt-Harangozó, Olivia Malotka, and Isabelle Mackuth for help with recruiting and testing participants, and Daniel Schad for consultancy regarding the statistical analysis. We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Principal Component Analysis was used to generate one composite score to reflect general cognitive ability on the basis of four tests from the WAIS. In the analyses, we used the first Principal Component as a factor to represent cognitive ability, which, as can be seen from the loadings in Table A1, represented all four variables to a comparable degree, and captured a proportion of 0.58 of their variance.

Table A1. Results of the Principal Component Analysis over the data of four variables relating to general cognitive ability.

Principal Component Analysis	Comp.1	Comp.2	Comp.3	Comp.4
Importance of Components				
Standard deviation	1.53	0.87	0.85	0.43
Proportion of Variance	0.58	0.19	0.18	0.05
Loadings				
Similarities (verbal comprehension)	0.45	0.57	0.60	0.33
Digit span (short-term memory)	0.40	−0.76	0.49	−0.13
Symbol search (processing speed)	0.58	0.26	−0.29	−0.72
Coding (processing speed)	0.54	−0.19	−0.57	0.60

Appendix B

Principal Component Analysis was used to generate one composite score to reflect musical experience on the basis of three questions from the questionnaire regarding the number of acquired musical instruments/activities, the earliest age of acquiring a musical instrument/activity, and the duration of musical training in years. In the analyses, we used the first Principal Component as a factor to represent musical experience, which, as can be seen from the loadings in Table A2, represented all three variables to a comparable degree, and captured a proportion of 0.82 of their variance.

Table A2. Results of the Principal Component Analysis over the data of three variables relating to general musical experience.

Principal Component Analysis	Comp.1	Comp.2	Comp.3
Importance of Components			
Standard deviation	1.57	0.52	0.51
Proportion of Variance	0.82	0.09	0.09
Loadings			
Duration of musical training	0.58	0.81	0.12
Number of musical instruments/activities	0.58	−0.50	0.64
Age of musical instrument/activity acquisition	0.58	−0.30	−0.76

Appendix C

For each group, we calculated a generalized linear mixed effects model (under use of the bobyqa optimizer) with condition as fixed factor, participants and items as random factors, but no random slopes because of false convergence. The intercept was set to zero, so that each of the three acoustic conditions (intensity, control, and duration) was compared to chance. Results are provided in Table A3.

Table A3. Results of two generalized mixed effects models (one for the group of adults with dyslexia, and one for the group of adults without dyslexia) to test the groups' preferences against chance in the three acoustic conditions reported in Section 3.3.

Fixed Effects	***β***	***SE***	***z***	***p***	***Sig.***
With Dyslexia					
Intensity	0.41	0.07	5.71	<0.001	***
Control	0.31	0.14	2.14	0.03	*
Duration	−0.25	0.07	−3.58	<0.001	***
Without Dyslexia					
Intensity	1.12	0.12	9.24	<0.001	***
Control	0.92	0.18	5.16	<0.001	***
Duration	−0.53	0.11	−4.68	<0.001 [1]	***

[1] Formula: Response ~ −1 + Condition + (1 | participant) + (1 | item). Each line shows the coefficients of the intercept of each of the separate models. Negative β estimates indicate more iambic responses, and positive β estimates indicate more trochaic responses. Level of significance: * $p < 0.05$, *** $p < 0.001$.

Appendix D

The results of the linear regression analysis for testing whether musical rhythm perception ability (MET scores) is predicted by dyslexia (group factor), cognitive ability (first principal component of the WAIS scores) and musical experience (first principal component combining the number of learned musical instruments/activities, age of musical acquisition, and duration of musical training) are provided in Table A4.

Table A4. Parameters of the regression analysis of the effects of dyslexia, musical experience and cognitive ability on musical rhythm perception ability.

Effects	***β***	***SE***	***T***	***p***	***Sign.***
Intercept	−0.01	0.02	−0.65	0.52	
Group	0.07	0.03	2.28	0.03	*
Musical experience	−0.01	0.01	−1.10	0.28	
Cognitive ability	0.03	0.01	3.27	0.002	**
Group* Musical rhythm perception ability	0.02	0.02	1.44	0.16	
Group*Cognitive ability	0.01	0.02	0.63	0.53 [1]	

[1] Formula: lm(Musical rhythm perception ability ~ Group*Musical experience + Group*Cognitive ability). Level of significance: * $p < 0.05$, ** $p < 0.01$.

Appendix E

For each group, we calculated a generalized linear mixed effects model (under use of the bobyqa optimizer) with condition as fixed factor, participants and items as random factors, but no random slopes because of false convergence. The intercept was set to zero, so that each of the three acoustic conditions (intensity, control, and duration) was compared to chance. Results are provided in Table A10. To test the effects of the factors group and musical rhythm perception ability on the three conditions, while controlling for cognitive ability and musical experience, generalized linear mixed-effects models were built that incrementally increased the number of predictors in a stepwise fashion. Model fits were compared by means of their loglikelihood using the anova() function from the LME4 package ([57]). Successive difference contrast coding was used for comparing groups and conditions. This contrast (coded with the contr. sdif() function from the MASS package, [85]) assigns the grand mean to the intercept, and beta coefficients indicate the difference scores between two compared levels. In the case of group, this contrast was specified as subtracting the group with dyslexia from the group without dyslexia. For condition, the contrasts were Duration-Intensity (β reflecting duration minus intensity) and Control-Duration (control minus duration). Both continuous predictors were centered around their mean to reduce collinearity, and z-transformed (using the scale() function) as models

with untransformed predictors did not converge (mathematically, z-transformation does not affect the results). The models were coded containing a fraction, with condition being the numerator, and the predictors group, musical rhythm perception ability, and/or cognitive ability being the denominator. By this, it is possible to assess the effects of the predictors group, musical rhythm perception ability and cognitive ability on each of the conditions in separation.

Random intercept for participants and items, and random slopes for the condition contrasts by participants were included. Correlations of the random slopes by participants were subtracted (by ||), as not all reported models would converge when including them and comparisons suggested that they did not significantly account for variance. Models including random slopes for cognitive ability, group, and/or musical rhythm perception ability by item did not improve the model fits.

Parameters of the tested models are reported in the tables below. To correct for multiple comparisons, adjusted *p*-values (Bonferroni correction) are reported in the last column (model 2, *p*-values multiplied by 2, model 3: *p*-values multiplied by 3, etc.). Level of significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5. Output of the first model exploring the effects of group on the three conditions.

Random Effects	**Groups**	**Name**		**Variance**	**Std.Dev.**
	Item	(Intercept)		0.15	0.38
	Participant	Control-Duration		1.07	1.03
	participant.1	Duration-Intensity		0.09	0.30
	participant.2	(Intercept)		0.00	0.00
Fixed Effects	***β***	***SE***	***z***	***p***	***Sign.***
(Intercept)	0.36	0.06	5.65	<0.001	***
Duration-Intensity	−1.26	0.17	−7.27	<0.001	***
Control-Duration	1.04	0.13	8.29	<0.001	***
Intensity*Group	0.75	0.25	3.00	0.003	**
Duration*Group	−0.28	0.17	−1.66	0.10	
Control*Group	0.60	0.25	2.42	0.02	* [1]

[1] Formula: *Response ~ Condition/(Group)* + (1 + *condL2v1* + *condL3v2* || *participant*) + (1 | *item*). Level of significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A6. Output of the second model exploring the effects of musical rhythm perception ability on the three conditions.

Random Effects	**Groups**	**Name**		**Variance**		**Std.Dev.**
	Item	(Intercept)		0.14		0.37
	Participant	Control-Duration		1.10		1.05
	participant.1	Duration-Intensity		0.10		0.32
	participant.2	(Intercept)		0.00		0.01
Fixed Effects	***β***	***SE***	***z***	***p***	***Sign.***	**Corrected *p***
(Intercept)	0.37	0.07	5.54	<0.001	***	<0.001
Duration-Intensity	−1.26	0.17	−7.21	<0.001	***	<0.001
Control-Duration	1.06	0.13	8.37	<0.001	***	<0.001
Intensity*Musical rhythm perception ability	0.29	0.13	2.34	0.02	*	0.04
Duration*Musical rhythm perception ability	−0.32	0.09	−3.72	<0.001	***	<0.001
Control*Musical rhythm perception ability	0.55	0.13	4.27	<0.001	***	<0.001 [1]

[1] Formula: *Response ~ Condition/(Musical rhythm perception ability)* + (1 + *Duration-Intensity* + *Control-Duration* || *participant*) + (1 | *item*). Level of significance: * $p < 0.05$, *** $p < 0.001$.

Table A7. Output of the third model exploring the effects of group and musical rhythm perception ability on the three conditions.

Random Effects	**Groups**		**Name**		**Variance**	**Std.Dev.**
	item		(Intercept)		0.14	0.38
	participant		Control-Duration		1.02	1.01
	participant.1		Duration-Intensity		0.01	0.11
	participant.2		(Intercept)		0.07	0.26
Fixed Effects	***β***	***SE***	***z***	***p***	***Sign.***	***Corr. p***
(Intercept)	0.40	0.07	5.53	<0.001	***	<0.001
Duration-Intensity	−1.20	0.20	−6.08	<0.001	***	<0.001
Control-Duration	0.91	0.14	6.41	<0.001	***	<0.001
Intensity*Group	0.60	0.28	2.12	0.03	*	0.09
Duration*Group	0.05	0.19	0.27	0.78		
Control*Group	0.07	0.28	0.24	0.81		
Intensity*Musical rhythm perception ability	0.14	0.14	0.97	0.33		
Duration* Musical rhythm perception ability	−0.34	0.10	−3.53	<0.001	***	<0.001
Control* Musical rhythm perception ability	0.55	0.15	3.74	<0.001	***	<0.001
Intensity*Group* Musical rhythm perception ability	−0.14	0.28	−0.52	0.61		
Duration*Group* Musical rhythm perception ability	−0.40	0.19	−2.09	0.04	*	0.12
Control*Group* Musical rhythm perception ability	0.20	0.29	0.67	0.50 [1]		

[1] Formula: *Response ~ Condition/*(*Group * Musical rhythm perception ability*) + (1 + *Duration-Intensity* + *Control-Duration* || *participant*) + (1| *item*).Level of significance: * $p < 0.05$, *** $p < 0.001$.

Table A8. Output of the fourth model that extends model 3 by cognitive ability as control variable, which did not improve the model fit.

Random Effects	**Groups**		**Name**		**Variance**	**Std.Dev.**
	Item		(Intercept)		0.15	0.38
	Participant		Control-Duration		0.91	0.96
	participant.1		Duration-Intensity		0.07	0.26
	participant.2		(Intercept)		0.01	0.08
Fixed Effects	***β***	***SE***	***z***	***p***	***Sign.***	***Corr. p***
(Intercept)	0.36	0.07	4.93	<0.001	***	<0.001
Duration-Intensity	−1.13	0.2	−5.77	<0.001	***	<0.001
Control-Duration	0.89	0.15	5.98	<0.001	***	<0.001
Intensity*Group	0.44	0.28	1.58	0.12		
Duration*Group	0.12	0.19	0.61	0.54		
Control*Group	0.02	0.29	0.07	0.94		
Intensity* Musical rhythm perception ability	0	0.15	0.01	0.99		
Duration* Musical rhythm perception ability	−0.29	0.1	−2.80	0.005	**	0.02
Control* Musical rhythm perception ability	0.49	0.16	2.97	0.003	**	0.01
Intensity*Cognitive ability	0.31	0.15	2.03	0.04	*	0.16
Duration*Cognitive ability	−0.12	0.11	−1.12	0.26		
Control*Cognitive ability	0.11	0.17	0.64	0.52		
Intensity*Group* Musical rhythm perception ability	−0.37	0.3	−1.22	0.22		
Duration*Group* Musical rhythm perception ability	−0.41	0.2	−2.01	0.04	*	0.16
Control*Group* Musical rhythm perception ability	0.07	0.33	0.21	0.84		
Intensity*Group*Cognitive ability	0.52	0.31	1.68	0.09		
Duration*Group*Cognitive ability	0.04	0.21	0.18	0.86		
Control*Group*Cognitive ability	0.24	0.34	0.73	0.47 [1]		

[1] Table A8 reports the results of Model 4, which extended Model 3 by adding cognitive ability as control variable (Formula: *Response ~ Condition/*(*Group*Musical rhythm perception ability + Cognitive ability*) + (1 + *Duration-Intensity* + *Control-Duration* || *participant*) + (1 | *item*). Level of significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Not different from Model 3, the results of Model 4 suggest highly significant effects of musical rhythm perception ability on the duration ($p = 0.02$) and the control condition ($p = 0.008$). Intensity was, just as in Model 3, not modulated by musical rhythm perception ability, and there were, again, no interactions of any condition and group, and neither any three-way interactions of any condition with group and musical rhythm perception ability. Altogether, there were no significant effects of cognitive ability. Model comparisons revealed that Model 4 was not better than Model 3 ($\chi^2 = 8.05$, $p < 0.23$).

Table A9. Output of the fifth model that extends model 3 by musical experience as control variable, which did not improve the model fit.

Random Effects	**Groups**	**Name**			**Variance**	**Std.Dev.**
	item	(Intercept)			0.14	0.38
	participant	Control-Duration			0.98	0.99
	participant.1	Duration-Intensity			0.06	0.25
	participant.2	(Intercept)			0.01	0.05
Fixed Effects	***β***	***SE***	***z***	***p***	***Sign.***	***Corr. p***
(Intercept)	0.38	0.07	5.39	<0.001	***	<0.001
Duration-Intensity	−1.14	0.2	−5.71	<0.001	***	<0.001
Control-Duration	0.85	0.15	5.76	<0.001	***	<0.001
Intensity*Group	0.62	0.28	2.19	0.02	*	0.1
Duration*Group	−0.06	0.19	−0.32	0.74		
Control*Group	0.04	0.29	0.17	0.87		
Intensity*Musical rhythm perception ability	0.12	0.14	0.88	0.38		
Duration*Musical rhythm perception ability	−0.32	0.09	−3.44	<0.001	***	0.003
Control*Musical rhythm perception ability	0.53	0.15	3.65	<0.001	***	0.001
Intensity*Musical experience	−0.03	0.13	0.23	0.81		
Duration*Musical experience	0.2	0.09	2.38	0.02	*	0.09
Control*Musical experience	0.09	0.13	0.73	0.46		
Intensity*Group*Rhythm perception ability	−0.13	0.28	−0.46	0.64		
Duration*Group*Rhythm perception ability	−0.46	0.19	−2.49	0.01	*	0.06
Control*Group*Rhythm perception ability	0.19	0.29	0.69	0.49		
Intensity*Group*Musical experience	0.14	0.25	0.56	0.58		
Duration*Group*Musical experience	−0.07	0.17	−0.43	0.66		
Control*Group*Musical experience	0.33	0.26	1.26	0.21 [1]		

[1] Formula: *Response ~ Condition/(Group * Muscial rhythm perception ability + Musical experience)* + (1 + *Duration-Intensity* + *Control-Duration* || *participant*) + (1 | *item*). Level of significance: * $p < 0.05$, *** $p < 0.001$.Table A9 reports the results of Model 5, which extended Model 3 by adding musical experience as control variable (Formula: *Response ~ Condition/(Group*Musical rhythm perception ability + Musical experience)* + (1 + *Duration-Intensity* + *Control-Duration* || *participant*) + (1 | *item*). The results of Model 5 match those of Model 3, suggesting the same highly significant effects of musical rhythm perception ability on the duration ($p = 0.003$) and the control condition ($p = 0.001$). Altogether, there were no significant effects of musical experience. Model comparisons revealed that Model 5 was not better than Model 3 ($\chi^2 = 8.82$, $p < 0.18$).

Appendix F

The results of the linear regression analysis for testing the effects of dyslexia (group factor), musical rhythm perception ability (MET scores), and cognitive ability (First principal component of the WAIS scores) on reading ability (SLRT nonword reading scores) are provided in Table A10.

Table A10. Parameters of the regression analysis of the effects of dyslexia, musical rhythm perception ability and cognitive ability on reading ability.

Effects	***β***	***SE***	***T***	***p***	***Sign.***
Intercept	1.84	3.58	0.52	0.61	
Group	31.48	7.15	4.40	<0.001	***
Musical rhythm perception ability	18.91	33.50	0.56	0.58	
Cognitive ability	3.75	2.55	1.47	0.15	
Musical experience	2.21	2.05	1.08	0.29	
Group*Musical rhythm perception ability	−132.09	66.99	−1.98	0.056	
Group*Cognitive ability	2.89	5.11	0.57	0.58	
Group*Musical experience	4.72	4.11	1.15	0.26 [1]	

[1] Formula: lm(Reading ability ~ Group*Musical rhythm perception ability + Group*Cognitive ability+Group*Musical experience). Level of significance: *** $p < 0.001$.

References

1. Bishop, D.V.M.; Snowling, M.J. Developmental dyslexia and specific language impairment: Same or different? *Psychol. Bull.* **2004**, *130*, 858–886. [CrossRef]
2. Ramus, F. Developmental dyslexia: Specific phonological deficit or general sen-Sorimotor dysfunction? *Curr. Opin. Neurobiol.* **2003**, *13*, 212–218. [CrossRef]
3. Snowling, M.J. *Dyslexia*, 2nd ed.; Blackwell: Oxford, UK, 2000.
4. Snowling, M.J. From language to reading and dyslexia. *Dyslexia* **2001**, *7*, 37–46. [CrossRef]
5. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef]
6. Hulme, C.; Snowling, M.J. Reading disorders and dyslexia. *Curr. Opin. Pediatr.* **2016**, *28*, 731. [CrossRef]
7. Noordenbos, M.W.; Serniclaes, W. The categorical perception deficit in dyslexia: A meta-Analysis. *Sci. Stud. Read.* **2015**, *19*, 340–359. [CrossRef]
8. Best, C.T. A direct realist perspective on cross-Language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*; Strange, W., Ed.; York Press: Timonium, MD, USA, 1995; pp. 171–204.
9. Hayes, B. *Metrical Stress Theory: Principles and Case Studies*; The University of Chicago Press: Chicago, IL, USA; London, UK, 1995.
10. Liberman, M.; Prince, A. On stress and linguistic rhythm. *Linguist. Inq.* **1977**, *8*, 249–336.
11. Selkirk, E.O. *Phonology and Syntax: The Relation between Sound and Structure*; MIT Press: Cambridge, MA, USA, 1984.
12. Goswami, U.; Gerson, D.; Astruc, L. Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Read. Writ.* **2010**, *23*, 995–1019. [CrossRef]
13. Goswami, U.; Mead, N.; Fosker, T.; Huss, M.; Barnes, L.; Leong, V. Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *J. Mem. Lang.* **2013**, *69*, 1–17. [CrossRef]
14. Leong, V.; Hämäläinen, J.; Soltész, F.; Goswami, U. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *J. Mem. Lang.* **2011**, *64*, 59–73. [CrossRef]
15. De Bree, E.; Van Alphen, P.M.; Fikkert, P.; Wijnen, F. Metrical stress in comprehension and production of Dutch children at risk of dyslexia. In *32nd Annual Boston University Conference on Language Development*; Cascadilla Press: Somerville, MA, USA, 2008; pp. 60–71.
16. Holliman, A.J.; Wood, C.; Sheehy, K. Sensitivity to speech rhythm explains individual differences in reading ability independently of phonological awareness. *Br. J. Dev. Psychol.* **2008**, *26*, 357–367. [CrossRef]
17. Wood, C. Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *J. Res. Read.* **2006**, *29*, 270–287. [CrossRef]
18. Wood, C.; Terrell, C. Poor readers' ability to detect speech rhythm and perceive rapid speech. *Br. J. Dev. Psychol.* **1998**, *16*, 397–413. [CrossRef]
19. Goswami, U. A neural oscillations perspective on phonological development and phonological processing in developmental dyslexia. *Lang. Linguist. Compass* **2019**, *13*, e12328. [CrossRef]
20. Hämäläinen, J.A.; Salminen, H.K.; Leppänen, P.H. Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/field evidence. *J. Learn. Disabil.* **2013**, *46*, 413–427. [CrossRef]
21. Power, A.J.; Colling, L.C.; Mead, N.; Barnes, L.; Goswami, U. Neural encoding of the speech envelopeby children with developmental dyslexia. *Brain Lang.* **2016**, *160*, 1–10. [CrossRef]
22. Power, A.J.; Mead, N.; Barnes, L.; Goswami, U. Neural entrainment to rhythmically-Presented auditory, visual and audio-Visual speech in children. *Front. Psychol.* **2012**, *3*. [CrossRef]
23. Molinaro, N.; Lizarazu, M.; Lallier, M.; Bourguignon, M.; Carreiras, M. Out-Of-Synchrony speechentrainment in developmental dyslexia. *Hum. Brain Mapp.* **2016**, *37*, 2767–2783. [CrossRef]
24. Power, A.J.; Mead, N.; Barnes, L.; Goswami, U. Neural entrainment to rhythmic speech in childrenwith developmental dyslexia. *Front. Hum. Neurosci.* **2013**, *7*. [CrossRef]
25. Goswami, U.; Huss, M.; Mead, N.; Fosker, T.; Verney, J.P. Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex* **2013**, *49*, 1363–1376. [CrossRef]

26. Holliman, A.J.; Wood, C.; Sheehy, K. The contribution of sensitivity to speech rhythm and non-Speech rhythm to early reading development. *Educ. Psychol.* **2010**, *30*, 247–267. [CrossRef]
27. Huss, M.; Verney, J.P.; Fosker, T.; Mead, N.; Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* **2011**, *47*, 674–689. [CrossRef] [PubMed]
28. Colling, L.J.; Noble, H.L.; Goswami, U. Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Front. Neurosci.* **2017**, *11*. [CrossRef] [PubMed]
29. Thomson, J.M.; Fryer, B.; Maltby, J.; Goswami, U. Auditory and motor rhythm awareness in adults with dyslexia. *J. Res. Read.* **2006**, *29*, 334–338. [CrossRef]
30. Thomson, J.M.; Goswami, U. Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *J. Physiol. Paris* **2008**, *102*, 120–129. [CrossRef] [PubMed]
31. Beckman, M.E.; Pierrehumbert, J. Intonational structure in Japanese and English. *Phonol. Yearb.* **1986**, *3*, 255–309. [CrossRef]
32. Klatt, D.H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* **1976**, *59*, 1208–1221. [CrossRef]
33. Bolton, T.L. Rhythm. *Am. J. Psychol.* **1894**, *6*, 145–238. [CrossRef]
34. Woodrow, H. A quantitative study of rhythm: The effect of variations in intensity, rate, and duration. *Arch Psychol.* **1909**, *14*, 1–66.
35. Woodrow, H. Time perception. In *Handbook of Experimental Psychology*; Stevens, S., Ed.; Wiley: Oxford, UK, 1951; pp. 1224–1236.
36. Rice, C.C. Binarity and Ternarity in Metrical Theory: Parametric Extensions. Ph.D. Thesis, University of Texas at Austin, Austin, TX, USA, 1992.
37. Bion, R.A.H.; Benavides, S.; Nespor, M. Acoustic markers of prominence influence adults' and infants' memory of speech sequences. *Lang. Speech* **2011**, *54*, 123–140. [CrossRef]
38. Bhatara, A.; Boll-Avetisyan, N.; Unger, A.; Nazzi, T.; Höhle, B. Native language affects rhythmic grouping of speech. *J. Acoust. Soc. Am.* **2013**, *134*, 3828–3843. [CrossRef] [PubMed]
39. Lerdahl, F.; Jackendoff, R. *Generative Theory of Tonal Music*; MIT Press: Cambridge, UK, 1983.
40. Narmour, E. *The Analysis and Cognition of Basic Melodic Structures*; University of Chicago Press: Chicago, IL, USA, 1990.
41. Todd, N. A model of expressive timing in tonal music. *Music Percept* **1985**, *3*, 33–58. [CrossRef]
42. Patel, A.D.; Iversen, J.R. The linguistic benefits of musical abilities. *Trends Cogn. Sci.* **2007**, *11*, 369–372. [CrossRef]
43. Patel, A.D. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* **2011**, *2*, 1–14. [CrossRef]
44. Strait, D.; Kraus, N. Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music Percept* **2011**, *29*, 133–146. [CrossRef]
45. Bishop-Liebler, P.; Welch, G.; Huss, M.; Thomson, J.M.; Goswami, U. Auditory temporal processing skills in musicians with Dyslexia. *Dyslexia* **2014**, *20*, 261–279. [CrossRef]
46. Protopapas, A. From temporal processing to developmental language disorders: Mind the gap. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2014**, *369*. [CrossRef]
47. Iversen, J.R.; Patel, A.D.; Ohgushi, K. Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* **2008**, *124*, 2263–2271. [CrossRef]
48. Crowhurst, M.J.; Teodocio Olivares, A. Beyond the Iambic-Trochaic Law: The joint influence of duration and intensity on the perception of rhythmic speech. *Phonology* **2014**, *31*, 51–94. [CrossRef]
49. Bhatara, A.; Boll-Avetisyan, N.; Agus, T.; Höhle, B.; Nazzi, T. Language experience affects grouping of musical instrument sounds. *Cog. Sci.* **2016**, *40*, 1816–1830. [CrossRef]
50. Langus, A.; Seyed-Allaei, S.; Uysal, E.; Pirmoradian, S.; Marino, C.; Asaadi, S.; Nespor, M. Listening natively across perceptual domains? *J. Exp. Psychol. Learn Mem. Cogn.* **2016**, *42*, 1127–1139. [CrossRef]
51. Boll-Avetisyan, N.; Bhatara, A.; Unger, A.; Nazzi, T.; Höhle, B. Effects of experience with L2 and music on rhythmic grouping by French listeners. *Biling* **2016**, *19*, 971–986. [CrossRef]
52. Boll-Avetisyan, N.; Bhatara, A.; Höhle, B. Effects of musicality on the perception of rhythmic structure in speech. *Lab. Phonol.* **2017**, *8*, 9. [CrossRef]

53. Wallentin, M.; Nielsen, A.H.; Friis-Olivarius, M.; Vuust, C.; Vuust, P. The Musical Ear Test, a new reliable test for measuring musical competence. *Learn Individ. Differ.* **2010**, *20*, 188–196. [CrossRef]
54. Levitin, D.J. What Does It Mean to Be Musical? *Neuron* **2012**, *73*, 633–637. [CrossRef]
55. Moll, K.; Landerl, K. *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT)*; Huber: Bern, Switzerland, 2010.
56. Westfall, J. PANGEA: Power ANalysis for GEneral Anova Designs. Available online: http://jakewestfall.org/publications/pangea.pdf (accessed on 19 March 2019).
57. Akeroyd, M.A. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* **2008**, *47*, S53–S71. [CrossRef]
58. Petermann, F. WAIS-IV. Wechsler Adult Intelligence Scale-Fourth Edition. Deutschsprachige Adaptation der WAIS-IV von D. Wechsler. Pearson Assessment: Frankfurt am Main, Germany, 2012. Available online: https://www.testzentrale.de/shop/wechsler-adult-intelligence-scale-fourth-edition.html (accessed on 11 March 2019).
59. Cunitz, K. Case-Marking and Animacy-Sentence Processing in Five- and Six-Year-Old Children. Ph.D. Thesis, University Leipzig, Leipzig, Germany, 2016.
60. Mottier, G. Mottier-Test. Über Untersuchungen zur Sprache lesegestörter Kinder. *Folia Phoniatr. Logop.* **1951**, *3*, 170–177. [CrossRef]
61. Baayen, R.H. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*; Cambridge University Press: Cambridge, UK, 2008.
62. Winter, B. Linear Models and Linear Mixed Effects Models in R with Linguistic Applications. Available online: https://arxiv.org/abs/1308.5499 (accessed on 11 March 2019).
63. Bates, D.M.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. S. Software* **2015**, *67*, 1–48. [CrossRef]
64. Rawlings, J.O.; Pantula, S.G.; Dickey, D.A. *Applied Regression Analysis: A Research Tool*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2001.
65. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 11 March 2019).
66. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-Effects models using lme4. *arXiv* **2014**, arXiv:1406.5823.
67. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009.
68. Fox, J.; Andersen, R. Effect displays for multinomial and proportional-Odds logit models. In *Sociology Methodology*; Stolzenberg, R.M., Ed.; American Sociological Association: Washington DC, USA, 2006; Volume 36, pp. 225–255.
69. Dupoux, E.; Pallier, C.; Sebastian-Gallés, N.; Mehler, J. A destressing 'deafness' in French? *J. Mem. Lang.* **1997**, *36*, 406–421. [CrossRef]
70. Dupoux, E.; Peperkamp, S.; Sebastián-Gallés, N. A robust method to study stress "deafness". *J. Acoust. Soc. Am.* **2001**, *110*, 1606–1618. [CrossRef] [PubMed]
71. Abboub, N.; Boll-Avetisyan, N.; Bhatara, A.; Höhle, B.; Nazzi, T. An exploration of rhythmic grouping of speech sequences by French-and German-learning infants. *Front. Hum. Neurosci.* **2016**, *10*, 292. [CrossRef] [PubMed]
72. Hay, J.F.; Saffran, J.R. Rhythmic grouping biases constrain infant statistical learning. *Infancy* **2012**, *17*, 610–641. [CrossRef] [PubMed]
73. De la Mora, D.M.; Nespor, M.; Toro, J.M. Do humans and nonhuman animals share the grouping principles of the Iambic-Trochaic Law? *Atten. Percept. Psychophys.* **2013**, *75*, 92–100. [CrossRef]
74. Toro, J.M.; Nespor, M. Experience-Dependent emergence of a grouping bias. *Biol. Lett.* **2015**, *11*, 20150374. [CrossRef]
75. Frost, R.L.A.; Monaghan, P.; Tatsumi, T. Domain-General mechanisms for speech segmentation: The role of duration information in language learning. *J. Exp. Psychol. Hum. Percept. Perform.* **2017**, *43*, 466–476. [CrossRef]
76. Boll-Avetisyan, N.; Bhatara, A.; Unger, A.; Nazzi, T.; Höhle, B. Rhythmic grouping biases in simultaneous bilinguals. *Biling* **2020**, 1–12, Published online by Cambridge University Press: 20 February 2020. [CrossRef]

77. Ho, Y.C.; Cheung, M.C.; Chan, A.S. Music training improves verbal but not visual memory: Cross-Sectional and longitudinal explorations in children. *Neuropsychology* **2003**, *17*, 439–450. [CrossRef]
78. Chan, A.S.; Ho, Y.; Cheung, M. Music training improves verbal memory. *Nature* **1998**, *396*, 128. [CrossRef]
79. Saito, S. The phonological loop and memory for rhythms: An individual differences approach. *Memory* **2001**, *9*, 313–322. [CrossRef]
80. Ramus, F.; Szenkovits, G. What phonological deficit? *Q. J. Exp. Psychol.* **2008**, *61*, 129–141. [CrossRef] [PubMed]
81. Lee, H.-Y.; Sie, Y.-S.; Chen, S.-C.; Cheng, M.-C. The music perception performance of children with and without dyslexia in Taiwan. *Psychol. Rep.* **2015**, *116*, 1–10. [CrossRef] [PubMed]
82. Thomson, J.M.; Leong, V.; Goswami, U. Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Read. Writ.* **2013**, *26*, 139–161. [CrossRef]
83. Bhide, A.; Power, A.; Goswami, U. A Rhythmic Musical Intervention for Poor Readers: A Comparison of Efficacy with a Letter-Based Intervention. *Mind Brain Educ.* **2013**, *7*, 113–123. [CrossRef]
84. Goswami, U. Sensory theories of developmental dyslexia: Three challenges for research. *Nat. Rev. Neurosci.* **2015**, *16*, 43–54. [CrossRef] [PubMed]
85. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 3rd ed.; Springer: New York, NY, USA, 2002.

brain sciences

MDPI

Communication

Cross-Modal Priming Effect of Rhythm on Visual Word Recognition and Its Relationships to Music Aptitude and Reading Achievement

Tess S. Fotidzis [1,*], Heechun Moon [2], Jessica R. Steele [3] and Cyrille L. Magne [3]

[1] Literacy Ph.D. Program, Middle Tennessee State University, Murfreesboro, TN 37132, USA
[2] Institutional Research, University of South Alabama, Mobile, AL 36688, USA; hmoon@southalabama.edu
[3] Psychology Department, Middle Tennessee State University, Murfreesboro, TN 37132, USA; jrs2bw@mtmail.mtsu.edu (J.R.S.); cyrille.magne@mtsu.edu (C.L.M.)
* Correspondence: tsf2m@mtmail.mtsu.edu; Tel.: +1-847-372-0370

Received: 22 October 2018; Accepted: 28 November 2018; Published: 29 November 2018

Abstract: Recent evidence suggests the existence of shared neural resources for rhythm processing in language and music. Such overlaps could be the basis of the facilitating effect of regular musical rhythm on spoken word processing previously reported for typical children and adults, as well as adults with Parkinson's disease and children with developmental language disorders. The present study builds upon these previous findings by examining whether non-linguistic rhythmic priming also influences visual word processing, and the extent to which such cross-modal priming effect of rhythm is related to individual differences in musical aptitude and reading skills. An electroencephalogram (EEG) was recorded while participants listened to a rhythmic tone prime, followed by a visual target word with a stress pattern that either matched or mismatched the rhythmic structure of the auditory prime. Participants were also administered standardized assessments of musical aptitude and reading achievement. Event-related potentials (ERPs) elicited by target words with a mismatching stress pattern showed an increased fronto-central negativity. Additionally, the size of the negative effect correlated with individual differences in musical rhythm aptitude and reading comprehension skills. Results support the existence of shared neurocognitive resources for linguistic and musical rhythm processing, and have important implications for the use of rhythm-based activities for reading interventions.

Keywords: implicit prosody; rhythm sensitivity; event related potentials; reading achievement; musical aptitude

1. Introduction

Music and language are complex cognitive abilities that are universal across human cultures. Both involve the combination of small sound units (e.g., phonemes for speech, and notes for music) which in turn, allow us to generate an unlimited number of utterances or melodies, in accordance with specific linguistic or musical grammatical rules (e.g., [1]). Of specific interest for the present study, is the notion of rhythm. In music, rhythm is marked by the periodic succession of acoustic elements as they unfold over time, and some of these elements may be perceived as stronger than others. Meter is defined as the abstract hierarchical organization of these recurring strong and weak elements that emerge from rhythm. It is this metrical structure that allows listeners to form predictions and anticipations, and in turn dance or clap their hands to the beat of the music [2].

Similarly, in speech, the pattern of stressed (i.e., strong), and unstressed (i.e., weak) syllables occurring at the lexical level contributes to the metrical structure of an utterance. Lexical stress is usually defined as the relative emphasis that one syllable, or several syllables, receive in a word [3].

Stress is typically realized by a combination of increased duration, loudness, and/or pitch change. In many languages, such as English, the salience of the stressed syllable is further reinforced by the fact that many unstressed syllables contain a reduced vowel [4]. Some languages are described as having fixed stress because the location of the stress is predictable. For instance, in French, the stress is usually on the final full syllable [5]. By contrast, several languages are considered to have variable stress because the position of the stress is not predictable. In such languages, like English, stress may serve as a distinctive feature to distinguish noun-verb stress homographs [6]. For example, the word "permit" is stressed on the first syllable when used as a noun, but stressed on the second syllable when used as a verb.

There is increasing support for the existence of rhythmic regularities in English, despite the apparent lack of physical periodicity of the stressed syllables when compared to the rhythmic structure of music (e.g., [7]). During speech production, rhythmic adjustments, such as stress shifts, may take place to avoid stress on adjacent syllables, and these stress shifts may give rise to a more regular alternating pattern of stressed and unstressed syllables [8]. For example, "thirteen" is normally stressed on the second syllable, but the stress can shift to the first syllable when followed by a word with initial stress (e.g., "thirteen people"). These rhythmic adjustments may play a role in speech perception, as suggested by findings showing that sentences with stress shifts are perceived as more natural than sentences with stress clashes, despite that words with shifted stress deviate from their default metrical structure [9].

In music, the Dynamic Attending Theory (DAT) provides a framework in which auditory rhythms are thought to create hierarchical expectancies for the signal as it unfolds over time [10,11]. According to the DAT, distinct neural oscillations entrain to the multiple hierarchical levels of the metrical structure of the auditory signal, and strong metrical positions act as attentional attractors, thus making acoustic events occurring at these strong positions easier to process. Similarly, listeners do not pay equal attention to all parts of the speech stream, and speech rhythm may influence which moments are hierarchically attended to in the speech signal. For instance, detection of a target phoneme was found to be faster if it was embedded in a rhythmically regular sequence of words (i.e., regular time interval between successive stressed syllables), thus suggesting that speech rhythm cues, such as stressed syllables, guide listeners' attention to specific portions of the speech signal [12]. Further evidence suggests that predictions regarding speech rhythm and meter may be crucial for language acquisition [13], speech segmentation [14], word recognition [15], and syntactic parsing [16].

Given the structural similarities between music and language, a large body of literature has documented which neuro-cognitive systems may be shared between language and music (e.g., [7,17,18]), and converging evidence support the idea that musical and linguistic rhythm perception skills partially overlap [19–21]. In line with these findings, several EEG studies revealed a priming effect of musical rhythm on spoken language processing. For instance, listeners showed a more robust neural marker of beat tracking and better comprehension when stressed syllables aligned with strong musical beats in sung sentences [22]. Likewise, EEG findings demonstrated that spoken words were more easily processed when they followed non-linguistic primes with a metrical structure that matched the word metrical structure [23]. A follow-up study using a similar design showed this benefit of rhythm priming on speech processing may be mediated by cross-domain neural phase entrainment [24].

The purpose of the present study was to shed further light on the effect of non-linguistic rhythmic priming on language processing (e.g., [22–24]). We specifically focused on words with a trochaic stress pattern (i.e., a stressed syllable followed by an unstressed syllable) because in the English lexicon, they constitute more than 85% of content words [25]. This high frequency of the trochaic pattern may play a particularly preponderant role in English language development, as infants seem to adopt a metrical segmentation strategy by considering a stress syllable as the beginning of a word in the continuous speech stream [26]. Evidence in support of this important role of the trochaic pattern comes from studies conducted with English speaking infants who develop a preference for the trochaic pattern as early as the age of 6 months [27]. By contrast, the ability to detect words with an iambic

pattern (i.e., an unstressed syllable followed by a stressed syllable) develops later, around 10.5 months, and seems to rely more on using additional sets of linguistic knowledge regarding phonotactic constraints (i.e., the sequences of phonemes that are allowed in a given language), and allophonic cues (i.e., the multiple phonetic variants of a phoneme, whose occurrences depend on their position in a word and their phonetic context), rather than stress cues [13].

The first specific aim was to examine whether the cross-domain rhythmic priming effect is also present when target words are visually presented. To this end, participants were presented with rhythmic auditory prime sequences (either a repeating pattern of long-short or short-long tone pairs), followed by a visual target word with a stress pattern that either matched, or mismatched, the temporal structure of the prime (See Figure 1). Based on previous literature (e.g., [20,23,28]), we predicted that words that do not match the temporal structure of the rhythmic prime would elicit an increased centro-frontal negativity.

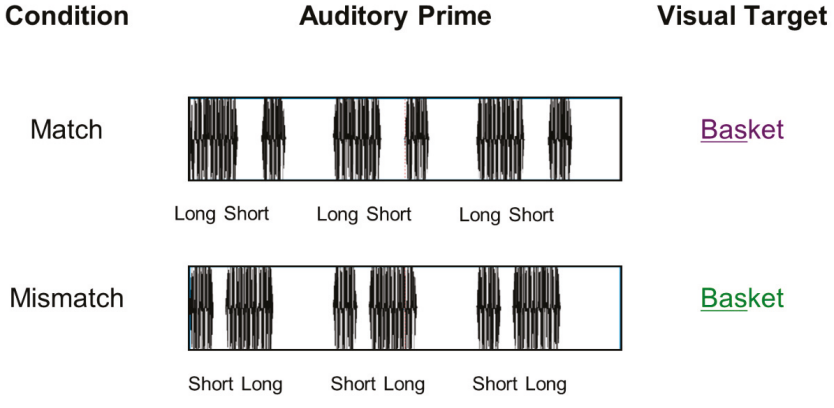


Figure 1. Rhythmic cross-modal priming experimental paradigm. The auditory prime (long-short or short-long sequence) is followed by a target visual word with a stress pattern that either match or mismatch the prime (Note: stressed syllable is underlined for illustration purposes only).

A second aim of the study was to determine whether such rhythmic priming effect would be related to musical aptitude. Musical aptitude has been associated with enhanced perception of speech cues that are important correlates of rhythm. For instance, individuals with formal musical training better detect violations of word pitch contours [29,30] and syllabic durations [31] than non-musicians. In addition, electrophysiological evidence shows that the size of a negative ERP component elicited by spoken words with an unexpected stress pattern correlates with individual differences in musical rhythm abilities [20]. Thus, in the present study, we expected the amplitude of the negativity elicited by the cross-modal priming effect to correlate with individual scores on a musical aptitude test, if the relationship between musical aptitude and speech rhythm sensitivity transfers to the visual domain.

Finally, the third study aim was to test whether the cross-modal priming effect present in the ERPs correlated with individual differences in reading achievement. Mounting evidence suggests a link between sensitivity to auditory rhythm skills (both linguistic and musical) and reading abilities (e.g., [32–35]). As such, we collected individuals' scores on a college readiness reading achievement test to examine whether the cross-modal ERP effect correlated with individual differences in reading comprehension skills. We expected the amplitude of the negativity elicited by the cross-modal priming effect to correlate with individual scores on the American College Testing (ACT) reading test, if rhythm perception skills relate to reading abilities as suggested by the current literature [32–35].

2. Materials and Methods

2.1. Participants

Eighteen first year college students took part in the experiment (14 females and 4 males, mean age = 19.5, age range: 18–22). All were right-handed, native English speakers with less than two years of formal musical training. None of the participants were enrolled in a Music major. The study was approved by the Institutional Review Board at Middle Tennessee State University, and written consent was obtained from the participants prior to the start of the experiment.

2.2. Standardized Measures

The Advanced Measures of Music Audiation (AMMA; [36]) was used to assess participants' musical aptitude. The AMMA has been used previously to measure the correlation between musical aptitude and index of brain activities (e.g., [20,37–39]). This measure was nationally standardized with a normed sample of 5336 U.S. students and offers percentile ranked norms for both music and non-music majors. Participants were presented with 30 pairs of melodies and asked to determine whether the two melodies of each pair were the same, tonally different, or rhythmically different. The AMMA provides separate scores for rhythmic and tonal abilities. For non-Music majors, reliability scores are 0.80 for the tonal score and 0.81 for the rhythm score [36].

The reading scores on the ACT exam were used to examine the relationship between reading comprehension and speech rhythm sensitivity. The ACT reading section is a standardized achievement test that comprises short passages from four categories (prose fiction, social science, humanities, and natural science) and 40 multiple-choice questions that test the reader's comprehension of the passages. Scores range between 1 and 36. The test was administered and scored by the non-profit organization of the same name (ACT, Inc., Iowa City, IA, USA) using a paper and pencil format.

2.3. EEG Cross-Modal Priming Paradigm

Prime sequences consisted of a rhythmic tone pattern of either a long-short or short-long structure repeated three times. The tones consisted of a 500 Hz sine wave with a 10 ms rise/fall, and a duration of either 200 ms (long) or 100 ms (short). In long-short sequences, the long tone and short tone were separated by a silence of 100 ms, and each of the three successive long-short tone pairs was followed by a silence of 200 ms. In short-long sequences, the short tone and long tone were separated by a silence of 50 ms, and each of the three successive short-long tone pairs was followed by a silence of 250 ms. Because previous research has shown that native speakers of English have a cultural bias toward grouping a sequence of tones differing in duration, into short-long patterns [40,41], a series of behavioral pilot experiments were conducted with different iterations of the tone sequences to determine which parameters would provide consistent perception of either long-short or short-long patterns.

Visual targets were composed of 140 English real-word bisyllabic nouns and 140 pseudowords, which were all selected from the database of the English Lexicon Project [42]. The lexical frequency of all the words was controlled using the log HAL frequency [43]. The mean log HAL frequency for each set of stress patterns was 10.28 (SD = 0.98) for trochaic sequences and 10.28 (SD = 0.97) for iambic sequences. Pseudowords were matched to the real words in terms of syllable count and word length and were used only for the purpose of the lexical decision task. Half of the real words (N = 70) had a trochaic stress pattern (i.e., stressed on the first syllable, for example, "basket"). The other half consisted of fillers with an iambic stress pattern (i.e., stressed on the second syllable, for example, "guitar").

Short-long and long-short prime sequences were combined with the visual target words to create two experimental conditions in which the stress pattern of the target word either matched or mismatched the rhythm of the auditory prime.

We chose to analyze only the ERPs elicited by trochaic words for several reasons. First, trochaic words comprise the predominant stress pattern in English (85–90% of spoken English words according to [34]), and consequently, participants will likely be familiar with their pronunciation.

Second, because stressed syllables correspond to word onset in trochaic words, this introduces fewer temporal jitters than for iambic words when computing ERPs across trials. This scenario is particularly problematic for iambic words during silent reading, because there is no direct way to measure when participants read the second syllable. Third, participants were recruited from a university located in the southeastern region of the United States, and either originated from this area, or have been living in the area for several years. It is well documented that the Southern American English dialect tends to place stress on the first syllable of many iambic words despite that these types of words are stressed on the second syllable in standard American English (e.g., [44]). As such, rhythmic expectations are less clear to predict for iambic words.

2.4. Procedure

Participants' musical aptitude was first measured using the AMMA [36]. Following administration of the AMMA test, participants were seated in a soundproofed and electrically shielded room. Auditory prime sequences were presented through headphones, and target stimuli were visually presented on a computer screen placed at approximately 3 feet in front of the participant. Words and pseudowords were written in black lowercase characters on a white background. No visual cue was provided to the participant regarding the location of the stress syllables in the target words. Stimulus presentation was controlled using the software E-prime 2.0 Professional with Network Timing Protocol (Psychology software tools, Inc., Pittsburgh, PA, USA). Participants were presented with 5 blocks of 56 stimuli. The trials were randomized within each block, and the order of the blocks was counterbalanced across participants. Each trial was introduced by a fixation cross displayed at the center of a computer screen that remained until 2 s after the onset of the visual target word. Participants were asked to silently read the target word and to press one button if they thought it was a real English word, or another button if they thought it was a nonword. The entire experimental session lasted 1.5 h.

2.5. EEG Acquisition and Preprocessing

EEG was recorded continuously from 128 Ag/AgCL electrodes embedded in sponges in a Hydrocel Geodesic Sensor Net (EGI, Eugene, OR, USA) placed on the scalp, connected to a NetAmps 300 amplifier, and using a MacPro computer. Electrode impedances were kept below 50 kΩ. Data was referenced online to Cz and re-referenced offline to the averaged mastoids. In order to detect the blinks and vertical eye movements, the vertical and horizontal electrooculograms (EOG) were also recorded. The EEG and EOG were digitized at a sampling rate of 500 Hz. EEG preprocessing was carried out with NetStation Viewer and Waveform tools. The EEG was first filtered with a bandpass of 0.1 to 30 Hz. Data time-locked to the onset of trochaic target words was then segmented into epochs of 1100 ms, starting with a 100 ms prior to the word onset and continuing 1000 ms post-word-onset. Trials containing movements, ocular artifacts, or amplifier saturation were discarded. ERPs were computed separately for each participant and each condition by averaging together the artifact-free EEG segments relative to a 100 ms pre-baseline.

2.6. Data Analysis

Statistical analyses were performed using MATLAB and the FieldTrip open source toolbox [45]. A planned comparison between the ERPs elicited by mismatching trochaic words and matching trochaic words was performed using a cluster-based permutation approach. This non-parametric data-driven approach does not require the specification of any latency range or region of interest a priori, while also offering a solution to the problem of multiple comparisons (see [46]).

To relate the ERP results to the behavioral measures (i.e., musical aptitude and reading comprehension), an index of sensitivity to speech rhythm cues was first calculated from the ERPs using the mean of the significant amplitude differences between ERPs elicited by matching and mismatching trochaic words at each channels, and time points belonging to the resulting clusters (see [20,47] for

similar approaches). Pearson correlations were then tested between the ERP cluster mean difference and the participants' scores on the AMMA and ACT reading section, respectively. A multiple regression was also computed with the ERP cluster mean difference as the outcome measure, and the AMMA Rhythm scores and ACT Reading scores as the predictor variables.

3. Results

3.1. Metrical Expectancy

Overall, participants performed well on the lexical decision task, as suggested by the mean accuracy rate (M = 98.82%, SD = 0.85). A paired samples t-test was computed to compare accuracy rates for real target words in the matching (M = 99.83%, SD = 0.70), and mismatching (M = 99.42%, SD = 1.40) rhythm conditions. No statistically significant differences were found between the two conditions, t (35) = 1.54, p = 0.13, two-tailed.

Analyses of the ERP data revealed that target trochaic words that mismatched the rhythmic prime elicited a significantly larger negativity from 300 to 708 ms over a centro-frontal cluster of electrodes ($p < 0.001$, See Figure 2).

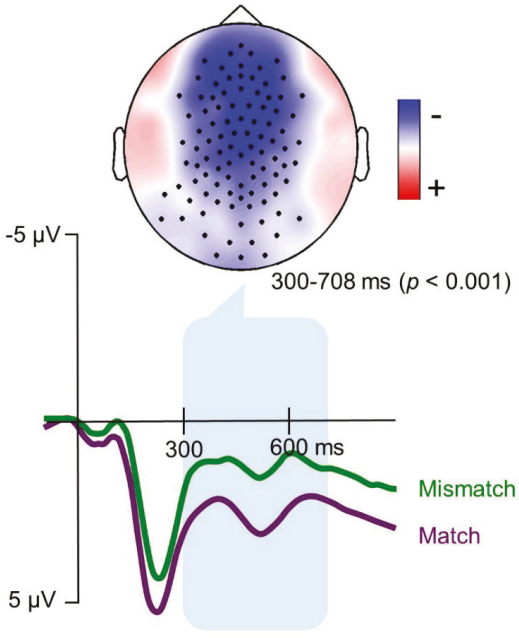


Figure 2. Rhythmic priming Event-related potential (ERP) effect. Grand-average event-related potentials (ERPs) recorded for matching (purple), and mismatching (green) trochaic target words, averaged for the significant group of channels in the cluster. The latency range of the significant clusters is indicated in blue. (Note: Negative amplitude values are plotted upward. The topographic map shows the mean differences in scalp amplitudes in the latency range of the significant clusters. Electrodes belonging to the cluster are indicated with a black dot).

3.2. Brain-Behavior Relationships

The negative ERP cluster mean difference was statistically significantly positively correlated with the AMMA Rhythm scores (r = 0.74, $p < 0.001$; see Figure 3A) and the ACT Reading scores (r = 0.60, p = 0.009; see Figure 3B). A statistically significant positive correlation was also found between

the AMMA Rhythm scores and ACT Reading scores (r = 0.55, p = 0.016; see Figure 3C). By contrast, no statistically significant correlation was found between the AMMA Tonal scores and the negative ERP cluster mean difference (r = 0.30, p = 0.23) or the ACT Reading scores (r = 0.09, p = 0.70). The maximum Cook's distance for the reported correlations indicated no undue influence of any data point on the fitted models (max Cook's d < 0.5).

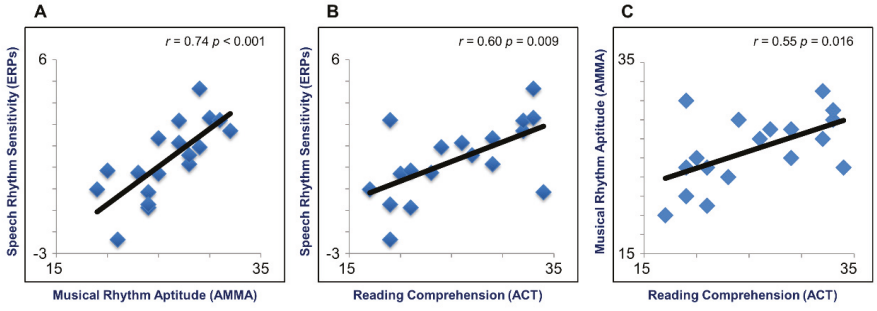


Figure 3. Brain-behavior correlations. (**A**) Correlation between speech rhythm sensitivity (as indexed by the negative ERP cluster mean difference) and musical rhythm aptitude; (**B**) correlation between speech rhythm sensitivity and reading comprehension; (**C**) correlation between musical rhythm aptitude and reading comprehension. (Note: The solid line represents a linear fit.)

A multiple regression was conducted to investigate whether AMMA Rhythm scores and ACT Reading scores predicted the size of the negative ERP cluster mean difference. Table 1 summarizes the analysis results. The regression model explained 59.9% of the variance and was a statistically significant predictor of the negative ERP cluster mean difference (R^2 = 0.599, F (2,15) = 11.2, p = 0.001). As can be seen in Table 1, AMMA Rhythm scores statistically significantly contributed to the model (β = 0.594, t (15) = 3.023, p = 0.009), but ACT Reading scores did not (β = 0.267, t (15) = 1.359, p = 0.194). The final predictive model was: Negative ERP Cluster Mean Difference = (0.281 × AMMA Rhythm) + (0.081 × ACT Reading) + 8.000.

Table 1. Multiple regression coefficients.[1]

Source	*B*	*SE*	*β*	*t*	*p*
Constant	8.000	2.033		3.935	0.001
ACT Reading	0.081	0.060	0.267	1.359	0.194
AMMA Rhythm	0.281	0.093	0.594	3.023	0.009

[1] Outcome: Negative ERP cluster mean difference; *B*: unstandardized coefficient; *SE*: standard error; *β*: standardized coefficient; *t*: *t*-value; *p*: *p*-value; ACT: American College Testing; AMMA: Advanced Measures of Music Audiation.

4. Discussion

The current study aimed to examine the cross-modal priming effect of non-linguistic auditory rhythm on written word processing and investigate whether such effect would relate to individual differences in musical aptitude and reading comprehension. As hypothesized, trochaic target words that did not match the rhythmic structure of the auditory prime were associated with an increased negativity over the centro-frontal part of the scalp. This finding is in line with previous ERP studies on speech rhythm and meter [6,15,20,28,31,48–50]. It has been generally proposed that this negative effect either reflects an increased N400 [15,49], or a domain-general rule-based error-detection mechanism [6,20,28,31,51,52]. The fact that similar negative effects have been reported in response to metric deviations in tone sequences (e.g., [53,54]) further supports the latter interpretation.

While the aforementioned studies were conducted either in the linguistic or musical domain, the negative effect observed for mismatching target words was generated by non-linguistic prime sequences in the present experiment. Cason and Schön [23] previously reported a cross-domain priming effect of music on speech processing, which was reflected by a similar increased negativity when the metrical structure of the spoken target word did not match the rhythmic structure of the musical prime. Several other findings have since shown that temporal expectancies generated by rhythmically regular non-linguistic primes can facilitate spoken language processing in typical adults (e.g., [24,55]), and children [56,57], as well as adults with Parkinson's disease [58], children with cochlear implants [59], and children with language disorders [60]. This beneficial effect may stem from the regular rhythmic structure of the prime, which provides temporally predictable cues to which internal neural oscillators can anchor [24]. The present findings support and extend this line of research by showing this negativity is elicited even when the target words were visually presented, thus suggesting that non-linguistic rhythm can not only induce metrical expectations across distinct cognitive domains, but also across different sensory modalities [61]. These findings also provide additional evidence in favor of the view that rhythm/meter processing relies on a domain-general neural system that is not specific to language [19,21,22].

We further investigated whether this cross-modal priming effect was related to individual differences in musical aptitude. Interestingly, our results showed a statistically significant correlation between the size of the brain response elicited by unexpected stress patterns and the AMMA rhythm subscore, but not the tonal subscore. In addition, musical rhythm aptitude was a statistically significant predictor of speech rhythm sensitivity, even after controlling for reading comprehension skills. This is in line with previous ERP studies showing that adult musicians performed better than non-musicians at detecting words pronounced with an incorrect stress pattern [31]. In addition, this enhanced sensitivity to speech meter was associated with larger electrophysiological responses to incorrectly pronounced words, which was interpreted as reflecting more efficient early auditory processing of the temporal properties of speech.

Robust associations have also been found between musical rhythm skills and speech prosody perception, even after controlling for years of music education [19]. Noteworthy for the present experiment, individual differences in brain sensitivity to speech rhythm variations can be explained by variance in musical rhythm aptitude in individuals with less than two years of musical training. For instance, in a recent experiment [20], participants' musical aptitude was assessed using the same standardized measure of musical abilities (i.e., AMMA) as in the present study. Participants listened to sequences consisting of four bisyllabic words for which the stress pattern of the final word either matched or mismatched the stress pattern of the preceding words. Words with a mismatching stress pattern elicited an increased negative ERP component with the same scalp distribution and latency as the one found in the current data. More importantly, participants' musical rhythm aptitude statistically significantly correlated with the size of the negative effect. Thus, in light of the aforementioned literature, the present results confirm and extend previous data suggesting a possible transfer of learning between the musical and linguistic domains (See [62] for a review).

Adding to the growing literature showing a relationship between sensitivity to speech rhythm and reading skills, our results revealed a statistically significant positive correlation between the scores on the ACT reading subtest and the size of the negative ERP effect elicited by mismatching stress patterns. Previous studies have mainly focused on typically developing young readers using several novel speech rhythm tasks in conjunction with standardized measures of reading abilities, and results consistently showed a correlation between performances on the speech rhythm tasks and individual differences in word reading skills [63–66]. It has been proposed that early sensitivity to speech rhythm cues may contribute to the development of phonological representations [32]. However, sensitivity to speech rhythm cues still explains unique variance in word reading skills after controlling for phonological processing skills [67], thus suggesting that it also makes a significant contribution to reading development independently of phonological awareness.

More directly related to the present study, research with older readers and adults suggests that knowledge of the prosodic structure of words continues to play a role in skilled reading. For instance, visual word recognition is facilitated when primed by word fragments with a matching stress pattern [68,69]. Two other studies conducted on typical adults focused on lexical stress perception in isolated multisyllabic words [70,71], and found a significant relationship with reading comprehension. Likewise, adult struggling readers usually show lower performance than their typical peers on tasks measuring perception of word stress patterns or auditory rhythms [72–75] (but see [74,76]).

Interestingly, the finding that reading comprehension was not a statistically significant contributor to speech rhythm sensitivity after controlling for musical rhythm aptitude supports the Temporal Sampling Framework (TSF) proposed by Goswami [32]. According to the TSF, the link between speech rhythm sensitivity and reading skills is mediated by domain-general neurocognitive mechanisms for processing acoustic information carrying rhythmic cues. In line with this interpretation, we found a statistically significant correlation between the AMMA rhythm scores and reading achievement scores.

The OPERA (overlap, precision, emotion, repetition, attention) hypothesis formulated by Patel [77,78] further provides a potential explanation of music-training driven plasticity in brain networks involved in language. OPERA offers a set of five optimal conditions that must be met for music training to drive plasticity: (1) music and language have overlapping anatomical substrates; (2) music activities require a greater level of precision compared to language; (3) music activities evoke strong emotions; (4) music training involves repeated practice; (5) music activities require sustained attention. In line with this framework, the Precise Auditory Timing Hypothesis (PATH) proposed by Tierney and Kraus [79] predicts that music programs that focus on rhythm activities, with an emphasis on entrainment and timing, will be more effective in improving reading-related skills, such as phonological processing skills, because there are overlaps between language and music networks processing rhythmic information, and music requires a higher level of auditory-motor timing precision than language. OPERA and PATH thus provide compelling explanations for the significant relationships we report here between musical rhythm aptitude, speech rhythm sensitivity, and reading achievement. While our present study was correlational (and conducted with non-musicians), data from recent longitudinal studies using randomized controlled trials indeed show promising results of rhythm-based intervention for the development of language skills in children with reading disorders [80], and typical peers [81].

Finally, the fact that we found a "metrical" negativity to visual targets, despite that participants were not allowed to sound out the words, further supports theories proposing that information about the metrical structure of a word is part of its lexical representation and automatically retrieved during silent reading [82,83]. This idea is in line with the Implicit Prosody Hypothesis (IPH) originally proposed by Fodor [84]. The IPH is closely related to the concept of verbal imagery or inner voice, which can be found in the literature throughout the 20th century [82]. According to the IPH, readers create a mental representation of the prosodic structure of the text while they are silently reading. Several studies have provided compelling evidence in support for the IPH, especially regarding lexical stress. For instance, eye-tracking studies showed that readers had longer reading times and more eye fixations for four-syllable words with two stressed syllables, than for one stressed syllable [85], and that expectations generated by the stress pattern of successive words may affect early stages of syntactic analysis of upcoming words in written sentences [82,86]. Taken together, these results and the present data provide compelling evidence for a role of prosodic representations regarding a word stress pattern during silent reading.

One potential limitation of the current research is the use of ACT reading scores, which may not be fully representative of the participants' reading skills. In particular, phonemic awareness, decoding, and fluency, which are components known to greatly contribute to reading comprehension [87], cannot be teased apart in the ACT reading subsets. Future research using a more comprehensive battery of language and reading assessments would better allow a more complete understanding of which reading components are more closely related to speech rhythm perception skills.

5. Conclusions

The present data confirm and extend previous studies showing facilitating effects of a regular non-linguistic rhythm on spoken language processing (e.g., [23,55,59]), by demonstrating this to also be the case for written language processing. We propose that this cross-modal effect of rhythm is mediated by the automatic retrieval of the word metrical structure (i.e., implicit prosody) during silent reading (i.e., implicit prosody generated through verbal imagery). Finally, because we found that the negativity associated with this cross-modal priming effect of rhythm correlated with individual differences in musical aptitude and reading achievement, this further supports the potential clinical and education implications of using rhythm-based intervention for populations with language or learning disabilities.

Author Contributions: T.S.F. collected the data and wrote the paper. H.M. collected and analyzed the data. J.R.S. wrote and edited the paper. C.L.M. conceived the idea, designed the experiments, and wrote the paper.

Funding: This study was funded by NSF Grant # BCS-1261460 awarded to Cyrille Magne and by the MTSU Foundation.

Conflicts of Interest: The authors declare no conflict of interest. The funding sources had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

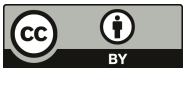
References

1. Lerdahl, F.; Jackendoff, R. An overview of hierarchical structure in music. *Music Percept. Interdiscip. J.* **1984**, *1*, 229–252. [CrossRef]
2. London, J. *Hearing in Time: Psychological Aspects of Musical Meter*, 1st ed.; Oxford University Press: Oxford, UK, 2004; ISBN 0-19-516081-9.
3. Fox, A. *Prosodic Features and Prosodic Structures: The Phonology of Suprasegmentals*; Oxford University Press: New York, NY, USA, 2000.
4. Nespor, M. On the rhythm parameter in phonology. In *Logical Issues in Language Acquisition*; Rocca, I., Ed.; Foris Publications: Dordrech, The Netherlands, 1990; pp. 157–175.
5. Delattre, P. *Studies in French and Comparative Phonetics*, 1st ed.; Mouton: The Hague, The Netherlands, 1966.
6. Moon, H.; Magne, C. Noun/verb distinction in English stress homographs: An ERP study. *Neuroreport* **2015**, *26*, 753–757. [CrossRef] [PubMed]
7. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: New York, NY, USA, 2008; ISBN 978-0-19-975530-1.
8. Liberman, M.; Prince, A. On stress and linguistic rhythm. *Linguist. Inq.* **1977**, *8*, 249–336.
9. Henrich, K.; Alter, K.; Wiese, R.; Domahs, U. The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain Lang.* **2014**, *136*, 19–30. [CrossRef] [PubMed]
10. Jones, M.R.; Boltz, M. Dynamic attending and responses to time. *Psychol. Rev.* **1989**, *96*, 459–491. [CrossRef] [PubMed]
11. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
12. Quené, H.; Port, R.F. Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* **2005**, *62*, 1–13. [CrossRef] [PubMed]
13. Jusczyk, P.W. How infants begin to extract words from speech. *Trends Cogn. Sci.* **1999**, *3*, 323–328. [CrossRef]
14. Mattys, S.L.; Samuel, A.G. How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *J. Mem. Lang.* **1997**, *36*, 87–116. [CrossRef]
15. Magne, C.; Astésano, C.; Aramaki, M.; Ystad, S.; Kronland-Martinet, R.; Besson, M. Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cereb. Cortex* **2007**, *17*, 2659–2668. [CrossRef] [PubMed]
16. Schmidt-Kassow, M.; Kotz, S.A. Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Res.* **2008**, *1226*, 144–155. [CrossRef] [PubMed]
17. Jantzen, M.G.; Large, E.W.; Magne, C. Editorial: Overlap of neural systems for processing language and music. *Front. Psychol.* **2016**, *7*, 876. [CrossRef] [PubMed]

18. Peretz, I.; Vuvan, D.; Lagrois, M.-E.; Armony, J.L. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. B* **2015**, *370*, 20140090. [CrossRef] [PubMed]
19. Hausen, M.; Torppa, R.; Salmela, V.R.; Vainio, M.; Särkämö, T. Music and speech prosody: A common rhythm. *Front. Psychol.* **2013**, *4*, 566. [CrossRef] [PubMed]
20. Magne, C.; Jordan, D.K.; Gordon, R.L. Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain Lang.* **2016**, *153*, 13–19. [CrossRef] [PubMed]
21. Peter, V.; Mcarthur, G.; Thompson, W.F. Discrimination of stress in speech and music: A mismatch negativity (MMN) study. *Psychophysiology* **2012**, *49*, 1590–1600. [CrossRef] [PubMed]
22. Gordon, R.L.; Magne, C.L.; Large, E.W. EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Front. Psychol.* **2011**, *2*, 352. [CrossRef] [PubMed]
23. Cason, N.; Schön, D. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* **2012**, *50*, 2652–2658. [CrossRef] [PubMed]
24. Falk, S.; Lanzilotti, C.; Schön, D. Tuning neural phase entrainment to speech. *J. Cogn. Neurosci.* **2017**, *29*, 1378–1389. [CrossRef] [PubMed]
25. Cutler, A.; Carter, D.M. The Predominance of Strong Initial Syllables in the English Vocabulary. *Comput. Speech Lang.* **1987**, *2*, 133–142. [CrossRef]
26. Cutler, A.; Norris, D. The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* **1988**, *14*, 113–121. [CrossRef]
27. Jusczyk, P.W.; Cutler, A.; Redanz, N.J. Infants' preference for the predominant stress patterns of English words. *Child Dev.* **1993**, *64*, 675–687. [CrossRef] [PubMed]
28. Rothermich, K.; Schmidt-Kassow, M.; Schwartze, M.; Kotz, S.A. Event-related potential responses to metric violations: Rules versus meaning. *Neuroreport* **2010**, *21*, 580–584. [CrossRef] [PubMed]
29. Magne, C.; Schön, D.; Besson, M. Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* **2006**, *18*, 199–211. [CrossRef] [PubMed]
30. Schön, D.; Magne, C.; Besson, M. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* **2004**, *41*, 341–349. [CrossRef] [PubMed]
31. Marie, C.; Magne, C.; Besson, M. Musicians and the metric structure of words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef] [PubMed]
32. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
33. Harrison, E.; Wood, C.; Holliman, A.J.; Vousden, J.I. The immediate and longer-term effectiveness of a speech-rhythm-based reading intervention for beginning readers. *J. Res. Read.* **2018**, *41*, 220–241. [CrossRef]
34. Holliman, A.J.; Williams, G.J.; Mundy, I.R.; Wood, C.; Hart, L.; Waldron, S. Beginning to disentangle the prosody-literacy relationship: A multi-component measure of prosodic sensitivity. *Read. Writ.* **2014**, *27*, 255–266. [CrossRef]
35. Thomson, J.; Jarmulowicz, L. *Linguistic Rhythm and Literacy*; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2016. [CrossRef]
36. Gordon, E.E. *Predictive Validity Study of AMMA: A One-Year Longitudinal Predictive Validity Study of the Advanced Measures of Music Audiation*; GIA Publications: Chicago, IL, USA, 1990.
37. Schneider, P.; Scherg, M.; Dosch, H.G.; Specht, H.J.; Gutschalk, A.; Rupp, A. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* **2002**, *5*, 688–694. [CrossRef] [PubMed]
38. Seppänen, M.; Brattico, E.; Tervaniemi, M. Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiol. Learn. Mem.* **2007**, *87*, 236–247. [CrossRef] [PubMed]
39. Vuust, P.; Brattico, E.; Seppänen, M.; Näätänen, R.; Tervaniemi, M. The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia* **2012**, *50*, 1432–1443. [CrossRef] [PubMed]
40. Hay, J.S.F.; Diehl, R.L. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Percept. Psychophys.* **2007**, *69*, 113–122. [CrossRef] [PubMed]
41. Iversen, J.R.; Patel, A.D.; Ohgushi, K. Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* **2008**, *124*, 2263–2271. [CrossRef] [PubMed]

42. Balota, D.A.; Yap, M.J.; Hutchison, K.A.; Cortese, M.J.; Kessler, B.; Loftis, B.; Neely, J.H.; Nelson, D.L.; Simpson, G.B.; Treiman, R. The English Lexicon Project. *Behav. Res. Methods* **2007**, *39*, 445–459. [CrossRef] [PubMed]
43. Lund, K.; Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* **1996**, *28*, 203–208. [CrossRef]
44. Thomas, E.R. Rural white southern accents. In *A Handbook of Varieties of English: A Multimedia Reference Tool*; Kortmann, B., Schneider, E.W., Eds.; Mouton de Gruyter: New York, NY, USA, 2004; pp. 300–324.
45. Oostenveld, R.; Fries, P.; Maris, E.; Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, *2011*, 1–9. [CrossRef] [PubMed]
46. Maris, E.; Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **2007**, *164*, 177–190. [CrossRef] [PubMed]
47. Lense, M.D.; Gordon, R.L.; Key, A.P.F.; Dykens, E.M. Neural correlates of cross-modal affective priming by music in Williams syndrome. *Soc. Cogn. Affect. Neurosci.* **2014**, *9*, 529–537. [CrossRef] [PubMed]
48. Bohn, K.; Knaus, J.; Wiese, R.; Domahs, U. The influence of rhythmic (IR) regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia* **2013**, *51*, 760–771. [CrossRef] [PubMed]
49. Domahs, U.; Wiese, R.; Bornkessel-Schlesewsky, I.; Schlesewsky, M. The processing of German word stress: Evidence for the prosodic hierarchy. *Phonology* **2008**, *25*, 1–36. [CrossRef]
50. McCauley, S.M.; Hestvik, A.; Vogel, I. Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Lang. Speech* **2012**, *56*, 23–44. [CrossRef] [PubMed]
51. Rothermich, K.; Schmidt-Kassow, M.; Kotz, S.A. Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia* **2012**, *50*, 232–244. [CrossRef] [PubMed]
52. Schmidt-Kassow, M.; Kotz, S.A. Attention and perceptual regularity in speech. *Neuroreport* **2009**, *20*, 1643–1647. [CrossRef] [PubMed]
53. Brochard, R.; Abecasis, D.; Potter, D.; Ragot, R.; Drake, C. The "Ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychol. Sci.* **2003**, *14*, 362–366. [CrossRef] [PubMed]
54. Ystad, S.; Magne, C.; Farner, S.; Pallone, G.; Aramaki, M.; Besson, M.; Kronland-Martinet, R. Electrophysiological study of algorithmically processed metric/rhythmic variations in language and music. *EURASIP J. Audio Speech, Music Process.* **2007**, *2007*, 03019. [CrossRef]
55. Cason, N.; Astésano, C.; Schön, D. Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychol.* **2015**, *155*, 43–50. [CrossRef] [PubMed]
56. Gordon, R.L.; Shivers, C.M.; Wieland, E.A.; Kotz, S.A.; Yoder, P.J.; Devin McAuley, J. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* **2015**, *18*, 635–644. [CrossRef] [PubMed]
57. Chern, A.; Tillmann, B.; Vaughan, C.; Gordon, R.L. New evidence of a rhythmic priming effect that enhances grammaticality judgments in children. *J. Exp. Child Psychol.* **2018**, *173*, 371–379. [CrossRef] [PubMed]
58. Kotz, S.A.; Gunter, T.C. Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 62–68. [CrossRef] [PubMed]
59. Cason, N.; Hidalgo, C.; Isoard, F.; Roman, S.; Schön, D. Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology* **2015**, *29*, 102–107. [CrossRef] [PubMed]
60. Przybylski, L.; Bedoin, N.; Herbillon, V.; Roch, D.; Kotz, S.A.; Tillmann, B. Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* **2013**, *27*, 121–131. [CrossRef] [PubMed]
61. Brochard, R.; Tassin, M.; Zagar, D. Got rhythm for better and for worse. Cross-modal effects of auditory rhythm on visual word recognition. *Cognition* **2013**, *127*, 214–219. [CrossRef] [PubMed]
62. Gordon, R.L.; Magne, C.L.; Magne, C.L. Music and the brain: Music and cognitive abilities. In *The Routledge Companion to Music Cognition*; Ashley, R., Timmers, R., Eds.; Routledge: New York, NY, USA, 2017; pp. 49–62.
63. Holliman, A.J.; Wood, C.; Sheehy, K. Sensitivity to speech rhythm explains individual differences in reading ability independently of phonological awareness. *Br. J. Dev. Psychol.* **2008**, *26*, 357–367. [CrossRef]
64. Holliman, A.J.; Wood, C.; Sheehy, K. A cross-sectional study of prosodic sensitivity and reading difficulties. *J. Res. Read.* **2012**, *35*, 32–48. [CrossRef]
65. Whalley, K.; Hansen, J. The role of prosodic sensitivity in children's reading development. *J. Res. Read.* **2006**, *29*, 288–303. [CrossRef]

66. Wood, C. Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *J. Res. Read.* **2006**, *29*, 270–287. [CrossRef]
67. Holliman, A.J.; Gutiérrez Palma, N.; Critten, S.; Wood, C.; Cunnane, H.; Pillinger, C. Examining the independent contribution of prosodic sensitivity to word reading and spelling in early readers. *Read. Writ.* **2017**, *30*, 509–521. [CrossRef]
68. Cooper, N.; Cutler, A.; Wales, R. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Lang. Speech* **2002**, *45*, 207–228. [CrossRef] [PubMed]
69. Friedrich, C.K. Neurophysiological correlates of mismatch in lexical access. *BMC Neurosci.* **2005**, *6*, 64. [CrossRef] [PubMed]
70. Chan, J.S.; Wade-Woolley, L. Explaining phonology and reading in adult learners: Introducing prosodic awareness and executive functions to reading ability. *J. Res. Read.* **2018**, *41*, 42–57. [CrossRef]
71. Heggie, L.; Wade-Woolley, L. Prosodic awareness and punctuation ability in adult readers. *Read. Psychol.* **2018**, *39*, 188–215. [CrossRef]
72. Leong, V.; Hämäläinen, J.; Soltész, F.; Goswami, U. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *J. Mem. Lang.* **2011**, *64*, 59–73. [CrossRef]
73. Leong, V.; Goswami, U. Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hear. Res.* **2014**, *308*, 141–161. [CrossRef] [PubMed]
74. Mundy, I.R.; Carroll, J.M. Speech prosody and developmental dyslexia: Reduced phonological awareness in the context of intact phonological representations. *J. Cogn. Psychol.* **2012**, *24*, 560–581. [CrossRef]
75. Thomson, J.M.; Fryer, B.; Maltby, J.; Goswami, U. Auditory and motor rhythm awareness in adults with dyslexia. *J. Res. Read.* **2006**, *29*, 334–348. [CrossRef]
76. Dickie, C.; Ota, M.; Clark, A. Revisiting the phonological deficit in dyslexia: Are implicit nonorthographic representations impaired? *Appl. Psycholinguist.* **2013**, *34*, 649–672. [CrossRef]
77. Patel, A.D. Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front. Psychol.* **2011**, *2*, 142. [CrossRef] [PubMed]
78. Patel, A.D. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* **2014**, *308*, 98–108. [CrossRef] [PubMed]
79. Tierney, A.; Kraus, N. Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* **2014**, *8*, 949. [CrossRef] [PubMed]
80. Gordon, R.L.; Fehd, H.M.; McCandliss, B.D. Does music training enhance literacy skills? A meta-analysis. *Front. Psychol.* **2015**, *6*, 1777. [CrossRef] [PubMed]
81. Francois, C.; Chobert, J.; Besson, M.; Schon, D. Music training for the development of speech segmentation. *Cereb. Cortex* **2013**, *23*, 2038–2043. [CrossRef] [PubMed]
82. Breen, M.; Clifton, C. Stress matters: Effects of anticipated lexical stress on silent reading. *J. Mem. Lang.* **2011**, *64*, 153–170. [CrossRef] [PubMed]
83. Magne, C.; Gordon, R.L.; Midha, S. Influence of metrical expectancy on reading words: An ERP study. In Proceedings of the Speech Prosody 2010 Conference, Chicago, IL, USA, 10–14 May 2010; pp. 1–4.
84. Fodor, J.D. Learning to parse? *J. Psycholinguist. Res.* **1998**, *27*, 285–319. [CrossRef]
85. Ashby, J.; Clifton, C., Jr. The prosodic property of lexical stress affects eye movements during silent reading. *Cognition* **2005**, *96*, B89–B100. [CrossRef] [PubMed]
86. Kentner, G.; Vasishth, S. Prosodic Focus Marking in Silent Reading: Effects of Discourse Context and Rhythm. *Front. Psychol.* **2016**, *7*, 319. [CrossRef] [PubMed]
87. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Available online: https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf (accessed on 1 November 2018).



brain sciences

MDPI

Article

Impaired Recognition of Metrical and Syntactic Boundaries in Children with Developmental Language Disorders

Susan Richards * and Usha Goswami

Centre for Neuroscience in Education, University of Cambridge, Cambridge CB2 3EB, UK; ucg10@cam.ac.uk (U.G.)

* Correspondence: susan.richards@cantab.net; Tel.: +44-1223-333550

Received: 19 December 2018; Accepted: 31 January 2019; Published: 5 February 2019

Abstract: In oral language, syntactic structure is cued in part by phrasal metrical hierarchies of acoustic stress patterns. For example, many children's texts use prosodic phrasing comprising tightly integrated hierarchies of metre and syntax to highlight the phonological and syntactic structure of language. Children with developmental language disorders (DLDs) are relatively insensitive to acoustic stress. Here, we disrupted the coincidence of metrical and syntactic boundaries as cued by stress patterns in children's texts so that metrical and/or syntactic phrasing conflicted. We tested three groups of children: children with DLD, age-matched typically developing controls (AMC) and younger language-matched controls (YLC). Children with DLDs and younger, language-matched controls were poor at spotting both metrical and syntactic disruptions. The data are interpreted within a prosodic phrasing hypothesis of DLD based on impaired acoustic processing of speech rhythm.

Keywords: language disorder; rhythm; prosody

1. Introduction

Children with Developmental Language Disorder (DLD) have persistent difficulties with learning language that are not associated with a known condition, such as sensori-neural hearing loss or Autism Spectrum Disorder [1]. Prevalence of the disorder is estimated at approximately 7% in primary school populations [2–4], and children with DLD can face a variety of challenges in accessing education and employment. Children with DLD typically have difficulty with the accurate processing and production of grammatical structures in speech [5–7].

Although the implications of having DLD are well-documented, and DLD is found across languages, the underlying causes are as yet unclear. A range of perceptual and cognitive hypotheses have been proposed, including impaired rapid auditory processing [8], impaired phonological memory [9] and genetically determined grammatical deficits [7]. One aspect of language processing that has not attracted significant research attention is the processing of language rhythm. The concept of language rhythm is not consistently defined in the literature and has often been regarded as a purely temporal phenomenon [10]. Others, however, have conceptualised linguistic rhythm in terms of the patterning of syllable prominence, with some syllables being acoustically more prominent than others [11]. Regarding the rhythm of spoken English, syllable prominence can be thought of in terms of strong or stressed syllables (the more prominent) and weak or unstressed syllables (the less prominent). For example, in the word baNAna, the second syllable 'NA' is more prominent than the first syllable 'ba' and third syllable'na'. Accordingly, this word has a weak-strong-weak rhythmic structure, with the second syllable 'NA' carrying the primary stress. The patterning of strong and weak syllables across words, phrases and sentences thus contributes to the perception of language rhythm in addition to temporal factors. These patterns, made up of strong and weak syllables, can

be grouped hierarchically into larger units via prosodic feet, in which one or more weak syllables is grouped with a strong syllable to form a temporal unit. This concept is a familiar one in certain kinds of poetry, in which patterns of recurring syllable rhythms are grouped to fit a higher-order temporal structure, for example via trochees (strong-weak syllable groupings repeating) or dactyls (strong-weak-weak groupings repeating). The pattern of groupings of strong and weak syllables into temporal units is commonly referred to as 'metre'. For the English nursery rhyme 'Jack and Jill went up the hill', a perfect metrical poem, a trochaic rhythm is used, whereas for the nursery rhyme 'Pussycat pussycat where have you been?' a dactyl structure is repeated.

There are sound theoretical reasons for regarding efficient rhythmic processing as a key foundation skill for language development, making the study of the potential impacts of an early difficulty with efficient rhythmic processing of interest in attempting to understand developmental language disorders. Infants are exposed to the rhythmic aspects of language before birth and as newborns are able to use rhythmic properties to differentiate between languages [12,13]. Rhythmic sensitivity is accordingly considered a precursor of language acquisition, with the earliest representations of the speech signal encoding its rhythmic structure. Subsequent aspects of language, such as semantics and syntax, may be scaffolded onto these rhythmic representations [14]. Infants have been shown to use rhythmic aspects of language to establish structured linguistic representations at the level of word boundaries [15], lexical representations [16,17] and larger-grained grammatical units, such as phrases and clauses [18,19]. If rhythm is able to act in a bootstrapping role for subsequent language, then a difficulty in processing rhythm at the earliest stages of development could have a significant impact on the child's subsequent trajectory of language development. In this study, we investigate the potential impact that a rhythmic processing difficulty might have at the interface of rhythm and syntactic structure. This is of particular interest since children with DLD typically present with difficulties in the accurate processing and production of linguistic syntax [20], and are known to have difficulties in processing linguistic stress patterns [21]. Accordingly, it is possible that difficulties in processing prosodic phrasing and prosodic hierarchies dependent on stress patterning may underlie their syntactic difficulties [22].

Impaired auditory sensory processing skills in children with DLD appear to contribute to their impaired processing of syllable stress patterns [21]. Four key acoustic parameters contribute to the perception of stress: frequency, intensity, duration and amplitude envelope rise time (AERT) [23]. Stressed syllables tend to be of a higher frequency than unstressed syllables, have longer durations and are of a higher intensity than unstressed syllables [23]. The fourth parameter, AERT, refers to the length of time between the between the onset of a sound and the point at which its amplitude reaches peak intensity. In speech, this is typically measured as the rise in amplitude from the beginning of a syllable until the speaker reaches the peak of the syllable nucleus (vowel). Stressed syllables have larger rise times, with a greater change in amplitude until the amplitude peaks at the syllable nucleus, whilst unstressed syllables have smaller changes in amplitude before the peak of the nucleus is reached. In order to speak deliberately to a rhythm, the speaker times their production of the rise times of the vowels in each stressed syllable. Children's sensory processing of frequency, duration, intensity and AERT are therefore likely to be central to their ability to differentiate syllable stress patterns and prosodic hierarchies.

Research into the frequency sensitivity of children with DLD has produced mixed results, with some cohorts of language-impaired children being found to have reduced frequency discrimination skills [22,24], whilst other groups have not differed from age-matched controls [21,25]. Duration discrimination has reliably been found to be poorer in children with DLD [21,22,26], whilst tests of intensity discrimination have found no difference between children with DLD and age-matched controls [24,26]. Several studies have shown impaired discrimination of AERT in children with DLD [21,22,26,27], leading to our first investigations into impaired speech rhythm in DLD [26]. Children with DLD also have difficulties with non-linguistic aspects of rhythmic processing. For example, Corriveau and Goswami [28] asked children with DLD to tap to a metronome beat and found that they were considerably poorer at synchronising their taps with the metronome than either

age-matched or language-matched control children at rates of 2 Hz and 1.5 Hz (rates that broadly correspond to typical inter-stress intervals found in speech, [29]). A widely-studied family, known as the KE family, some members of whom display a hereditary form of DLD characterised by articulation difficulties, have also been tested with tasks measuring sensitivity to non-speech pitch and rhythm. Affected members performed more poorly on tests of rhythmic perception and production, indicating a level of rhythmic difficulty for those who also displayed language difficulties [30]. Indeed, tapping to a beat is also impaired in children who stutter [31].

Regarding relations with linguistic processing, in their study of children with DLD, Cumming et al. [32] reported that individual differences in a speech rhythm matching task and a musical beat perception task were significant predictors of children's scores in standardised measures of receptive and expressive language development. Those children with DLD who had better rhythm matching or better musical beat perception had better language scores than those with poorer rhythmic skills. Both Corriveau and Goswami [28] and Cumming et al. [32] reported that individual differences in beat synchronisation contributed unique variance to measures of language and literacy. Weinert [33] also linked rhythmic processing with language ability, finding that children with DLD who did more poorly in a rhythm discrimination task were also poorer at learning an artificial language. Finally, links between rhythmic processing and language skills have also been reported for typically developing children. Gordon et al. [34] found that performance in a test of rhythm discrimination correlated significantly with scores in expressive morpho-syntax in 6-year-old children with no language impairments, accounting for 48% of variance in scores. Accordingly, proficiency in rhythmic processing may be linked to better syntactic skills across the ability range.

One plausible reason for a relationship between rhythm and syntax could be that children with better rhythmic skills may be better at exploiting prosodic phrasing in order to bootstrap language learning [21,22,32]. There is evidence that prosodic phrasing contains cues to syntactic structure, and that both adult language-listeners and infant language-learners make use of these cues in order to parse the speech stream and comprehend language. For example, Price, Ostendorf, Shattuck-Hufnagel and Fong [35] found that adult listeners were able to disambiguate between two possible syntactic parsings of phonologically identical sentences by using prosodic features, such as intonational phrase boundaries and size of prosodic breaks (duration of pauses). Infant experiments have employed preference paradigms, in which infants aged between 7 and 10 months are played recordings with pauses inserted either clause-finally (i.e., coinciding with a syntactic boundary) or mid-clause [18]. The infants demonstrated a preference for stimuli where the pauses were clause-final. A similar preference was also demonstrated by 9-month-old infants for phrase-final pauses [19]. This indicates that, before the end of their first year, infants are already sensitive to the typical coincidence of prosodic and syntactic cues found in the language environment. Jusczyk et al. describe this use of prosodic cues as a 'perceptual precategorisation' [19] (p. 287), thought to enable a more detailed analysis of each resulting perceptual grouping. By aligning the segmentation of perceptual groups with meaningful grammatical units, this precategorisation process would serve perceptually to delimit alternatives, effectively chunking the continuous incoming speech stream and consequently enabling a more nuanced grammatical analysis to take place. By this means, efficient processing of the prosodic structure of speech could pave the way for efficient learning of syntactic organisation.

Whilst much research has been conducted on the nature of the grammatical deficit in DLD, little attention has been paid to the role that prosodic factors may play in the development of grammatical competence, and hence to the role that a difficulty with processing prosodic phrasing might have in the trajectory of the disorder. However, the infant work outlined above indicates that prosodic processing of rhythm patterns may lay the foundations for the discovery of grammatical units at an early stage of language development. In line with this perspective, Demuth has demonstrated that young typically developing children will vary their production of grammatical morphemes depending on the prosodic context. Accordingly, she has argued for a 'Prosodic Licensing' approach to syntactic development, in which the prosodic structure of a given language and the location of a particular

grammatical morpheme in the prosodic contexts afforded by that language will interact to 'license' the use of particular morphemes by the young child [36,37]. Demuth and Tomas [37] argued that an understanding of how prosodic phonology operated to support morphological development in typical development could help to illuminate morpho-syntactic errors by children with DLD. Given our perceptual studies showing that children with DLD have difficulties in processing both speech and non-speech rhythm [21,22,32], children with DLD may also have difficulties in processing the rhythmic aspects of speech that can facilitate the overall acquisition of grammatical structure. If so, this could provide an acoustic, stimulus-driven account of the grammatical difficulties that typify the receptive and expressive language of children with DLD.

The current investigation explores children's sensitivity to prosodic phrasing as a cue to the parsing of the speech stream into smaller, more manageable, grammatical units. Whilst prosodic and syntactic structures do not always coincide in natural speech, there is nonetheless a core area of children's typical language exposure in which the two levels are tightly integrated, namely the realm of children's oral and textual culture. Children's stories and nursery routines draw heavily on rhythmic devices to structure language, as aspects of children's linguistic life, such as nursery rhymes and clapping games, depend on the integration of repetitive language and repetitive rhythm. A further aspect of a typical child's linguistic environment is children's literature, which frequently relies heavily upon rhythm and rhyme. Many successful children's authors build upon the playfulness of oral rhymes, with writing characterised by strong, repetitive rhyme and rhythm frameworks. We hypothesised that the predominance of rhythm and rhyme in these texts may serve a scaffolding function in developing children's awareness of prosodic-syntactic units. Accordingly, we selected a representative story by former UK children's laureate Julia Donaldson called '*Room on the Broom*': a story with a strong rhythmic format [38].

The rhythmic format of *Room on the Broom* creates a tight integration of prosody and syntax and hence contains rich structural cues to grammar. The child is exposed to cues at multiple hierarchical layers, drawing their attention simultaneously to the phonological, prosodic and syntactic structure of the language. The property of rhyme emphasises the phonological structure of words by drawing attention to the onset-rime division, whilst also providing a guide to linguistic structure since each rhyme occurs at the end of a syntactic unit (be that clause or phrase). The overarching metrical structure also draws attention to the rhyme boundary point, since it occurs at regular intervals every four metrical feet. Within that metrical structure, there are further subdivisions into pairs of metrical feet, each of which also generally represents a complete syntactic unit. The metrical structure is therefore not an arbitrary form superimposed on the syntax of the text, but the two structures form a rich and highly integrated input which serves to highlight and reinforce the rhythmic and syntactic properties of language.

An illustration is provided as Figure 1, which decomposes the structural embedding in the opening sentence of this popular children's book. The figure marks out the major syntactic structures (shown above the text in green) and the major prosodic structures (shown below the text in blue). The figure shows that the major groupings in the syntactic structure are mirrored by major prosodic boundaries (the dashed red lines) in the prosodic structure. The prosodic boundaries are hierarchically nested such that the larger the prosodic-syntactic unit, the greater the overlap of boundary cues. Accordingly, the end of each rhyme line represents the combined boundary of four different levels of metrical analysis, as well as the boundary of a major syntactic unit. The prosodic structure is built around the stressed syllables, which serve to demarcate the end of a metrical foot (predominantly anapaest; i.e., weak-weak-Strong (wwS)). The symmetry is not faultless, as can be seen from 'a very tall hat', in which the lexical word 'very' crosses the boundary of the metrical foot; however, for the majority of the couplet, there is a strong coincidence of prosodic and syntactic boundaries. Given this level of dovetailing between the prosodic and syntactic structures, our study aimed to measure to what extent the children with DLD were able to integrate these two systems of representation.

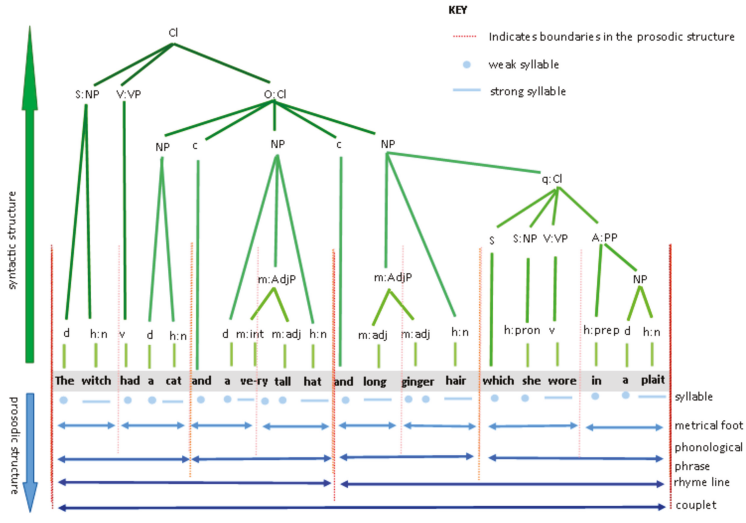


Figure 1. A diagram to illustrate the syntactic and prosodic structure of a line from Room on the Broom. Abbreviations: d-determiner; h:n-noun, head of noun phrase; v-verb; m:int-modifier:intensifier; m:adj-modifier:adjective; h:pron-pronoun, head of noun phrase; h:prep-preposition, head of prepositional phrase; c-conjunction, q-qualifier; Cl-clause; S-subject; O-object; A-adverb; NP-noun phrase; VP-verb phrase; PP-prepositional phrase.

2. Materials and Methods

2.1. Participants

Fifty-nine (59) children took part in the study. Children were divided into three groups: 13 had developmental language disorder (DLD group) Mean (*M*) age 102 months, range 77–140; 24 were age-matched typically developing controls (AMC group) *M* 107 months, range 77–132; and 22 were younger, language-matched controls (YLC group) *M* 66 months, range 57–74. All of the children were attending mainstream schools across state and private sectors in the East of England.

Children with DLD were recruited via their schools by asking teachers to nominate pupils whom they considered displayed difficulties with language. Those children identified by their teachers then completed four standardised language tests, the British Picture Vocabulary Scales-2nd Edition (BPVS II) [39], and three subtests of the Clinical Evaluation of Language Fundamentals UK-3rd Edition (CELF3^{UK}) [40]: the Recalling Sentences, Concepts & Directions and Formulated Sentences subtests. Children who scored at or below −1.33 SD on at least two of the four tests went on to be included in the DLD group. Age-matched children (AMC group) were largely recruited from the same schools as the children with DLD and also completed the four standardised language tests. Only children scoring higher than −1 SD on all four tests were included in the study as part of the AMC group. The younger children (YLC group) all attended a single school who agreed to take part for this purpose. Children in the YLC group completed the BPVS II and the CELF3^{UK} Recalling Sentences subtest only. All children also completed the Block Design, Picture Completion and Digit Span subtests of the Wechsler Intelligence Scale for Children 3rd Edition (WISC III) [41] as measures of phonological memory and non-verbal intelligence quotient (IQ). Results of the standardised tests are displayed in Table 1.

As different groups did different tasks, one-way ANOVAs by group or independent samples t-tests were used to assess group differences. The matching was confirmed as the DLD group did not differ significantly from the AMC group on Age (months) (p = 0.675), whilst both the DLD and AMC groups were significantly older than the YLC group (p = 0.000). The DLD and YLC groups did not differ significantly from each other on measures of language (Recalling Sentences p = 0.434; BPVS II

$p = 0.641$), whilst both groups were significantly different from the AMC group ($p = 0.000$). The DLD group were also significantly different from the AMC group on the additional language measures of Formulated Sentences and Concepts & Directions ($p = 0.000$).

For the IQ measures, the DLD group scored within one standard deviation of the standardised mean for both tasks, indicating that their non-verbal IQ was within typical norms; however, their scores as a group were nonetheless significantly lower than those of the AMC group (Picture Completion $p = 0.017$; Block Design $p = 0.014$). The DLD group also had a significantly lower Digit Span score than the AMC group ($p = 0.000$).

Table 1. Results of standardized tests by group (Language: raw scores; intelligence quotient (IQ): scaled scores): one-way ANOVAs and independent samples *t*-tests.

Test	AMC Mean (SD)	DLD Mean (SD)	YLC Mean (SD)	*df*	*F*	*p*
Age (months) [b,c]	107.33 (16.447)	102.31 (17.504)	66.18 (4.532)	2, 23.104 [d]	89.068	0.000
Language						
Recalling Sentences [a,b]	46.38 (12.917)	21.15 (6.950)	24.29 (7.309)	2, 33.565 [d]	32.388	0.000
BPVSII [a,b]	98.92 (21.279)	71.54 (11.163)	67.73 (13.364)	2, 34.879 [d]	18.686	0.000
				df	***t***	***p***
Formulated Sentences [a,]	31.58 (7.638)	16.08 (5.499)	-	35	6.452	0.000
Concepts & Directions [a,]	25.00 (4.283)	14.54 (6.293)	-	18.181 [d]	5.359	0.000
IQ [e]						
Picture Completion [a,b]	10.42 (2.518)	8.46 (1.664)	-	35	2.510	0.017
Block Design [a,b]	10.42 (2.888)	8.15 (1.725)	-	35	2.577	0.14
Digit Span [a,b]	9.96 (1.574)	6.62 (1.193)	-	35	6.675	0.000

[a] Aged-matched children (AMC) > Developmental Language Disorder (DLD); [b] AMC > younger, language-matched control (YLC); [c] DLD > YLC; [d] adjusted *F* and *df* used due to significant Levene's test; [e] IQ subtests are scaled scores: $M = 10$, $SD = 3$. BPVSII, British Picture Vocabulary Scales-2nd Edition.

2.2. Materials

The aim of the experimental task was to investigate whether children were sensitive to the coincident boundaries of prosodic and syntactic units as exemplified in the rhythmic texts that form a central part of children's literature. The rhyming couplets in the chosen text consisted of two lines, each of which contained four stressed syllables (in capitals):

the WITCH had a CAT and a VEry tall HAT

Each rhyme line was also composed of two syntactic units, each of which contained two stressed syllables (i.e., two metrical feet):

the WITCH had a CAT and a VEry tall HAT

This clear and regular correspondence between metrical and syntactic units continues throughout the text. From an analysis of the whole book, 10 couplets were chosen to form the stimulus set. Five couplets had the regular pattern:

__SwwS; wwSwwS e.g., 'Down!' cried the witch, and they flew to the ground,
_wSwwS; wwSwwS They searched for the hat, but no hat could be found.

The other five couplets had the regular pattern:

wSwwS; wwSwwS e.g., Then out from a tree, with an ear-splitting shriek,
wSwwS; wwSwwS There flapped a green bird, with the bow in her beak.

To investigate whether metrical groupings influence detection of syntactic-prosodic units, three conditions were created: Metrical-Coincident; Metrical-NonCoincident; and NonMetrical-NonCoincident. A pause was created in the spoken recordings of the couplets to create the three different conditions, as detailed in Table 2.

Table 2. Condition Characteristics with example stimuli.

Condition	Description	Example Couplet	Metrical Structure	
Metrical: Coincident (Met-Co)	A pause at the end of every two metrical feet. This pause therefore coincided with natural prosodic and syntactic groupings.	'Down!' cried the witch / and they flew to the ground / they searched for the hat / but no hat could be found	__S wwS _wS wwS	wwS wwS wwS wwS
Metrical: NonCoincident (Met-NonCo)	A pause that created regular metrical groupings but that did not coincide with syntactic boundaries.	'Down!' cried / the witch and they flew to / the ground they searched for / the hat but no hat could / be found	__S wSw wS_ wSw wS	w wSw wSw wSw
Non-Metrical: Non-Coincident (NonMet-NonCo)	A pause that created irregular rhythmic groupings and also did not coincide with syntactic boundaries	'Down!' cried the witch and they flew/ to the ground they searched / for the hat but no / hat could be / found	__S wwS wwS Sww S	wwSwwS wS ww

It should be noted that the syntax in each version remains identical, only the prosodic grouping is altered. Accurate judgements therefore would not reflect syntactic knowledge per se, but rather intuitive knowledge of how prosody and syntax typically interact.

2.3. Recording

All stimuli were recorded in a soundproof booth by a female speaker of British English using a TASCAM DR-100 recorder via a SHURE SM58 condenser microphone. A regular beat was induced in the speaker using a priming metronome beat in one ear (not audible on the recording) with an inter-beat interval of 750 ms. The stimulus was then spoken so as to align the stressed syllables of the recording with the beats at 750 ms intervals. The precision of this timing was then verified and adjusted as necessary with Audacity software. The inserted pause was equivalent to the insertion of one silent stressed syllable interval, such that there was 1500 ms between the preceding and the following stressed syllable.

Each couplet was recorded in three different versions: Met-Co, Met-NonCo and NonMet-NonCo. The couplets were then arranged in three blocks of 10 couplets, with each block containing a counterbalanced mix of all three conditions (e.g., four Met-Co, four Met-NonCo and three NonMet-NonCo). Each couplet occurred only once in each block, and the order of couplets was fixed across blocks. Each block was listened to in a separate session, with the order of presentation of blocks across the three sessions randomised across participants. Each child ultimately listened to each block and therefore recorded scores for all three versions of each couplet.

2.4. Procedure

Each child completed the task individually in a quiet area at school. In the first testing session, the experimenter read the entire storybook to the child so that each child was familiar with the text as a whole. Each task block was then presented as part of a wider set of experimental tasks in subsequent sessions.

The task was contextualised by talking about how when reading out loud it was important to take a breath in a 'sensible place, where it fits with the words' because otherwise 'it . . . sounds interrupted . . . like . . . this.' It was then explained that they were going to hear someone reading the words from 'Room on the Broom' but that sometimes the reader would take a breath in a 'funny place; where it sounds wrong; like it doesn't fit'. The task was presented using a laptop computer running Presentation software with the children listening through Sennheiser HD650 headphones via a UGM96

soundcard. The corresponding picture from the book was displayed during the playback of each stimulus. Responses and Response Times were recorded using key presses on the laptop keyboard. Children were asked to press the key with the green 'tick' sticker if they thought the breath sounded like it was in a sensible place which fitted with the words, or the key with the pink 'cross' sticker if they thought it sounded wrong or interrupted. These buttons corresponded to the 'L' and 'A' buttons on the keyboard.

Each presentation of a block of 10 trials was preceded by three practice trials, during which children were given feedback to ensure they understood the task. This was followed by presentation of the 10 experimental stimuli, during which children were given only generic encouragement.

2.5. Auditory Threshold Estimation Tasks

Children in the AMC and DLD groups also completed four auditory threshold (AT) estimation tasks designed to probe sensitivity to four key acoustic indicators of stress in speech: Amplitude Envelope Rise Time (AERT); Frequency; Duration and Intensity. These were presented via the laptop computer using the Dino software program.

The AT tasks all followed a similar format in which, for each trial, the child heard three tones and was asked to choose which tone was different from the other two. Presentation was always in an AXB format where the middle tone (X) was always the reference tone, one of A and B was also the reference tone whilst the other differed from the reference by a stipulated amount (see below). Children were shown a picture of three cartoon animals and were told that each animal would make a noise and jump at the same time. Their job was to choose the animal that made the different sound. Responses were through mouse click or by pointing. The program provided continuous feedback, with correct answers rewarded with a colourful icon and incorrect answers indicated by an auditory sigh. Each block was preceded by five practice trials during which children received live feedback and further explanation of the task. Tasks were presented in a fixed order of Frequency, Intensity, AERT, Duration.

Frequency: Stimuli consisted of 200 ms tones played at 80.95 dB. The minimum frequency was 250 Hz (reference tone) and the maximum was 279.92 Hz. Increments between tones were of 0.0513 semitones. Children were asked to choose the tone with the different, higher sound.

Intensity: Stimuli consisted of 200 ms tones at a frequency of 250 Hz. The minimum intensity was 61.472 dB and the maximum was 80.95 dB (reference tone). Intensity intervals between levels were of 0.5128 dB. Children were asked to choose the tone with the different, quieter sound.

AERT: Stimuli consisted of 800 ms tones played at 80.95 dB at a frequency of 531.25 Hz. The minimum rise time was a 15 ms slope (reference tone) and the maximum was a 300-ms slope. Fall-off was consistent at 50 ms. Increments to the slope between levels were of 7.0377 ms. Children were asked to choose the tone with the different, gentler beginning.

Duration: Stimuli consisted of tones played 80.95 dB at a frequency of 250 Hz. The minimum duration was 400 ms (reference tone) and maximum duration was 595 ms. Increments in duration between levels were of 5.1282 ms. Children were asked to choose the tone with the different, longer sound.

The Dino program uses a staircasing procedure in order to estimate the auditory threshold. Trials begin with the maximum difference between stimuli (i.e., levels 1 and 40) and initially use a two-up, one-down procedure. This means that two correct answers result in a narrowing of the difference between stimuli, whilst one incorrect answer results in a widening of the difference between stimuli. After four reversals, the procedure is three-up, one-down. Initially, stimuli pairings change by eight levels in each stepchange (e.g., moving from levels 1:40 to levels 1:32); after four reversals, this becomes progressively four-, two- and one-level stepchanges. The final threshold figure is taken as the mean level from the fourth reversal.

Ethical approval for the study was obtained from the Cambridge Psychology Research Ethics Committee reference PRE.2009.02.

3. Results

3.1. Accuracy

Children's scores were summed across blocks and calculated according to number of responses correct (i.e., identifying stimuli in condition Met-Co as correct with a 'tick' press and those in conditions Met-NonCo/NonMet-NonCo as incorrect with a 'cross' press). The maximum score was therefore 30, with a maximum score of 10 for each condition.

Due to software errors, two children from each group unintentionally listened to the same block presentation twice. These children's scores were removed from the summary analysis. From a boxplot of scores by group, one AMC child appeared as an outlier. This was confirmed by calculating this child's z-score; this child was still an outlier and so these scores were also removed. Scores for each of the conditions by group are given in Table 3.

Table 3. Accuracy scores by condition and group.

Score	AMC Mean (SD)	DLD Mean (SD)	YLC Mean (SD)
Overall Score	26.14 (3.692)	20.36 (5.519)	18.60 (5.995)
Met-Co	7.71 (3.002)	8.00 (2.449)	7.65 (2.183)
Met-NonCo	9.1 (1.261)	5.82 (2.857)	5.5 (3.456)
NonMet-NonCo	9.33 (1.278)	6.55 (3.236)	5.45 (3.268)

As will be recalled, the different conditions were mixed together during task presentation to the child; however, in order to judge whether the groups differed in sensitivity to the task, d' was calculated for each group. Hits were defined as selecting the 'tick response for target tick and the cross response for target cross'. The resulting mean group values were AMC d' = 2.442, DLD d' = 1.395, YLC d' = 1.045. A one-way ANOVA (DV d') revealed that the AMC group was significantly more sensitive than the DLD group ($p = 0.033$) and the YLC group ($p = 0.001$) (Games–Howell corrections). The sensitivity of the DLD and YLC groups did not differ from each other. Accordingly, the DLD children were less sensitive to prosodic-syntactic groupings than would be expected for their age, but were not less sensitive to these groupings than would be expected for their language attainment levels.

In order to compare the groups in terms of accuracy of performance, a 3 × 3 repeated-measures ANOVA (Group: AMC, DLD, YLC; Condition: Met-Co, Met-NonCo, NonMet-NonCo) was conducted. The ANOVA showed no significant main effect of Condition, $F(1.193,58.473) = 2.004$, $p = 0.160$ (Greenhouse–Geisser correction); however, there was a significant effect of Group, $F(2,49) = 12.077$, $p = 0.000$ and the Condition*Group interaction was also significant, $F(4,98) = 4.465$, $p = 0.002$. Pairwise comparisons (Bonferroni) indicated that the AMC group scored more highly than both the DLD group ($p = 0.011$) and the YLC group ($p = 0.000$), whilst there was no significant difference in score between the DLD and YLC groups. This is consistent with the d' analysis.

The significant Group*Condition interaction was explored by running a series of one-way ANOVAs for each condition. The ANOVAs revealed no main effect of group for the Met-Co condition, but a significant group effect for the Met-NonCo, $F(2,20.810) = 14.243$, $p = 0.000$ and for the NonMet-NonCo, $F(2,20.556) = 14.435$, $p = 0.000$ (Welch's F) conditions. Post-hoc tests (Games–Howell) showed that the AMC group were more accurate than the DLD and YLC groups in both of these conditions (Met-NonCo $p = 0.009$, DLD, $p = 0.001$, YLC; NonMet-NonCo $p = 0.044$, DLD, $p = 0.000$ YLC). The DLD and YLC groups did not differ significantly from each other in either condition.

Inspection of the graphed results (Figure 2) helps to illustrate the differing effect of condition for the three groups. As the graph shows, the AMC group scored more highly for the non-coincident Met-NonCo and NonMet-NonCo conditions than for the Met-Co condition, whilst the YLC group's scores were lower for the non-coincident stimuli than the coincident Met-Co type. An unexpected

pattern in the results was the relatively poor performance of the AMC group in the Met-Co condition. When compared to their performance in the non-coincident conditions, this suggests that the AMC children were slightly more likely to reject a correct rendition than to accept an incorrect one. The graph also indicates that the children with DLD were as accurate at identifying when the coincidence of prosodic-syntactic cues was correct (Met-Co) as were the AMC and YLC children. However, once these structures were disrupted, their performance fell markedly, suggesting poor sensitivity both to regular metrical groupings that did not coincide with a syntactic boundary (the Met-NonCo condition) and poor sensitivity to irregular groupings that did not coincide with a syntactic boundary. If the children with DLD were sensitive to metrical structure but did not relate this to syntactic structures, then we would expect lower accuracy for Met-NonCo than NonMet-NonCo. However, the performance of the DLD children in both disrupted conditions was statistically equivalent, suggesting that they were insensitive to both speech rhythm and its relationship to syntax.

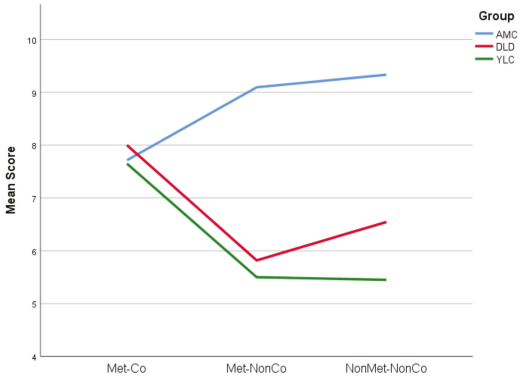


Figure 2. A graph showing the mean score for each condition by group. AMC: Aged-matched children; DLD: Developmental Language Disorder; YLC: younger, language-matched control.

3.2. Reaction Times

In order to explore group performance in more detail, reaction time data (RT) were also analysed. The mean RT was calculated for each child for each Condition (regardless of correctness of response). Data are shown in Figure 3 and Table 4.

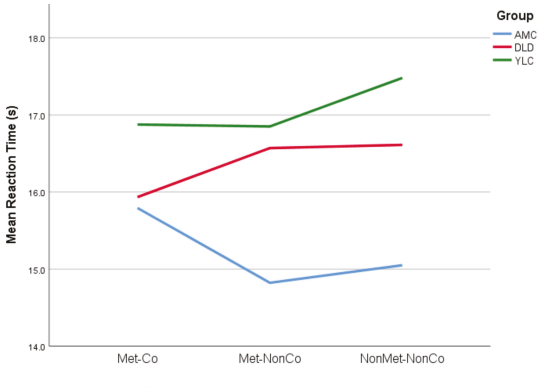


Figure 3. A graph of mean reaction times to each condition by group.

A repeated-measures 3 × 3 ANOVA (Group: (AMC, DLD, YLC) × Condition: (Met-Co, Met-NonCo, NonMet-NonCo)) showed no significant effect of condition on Reaction Times,

$F(2,98) = 0.796$, $p = 0.454$, nor of group, $F(2,49) = 2.968$, $p = 0.061$. However, there was a significant Group*Condition interaction, $F(4,98) = 2.877$, $p = 0.027$.

Table 4. Reaction Times (s) by Condition and Group.

	Met-Co Mean (SD)	Met-NonCo Mean (SD)	NonMet-NonCo Mean (SD)
AMC	15.7932 (1.4338)	14.8225 (3.0666)	15.0502 (2.2551)
DLD	15.9342 (2.4686)	16.5701 (3.3560)	16.6115 (2.6472)
YLC	16.8754 (2.5224)	16.8506 (2.9466)	17.4805 (2.8310)

In order to explore the Group*Condition interaction, a series of one-way ANOVAs were run for each Condition. There was no significant effect of group for the Met-Co and the Met-NonCo conditions. There was, however, a significant effect of group for the NonMet-NonCo task, $F(2,49) = 4.667$, $p = 0.014$. Pairwise comparison (Games–Howell) of the means for each group showed a significant difference between the YLC and AMC groups in the NonMet-NonCo condition: the younger children were significantly slower ($p = 0.012$). Furthermore, inspection of the graph in Figure 3 shows that both the YLC and DLD groups tended to be slower in response than the AMC group. Accordingly, for the two non-coincident tasks, the DLD children appeared to respond within a similar timeframe to the younger YLC children.

For completeness, the group*condition interaction was also explored using one-way repeated-measures ANOVAs by group. There were significant effects of condition for the AMC children, $F(1.339,26.774) = 4.181$, $p = 0.04$ (Greenhouse–Geisser correction) but no significant effect of condition for the DLD children, $F(2,20) = 1.286$, $p = 0.298$ nor for the YLC group, $F(2,38) = 1.687$, $p = 0.199$. For the AMC group, despite the significant overall effect of condition, post-hoc pairwise comparisons (Bonferroni) showed no significant differences between conditions, although there was a trend for the responses in NonMet-NonCo to be quicker than those of the Met-Co ($p = 0.077$). In other words, there was a tendency for the AMC group to take longer to decide that the coincident stimulus was correct than to decide that the non-coincident stimulus was incorrect, even though they performed well in both conditions. This result tallies with observations during testing: AMC children often pressed the [x] button as soon as they heard the first non-coincident boundary, immediately confident that this presentation was 'wrong'. In order to be sure that the coincident stimulus was correct, however, the stimulus had to be listened to in its entirety. This may explain this difference in response time trends for the AMC group.

A different effect was observed for the DLD group, who rarely pressed the response buttons before the full stimulus was played. This is reflected in the lack of variation in their response times. It seems that, for the DLD children, in marked contrast to the AMC children, there was no confident decision-making about aberrant prosodic-syntactic groupings reflected in quicker response times. The DLD children puzzled for an equally long time over the coincident stimuli as they did over the non-coincident stimuli. In doing so, they presented a response profile that was statistically comparable to the younger language-matched children.

3.3. Acoustic Threshold Estimation Tasks

The AMC and DLD groups both completed the four AT tasks in AERT, Frequency, Duration and Intensity. Two scores were not recorded by the software: one Frequency score (one DLD child) and one Intensity score (one AMC child). A series of independent samples t-tests was conducted to examine any differences in acoustic sensitivity between the two groups (see Table 5).

Table 5. Results of *t*-tests by group for auditory threshold (AT) tasks.

Task	AMC Mean (SD)	DLD Mean (SD)	*df*	*t*	*p*
AERT (ms) [a]	114.628 (76.0437)	207.193 (69.971)	30	−3.357	0.002
Frequency (Semitones) [a]	0.5681 (0.4428)	1.3747 (0.5117)	29	−4.512	0.000
Duration (ms) [a]	79.8949 (39.1790)	138.4288 (47.8637)	30	−3.720	0.001
Intensity (dB)	−2.9863 (1.51905)	−3.291244 (1.5838)	29	0.527	0.602

[a] AMC < DLD. AERT, Amplitude Envelope Rise Time.

The AMC group had significantly lower thresholds (i.e., were able to discriminate more fine-grained differences between stimuli) than the DLD group for the conditions of AERT, Frequency and Duration, whilst there was no significant difference between groups for Intensity. The finding that the AMC and DLD groups performed the Intensity threshold task at equivalent levels shows that the attentional load of the task is not a factor in explaining group performance.

A correlation analysis between acoustic threshold and accuracy score on the experimental task revealed significant correlations between task performance and sensitivity to the acoustic features of AERT ($p = 0.027$), Duration ($p = 0.004$) and Frequency ($p = 0.000$), with the greater the sensitivity to acoustic differences, the more accurate the performance on the task (see Table 6).

Table 6. Correlation coefficients (Pearson one-tailed) for AT tasks and accuracy score.

Task	Duration	Frequency	Intensity	Task Score
AERT	0.458 **	0.794 ***	−0.123	−0.343 *
Duration		0.530 **	−0.084	−0.524 **
Frequency			−0.107	−0.553 **
Intensity				−0.077

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The children with DLD therefore had larger thresholds (i.e., required a greater difference between stimuli in order to discriminate) for the acoustic measures of Frequency, Duration and AERT and all three of these measures were significantly correlated with success on the experimental task.

A series of three-step fixed order multiple regressions were carried out to explore the unique contributions of each of the acoustic parameters of Frequency, Duration and AERT to success on the task once age and non-verbal IQ (NVIQ) were controlled. NVIQ was taken as the mean score across the two subtests of Picture Completion and Block Design. Table 7 shows the results of the equations with accuracy as the dependent variable (i.e., the child's overall score summed across all three conditions).

Table 7. Stepwise regressions showing the unique variance in children's accuracy in judging metrical and syntactic boundaries contributed by the different auditory processing measures.

	b	SEb	β	*t*	*p*	ΔR^2	*p*
Model 1							
Age	−0.011	0.047	−0.036	−0.237	0.814	0.037	0.303
NVIQ	1.030	0.378	0.468	2.722	0.011	0.351	0.000
Frequency	−2.431	1.293	−0.314	−1.880	0.071	0.071	0.071
Model 2							
Age	0.008	0.054	0.022	0.140	0.889	0.063	0.165
NVIQ	1.076	0.433	0.435	2.482	0.019	0.293	0.001
Duration	−0.029	0.018	−0.288	−1.673	0.105	0.058	0.105
Model 3							
Age	0.014	0.056	0.041	0.255	0.801	0.063	0.165
NVIQ	1.309	0.413	0.529	3.168	0.004	0.293	0.001
AERT	−0.009	0.010	−0.145	−0.903	0.374	0.018	0.374

Note: b = unstandardised beta; SEb = standard error of b; β = standardized beta; ΔR^2 = change in R^2.

Age was not a significant predictor of performance in this task (p range = 0.165–0.303); however, NVIQ contributed significant amounts of unique variance for all three equations (range 29.3–35.1%, p range = 0.000–0.001). The greatest unique variance accounted for by the AT tasks was for Frequency (7.1%) followed by Duration (5.8%), although neither was significant once NVIQ was controlled (p = 0.071, 0.105, respectively). AERT also failed to make a significant contribution to overall accuracy once NVIQ was controlled (p = 0.374) contributing 1.8% of unique variance, the smallest amount. The significant correlations between sensitivity to AERT, Duration and Frequency and overall score may therefore have been partly mediated by NVIQ, with the acoustic cues of Frequency and Duration providing smaller (non-significant) additional contributions to the variance in score.

A further set of three-step fixed order regressions was calculated to explore the relationship between the different acoustic parameters and the overall mean response time (calculated by summing all RT values for each child and dividing by 30). These are shown in Table 8.

Table 8. Stepwise regressions showing the unique variance in children's response time in judging metrical and syntactic boundaries contributed by the different auditory processing measures.

	b	SEb	β	t	p	ΔR^2	p
Model 1							
Age	−0.059	0.029	−0.360	−2.016	0.054	0.188	0.015
NVIQ	−0.291	0.234	−0.254	−1.243	0.225	0.048	0.193
Frequency	−0.166	0.801	−0.041	−0.207	0.837	0.001	0.837
Model 2							
Age	−0.059	0.028	−0.367	−2.083	0.047	0.193	0.012
NVIQ	−0.301	0.228	−0.263	−1.321	0.197	0.048	0.186
Duration	−0.003	0.009	−0.062	−0.316	0.754	0.003	0.754
Model 3							
Age	−0.058	0.028	−0.360	−2.046	0.050	0.193	0.012
NVIQ	−0.269	0.211	−0.235	−1.275	0.213	0.048	0.186
AERT	0.000	0.005	−0.005	−0.029	0.977	0.000	0.977

In these regressions, Age was a significant predictor of RT, explaining between 18.8% and 19.3% of unique variance. NVIQ did not contribute significantly to the model, and nor did any of the three acoustic parameters. This suggests that the speed of responding was primarily mediated by the age of the participants, even though the younger LMC children were not included in these analyses. The negative b values indicate that the older the participants were, the faster their responses.

4. Discussion

This study set out to investigate the influence of phrasal metrical hierarchies of acoustic stress patterns on children's capacity to recognise appropriate prosodic-syntactic groupings. In children's literature, the metrical regularities of the texts typically serve to emphasise the syntactic structures of language through the coincidence of prosodic and syntactic boundaries. Here, children listened to different readings of phrases from a children's story, and were asked to indicate when the reader took a breath in a 'funny place, where it sounds wrong, like it doesn't fit'. Two types of disruption were tested, breaths that created metrical groupings that conflicted with syntactic groupings (Met-NonCo) and breaths that violated both metre and syntax (NonMet-NonCo). As children with DLD have known acoustic difficulties with stress, indexed by their insensitivity to amplitude envelope rise time (AERT), and acoustic difficulties with grouping, indexed by their insensitivity to duration [21,22], this may affect their ability to use the prosodic-syntactic grouping typical of representative texts in children's literature to aid grammatical learning. If children have robust knowledge of how prosody and syntax interact, then any violation of these coinciding units should be readily identified. Alternatively, if children are able to detect metrical patterns but are unable to relate these to the overall prosodic-syntactic structure,

then phrases in the condition in which there is an acoustic metrical rhythmic structure that does not coincide with the syntax (condition Met-NonCo) should prove more difficult to reject than phrases in the condition in which there is no consistent metrical acoustic pattern (condition NonMet NonCo). If the children with DLD are insensitive to both acoustic prosody and its relationship with syntax, then there should be no difference in performance between the three different conditions.

Overall, the DLD children were indeed less sensitive to the prosodic-syntactic groupings that we used, as their d' scores were significantly lower than that of the age-matched control (AMC) children. Nevertheless, the DLD children were as sensitive to the prosodic-syntactic groupings as the younger language-matched control (YLC) children, as their d' scores were statistically equivalent. However, their reduced sensitivity did not reflect a lack of effort in the task. The slowed response times of the DLD and YLC children showed that they were analyzing the non-coincident phrases, as they were slower to respond in the two non-coincident conditions. The DLD group were impaired compared to the AMC group at detecting violations of prosodic-syntactic units, and this was shown by both their response time data and their accuracy data. The accuracy of the DLD group was comparable to that of the younger children. This is the pattern that we would expect if DLD children have difficulty in processing language metre and its relationship with syntactic structures. From the accuracy scores, it seems that the most significant impact of metrical grouping appears to be the attention that it draws to the prosodic-syntactic unit, rather than its temporal regularity per se. If metrical regularity alone (regardless of syntax) were influencing responses, we would expect a higher error rate in the Met-NonCo condition than NonMet-NonCo, with DLD children responding to the regular metre in the former condition and judging the stimulus phrases as acceptable. The data did not support this explanation, as the slight difference visible in Figure 2 was not significant.

Indeed, the response time data showed that the DLD and YLC children did not differ in their response times across conditions. If DLD children were confident in using prosody to detect syntactic boundaries regardless of metre, we would expect swift responses to these violations (fast responding for both Met-NonCo and NonMet-NonCo conditions, as found for AMC children). Alternatively, if the DLD children were able to detect the metrical grouping but could not readily relate that perceptual structure to the syntactic groupings, then we would expect Met-NonCo stimuli to produce slower responses due to the conflicting information. This was not the case, suggesting that the DLD children could neither detect the metrical grouping nor integrate it with their expectations of prosodic-syntactic groupings. Inspection of the response time data showed that the children with DLD did not make early detections of errors, almost always choosing to listen to the whole recording. Accordingly, the children with DLD were unable to systematically determine whether structural boundaries had been violated. Their performance in both accuracy and speed of response resembled that of the younger children (YLC group), suggesting a developmental delay in their ability to integrate prosodic and syntactic structures.

Overall, the data suggest that younger children and DLD children have less well-developed schema for how prosody and syntax interact compared to older typically developing children and therefore that this is an aspect of language processing that continues to develop throughout childhood. The DLD children appear to have underdeveloped schema for the interaction of prosody and syntax for their age. This suggests that they may not be processing all of the cues available to them in segmenting the speech stream into prosodic units and grammatical clauses. Instead, they responded similarly to the younger children. Note that the syntax itself was identical in all three conditions, so the test is not one of grammatical structures per se, but of how these structures interact with prosodic units in typical speech.

The range of responses here sits interestingly between experiments with infants, which have shown that infants are sensitive discriminators of pauses inserted within clauses or at clause boundaries [42], and those with adults, who also judge sentences where pauses coincide with phrasal boundaries to be more natural [19]. The older AMC children were more adult-like in their responses, being able to judge both the Met-Co sentences as being natural and the Met-NonCo

and the NonMet-NonCo sentences as 'sounding funny'. The question, however, is why the DLD and, particularly, the YLC children have relatively poor accuracy for the non-coincident stimuli if 9-month-old infants are sensitive to these boundaries. One explanation could lie in task demands. In our experiment, the children were asked to decide explicitly which was the 'correct' version, and so this required a greater degree of metalinguistic awareness than the infants in Jusczyk et al.'s passive listening study [19]. This raises an important conceptual difference between 'sensitivity' in the sense of discriminating between prosodic structures and 'awareness' in the sense of consciously noticing the significance of any discriminated difference. It is possible that the children with DLD were sensitive to the differences between the prosodic structures in the stimuli but were unable to determine whether this observed difference resulted in a pragmatic difference: that the breaks were in a 'funny place'. As we also collected acoustic data, however, and found reduced sensitivity for DLD children in three of the four acoustic parameters that contribute to stress perception, this seems unlikely. It seems more likely that the children's lack of pragmatic awareness stemmed from their poorer discrimination of the acoustic parameters that enabled reliable identification of the prosodic structures. This perceptual difficulty then reduced the ability of the DLD children to integrate information about perceptual prominence with syntactic expectations.

On the basis of our previous acoustic work [21,22,26,28], we proposed that children with DLD may be delayed in developing schema for prosodic-syntactic hierarchies because of impaired sensory processing. In English, the four acoustic parameters that we measured of AERT, duration, frequency and intensity combine to give the percept of stress and thereby linguistic rhythm [23]. Accordingly, acoustic sensitivity to these parameters is likely to influence linguistic rhythm development. The DLD group had significantly higher thresholds for AERT, duration and frequency than did the AMC group, whilst they did not differ in intensity thresholds. This pattern of responses accords with previous studies that have found that children with DLD have impairments in discriminating AERT [21,26,27,43] and duration [21,22,26]. Some studies have also found a frequency impairment [22,24], but see also [21,25]. Poorer sensitivity in these acoustic tasks has previously been associated with poorer performance on tasks probing linguistic stress [21]. Here, we also found a significant association between the acoustic thresholds for AERT, duration and frequency and the capacity to detect violations of prosodic-syntactic boundaries. This suggests that the less sensitive auditory systems of children with DLD may be impacting upon their perception and subsequent integration of prosodic cues with larger-grained syntactic units. However, once non-verbal IQ was controlled in a series of regression equations, none of the acoustic parameters measured accounted for unique variance in task performance. This contrasts with some of our earlier DLD studies, where both AERT and duration measures have explained unique variance in language tasks even after NVIQ has been controlled [22,26]. Processing prosody across larger phrasal units requires the tracking of relative acoustic hierarchies across time and then integrating this lower-level phrasing into the overall acoustic hierarchy. If children with DLD need greater acoustic differentiation between phrases in order to discriminate the overall acoustic hierarchy, then the less salient stress cues available in natural language may result in demarcations in the signal being missed. This may in turn lead to a failure to establish schema (relative stress templates) for the hierarchical relationships necessary for interpreting prosodic structure at a phrasal and clausal level.

The coincidence of the acoustic cues that create the prosodic-syntactic structure is particularly salient in repetitive and rhythmic children's literature, which was used to generate the stimuli used in this experiment. In stories such as *Room on the Broom*, there is reciprocal cuing of prosodic and syntactic elements such that sensitivity to one facet of the structure should facilitate processing of the other facet. However, our results suggest that this was not the case for the DLD children. Unlike typically developing children of the same age, they were unable to detect the prosodic-syntactic mismatches, suggesting that they are not yet proficient in integrating the two structural systems. Instead, they performed like the younger children in the experiment. This suggests a developmental trajectory for the development of prosodic-syntactic schemata in which the children with DLD exhibit delayed

development. If children with DLD are in general slower to establish prosodic hierarchies, possibly due to poorer acoustic sensitivity, then it could be that they require greater exposure to structured linguistic input than do typically developing children to attain a similar developmental level.

It has previously been found that infant-directed speech (IDS) is much richer in acoustic cues to linguistic features than adult-directed speech (ADS) and that IDS can facilitate syntactic boundary detection in infants when compared with ADS [19,42]. Studies of the acoustic characteristics of IDS have found that the rhythmic focus rapidly shifts as the infant ages [44]. It could be therefore that for children with DLD, a longer period of structured prosodic input is required if sensitivity to the meaning of the units is to develop. If children with DLD are less efficient at discovering these acoustic cues to syntactic boundaries, and as IDS changes rapidly with the age of the child, it could be that children with DLD end up 'missing out' on this crucial early aspect of language acquisition: the incoming signal 'moves on' before their system is ready to cope with a less-structured and salient input. Such a scenario would have significant implications for language development. Morgan and Saffran [45] argued that prosody should be regarded as a kind of parameter-setting device, providing a rough categorisation of the input into smaller units and thereby constraining the amount of input, which is then subject to further analysis, for example by statistical learning. If this is the case, then sensitivity to prosodic units would be a powerful tool in the process of discovering grammatical units. In this view of language acquisition, poorer sensitivity (in terms of acoustic, stimulus-driven sensitivity) to metrical hierarchies of stress patterns would mean that constraining parameters fail to be set in chunking the input stream. Accordingly, a subsequent analysis would be carried out across much greater chunks of input, resulting in a far more unwieldy task. This in turn would lead to difficulty in segmenting language into grammatical units, such as clauses and phrases, with knock-on implications for acquiring smaller-grained aspects of morphology: exactly the kinds of linguistic difficulties that characterise children with DLD.

5. Conclusions

In conclusion, our results suggest that children with DLD have poorer sensitivity to the acoustic cues to linguistic rhythm that enable the creation of prosodic-syntactic schemata. This difficulty in recovering metrical hierarchies of acoustic stress patterns impairs their ability to capitalise on the prosodic cues to syntax present in speech which bootstrap grammatical competence. If prosodic cues enable more efficient parsing of the speech stream, then explicitly teaching children to listen for these acoustic stress cues may increase their ability to integrate prosody and syntax. Via such instruction, children's capacity to derive grammatical structure from prosodically driven input could be increased. Accordingly, interventions using rhythmic children's texts to highlight this congruence of prosody and syntax could theoretically be of great value in scaffolding grammatical development in children with DLD.

Author Contributions: Conceptualization, S.R. & U.G.; Investigation, S.R.; Supervision, U.G.; Writing (original draft), S.R.; Writing (review & editing), S.R. & U.G.

Funding: This research was funded by an Economic and Social Research Council PhD studentship awarded to S.R. and supervised by U.G.

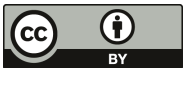
Acknowledgments: We would like to thank all the teachers, parents and children who enabled this research to take place.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; Consortium, C. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child Psychol. Psychiatry Allied Discip.* **2017**, *58*, 1068–1080. [CrossRef] [PubMed]
2. Dockrell, J.E.; Lindsay, G.; Palikara, O. Explaining the academic achievement at school leaving for pupils with a history of language impairment: Previous academic achievement and literacy skills. *Child Lang. Teach. Ther.* **2011**, *27*, 223–237. [CrossRef]
3. Norbury, C.F.; Gooch, D.; Wray, C.; Baird, G.; Charman, T.; Simonoff, E.; Vamvakas, G.; Pickles, A. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *J. Child Psychol. Psychiatry Allied Discip.* **2016**, *57*, 1247–1257. [CrossRef] [PubMed]
4. Tomblin, J.B.; Records, N.L.; Buckwalter, P.; Zhang, X.; Smith, E.; O'Brien, M. Prevalence of Specific Language Impairment in Kindergarten Children. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1245. [CrossRef] [PubMed]
5. Marchman, V.A.; Wulfeck, B.; Weismer, S.E. Morphological productivity in children with normal language and SLI: A study of the English past tense. *J. Speech Lang. Hear. Res.* **1999**, *42*, 206–219. [CrossRef] [PubMed]
6. Bishop, D.V.M.; Bishop, S.J.; Bright, P.; James, C.; Delaney, T.; Tallal, P. Different Origin of Auditory and Phonological Processing Problems in Children With Language Impairment. *J. Speech Lang. Hear. Res.* **1999**, *42*, 155–168. [CrossRef] [PubMed]
7. Van Der Lely, H.K.J.; Rosen, S.; McClelland, A. Evidence for a grammar-specific deficit in children. *Curr. Biol.* **1998**, *8*, 1253–1258. [CrossRef]
8. Tallal, P.; Piercy, M. Defects of Non-Verbal Auditory Perception in Children with Developmental Aphasia. *Nature* **1973**, *241*, 468–469. [CrossRef]
9. Gathercole, S.E.; Baddeley, A.D. Phonological Memory Deficits in Language Disordered Children: Is there a Causal Connection? *J. Mem. Lang.* **1990**, *29*, 336–360. [CrossRef]
10. Ramus, F.; Nespor, M.; Mehler, J. Correlates of linguistic rhythm in the speech signal. *Cognition* **1999**, *73*, 265–292. [CrossRef]
11. Arvaniti, A. Rhythm, timing and the timing of rhythm. *Phonetica* **2009**, *66*, 46–63. [CrossRef] [PubMed]
12. Mehler, J.; Jusczyk, P.; Lambertz, G.; Halsted, N.; Bertoncini, J.; Amiel-Tison, C. A precurson of language acquisition in young infants. *Cognition* **1988**, *29*, 143–178. [CrossRef]
13. Nazzi, T.; Bertoncini, J.; Mehler, J. Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* **1998**, *24*, 756–766. [CrossRef] [PubMed]
14. Mehler, J.; Dupoux, E.; Nazzi, T.; Dehaene-Lambertz, G. Coping with Linguistic Diversity: The Infant's Viewpoint. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*; Morgan, J.L., Demuth, K., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1996; pp. 101–117.
15. Jusczyk, P.W.; Houston, D.M.; Newsome, M. The Beginnings of Word Segmentation in English-Learning Infants. *Cogn. Psychol.* **1999**, *39*, 159–207. [CrossRef] [PubMed]
16. Curtin, S.; Mintz, T.H.; Christiansen, M.H. Stress changes the representational landscape: Evidence from word segmentation. *Cognition* **2005**, *96*, 233–262. [CrossRef] [PubMed]
17. Jusczyk, P.W.; Cutler, A.; Redanz, N.J. Infants' preference for the predominant stress patterns of English words. *Child Dev.* **1993**, *64*, 675–687. [CrossRef] [PubMed]
18. Hirsh-Pasek, K.; Kemler Nelson, D.G.; Jusczyk, P.W.; Cassidy, K.W.; Druss, B.; Kennedy, L. Clauses are perceptual units for young infants. *Cognition* **1987**, *26*, 269–286. [CrossRef]
19. Jusczyk, P.W.; Hirsh-Pasek, K.; Kemler Nelson, D.G.; Kennedy, L.J.; Woodward, A.; Piwoz, J. Perception of acoustic correlates of major phrasal units by young infants. *Cogn. Psychol.* **1992**, *24*, 252–293. [CrossRef]
20. Bishop, D.V.M.; Snowling, M.J. Developmental dyslexia and specific language impairment: Same or different? *Psychol. Bull.* **2004**, *130*, 858–886. [CrossRef]
21. Richards, S.; Goswami, U. Auditory Processing in Specific Language Impairment (SLI): Relations with the Perception of Lexical and Phrasal Stress. *J. Speech Lang. Hear. Res.* **2015**, *58*, 1292–1305. [CrossRef]
22. Cumming, R.; Wilson, A.; Goswami, U. Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: A new look at a perceptual hypothesis. *Front. Psychol.* **2015**, *6*, 972. [CrossRef] [PubMed]

23. Greenberg, S.; Carvey, H.; Hitchcock, L.; Chang, S. Temporal properties of spontaneous speech—A syllable-centric perspective. *J. Phon.* **2003**, *31*, 465–485. [CrossRef]
24. Mengler, E.D.; Hogben, J.H.; Michie, P.; Bishop, D.V.M. Poor frequency discrimination is related to oral language disorder in children: A psychoacoustic study. *Dyslexia* **2005**, *11*, 155–173. [CrossRef] [PubMed]
25. McArthur, G.M.; Bishop, D.V.M. Frequency Discrimination Deficits in People With Specific Language Impairment. *J. Speech Lang. Hear. Res.* **2004**, *47*, 527–541. [CrossRef]
26. Corriveau, K.; Pasquini, E.; Goswami, U. Basic Auditory Processing Skills and Specific Language Impairment: A New Look at an Old Hypothesis. *J. Speech Lang. Hear. Res.* **2007**, *50*, 647–666. [CrossRef]
27. Beattie, R.L.; Manis, F.R. Rise Time Perception in Children With Reading and Combined Reading and Language Difficulties. *J. Learn. Disabil.* **2013**, *46*, 200–209. [CrossRef] [PubMed]
28. Corriveau, K.H.; Goswami, U. Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex* **2009**, *45*, 119–130. [CrossRef]
29. Dauer, R.M. Stress-timing and syllable-timing reanalysed. *J. Phon.* **1983**, *11*, 51–62.
30. Alcock, K.J.; Passingham, R.E.; Watkins, K.; Vargha-Khadem, F. Pitch and timing abilities in inherited speech and language impairment. *Brain Lang.* **2000**, *75*, 34–46. [CrossRef]
31. Falk, S.; Müller, T.; Dalla Bella, S. Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Front. Psychol.* **2015**, *6*, 847. [CrossRef]
32. Cumming, R.; Wilson, A.; Leong, V.; Colling, L.J.; Goswami, U. Awareness of Rhythm Patterns in Speech and Music in Children with Specific Language Impairments. *Front. Hum. Neurosci.* **2015**, *9*, 672. [CrossRef] [PubMed]
33. Weinert, S. Deficits in acquiring language structure: The importance of using prosodic cues. *Appl. Cogn. Psychol.* **1992**, *6*, 545–571. [CrossRef]
34. Gordon, R.L.; Shivers, C.M.; Wieland, E.A.; Kotz, S.A.; Yoder, P.J.; Devin Mcauley, J. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* **2015**, *18*, 635–644. [CrossRef]
35. Price, P.J.; Ostendorf, M.; Shattuck-Hufnagel, S.; Fong, C. The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* **1991**, *90*, 2956–2970. [CrossRef] [PubMed]
36. Demuth, K. Prosodic Constraints on Morphological Development. In *Approaches to Bootstrapping: Phonological, Syntactic and Neurophysiological Aspects of Early Language Acquisition*; Weissenborn, J., Höhle, B., Eds.; John Benjamins: Amsterdam, The Netherlands, 2001; pp. 3–21.
37. Demuth, K.; Tomas, E. Understanding the contributions of prosodic phonology to morphological development: Implications for children with Specific Language Impairment. *First Lang.* **2016**, *36*, 265–278. [CrossRef]
38. Donaldson, J. *Room on the Broom*; Pan Macmillan: London, UK, 2002.
39. Dunn, L.; Dunn, L.; Whetton, C.; Burley, J. *British Picture Vocabulary Scales*, 2nd ed.; NFER-Nelson: London, UK, 1997.
40. Semel, E.; Wiig, E.H.; Secord, W.A. *Clinical Evaluation of Language Fundamentals*, 3rd ed.; UK Version; The Psychological Corporation: San Antonio, TX, USA, 2000.
41. Wechsler, D. *Wechsler Intelligence Scale for Children*, 3rd ed.; The Psychological Corporation: San Antonio, TX, USA, 1992.
42. Kemler Nelson, D.G.; Hirsh-Pasek, K.; Jusczyk, P.W.; Cassidy, K.W. How the prosodic cues in motherese might assist language learning. *J. Child Lang.* **1989**, *16*, 55–68. [CrossRef]
43. Fraser, J.; Goswami, U.; Conti-Ramsden, G. Dyslexia and specific language impairment: The role of phonology and auditory processing. *Sci. Stud. Read.* **2010**, *14*, 8–29. [CrossRef]
44. Leong, V.; Kalashnikova, M.; Burnham, D.; Goswami, U. Infant-Directed Speech Enhances Temporal Rhythmic Structure in the Envelope. In Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014.
45. Morgan, J.L.; Saffran, J.R. Emerging Integration of Sequential and Suprasegmental Information in Preverbal Speech Segmentation. *Child Dev.* **1995**, *66*, 911–936. [CrossRef]



brain sciences

MDPI

Article

Event-Related Potential Evidence of Implicit Metric Structure during Silent Reading

Mara Breen [1,*], Ahren B. Fitzroy [1,2] and Michelle Oraa Ali [1,3]

1 Department of Psychology and Education, Mount Holyoke College, South Hadley, MA 01075, USA
2 Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA 01003, USA
3 Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912, USA
* Correspondence: mbreen@mtholyoke.edu; Tel.: +1-413-538-2067

Received: 17 July 2019; Accepted: 5 August 2019; Published: 8 August 2019

Abstract: Under the Implicit Prosody Hypothesis, readers generate prosodic structures during silent reading that can direct their real-time interpretations of the text. In the current study, we investigated the processing of implicit meter by recording event-related potentials (ERPs) while participants read a series of 160 rhyming couplets, where the rhyme target was always a stress-alternating noun–verb homograph (e.g., permit, which is pronounced PERmit as a noun and perMIT as a verb). The target had a strong–weak or weak–strong stress pattern, which was either consistent or inconsistent with the stress expectation generated by the couplet. Inconsistent strong–weak targets elicited negativities between 80–155 ms and 325–375 ms relative to consistent strong–weak targets; inconsistent weak–strong targets elicited a positivity between 365–435 ms relative to consistent weak–strong targets. These results are largely consistent with effects of metric violations during listening, demonstrating that implicit prosodic representations are similar to explicit prosodic representations.

Keywords: implicit prosody; reading; meter; rhythm; lexical stress; event-related potentials; poetry

1. Introduction

According to the Implicit Prosody Hypothesis [1–3], readers generate imagined representations of prosodic structure during silent reading that are similar to the explicit prosodic representations that readers produce when reading aloud. This hypothesis has been supported by behavioral evidence demonstrating similarity between real and imagined representations of a variety of prosodic phenomena, including intonation, phrasing, stress, and meter [4,5]. For example, evidence for implicit intonational structure is provided by the fact that readers are faster to recognize target words that are produced aloud with a previously imagined intonation contour [6,7]. Readers impose implicit phrase boundaries in sentences that are long enough to have a phrase break [8] and tend to balance the size of adjacent phrases even during silent reading [9,10], providing evidence for implicit prosodic phrasing. Readers take longer to silently read words with two stressed syllables than words with one stressed syllable [11], and take longer to read sentences in which a local lexical stress pattern mismatches the predicted metric structure as determined by prior sentence material [12–16], providing evidence for an implicit metric structure. Although these behavioral similarities between patterns associated with explicit and implicit prosody provide indirect support for implicit prosodic representations, they cannot tell us to what extent implicit prosodic representations are processed similarly to explicit prosodic representations. In the current study, we used event-related potentials (ERPs) to investigate the processing of implicit prosodic representations, and how it compares to that of explicit prosody.

1.1. Behavioral Studies of Explicit and Implicit Linguistic Metric Representation

The specific focus of the current study is the similarity between implicit and explicit metric processing. For this investigation, we exploit metrical regularity in English; English is a stress-timed language, meaning that speakers produce temporally regularized sequences of strong (stressed) and weak (unstressed) syllables. The metric structure in stress-timed languages is conveyed by the timing of strong syllables [17,18]. There are constraints on the ordering of strong and weak beats in stress-timed languages, as strong beats tend to occur at regular intervals [19], speakers avoid clashes of strong beats and lapses of weak beats [20], and under some circumstances, speakers shift the location of stress on words to maintain metric regularity (e.g., thirTEEN MEN → THIRteen MEN) [18]. Speakers signal strong syllables in speech with a variety of acoustic cues, including longer duration and higher intensity [21–24]. Strong syllables also hold a privileged position in auditory language comprehension; listeners are faster to detect phonemes in stressed syllables [25], lexical access is more disrupted by the mispronunciation of stressed syllables than unstressed syllables [26], and listeners tend to interpret stressed syllables as word onsets [27,28]. Moreover, listeners use the pattern of strong and weak syllables to predict what words will come next [29], and to resolve lexical ambiguity [30–33].

Like speakers and listeners, there is evidence that readers are also sensitive to a metric structure. For example, readers spend more time fixating four-syllable words with two stressed syllables (e.g., RAdiAtion) than four-syllable words with one stressed syllable (e.g., geOmetry) [11]. In silent reading, syntactically ambiguous sentences are more likely to be resolved in ways that maintain alternating strong and weak syllables [14,15]. In the study that serves as the inspiration for the current study, Breen and Clifton tracked participants' eye movements as they read limericks designed to induce readers to generate strong expectations about the stress pattern of upcoming words [13]. The target word in the critical items, which was always the final word of the second line of the limerick, was a stress-alternating noun–verb homograph; these words are realized with strong–weak (SW) stress as a noun (e.g., PERmit), but weak–strong (WS) stress as a verb (e.g., perMIT) [29]. In this way, the target was either SW or WS, and this lexical stress pattern was either consistent or inconsistent with the metric structure of the limerick (see Table 1). Throughout this paper, we will refer to the occurrence of an inconsistent SW word when a WS word is predicted as a strong–weak (SW) violation, and to the occurrence of an inconsistent WS word when a SW word is predicted as a weak–strong (WS) violation.

Table 1. Metric structure of experimental couplets in each of the four conditions for the target word 'permit'.

		W	S	W	W	S	W	W	S	W
A.	Strong–weak, consistent	There	*once*	was	an	*old*	man	named	*Ker–*	mit
		who	*hunt–*	ed	with-	*out*	an-	y	***PER-***	**mit**
B.	Strong–weak, inconsistent	There	*once*	was	an	*old*	man	named	*Britt*	
		who	*hunt–*	ed	with-	*out*	a	* **PER-**	***mit***	
C.	Weak–strong, consistent	There	*once*	was	an	*old*	man	named	*Britt*	
		whose	*vic–*	es	no	*wife*	could	**per-**	***MIT***	
D.	Weak–strong, inconsistent	There	*once*	was	an	*old*	man	named	*Ker–*	mit
		Whose	*gamb–*	ling	his	*wife*	would	not	**** per-***	**MIT**

Italics and underlines indicate metrically strong syllables, bold indicates target words, and capital letters indicate lexical stressed syllables within target words. Asterisks (*) indicate metrically inconsistent targets. Screen breaks are indicated with solid vertical lines. Text emphasis is for descriptive purposes only; in the experiment, all words were presented in plain text (see Figure 1).

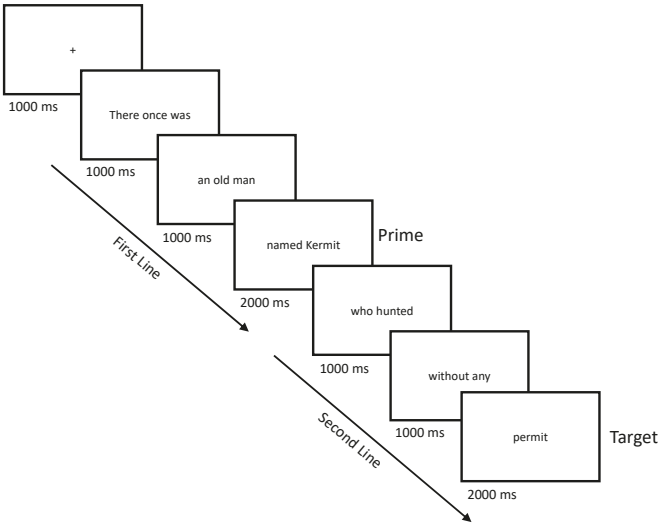


Figure 1. Presentation times in milliseconds of each region of the limerick couplets.

Breen and Clifton predicted that readers would encounter difficulty whenever the stress pattern of the target word mismatched the pattern of the limerick. However, they only observed an effect of metric mismatch for WS violations (e.g., Table 1D); reading times for SW violations (e.g., Table 1B) did not differ from those of consistent SW words. Breen and Clifton argued that these results reflect the uneven distribution of SW and WS words in the English lexicon; 85–90% of content words in English have an initial stressed syllable [34]. Specifically, there is a minimal cost to encountering a SW word in a context where a WS word is predicted because SW is the default stress pattern. Identifying a WS word in a context that predicts SW, on the other hand, is costly because of both the conflict with context and the lower base frequency of the WS pattern. This interpretation is supported by previous work showing that auditory word identification is more disrupted when a canonically SW word is pronounced as WS, than when a canonically WS word is pronounced as SW [35]. Moreover, the observed effect was not on initial reading times, but only on the combined duration of fixations on the target word and time spent rereading earlier sentence material. The latency of this effect, therefore, suggests that the WS violation did not disrupt initial reading times but required later reanalysis.

1.2. Event-Related Potential Studies of Explicit Linguistic Metric Processing

In ERP investigations of explicit metric processing during speech perception, multiple methods have been used to investigate metric violations. One major source of variation among these studies is whether the metric violation is determined by the lexical stress pattern of the word in isolation or only by the context in which the word occurs. In studies of the first variety, researchers presented multisyllabic words auditorily with the correct or incorrect stress pattern either in isolation [36–38] or in a sentence context [39,40]. In studies of the second variety, researchers established a context that created an expectation of a specific metric pattern, then presented a target that had the correct metric pattern in isolation but was consistent or inconsistent with the expected pattern created by the context. One such paradigm used word strings to create metric context: listeners heard a string of three or four prime words with the same lexical stress pattern (all SW or all WS, e.g., BANKer, HELPful, PARty or moRALE, emBRACE, deLIGHT) followed by a target word with the same stress pattern as the primes or the opposite pattern [41,42]. In another such paradigm, participants heard sentences with a consistent metric structure including a target which was either consistent or inconsistent with

the established pattern [43–48] (e.g., stress clash in "The chamPAGNE COCKtails are very delicious"). A final method used cross-modal information to inform prosodic interpretation, as in [49] where participants viewed pictures which disambiguated the meaning of semantically ambiguous two-syllable strings like greenhouse, which are disambiguated by stress patterns (GREENhouse vs. green HOUSE).

Regardless of the type of manipulation, these ERP studies demonstrate that encountering metric violations while listening generally gives rise to an early negativity between 250 and 500 ms [36–48]. However, this early effect is not consistent across studies, in terms of timing and polarity. Some of the variance can be explained by the different responses to SW violations and WS violations in two-syllable words; SW violations, where a SW word appears when a WS word is predicted, typically elicit an early negativity [41–47,49]. The results are more mixed for WS violations, where a WS word appears when a SW word is predicted, which elicit an early negativity in some cases [41,42] but has also been shown to elicit an early positivity relative to predicted metric patterns [36,37,40,42,48]. In two studies, both SW and WS violations elicited an early negativity, but the negativity to SW violations peaked earlier [41,49].

Additionally, explicit metric violations have often been shown to elicit a late positivity between 500 and 1000 ms [36–40,42–44,46,48]. In contrast to the early time window, this later effect does not seem to differ in polarity or timing as a function of target lexical stress pattern. However, its presence is dependent on the experimental task; in cases where the participants' task is to make an explicit assessment of the accuracy of the metric structure of the target, that target usually elicits a late positivity [36–38,46,48,49], though this is not always the case [45,47]. In contrast, if the participants' task does not include a specific assessment of the metric structure, a late positivity is absent [41]. Indeed, in cases where the explicitness of a metric judgment is varied within the experiment, a late positivity is generally evident only when the task requires this judgment [39,40,42–44].

Despite some variation across studies, these neural effects of metric inconsistency appear to be distinct from the neural effects of either syntactic or semantic violations. Syntactic violations typically elicit a biphasic response consisting of a left-lateralized anterior negativity peaking around 300 ms (LAN) and a posterior positivity peaking around 600 ms after stimulus onset (P600/LPC) [50]. A simultaneous test of metric and syntactic violations reported distinct negativities for each violation type however, with the negativity to metric violations occurring earlier than the negativity evoked by syntactic violations (which was interpreted as a LAN) [43]. Semantic violations typically elicit a parietally-maximal negativity around 400 ms (N400) [51]. Although some authors have interpreted the early negativity elicited by metric violations as an N400 [39,40], this metric negativity has been observed in response to illegal stress shifts in pseudowords which have no lexico-semantic content and should not result in an N400 [45]. Further, semantic incongruity and metric incongruity have been shown to modulate the amplitude of an early negativity differently when considered in the same design, even by authors who categorize deviations from a predicted metric structure as N400 effects [39,40]. Finally, [44] observed that simultaneous metric and semantic violations lead to a larger negativity than that observed for semantic violation alone, and [52] used neuroimaging to demonstrate that the responses to semantic and metric violations have different neural generators, providing evidence that metric violations are not simply processed as semantic violations.

1.3. Event-Related Potential Studies of Implicit Linguistic Metric Processing

ERPs have also been used to explore implicit metric representations during silent reading. In one study, readers were presented with strings of four two-syllable English prime words with consistent lexical stress patterns, followed by a target word that was consistent or inconsistent with the stress pattern of the previous words [53]. Both SW and WS violations resulted in a larger fronto-central negativity from 250–400 ms after word onset, relative to words with a predicted stress pattern. In addition, all SW targets, whether consistent or inconsistent with the context, elicited a larger negativity (350–450 ms after word onset) than WS targets. In another study exploring silent metric processing in word lists, readers were presented with strings of three two-syllable German prime words

followed by a SW or WS target. In this case, there were no observable ERP differences for SW violations, but WS violations were more positive than correct WS targets in three time windows: between 250–400, 400–600, and 600–800 ms after target onset [54]. A final study presented participants with an auditory tone sequence with a SW or WS pattern followed by a visually presented two-syllable English word which was consistent or inconsistent with the tone sequence stress pattern [55]. The results demonstrated a larger negativity from 300–700 ms after target presentation for SW violations compared to correct SW targets, but no significant ERP effect for WS violations. In general, these studies demonstrate that, similar to explicit metric violations, implicit metric violations often evoke an early negativity that is more reliably observed for SW than WS violations. Moreover, two of these studies are consistent with results from explicit meter studies in that when the task does not require an explicit metric judgment (and none of these did; rather, participants' task was to make an old/new judgment of the target [53], a lexical decision judgment [55], or answer a semantic question about the word strings [54]), there is no late positivity.

Multiple factors could be contributing to the variability in results observed across previous investigations of ERP responses to implicit metric violations. First, these studies have used different target words in the SW and WS conditions, meaning that the observed results may reflect differences beyond prosody, including phonetic, orthographic, or lexical differences between conditions. Second, these studies used single words or word lists to create metric expectations but, in these contexts, readers are not required to fully process the syntactic and semantic structure of the targets; this variability could lead to heterogenous depth of processing across conditions. Therefore, in the current study, we implemented metric expectations using metrically regular rhyming couplets, which encourage readers to make strong predictions about when strong and weak syllables will occur but also require deep linguistic processing. Moreover, our target words are stress-alternating noun–verb homographs, which can have SW or WS stress depending on the syntactic category. In this way, readers are exposed to the same visual, orthographic, and segmental input across all conditions.

If readers are generating implicit metric predictions during silent reading, we predict that targets which are inconsistent with the metric context will result in early differences in the ERP waveform compared to metrically consistent targets. However, based on prior work, we predict that this early effect may differ depending on the type of violation. Specifically, we predict SW violations will elicit an early negativity relative to consistent SW words. Conversely, WS violations may result in either a reduced negativity, or a positivity, relative to WS consistent targets. Moreover, we predict the absence of a late positivity in response to metric violations, as participants are not making explicit judgments about the metric structure.

2. Materials and Methods

2.1. Participants

Eighteen participants from Mount Holyoke College with an average age of 20 years (SD = 1.57 years) contributed data to the analyses. Seventeen participants identified as female and one identified as nonbinary/genderqueer. All participants were right-handed native speakers of American English, meaning they had been speaking English in the US since at least the age of three. One participant was born outside the US to English-speaking parents and moved to the US at age three. Five participants identified as bilingual as they had acquired high proficiency in another language starting before the age of three. All participants reported having normal or corrected-to-normal vision and had not taken psychoactive medications in the 24 h prior to the experiment. For the two-hour experiment, participants received compensation in the form of Psychology course research credit or $20. Data were collected but discarded from an additional four female participants due to voluntary withdrawal from the experiment (n = 1), recording equipment malfunction (n = 1), or excessive noise in the EEG (exclusion of more than 50% of trials from one or more conditions due to artifact; n = 2).

2.2. Materials

Experimental materials consisted of 160 limerick couplets (i.e., the first two lines of the limerick) adapted from the stimuli in [13], in which the final word of the second line was one of 40 stress-alternating noun–verb homograph targets (see Table 1). The stress pattern of these targets varies depending on the syntactic category: the noun form has a strong–weak pattern (e.g., PERmit), the verb (or adjective) form has a weak–strong pattern (e.g., perMIT). The homographs were selected from [29] and the Kucera–Francis corpus [56]. The frequency of occurrence of each target as a noun or verb/adjective in the Kucera–Francis corpus did not differ, as measured by a paired *t*-test, $t(39) = 0.28$, $p = 0.78$.

For each target homograph, four couplets were constructed, crossing the factors stress pattern (SW vs. WS) and metric consistency (consistent vs. inconsistent). All experimental couplets can be found in the Appendix A. The first line of each couplet established the metric and rhyming context for the target word. The stress pattern manipulation was implemented such that for half of the couplets, the target (e.g., permit in Table 1) was the noun form with a SW pattern (Table 1A,B). For the other half, the target was the verb/adjective form (Table 1C,D). The metric consistency manipulation meant that for half of the couplets, the stress pattern of the target homograph was consistent with the stress pattern predicted by the couplet (Table 1A,C). For the other half, the stress pattern of the target was inconsistent with the established pattern (Table 1B,D). The occurrence of an inconsistent SW word when a WS word is predicted is a strong–weak (SW) violation (Table 1B), and the occurrence of an inconsistent WS word when a SW word is predicted is a weak–strong (WS) violation (Table 1D).

In addition to the 160 experimental couplets, participants read 160 filler couplets which were always metrically consistent but varied in the stress pattern of the target regions (see Table S1 and examples (1), (2)). In this way, participants read a total of 80 rhythmically inconsistent items in a pool of 320 (25% of the total).

Examples:

1. There once was a man from Peru, who dreamt about eating his **shoe**
2. There once was a young man named Randy, who loved to eat all kinds of **candy**

2.3. Procedure

After providing informed consent, participants were seated comfortably in a sound-isolated room where they viewed the couplets on a computer screen located approximately 90 cm away. The 320 experimental and filler couplets were presented in a different randomized order for each participant. Each trial began with the presentation of the word "Ready?" which stayed on the screen until the participant responded with a keypress. The word was then replaced by a fixation cross, which remained on the screen for 1000 ms (Figure 1). Following the fixation cross, couplets were presented in six one-to-four-word (one-to-five-syllable) segments in the center of the screen (see Table S1). The 1st, 2nd, 4th and 5th segments were presented for 1000 ms each; the 3rd and 6th segments, corresponding to the end of the first and second lines, respectively, were presented for 2000 ms each. In the experimental couplets, the 3rd and 6th segments always contained two-to-three syllables, one of which was strong. The 1st, 2nd, 4th and 5th segments were more variable, but constrained so that each contained one strong syllable, and one-to-four weak syllables. The number of words and syllables varied across these segments because the couplets varied widely in terms of the number of words and stress patterns of the words that made them up. However, segments were consistently defined based on the first author's intuition of natural syntactic and prosodic breaks in limerick structure.

To ensure that participants were reading for meaning, 25% of the filler trials (12.5% of all trials) were followed by a yes/no comprehension question about the semantic content. Participants held a response box in their lap for the duration of the experiment which they used to answer comprehension questions, and to advance the presentation of trials. Participants were given breaks between trials to allow time for blinking, as well as a longer break after every 40 trials; the length of these breaks were determined by the participant. The entire experimental session lasted approximately 2 h. All experimental procedures were approved in advance by the Institutional Review Board of Mount Holyoke College.

Reference-free electroencephalogram (EEG) data were collected using 64 active Ag/AgCl electrodes placed in an elastic cap and connected to a BioSemi Active-Two system, which digitized the EEG at a sampling rate of 2048 Hz and employed a hardware lowpass filter reaching −3 dB at 409.6 Hz. Reference-free EEG was also collected from two active electrodes attached bilaterally to the participant's mastoids, and from four active electrodes placed above and below the left eye and bilaterally outside the outer canthi. All electrode offsets were brought below 20 mV at the start of the recording and kept below 50 mV throughout the recording. Continuous EEG data were referenced offline to the averaged mastoid recording, downsampled to 512 Hz, and filtered at 60 Hz using a Parks–McClellan notch filter. Bipolar vertical and horizontal electrooculogram (VEOG, HEOG) signals were derived by subtracting the above eye signal from the below eye signal, and the left from the right eye signal, respectively. Continuous EEG was segmented into epochs from 100 ms prior to target word onset to 800 ms following target word onset, and baseline-corrected to the 100 ms prestimulus period. Electrodes Oz and Iz were each identified as unusable for at least one participant and were excluded from further processing and analysis. Epochs containing eyeblinks or eye movements were identified algorithmically using moving window peak-to-peak voltage deflection detection on the VEOG channel (threshold = 150 μV, window size = 200 ms, window step = 25 ms) and step-like artifact detection on the HEOG channel (threshold = 100 μV, window size = 400 ms, window step = 25 ms), respectively. Additionally, epochs exceeding ±170 μV in any EEG channel were marked as artifact. The results of automatic artifact detection were then manually inspected and if needed, adjusted, and trials found to contain artifacts were excised. Artifact-free trials were then averaged by participant and condition; participants included in the analysis contributed data from at least 20 out of 40 trials (M = 31; SD = 6) in every condition. EEG data processing was performed in MATLAB using the EEGLAB [57] and ERPLAB [58] analysis packages.

2.4. Analysis

Previous ERP investigations of implicit linguistic metric processing have revealed effects across multiple time windows, with inconsistent time windows observed across studies [36–49]. We therefore opted to define our temporal regions of interest using a data-driven approach. To minimize implicit multiple comparisons when selecting time windows [59], we performed a series of cluster-based permutation tests over a moving 50 ms window (5 ms step) using the Mass Univariate ERP Toolbox [60]. These tests were performed separately for SW and WS violations. Within each moving window, electrodes at which the inconsistent vs. consistent *t*-test of mean window amplitude resulted in $p \leq 0.01$ were identified, then clustered if they were within 5.44 cm of one another. Cluster magnitudes were then calculated as the sum of all *t*-scores for electrodes contained within a cluster. Lower-tailed *t*-tests were used for the SW comparisons based on prior findings that SW violations consistently elicit relative negativities [36,41,44,45,49,53,55], whereas two-tailed *t*-tests were used for the WS comparisons based on prior findings that WS violations elicit both relative negativities and positivities [40–42,54]. This process was replicated over 5000 shuffled iterations, and a cluster magnitude threshold was defined as the magnitude that clusters met or exceeded on only 5% of the shuffled (i.e., chance) iterations. Moving windows within which any clusters identified in the experimental data met or exceeded the cluster magnitude threshold were defined as temporal regions of interest (see Figure S1). This approach revealed three regions of interest, which were further investigated using conventional, ANOVA-based ERP analyses: 80–155 ms (SW), 325–375 ms (SW), and 365–435 ms (WS).

We selected 49 electrodes for conventional, ANOVA-based ERP analysis (Figure 2). Scalp position was treated as two factors in the statistical model: electrode anteriority had seven levels ranging from most anterior to most posterior electrodes, and electrode laterality had seven levels, ranging from left to right. Based on our cluster-based permutation tests, we assessed SW ERP amplitudes in two time windows (80–155 ms and 325–375 ms), and WS ERP amplitudes in one time window (365–435 ms). Mean amplitudes from each participant in each time window were entered into a 2 (metric consistency) × 7 (anteriority) × 7 (laterality) repeated-measures ANOVA. Significant and marginal interactions of

metric consistency with electrode position in the absence of a main effect of metric consistency were further investigated with follow-up ANOVAs over constrained scalp regions. Only main effects and interactions which involve metric consistency will be discussed. Whenever Mauchly's Test indicated that the assumption of sphericity had been violated for comparisons with more than two levels, Huynh–Feldt-corrected degrees of freedom were used to compute statistical significance. All statistical analyses were implemented in the R statistical framework [61] with the ez package [62].

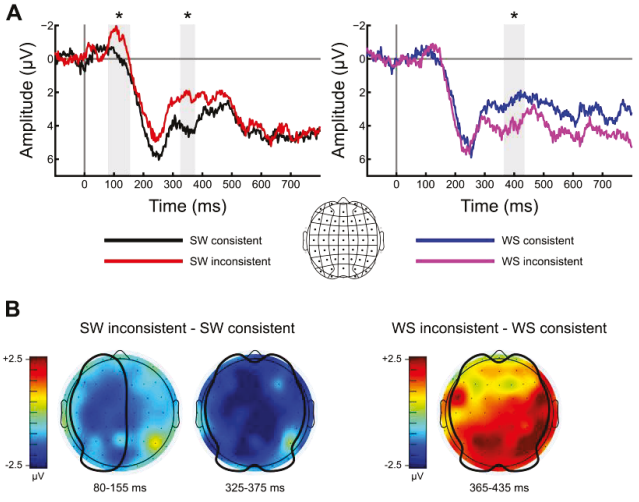


Figure 2. Event-related potential (ERP) results. (**A**) Effects of metric predictability on a grand average (n = 18) waveform amplitude for strong–weak (SW) targets (left) and weak–strong (WS) targets (right). Temporal regions of interest identified in the cluster-based permutation analyses are highlighted in grey. Temporal regions of interest that revealed a significant ($p < 0.05$) main effect of metric consistency in conventional ANOVA analyses are indicated with an asterisk. Waveforms are averaged over the 49 electrodes included in the ANOVA analyses; the 7 (anteriority) × 7 (laterality) grid arrangement used to model electrode position in all ANOVAs is shown in the inset. (**B**) Scalp maps showing the topography of mean amplitude differences between the inconsistent and consistent conditions within the two temporal regions of interest identified for SW targets (left), and the one temporal region of interest identified for WS targets (right). The scalp region over which a significant ($p < 0.05$) main effect of metric consistency was observed within the specified time window is outlined in black for each scalp map.

3. Results

3.1. Behavioral

Participants answered comprehension questions with an average accuracy rate of 96.25% (SD = 4.2%), demonstrating that they were attending to the couplets and engaged with the task.

3.2. Event-Related Potentials

3.2.1. SW Violations

SW violations elicited a negativity from 80–155 ms relative to predicted SW targets over left and medial scalp positions (Figure 2). An overall 2 × 7 × 7 ANOVA looking only at SW targets revealed a marginal interaction of metric consistency and electrode laterality, $F(2.74,46.55) = 2.38$, $p = 0.087$, $\eta^2 = 0.007$. A follow-up 2 × 7 ANOVA looking only at SW targets over left and medial scalp positions

revealed a negativity elicited by SW violations relative to predicted SW targets, $F(1,17) = 4.75$, $p = 0.044$, $\eta^2 = 0.14$.

SW violations also elicited a negativity from 325–375 ms relative to predicted SW targets over the entire scalp, that was largest over left and medial scalp positions (Figure 2). An overall $2 \times 7 \times 7$ ANOVA looking only at SW targets revealed a main effect of metric consistency, $F(1,17) = 5.32$, $p = 0.034$, $\eta^2 = 0.09$, and a marginal interaction of metric consistency and electrode laterality indicated that this effect was largest over left and medial scalp positions, $F(3.38,57.45) = 2.51$, $p = 0.06$, $\eta^2 = 0.004$.

3.2.2. WS Violations

WS violations elicited a positivity from 365–435 ms relative to predicted WS targets over the entire scalp, that was largest over central and posterior scalp positions (Figure 2). An overall $2 \times 7 \times 7$ ANOVA looking only at WS targets revealed a main effect of metric consistency, $F(1,17) = 4.99$, $p = 0.039$, $\eta^2 = 0.06$, and a marginal interaction of metric consistency and electrode anteriority indicated that this effect was largest over central and posterior scalp positions, $F(2.68,45.54) = 2.78$, $p = 0.06$, $\eta^2 = 0.01$.

4. Discussion

The goal of the current study was to investigate the realization of metric representations during silent reading using ERPs. Participants silently read metrically regular rhyming couplets in which the final target word had a strong–weak (SW) or weak–strong (WS) lexical stress pattern that was either consistent or inconsistent with the metric stress pattern predicted by the couplet. The results demonstrated that SW targets which were inconsistent with the stress pattern of the couplet (i.e., SW violations) elicited two separate negativities (80–155 ms and 325–375 ms after word onset) relative to SW targets which were consistent with the predicted stress pattern. Conversely, WS targets inconsistent with the stress pattern of the couplet (i.e., WS violations) elicited an early positivity (365–435 ms after word onset) relative to WS targets which were consistent with the predicted stress pattern. Neither SW nor WS violations elicited a late positivity. Together with prior results, the current results support the Implicit Prosody Hypothesis, which maintains that readers are generating implicit versions of prosodic structure even when reading silently, and that these representations are similar to explicit ones.

The observation of a significant negative left-lateralized deflection from 80–155 ms in response to SW violations is an unexpected result based on prior work on explicit and implicit linguistic metric processing. Few studies of linguistic meter have reported consistent differences in components this early, though one study demonstrated a significant negativity between 100–320 ms in response to an inappropriate stressed syllable [46]. However, negativities in the 100–200 ms time window have been widely observed in response to metric violations in musical studies. This effect, termed the metric mismatch negativity (MMN), has been observed when a strong tone occurs at an unpredicted temporal location (i.e., when a weak tone is predicted) [63–65]. This situation is analogous to the circumstance under which we observed the early negativity in the current study, such that a strong beat at a predicted weak time elicits the early negativity (SW violation), whereas a weak beat at a predicted strong time does not (WS violation). Importantly, as this effect was detected based on a marginal interaction of metric consistency with electrode position and this is the first study we are aware of to report this early negativity in response to an implicit strong beat occurring at a predicted weak time, additional experiments will be required to determine the reliability and meaning of this component.

The negativity between 325–375 ms observed for SW violations is consistent with results from previous investigations of both explicit and implicit violations of metric structure. Specifically, previous studies have demonstrated that SW metric violations result in a negative deflection in the 250–500 ms range relative to metrically consistent targets [36,41,44,45,49]. Moreover, a similar effect has also been shown in a small set of studies investigating metric structure in silent reading of single words [53,55]. The current study extends this finding to silent reading of metric violations in sentence contexts using orthographically identical items across all conditions. The observation of a positivity for WS

violations from 365–435 ms after word onset is also consistent with both prior listening and reading studies. Two prior studies of explicit metric violations [40,42] and one prior study of implicit metric processing [54] have observed positive deflections for consistent WS targets relative to inconsistent WS targets. Our results therefore suggest that prior findings of different responses to SW and WS violations are not simply due to idiosyncratic differences between the SW and WS target items chosen for these prior experiments, but do indeed reflect the activation of abstract metric representations during silent reading.

The different results observed across multiple studies for SW vs. WS violations may be due to differences in the underlying phonological structure of the target words. According to [17], the trochaic foot (SW) is the default phonological structure in Germanic languages, including English. This phonological constraint is realized in the lexical stress patterns of words, such that most two-syllable words begin with a stressed syllable (85–90% of the time in English [34]; 73% of the time in German [66]). This asymmetry means that accessing a SW (trochaic) representation of a target is globally easier than accessing a WS (iambic) representation, irrespective of the context in which the target occurs. Therefore, the lexical representation of a SW target is harder to access when its stress pattern conflicts with the local metric context, than when its stress pattern is consistent with the local context. Conversely, resolving WS violations is more challenging for readers, due to conflicting cues in both the local environment and the global environment.

Under this view of phonological asymmetry, the negativity observed between 325–375 ms for SW violations in the current study, and in a similar time window in other studies, could be related to the N400, which reflects the ease with which lexical access is achieved. The negativity for SW violations could reflect either additional lexical processing due to the added difficulty of accessing the appropriate lexical content in the presence of lexical stress mismatch, or lexical repair processes due to automatic activation of the metrically consistent, but semantically inconsistent, alternate form of the noun/verb homograph. However, it is important to note that this interpretation of the negativity as indexing lexical processing is challenged by previous work exploring simultaneous violations of metric and semantic structure, in which the latency and distribution of the negativity differs across violation types [39,40], as well as evidence that metrically inconsistent pseudowords also elicit such negativities, even though they lack semantic context [45]. Alternatively, it could be that the negativity we observed in the current experiment indicates the violation of a consistent, rule-based sequence, in this case realized as the metric structure [45].

In contrast, the positivity observed between 365–435 ms for WS violations in the current study, and in a similar time window in other studies, could be related to conflict processing. When a WS violation occurs, the reader must resolve the conflict between a metric context which leads them to predict a SW target and a semantic context which leads them to predict a WS target. In addition, there is the added conflict that WS two-syllable words are phonologically marked in the language. These factors together may lead to the observed positivity, which is signaling an error in processing that is harder for readers to recover from. This interpretation is consistent with previous ERP research of the metric structure in German, where metric violations in three-syllable words that did not violate metric foot structure led to an early negativity, whereas violations that also conflicted with foot structure resulted in an early positivity [36], similar to the results in the current study.

Consistent with other explicit and implicit metric processing studies that do not involve an explicit metric task, we did not observe evidence of a late positivity for metric violations relative to consistent metric conditions. Previous studies of both explicit and implicit metric processing demonstrate that late positivities in response to metric violation are most likely observed when the participant's task is to assess the metric structure. Indeed, only one previous study of implicit metric processing observed a late positivity in response to metric violations [54] while two others did not [53,55], and none of these studies required an explicit metric judgment. This interpretation is in line with previous work showing a dissociation between early and late ERP effects of syntactic violations, where early negativities are thought to reflect automatic processing and late positivities are thought to reflect controlled processes

of repair [67,68] and the difficulty of the required repair process [69]. The current results suggest that although both implicit and explicit metric violations are automatically detected, as evidenced by early (<500 ms) waveform differences, only violations that rise to the level of awareness give rise to a late positivity.

It is also possible that the lack of a late positivity in the current study reflects a lack of power; our choice to present the same orthographic information across conditions meant that the number of items in the experiment was limited by the number of two-syllable stress-alternating noun–verb homographs in English that were known to our participants and could be embedded in rhyming couplets. Moreover, compared to previous studies using word lists, the stress pattern of the target in the current study was locally ambiguous, and only disambiguated by the implicit metric structure provided by the context. Although this manipulation is a better test of the abstract metric structure compared to other studies that used different items across SW and WS conditions, it produces a less clearly defined metric violation than paradigms employing single target words with unambiguous stress patterns.

Although the current results are generally consistent with prior ERP work on explicit and implicit linguistic metric structures, they are inconsistent with results observed in a previous eye-tracking experiment using the same materials. Recall that Breen and Clifton observed inflated reading times only for WS violations, and not for SW violations [13]; moreover, these effects were observed only in relatively later reading time measures. Conversely, our results demonstrate significant early ERP differences for both SW and WS violations, though they differ in polarity, timing, and topography. These differential effects are likely due to differences in the temporal control of stimulus presentation between the studies. In Breen and Clifton's experiment, participants read normally at their own pace, meaning they could take as much time as needed to process material in advance of the critical word, and could look back to prior sentence material to resolve difficulty generated at the target word. In contrast, materials in the current study were presented in a region-by-region segmented fashion, giving participants less time to generate predictions about upcoming material, and disallowing regressions. Moreover, the fact that the current materials were presented in a time-controlled manner means that the metric structure of the sentence materials was more obvious for readers, making the metric inconsistency more explicit, resulting in significant ERP effects of both types of metric violations.

Future work could directly investigate the role of temporal stimulus control on implicit metric violation processing by replicating the current paradigm using simultaneous collection of eye-tracking data and ERPs, a method which has already been used to successfully adjudicate debates about linguistic processing in eye movements [70,71]. In this way, the role of metric inconsistency in silent reading could be assessed without explicitly controlling the timing of materials. Additionally, while current results demonstrate that readers engage in implicit prosody during silent reading of poetry, it is an open question to what extent these findings generalize to normal reading. The couplets used in the current study were designed to have strict metric and rhyming structure, which is rare in non-poetic language. However, our study does provide an insight to the role of meter in implicit prosody. To determine whether our result can be replicated in non-poetic contexts which do not have concomitantly high metrical expectancies, future work will explore differences in brain activity in response to metric violations in silently-read prose sentences.

5. Conclusions

The current results provide further evidence of an intimate link between metric processing during listening and metric processing during silent reading, which may help inform our understanding of previously described relationships between children's sensitivity to an auditory metric structure and silent reading comprehension. For example, the ability of older children to track a perceived metric structure predicts phonological awareness and reading outcomes [72,73], and children's ability to detect a mis-stressed word predicts phonological awareness and word knowledge [74]. It may be the case that these reading abilities are facilitated by implicit metric structure representations. This claim is

further bolstered by a relationship between prosodic fluency and reading comprehension in high school students—those who demonstrate higher prosodic fluency also showed an increased comprehension ability [75,76]. Research about implicit prosody and the underlying neurocognitive processes occurring during silent reading may, therefore, inform future work designing prosodic interventions to improve children's reading comprehension abilities.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3425/9/8/192/s1, Table S1: Distribution of stress patterns across segments for experimental and filler couplets, Figure S1: Moving window cluster-based permutation test results.

Author Contributions: Conceptualization, M.B.; methodology, M.B., A.B.F.; software, M.B., A.B.F.; validation, M.B., A.B.F.; formal analysis, M.B., A.B.F., M.O.A.; investigation, M.O.A.; resources, M.B.; data curation, M.B., A.B.F., M.O.A.; writing—original draft preparation, M.B.; writing—review and editing, A.B.F., M.O.A.; visualization, M.B., A.B.F., M.O.A.; supervision, M.B.; project administration, M.B., A.B.F.; funding acquisition, M.B., M.O.A.

Funding: This research was funded in part by a James S. McDonnell Foundation Scholar Award in Understanding Human Cognition to author M.B. and a Harap Scholarship Fund Award to author M.O.A.

Acknowledgments: The authors would like to acknowledge the assistance of Charles Clifton with materials development, Elizabeth Brija with experimental coding, and Xuefei Chen, Hannah Galloway, Margaret Golder, Johanna Kneifel, Priscilla Lopez, and Corrin Moss with data collection for the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

1a. SW/consistent

You must hear/my story,/your highness.
I have the/young princess's/address

1b. SW/inconsistent

My workspace/is such a/big mess.
I lost an/important/address

1c. WS/consistent

My workspace/is such a/big mess.
My clutter/I have to/address

1d. WS/inconsistent

You must hear/my story,/your highness.
Your habits I/find I must/address

2a. SW/consistent

The guy who/got lost on/a flyby
Dropped all of/his bombs on an/ally

2b. SW/inconsistent

I know a/young woman/from Rye,
Who'd make such/a lovely/ally

2c. WS/consistent

I know a/young woman/from Rye,
With whom I/would like to/ally

2d. WS/inconsistent

The guy who/got lost on/a flyby
Killed folks with whom/we want to/ally

3a. SW/consistent

I just saw/a dog and/a tomcat,
Engaged in/some furious/combat

3b. SW/inconsistent

I witnessed/a dog and/a cat,
Engaged in/some angry/combat

3c. WS/consistent

I witnessed/a dog and/a cat,
Who seemingly/tried to/combat

3d. WS/inconsistent

I just saw/a dog and/a tomcat,
That we must/be ready to/combat

4a. SW/consistent

I heard someone/say through/the grapevine:
The farmer is/driving his/combine

4b. SW/inconsistent

The farmer/got caught/drinking wine,
Then harvesting/in his/combine

4c. WS/consistent

The farmer/got caught/drinking wine,
And shotguns/and booze don't/combine

4d. WS/inconsistent

I heard someone/say through/the grapevine:
That farmer is/hoping to/combine

5a. SW/consistent

I processed/some prints in/the darkroom
Of people I'd/met on a/commune

5b. SW/inconsistent

I know some/who worship/the moon,
And live/in a hippie/commune

5c. WS/consistent

I know some/who worship/the moon,
With nature/they like to/commune

5d. WS/inconsistent

I processed/some prints in/the darkroom
Of folks who/just wanted to/commune

6a. SW/consistent

If out in/the mountains/you backpack,
Your team must/agree to this/compact

6b. SW/inconsistent

Before you/head out with/that pack,
Your team has/to sign this/compact

6c. WS/consistent

Before you/head out with/that pack,
Be sure that/your gear is/compact

6d. WS/inconsistent

If out in/the mountains/you backpack,
Your gear must/be basic and/compact

7a. SW/consistent

The crew worked/so hard for/their paychecks
They thought they'd/develop a/complex

7b. SW/inconsistent

There once was/a young man/named Rex
Who owned/an apartment/complex

7c. WS/consistent

There once was/a young man/named Rex
Whose theories/were big and/complex

7d. WS/inconsistent

The crew worked/so hard for/their paychecks
Their work was/so terribly/complex

8a. SW/consistent

We stayed in/the woods at/a campground,
Which wasn't too/far from a/compound

8b. SW/inconsistent

We got that/old dog at/the pound
He came from/a private/compound

8c. WS/consistent

We stayed in/the woods at/a campground,
Our pleasure in/nature to/compound

8d. WS/inconsistent

We got that/old dog at/the pound
Our sadness/will surely/compound

9a. SW/consistent

There was a/young heroin/addict,
Who ended up/causing a/conflict

9b. SW/inconsistent

My parents/are being/quite strict.
Our views are/in open/conflict

9c. WS/consistent

My parents/are being/quite strict.
Their wishes/and mine do/conflict

9d. WS/inconsistent

There was a/young heroin/addict,
Whose habits and/others did/conflict

10a. SW/consistent

The athlete/who just failed/a drugtest,
Will soon face/a challenging/contest

10b. SW/inconsistent

The athlete/who thinks he's/the best
Just lost an/important/contest

10c. WS/consistent

The athlete/who thinks he's/the best
Holds titles/that others/contest

10d. WS/inconsistent

The athlete/who just failed/a drugtest,
Is planning the/charges to/contest

11a. SW/consistent

Although that/young man is/an addict,
He really should/not be a/convict

11b. SW/inconsistent

I think that/the judge was/too strict
In sentencing/that young/convict

11c. WS/consistent

I think that/the judge was/too strict,
The jury too/quick to/convict

11d. WS/inconsistent

Although that/young man is/an addict,
I think that the/judge shouldn't/convict

12a. SW/consistent

That man applies/way too much/hair grease.
A friend should/suggest a big/decrease

12b. SW/inconsistent

Forgive me/for stating/my peace,
But you must/commence a/decrease

12c. WS/consistent

Forgive me/for stating/my peace,
Your appetite/you must/decrease

12d. WS/inconsistent

That man applies/way too much/hair grease.
I think the/amount he should/decrease

13a. SW/consistent

The Soviet/spy is a/suspect.
The case has/but one major/defect

13b. SW/inconsistent

The Soviet/spy they/suspect,
Has plans with/a major/defect

13c. WS/consistent

The Soviet/spy they/suspect,
Is planning/quite soon to/defect

13d. WS/inconsistent

The Soviet/spy is/a suspect.
I heard that/he's planning to/defect

14a. SW/consistent

In nothing/but jeans and/a t-shirt,
That man took/a trip 'cross the/desert

14b. SW/inconsistent

The fighting/he tried/to avert,
By running off/through the/desert

14c. WS/consistent

The fighting/he tried to/avert,
By choosing/his squad to/desert

14d. WS/inconsistent

In nothing but/jeans and a/t-shirt,
A soldier his/squad chose to/desert

15a. SW/consistent

I know of/an elegant/female
Her outfits lack/no fashion/detail

15b. SW/inconsistent

There once was/a woman/named Gail
Whose fashion/had every/detail

15c. WS/consistent

I know of/an elegant/female
Who wanted/her auto to/detail

15d. WS/inconsistent

There once was/a woman/named Gail
Who wanted/her car to/detail

16a. SW/consistent

We once had/a tiresome/house guest,
Who loved to/read Birdwatcher's/Digest

16b. SW/inconsistent

We once had/a friend as/a guest,
Who loved to/skim Reader's/Digest

16c. WS/consistent

We once had/a friend as/a guest,
Whose cooking/we could not/digest

16d. WS/inconsistent

We once had/a tiresome/house guest,
Whose humor/was painful to/digest

17a. SW/consistent

The gymnast/requested/a recount
Her score, she/thought, rated no/discount

17b. SW/inconsistent

He could not/afford the/amount,
And asked for/a modest/discount

17c. WS/consistent

He could not/afford the/amount.
The invoice they/would not/discount

17d. WS/inconsistent

The gymnast/requested/a recount
She thought it/was wrongful to/discount

18a. SW/consistent

In order to/prove your/attendance
You'll have to/check in at the/entrance

18b. SW/inconsistent

This gorgeous/young woman/from France
Made everyone/jam the/entrance

18c. WS/consistent

This gorgeous/young woman/from France
Would often/the young men/entrance

18d. WS/inconsistent

There was a/young woman/whose nude dance
Would always/the gentlemen/entrance

19a. SW/consistent

He tried not/to get badly/sidetracked.
He needed/some raspberry/extract

19b. SW/inconsistent

The recipe/seemed quite/exact.
It called for/some almond/extract

19c. WS/consistent

The recipe/seemed quite/exact.
Some essence/you had to/extract

19d. WS/inconsistent

He tried not/to get badly/sidetracked
Some essence/he wanted to/extract

20a. SW/consistent

The city/must safeguard/the seaports,
To save us/from dangerous/imports

20b. SW/inconsistent

The panel/is set to/report
On how much/we pay for/imports

20c. WS/consistent

The panel/is set to/report
On how much/the city/imports

20d. WS/inconsistent

The city/must safeguard/the seaports,
Because of how/much it now/imports

21a. SW/consistent

The man who/asked you for/a consult
Was given/a horrible/insult

21b. WS/consistent

That woman/who likes the/occult,
Is very/unsafe to/insult

21c. WS/inconsistent

The man who/asked you for/a consult
Is no-one/you wanted to/insult

21d. SW/inconsistent

That woman/who likes/the occult,
Will tolerate/no more/insults

22a. SW/consistent

The teacher/assigned them/a project
To find an/unusual/object

22b. SW/inconsistent

The winners/will get to/select
A shiny/expensive/object

22c. WS/consistent

The mayor/that we might/elect
Has views to/which others/object

22d. WS/inconsistent

The teacher/assigned them/a project
That forced many/parents to/object

23a. SW/consistent

There once was/an old man/named Kermit,
Who hunted/without any/permit

23b. SW/inconsistent

There once was/an old man/named Britt
Who hunted/without a/permit

23c. WS/consistent

There once was/an old man/named Britt
Whose vices/no wife could/permit

23d. WS/inconsistent

There once was/an old man/named Kermit
Whose gambling his/wife would not/permit

24a. SW/consistent

I know of/an old man/named Herbert,
Who's known around/town as a/pervert

24b. SW/inconsistent

The nun/did her best/to convert
A man whom/they call a/pervert

24c. WS/consistent

That nun/did her best/to convert
Young kids who/the truth do/pervert

24d. WS/inconsistent

I know of/an old man/named Herbert
Who always the/truth tries to/pervert

25a. SW/consistent

There once was/a penniless/peasant,
Who couldn't/afford a nice/present

25b. SW/inconsistent

There once was/a clever/young gent,
Who bought for/his girl a/present

25c. WS/consistent

There once was/a clever/young gent,
Who had a/nice talk to/present

25d. WS/inconsistent

There once was/a penniless/peasant,
Who went to/his master to/present

26a. SW/consistent

He couldn't/hide all of/his misdeeds,
But made off/with all of the/proceeds

26b. SW/inconsistent

In light/of the man's/dirty deeds,
He won't/receive any/proceeds

26c. WS/consistent

In light/of the man's/dirty deeds,
On Monday/his trial/proceeds

26d. WS/inconsistent

He couldn't/hide all of/his misdeeds
On Monday/his retrial/proceeds

27a. SW/consistent

There once was/a crusty old/recluse,
Who grew the/most wonderful/produce

27b. SW/inconsistent

There simply/is no good/excuse
For failing to/eat your/produce

27c. WS/consistent

There simply/is no good/excuse
For failing/to work and/produce

27d. WS/inconsistent

There once was/a crusty/old recluse,
Whose garden great/harvests would/produce

28a. SW/consistent

With all of/their time spent/at recess,
The children/make no forward/progress

28b. SW/inconsistent

The efforts/at peace,/I confess,
Are making/no forward/progress

28c. WS/consistent

The efforts/at peace,/I confess,
Will simply/no longer/progress

28d. WS/inconsistent

With all of/their time spent/at recess,
The children will/soon fail to/progress

29a. SW/consistent

I noticed/a ruinous/defect
In part of/the candidate's/project

29b. SW/inconsistent

The man we/will likely/elect
Endorses/this wacky/project

29c. WS/consistent

The mayor that/folks will/elect
According/to what polls/project

29d. WS/inconsistent

I noticed/a ruinous/defect
In what that/new candidate/projects

30a. SW/consistent

There once was/a young man/named Ernest,
Who sponsored/a violent/protest

30b. SW/inconsistent

They put the/man under/arrest
For leading/an angry/protest

30c. WS/consistent

They put the/man under/arrest,
And gave him/no time to/protest

30d. WS/inconsistent

There once was/a young man/named Ernest,
Who rounded up/people to/protest

31a. SW/consistent

In a voice/that was piercing/and treble,
The serfs were/inspired by a/rebel

31b. SW/inconsistent

The infantry/failed to/repel
The followers/of the/rebel

31c. WS/consistent

The infantry/failed to/repel
The fighters/who want to/rebel

31d. WS/inconsistent

In a voice/that was piercing/and treble,
The leader urged/peasants to/rebel

32a. SW/consistent

That basketball/star's like a/bloodhound.
He seeks out/and catches each/rebound

32b. SW/inconsistent

The basketball/star turned/around,
and caught an/amazing/rebound

32c. WS/consistent

The basketball/star turned/around,
And watched for/the shot to/rebound

32d. WS/inconsistent

That basketball/star's like a/bloodhound.
He waits for/each jumpshot to/rebound

33a. SW/consistent

I met an/old friend who/played baseball,
Who warned of/a new safety/recall

33b. SW/inconsistent

I met an/old friend at/the mall,
Who warned of/a safety/recall

33c. WS/consistent

I met an/old friend at/the mall,
Whose name I/just could not/recall

33d. WS/inconsistent

I met an/old friend who/played baseball,
But what his/name was I can't/recall

34a. SW/consistent

There once was/a young man/named Eckerd
Who broke an old/pole-vaulting/record

34b. SW/inconsistent

The athlete/won quite an/award
For breaking/the scoring/record

34c. WS/consistent

The athlete/won quite an/award
The cameras/were there to/record

34d. WS/inconsistent

There once was/a young man/named Eckerd
Whose pole-vaulting/feats they did/record

35a. SW/consistent

Last year I/created a/stock fund.
And managed to/get a big/refund

35b. SW/inconsistent

I have to/admit I/am stunned,
You didn't/give me my/refund

35c. WS/consistent

I have to/admit I/am stunned,
My payments/you will not/refund

35d. WS/inconsistent

Last year I/created a/stock fund.
The fees they/would happily/refund

36a. SW/consistent

The judges must/all watch/the replay
To find out which/team won the/relay

36b. SW/inconsistent

A messenger/came by/today
To find out/who won the/relay

36c. WS/consistent

A messenger/came by/today;
A message/he had to/relay

36d. WS/inconsistent

The judges must/all watch/the replay.
Results to the/coach they will/relay

37a. SW/consistent

I read an/unusual/essay
'Bout how they/conducted a/survey

37b. SW/inconsistent

A lovely/young woman/named Fay
Was asked to/complete a/survey

37c. WS/consistent

A lovely/young woman/named Fay
The future/she liked to/survey

37d. WS/inconsistent

I read an/unusual/essay
Describing how/folks tried to/survey

38a. SW/consistent

The cops are/an interesting/subject
They bullied their/most recent/suspect

38b. SW/inconsistent

The cops/didn't try/to protect
A recently/collared/suspect

38c. WS/consistent

The cops/didn't try/to protect
The people/they chose to/suspect

38d. WS/inconsistent

The cops are/an interesting/subject
They bully/the people they/suspect

39a. SW/consistent

A striking young/woman named/Rembrandt,
From Portugal,/she was a/transplant

39b. SW/inconsistent

A striking young/dame named/van Zandt,
From Spain was/a recent/transplant

39c. WS/consistent

A striking young/dame named/van Zandt
Had roses/she hoped to/transplant

39d. WS/inconsistent

A striking young/woman named/Rembrandt,
Had roses she/wanted to/transplant

40a. SW/consistent

To get to/the local gym's/squash court,
You must take/municipal/transport

40b. SW/inconsistent

The mafia/tried to/extort
The captain/of public/transport

40c. WS/consistent

The mafia/tried to/extort
A man who/had tried to/transport

40d. WS/inconsistent

To get to/the local/gym's squash court,
Your gear should/be ready to/transport

References

1. Fodor, J.D. Learning To Parse? *J. Psycholinguist. Res.* **1998**, *27*, 285–319. [CrossRef]
2. Fodor, J.D. Psycholinguistics cannot escape prosody. In Proceedings of the Speech Prosody 2002, International Conference, Aix-en-Provence, France, 11–13 April 2002.
3. Bader, M. Prosodic Influences on Reading Syntactically Ambiguous Sentences. In *Reanalysis in Sentence Processing*; Fodor, J.D., Ferreira, F., Eds.; Studies in Theoretical Psycholinguistics; Springer: Dordrecht, The Netherlands, 1998; pp. 1–46. ISBN 978-90-481-5037-3.
4. Breen, M. Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Lang. Linguist. Compass* **2014**, *8*, 37–50. [CrossRef]
5. Breen, M. Empirical Investigations of Implicit Prosody. In *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor*; Frazier, L., Gibson, E., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 177–192. ISBN 978-3-319-12960-0.
6. Abramson, M. The Written Voice: Implicit Memory Effects of Voice Characteristics following Silent Reading and Auditory Presentation. *Percept. Mot. Skills* **2007**, *105*, 1171–1186. [CrossRef] [PubMed]

7. Speer, S.R.; Foltz, A. The implicit prosody of corrective contrast primes appropriately intonated probes (for some readers). In *Explicit and Implicit Prosody in Sentence Processing: Studies in Honor of Janet Dean Fodor*; Frazier, L., Gibson, E., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 263–285. ISBN 978-3-319-12960-0.
8. Hwang, H.; Schafer, A.J. Constituent Length Affects Prosody and Processing for a Dative NP Ambiguity in Korean. *J. Psycholinguist. Res.* **2009**, *38*, 151. [CrossRef] [PubMed]
9. Augurzky, P. *Attaching Relative Clauses in German: The Role of Implicit and Explicit Prosody in Sentence Processing*; MPI Series in Human Cognitive and Brain Science: Leipzig, Germany, 2006; Volume 77.
10. Hirose, Y. Recycling Prosodic Boundaries. *J. Psycholinguist. Res.* **2003**, *32*, 167–195. [CrossRef] [PubMed]
11. Ashby, J.; Clifton, C., Jr. The prosodic property of lexical stress affects eye movements during silent reading. *Cognition* **2005**, *96*, B89–B100. [CrossRef] [PubMed]
12. Breen, M.; Clifton, C. Stress matters revisited: A boundary change experiment. *Q. J. Exp. Psychol.* **2013**, *66*, 1896–1909. [CrossRef]
13. Breen, M.; Clifton, C., Jr. Stress matters: Effects of anticipated lexical stress on silent reading. *J. Mem. Lang.* **2011**, *64*, 153–170. [CrossRef]
14. Kentner, G. Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition* **2012**, *123*, 1–20. [CrossRef]
15. Kentner, G.; Vasishth, S. Prosodic Focus Marking in Silent Reading: Effects of Discourse Context and Rhythm. *Front. Psychol.* **2016**, *7*, 319. [CrossRef]
16. McCurdy, K.; Kentner, G.; Vasishth, S. Implicit prosody and contextual bias in silent reading. *J. Eye Mov. Res.* **2013**, *6*. [CrossRef]
17. Hayes, B. *Metrical Stress Theory: Principles and Case Studies*; University of Chicago Press: Chicago, IL, USA, 1995; ISBN 978-0-226-32103-5.
18. Liberman, M.; Prince, A. On stress and linguistic rhythm. *Linguist. Inq.* **1977**, *8*, 249–336.
19. Selkirk, E. References—Scientific Research Publish. In *Phonology and Syntax: The Relation between Sound and Structure*; The MIT Press: Cambridge, MA, USA, 1984; Available online: http://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/ReferencesPapers.aspx?ReferenceID=918312 (accessed on 30 June 2017).
20. Nespor, M.; Vogel, I. *Prosodic Phonology: With a New Foreword*; Walter de Gruyter: Berlin, Germany, 2007; ISBN 978-3-11-019790-7.
21. Beckman Mary, E. *Stress and Non-Stress Accent*; De Gruyter Mouton: Berlin, Germany; Boston, MA, USA, 2012; ISBN 978-3-11-013729-3.
22. Fry, D.B. Duration and Intensity as Physical Correlates of Linguistic Stress. *J. Acoust. Soc. Am.* **1955**, *27*, 765–768. [CrossRef]
23. Breen, M. Effects of metric hierarchy and rhyme predictability on word duration in The Cat in the Hat. *Cognition* **2018**, *174*, 71–81. [CrossRef] [PubMed]
24. Fitzroy, A.B.; Breen, M. Metric Structure and Rhyme Predictability Modulate Speech Intensity During Child-Directed and Read-Alone Productions of Children's Literature. *Lang. Speech* **2019**. [CrossRef]
25. Shields, J.L.; McHugh, A.; Martin, J.G. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *J. Exp. Psychol.* **1974**, *102*, 250–255. [CrossRef]
26. Mattys, S.L.; Samuel, A.G. How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *J. Mem. Lang.* **1997**, *36*, 87–116. [CrossRef]
27. Cutler, A.; Norris, D. The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* **1988**, *14*, 113–121. [CrossRef]
28. Cutler, A.; Dahan, D.; van Donselaar, W. Prosody in the Comprehension of Spoken Language: A Literature Review. *Lang. Speech* **1997**, *40*, 141–201. [CrossRef]
29. Pitt, M.A.; Samuel, A.G. The use of rhythm in attending to speech. *J. Exp. Psychol. Hum. Percept. Perform.* **1990**, *16*, 564–573. [CrossRef]
30. Breen, M.; Dilley, L.C.; McAuley, J.D.; Sanders, L.D. Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Lang. Cogn. Neurosci.* **2014**, *29*, 1132–1146. [CrossRef] [PubMed]
31. Dilley, L.C.; Mattys, S.L.; Vinke, L. Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *J. Mem. Lang.* **2010**, *63*, 274–294. [CrossRef]

32. Brown, M.; Salverda, A.P.; Dilley, L.C.; Tanenhaus, M.K. Expectations from preceding prosody influence segmentation in online sentence processing. *Psychon. Bull. Rev.* **2011**, *18*, 1189–1196. [CrossRef] [PubMed]
33. Dilley, L.C.; McAuley, J.D. Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* **2008**, *59*, 294–311. [CrossRef]
34. Cutler, A.; Carter, D.M. The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* **1987**, *2*, 133–142. [CrossRef]
35. Cutler, A.; Clifton, C., Jr. The use of prosodic information in word recognition. In *Attention and Performance X: Control of Language Processes*; Erlbaum: Hillsdale, NJ, USA, 1984; pp. 183–196.
36. Domahs, U.; Wiese, R.; Bornkessel-Schlesewsky, I.; Schlesewsky, M. The Processing of German Word Stress: Evidence for the Prosodic Hierarchy. *Phonology* **2008**, *25*, 1–36. [CrossRef]
37. Domahs, U.; Genc, S.; Knaus, J.; Wiese, R.; Kabak, B. Processing (un-)predictable word stress: ERP evidence from Turkish. *Lang. Cogn. Process.* **2013**, *28*, 335–354. [CrossRef]
38. Molczanow, J.; Domahs, U.; Knaus, J.; Wiese, R. The lexical representation of word stress in Russian: Evidence from event-related potentials. *Ment. Lex.* **2013**, *8*, 164–194.
39. Magne, C.; Astésano, C.; Aramaki, M.; Ystad, S.; Kronland-Martinet, R.; Besson, M. Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cereb. Cortex* **2007**, *17*, 2659–2668. [CrossRef]
40. Marie, C.; Magne, C.; Besson, M. Musicians and the metric structure of words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef]
41. Magne, C.; Jordan, D.K.; Gordon, R.L. Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain Lang.* **2016**, *153–154*, 13–19. [CrossRef] [PubMed]
42. Böcker, K.B.E.; Bastiaansen, M.C.M.; Vroomen, J.; Brunia, C.H.M.; Gelder, B.D. An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology* **1999**, *36*, 706–720. [CrossRef] [PubMed]
43. Schmidt-Kassow, M.; Kotz, S.A. Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* **2009**, *21*, 1693–1708. [CrossRef] [PubMed]
44. Rothermich, K.; Schmidt-Kassow, M.; Kotz, S.A. Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia* **2012**, *50*, 232–244. [CrossRef] [PubMed]
45. Rothermich, K.; Schmidt-Kassow, M.; Schwartze, M.; Kotz, S.A. Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport* **2010**, *21*, 580–584. [CrossRef]
46. Bohn, K.; Knaus, J.; Wiese, R.; Domahs, U. The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia* **2013**, *51*, 760–771. [CrossRef]
47. Henrich, K.; Wiese, R.; Domahs, U. How information structure influences the processing of rhythmic irregularities: ERP evidence from German phrases. *Neuropsychologia* **2015**, *75*, 431–440. [CrossRef]
48. Henrich, K.; Alter, K.; Wiese, R.; Domahs, U. The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain Lang.* **2014**, *136*, 19–30. [CrossRef]
49. McCauley, S.M.; Hestvik, A.; Vogel, I. Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Lang. Speech* **2013**, *56*, 23–44. [CrossRef]
50. Friederici, A.D.; Kotz, S.A. The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* **2003**, *20*, S8–S17. [CrossRef]
51. Kutas, M.; Federmeier, K.D. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* **2011**, *62*, 621–647. [CrossRef] [PubMed]
52. Rothermich, K.; Kotz, S.A. Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *NeuroImage* **2013**, *70*, 89–100. [CrossRef] [PubMed]
53. Magne, C.; Gordon, R.L.; Midha, S. Influence of metrical expectancy on reading words: An ERP study. In Proceedings of the Speech Prosody 2010-Fifth International Conference, Chicago, IL, USA, 10–14 May 2010.
54. Kriukova, O.; Mani, N. Processing metrical information in silent reading: An ERP study. *Front. Psychol.* **2016**, *7*, 1432. [CrossRef] [PubMed]
55. Fotidzis, T.; Moon, H.; Steele, J.; Magne, C. Cross-Modal Priming Effect of Rhythm on Visual Word Recognition and Its Relationships to Music Aptitude and Reading Achievement. *Brain Sci.* **2018**, *8*, 210. [CrossRef] [PubMed]
56. Francis, W.; Kucera, H. *Frequency Analysis of English Usage*; Houghton Mifflin: Boston, MA, USA, 1982.
57. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef] [PubMed]

58. Lopez-Calderon, J.; Luck, S.J. ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* **2014**, *8*, 213. [CrossRef] [PubMed]
59. Luck, S.J. *An Introduction to the Event-Related Potential Technique*; MIT Press: Cambridge, MA, USA, 2014; ISBN 978-0-262-32406-9.
60. Groppe, D.M.; Urbach, T.P.; Kutas, M. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology* **2011**, *48*, 1711–1725. [CrossRef]
61. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011.
62. Lawrence, M.A. ez: Easy Analysis and Visualization of Factorial Experiments. R Package Version 4.4-0. 2016. Available online: http://CRAN.R-project.org/package=ez (accessed on 30 June 2017).
63. Zhao, T.C.; Lam, H.T.G.; Sohi, H.; Kuhl, P.K. Neural processing of musical meter in musicians and non-musicians. *Neuropsychologia* **2017**, *106*, 289–297. [CrossRef]
64. Vuust, P.; Pallesen, K.J.; Bailey, C.; van Zuijen, T.L.; Gjedde, A.; Roepstorff, A.; Østergaard, L. To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage* **2005**, *24*, 560–564. [CrossRef]
65. Geiser, E.; Ziegler, E.; Jancke, L.; Meyer, M. Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* **2009**, *45*, 93–102. [CrossRef]
66. Féry, C. German Word Stress in Optimality Theory. *J. Comp. Ger. Linguist.* **1998**, *2*, 101–142. [CrossRef]
67. Hahne, A.; Friederici, A.D. Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *J. Cogn. Neurosci.* **1999**, *11*, 194–205. [CrossRef] [PubMed]
68. Kuperberg, G.R. Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res.* **2007**, *1146*, 23–49. [CrossRef] [PubMed]
69. O'Rourke, P.L.; Van Petten, C. Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Res.* **2011**, *1392*, 62–79. [CrossRef] [PubMed]
70. Kretzschmar, F.; Schlesewsky, M.; Staub, A. Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *J. Exp. Psychol. Learn. Mem. Cogn.* **2015**, *41*, 1648–1662. [CrossRef] [PubMed]
71. Henderson, J.M.; Luke, S.G.; Schmidt, J.; Richards, J.E. Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Front. Syst. Neurosci.* **2013**, *7*, 28. [CrossRef] [PubMed]
72. Gordon, R.L.; Shivers, C.M.; Wieland, E.A.; Kotz, S.A.; Yoder, P.J.; Devin McAuley, J. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* **2015**, *18*, 635–644. [CrossRef] [PubMed]
73. Huss, M.; Verney, J.P.; Fosker, T.; Mead, N.; Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* **2011**, *47*, 674–689. [CrossRef]
74. Wood, C. Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *J. Res. Read.* **2006**, *29*, 270–287. [CrossRef]
75. Breen, M.; Kaswer, L.; Van Dyke, J.A.; Krivokapić, J.; Landi, N. Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers. *Front. Psychol.* **2016**, *7*, 1026. [CrossRef]
76. Benjamin, R.G.; Schwanenflugel, P.J. Text Complexity and Oral Reading Prosody in Young Readers. *Read. Res. Q.* **2010**, *45*, 388–404. [CrossRef]

brain sciences

MDPI

Review

Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception

Brett R. Myers [1,2,*], Miriam D. Lense [1,3,4,5] and Reyna L. Gordon [1,4,5,6,*]

1 Department of Otolaryngology, Vanderbilt University Medical Center, 1215 21st Ave S, Nashville, TN 37232, USA
2 Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, USA
3 Vanderbilt Kennedy Center, 110 Magnolia Circle, Nashville, TN 37203, USA; miriam.lense@vumc.org
4 Vanderbilt Brain Institute, Vanderbilt University, 2215 Garland Ave, Nashville, TN 37232, USA
5 The Curb Center for Art, Enterprise, and Public Policy, Vanderbilt University, 1801 Edgehill Avenue, Nashville, TN 37212, USA
6 Department of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37240, USA
* Correspondence: brett.myers@vanderbilt.edu (B.R.M.); reyna.gordon@vanderbilt.edu (R.L.G.)

Received: 31 December 2018; Accepted: 15 March 2019; Published: 22 March 2019

Abstract: Prosodic cues in speech are indispensable for comprehending a speaker's message, recognizing emphasis and emotion, parsing segmental units, and disambiguating syntactic structures. While it is commonly accepted that prosody provides a fundamental service to higher-level features of speech, the neural underpinnings of prosody processing are not clearly defined in the cognitive neuroscience literature. Many recent electrophysiological studies have examined speech comprehension by measuring neural entrainment to the speech amplitude envelope, using a variety of methods including phase-locking algorithms and stimulus reconstruction. Here we review recent evidence for neural tracking of the speech envelope and demonstrate the importance of prosodic contributions to the neural tracking of speech. Prosodic cues may offer a foundation for supporting neural synchronization to the speech envelope, which scaffolds linguistic processing. We argue that prosody has an inherent role in speech perception, and future research should fill the gap in our knowledge of how prosody contributes to speech envelope entrainment.

Keywords: prosody; speech envelope; neural entrainment; rhythm; EEG

> "In a house constructed of speech, the bricks are phonemes, and the mortar is prosody. Without the latter, we'd simply live under a pile of rocks". —B.R.M.

1. Prosody Perception

Prosody is the stress, intonation, and rhythm of speech, which provides suprasegmental linguistic features across phonemes, syllables, and phrases [1–3]. Prosodic cues contribute affect and intent to an utterance [4] as well as emphasis, sarcasm, and more nuanced emotional states [5,6]. Certain prosodic cues are universal and can be interpreted cross-culturally even in an unfamiliar language [7,8]. Prosody also provides valuable markers for parsing a continuous speech stream into meaningful segments such as intonational phrase boundaries [9], dynamic pitch changes [10], and metrical information [11]. Parsing speech units based on prosodic perception is an imperative early stage in language acquisition, and it is considered a precursor to vocabulary and grammar development [12–14]. In addition, prosody can convey semantic information for context in a message [15,16]. Deficits in prosody perception have a negative downstream impact on linguistic abilities, literacy, and social interactions, e.g., [17–20].

Prosodic fluctuations are responsible for communicating a wealth of information, primarily through acoustic correlates such as duration, amplitude, and fundamental frequency. As any of these parameters changes, it influences the expression of stress, intonation, and rhythm of the spoken message [21,22]. One illustration of the dynamic and multidimensional nature of prosody is "motherese" or infant-directed speech, which is characterized by exaggerations in duration and fundamental frequency [23]. The exaggerated speech signal creates louder, longer, and higher pitch stressed syllables [24], which facilitates segmenting the speech into syllable components and disentangling word boundaries [25,26]. The modified prosodic qualities of infant-directed speech make the signal acoustically salient and engaging for infants [23,27], which yield later linguistic benefits such as boosts in vocabulary acquisition [28] and accessing syntactic structures [29]. This is one example of how prosody plays an important role in speech communication.

The importance of prosody to speech perception is widely acknowledged, yet it has been underrated in many studies examining neural entrainment to the speech envelope. The purpose for the current review is to demonstrate that prosodic processing is engrained in investigations of neural entrainment to speech and to encourage researchers to explicitly consider the effects of prosody in future investigations. We will review the speech envelope and its relation to the prosodic features of duration, amplitude, and fundamental frequency, and we will discuss electrophysiological methods for measuring speech envelope entrainment in neural oscillations. We will then highlight some previous research using these methods in typical and atypical populations with an emphasis on how the findings may be connected to prosody. Finally, we propose directions for future research in this field. It is our hope to draw attention to the role of prosody processing in neural entrainment to speech and to encourage researchers to examine the neural underpinnings of prosodic processing.

2. Amplitude Modulation

Prosody is determined by a series of acoustic correlates—duration, amplitude, and fundamental frequency—which can be represented in a number of ways, including the amplitude modulation (AM) envelope (also known as the temporal envelope) [30,31]. It is important to mention that a temporal waveform is composed of a "fine structure" and an "envelope". Fine structure consists of fast-moving spectral content (e.g., frequency characteristics of phonemes), while the envelope captures the broad contour of pressure variations in the signal (e.g., amplitude over time) [32]. In other words, the envelope is superimposed over the more rapidly oscillating fine structure. Both envelope and spectral components are important for speech comprehension—i.e., to "recognize speech" rather than "wreck a nice beach" (Figure 1) (see [33] but also [34]).

It has been suggested that the extraction of AM information is a fundamental procedure within the neural architecture of the auditory system [35]. The auditory cortex is particularly adept at rapidly processing spectro-temporal changes in the temporal fine structure [36], and it is possible that this processing is aided by the amplitude envelope first laying the foundation for more narrow linguistic structure [37]. For example, fine structure cues play an important role in speech processing, yet normal-hearing listeners are able to detect these cues from envelope information alone [38]. Even when spectral qualities are severely degraded, speech processing can be achieved with primarily envelope information [39], as the envelope provides helpful cues for parsing meaningful segments in speech [40,41]. Additionally, the temporal characteristics of an auditory object allow us to focus attention on the source and segregate it from competing sources [42], which makes detection of envelope cues essential in speech communication. Because the amplitude envelope captures suprasegmental features across the speech signal, it lends itself to being an excellent proxy for prosodic information, and we argue that studies that use the speech envelope are inherently targeting a response to prosody.

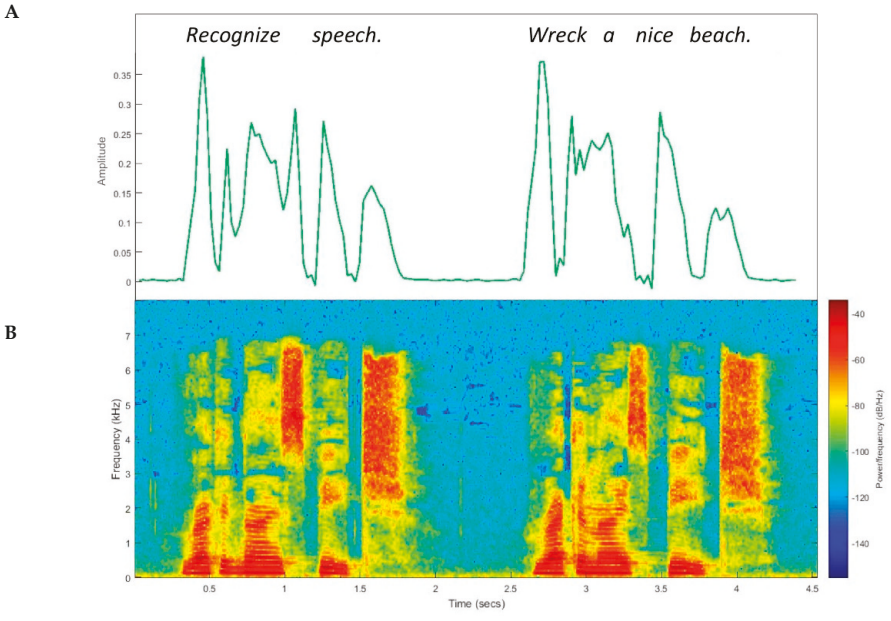


Figure 1. Two representations of an acoustic speech signal: Amplitude envelope (**A**) and spectrogram (**B**). Subtle differences between the phrases "recognize speech" and "wreck a nice beach" can be detected in both representations.

The speech amplitude envelope provides a linear representation of AM fluctuations over time. Acoustic stimuli are constructed of multiple temporal dimensions [31], and modulation energy varies based upon the selected band of carrier frequencies in the signal [35]. Speech can be portrayed through a hierarchical series of AM frequency scales [43]; that is, stress placement occurs at a rate of ~2 Hz [44], syllable rate occurs around 3–5 Hz [24], and phonemic structure has a faster rate of 8–50 Hz [31] (see Figure 2). Liss et al. [45] found that energy in the frequency bands below 4 Hz was intercorrelated, and energy above 4 Hz (up to 10 Hz) was separately intercorrelated. The frequency range between 4 and 16 Hz primarily affects speech intelligibility [46], while frequencies below 4 Hz strictly reflect prosodic variations, such as stress and syllable rate [47]. These multiple timescales of modulation energy within the speech envelope have been shown to elicit corresponding modulations in cortical activity during speech processing [48]. This correspondence appears to play a role in potentially challenging listening situations, such as: speech in noise [49], multiple speakers [50], complex auditory scenes [51], conflicting visual information [52], and divided attention [53,54]. Each of these situations (discussed in more detail later) requires the listener to exploit the natural timing of speech using prosodic cues, which are provided in the amplitude envelope [55].

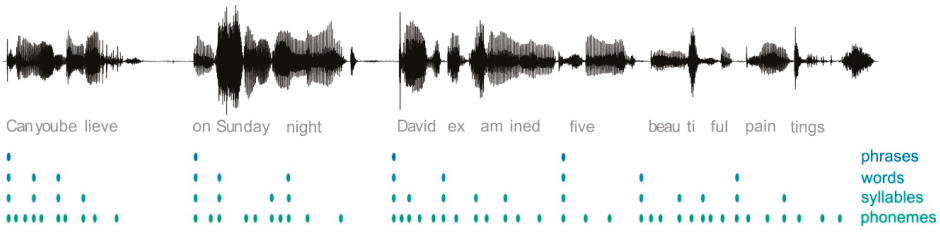


Figure 2. Acoustic waveform with its segmentation into phrases, words, syllables, and phonemes. Figure reproduced from [48].

3. Neural Entrainment to the Speech Envelope

Neural entrainment to the speech envelope has been a notoriously complex topic of study for several decades. In this section we will provide a broad overview of some investigative strides in this area. It is well known that neural oscillatory activity occurs in a constant stream of peaks and troughs while at rest and during cognitive processes. This stream becomes an adaptive spike train in response to environmental stimuli, such as the acoustic signal of speech. Numerous studies have shown that neural oscillatory activity in specific frequency bands is related to specific linguistic functions; for example, lower-level linguistic processing, such as detection of stress and syllable segmentation, occurs in lower frequencies (<4 Hz) [47], and semantic/syntactic processing may occur in higher frequencies (13–50 Hz) [56].

Traditional EEG approaches to prosody perception include analyzing event-related potential (ERP) activity at key events in the speech signal [57], such as stressed syllables [58], metric structure violations [59], pitch violations [60], and duration violations [11]. While these techniques are important for determining brain responses to prosodic features, they do not provide a comprehensive measure of how the brain tracks and encodes the multidimensional aspects of prosody over time. Because prosody refers to suprasegmental features (duration, amplitude, fundamental frequency), which vary throughout an utterance, it is useful to analyze prosody across the temporal domain rather than at one point in time. For this we turn to the speech amplitude envelope as a representation of suprasegmental information.

Recent developments in the literature have explored ways to measure continuous neural entrainment, which is a phenomenon where neuronal activity synchronizes to the periodic qualities of the incoming stimuli [61]. The oscillations of the auditory cortex reset their phase to the rhythm of the speech signal, which is an essential process for speech comprehension [33]. This is known as phase-locking, which can be measured with a cross-correlation procedure between the speech stimulus and the resultant M/EEG signal [62]. Cross-correlation uncovers similarities between two time series using a range of lag windows [63]. This is an efficient method for observing the response to continuous speech without requiring a large number of stimulus repetitions, since this analysis inherently increases the signal-to-noise ratio [64].

Speech processing occurs through a large network of cortical sources [65], and phase-locking can be measured to locate functionally independent sources [66]. These sources may occur bilaterally depending on the timescale [67], such that the left hemisphere favors rapid temporal features of speech, and the right hemisphere tracks slower features. The right hemisphere generally shows stronger tracking of the speech envelope [68,69]; however, envelope tracking has also been shown to be a bilateral process [62,70].

When measuring how the speech envelope is represented in neural data, one issue with a simple cross-correlation between envelope and neural response is that temporal smearing (from averaging across time points) will create noise in the correlation function [71]. A solution to this is to use a modeling approach, known as a temporal response function (TRF) [68], to describe the linear mapping between stimulus and response. This approach stems from a system identification technique [72] that

models the human brain as a linear time-invariant system. Of course, the brain does not operate on a linear or time-invariant schedule, but these assumptions are commonly accepted in neurophysiology research for characterizing the system by its impulse response [73,74].

The modeling approach can operate in either the forward or backward direction. Forward modeling describes the mapping of a speech stimulus to a neural network [68,75,76] using a TRF that represents the linear transformation that generated the observed neural signal [77] (Figure 3). When using the envelope representation of speech, the forward model treats the stimulus as a univariate input affecting each recording channel separately. However, since the speech signal is transformed in the auditory pathway into multiple frequency bands [78], the forward modeling procedure may benefit from a multivariate temporal response function (mTRF) [71], which uses the spectrogram representation to evaluate speech encoding. Even in the multivariate domain, forward modeling still maps the stimulus to each response channel independently [79].

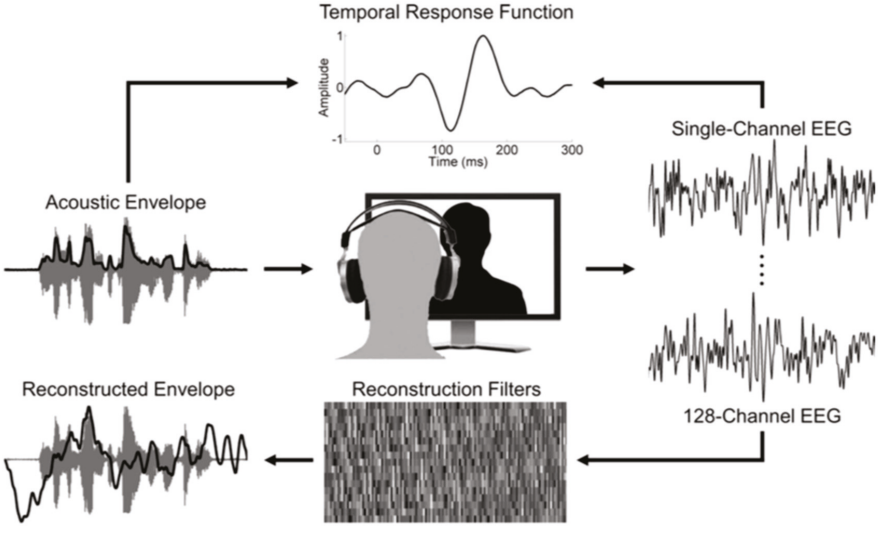


Figure 3. The temporal response function (TRF)—calculated with a linear least squares approach—represents the mapping from acoustic envelope onto each channel of EEG data (forward modeling). A multivariate reconstruction filter can be applied to data from all channels to estimate the acoustic envelope (backward modeling). Reconstruction accuracy can be measured by Pearson correlation between original and reconstructed envelopes. Figure reproduced from [75].

Backward modeling is a mathematical representation of the linear mapping from the multivariate neural response back to the stimulus [71]. This modeling approach yields a decoder that attempts to reconstruct a univariate stimulus feature, such as the speech envelope. As described in [71], this decoder function is derived by minimizing the mean squared error between the stimulus and reconstruction. In the backward direction, recording channels are weighted based on the information that they provide for the reconstruction [77], which removes inter-channel redundancies—an advantage over forward modeling. By modeling in the backward direction, researchers are able to compare stimulus reconstructions to the original stimulus, for instance with a correlation coefficient as a marker of reconstruction accuracy [80]. This provides a reliable index for the degree to which the envelope is encoded in the neural network. While other methods—such as cross-correlations and inter-trial phase coherence—are adequate for measuring phase-locking in speech comprehension, the modeling approach has been gaining attention as an attractive analysis method in recent years. Regardless of

the method used, measuring neural entrainment to the speech envelope is an excellent way to target prosodic processing, yet this has been underutilized in the literature.

4. Selected Findings in Envelope Entrainment

Many questions about speech processing can be investigated by looking at neural entrainment to the speech envelope, though we must be careful about how we interpret the results (see [34]; Table 1 provides a summary of selected studies). Peelle et al. [32] compared intelligible speech with unintelligible noise-vocoded speech, and they found that cortical oscillations in the theta (4–7 Hz) band are more closely phase-locked to intelligible speech. This may suggest that linguistic information and contextual associations enhance phase-locking to the envelope. However, others have measured envelope tracking in the auditory cortex even when the signal is devoid of communicative value. For example, Nourski et al. [81] found envelope entrainment even when speech rate was compressed to an unintelligible degree; Howard and Poeppel [82] found envelope entrainment to time-reversed speech stimuli; Mai et al. [56] found envelope entrainment to pseudo-word utterances. We acknowledge that envelope entrainment is often enhanced by intelligibility [49], but given the conflicting results described here, it is difficult to say whether intelligibility predicts entrainment or vice versa. What we can take away from these findings is that acoustic features of the stimulus—such as suprasegmental cues—seem to contribute to the neural entrainment effect and that the effect of neural entrainment on speech intelligibility warrants further investigation.

Attention has also been shown to influence envelope entrainment, and selective attention in a multi-speaker environment can be observed by the degree to which neural oscillations entrain to a given speech envelope [54,61,83]. The classic cocktail party situation has been studied for decades [84] and continues to be of interest today, e.g., [85]. In a natural auditory environment, many sounds are merged together and presented to the ear simultaneously, and the listener is tasked with segregating the sounds and attending to a particular source while ignoring the others [86]. By analyzing speech envelope representations, we can determine how the neural circuit parses and segregates these auditory objects. Ding and Simon [54] demonstrated that when a listener hears two speakers simultaneously, the neural decoding process is able to reconstruct the stimulus envelopes of both speech streams. The stimulus reconstruction is more strongly correlated to the envelope of the attended speaker (also [61,62,83]). Similar results have also been shown with invasive electrodes in electrocorticography (ECoG) research [53,87]. Despite the methodology used, these studies have suggested that neural encoding of an auditory scene involves selective phase-locking to specific auditory objects that are presented concurrently in a single auditory mixture. As mentioned previously, prosodic features of a speech stream help a listener to parse speech and attend to it, so prosody likely plays an important role in multi-speaker envelope entrainment, yet manipulations of the prosodic features of speech are rarely included as a variable in multi-speaker entrainment studies.

Speech envelopes are also of interest in studies examining audiovisual presentation of speech. Visual speech provides critical information regarding the timing and content of the acoustic signal [88]. It has long been acknowledged that listeners perceive speech better when they can both see and hear the individual speaking [89]. Articulatory and facial movements provide visual temporal cues that complement meaningful markers in the auditory stream. Visual rhythmic movements help parse syllabic boundaries [90], a wider mouth opening indicates louder amplitude [91], and seeing a conversational partner assists in segregating a speech stream from overlapping speakers [92]. Visual cues and gestures are tightly linked to speech prosody [93–95], and this alignment emphasizes suprasegmental features of the speech signal.

When auditory and visual information are incongruous, speech perception may be hindered and even lead the listener to falsely perceive a sound that was not presented in either modality (à la "The McGurk Effect") [96]. Congruent audiovisual speech enhances envelope tracking compared to incongruent information and also shows greater envelope encoding than auditory only speech, visual only speech, or the combination of the two unisensory modalities [97]. Audiovisual speech

also has marked benefits for neural tracking when presented in noisy conditions [98] (also [99,100]). This is indicative of multisensory enhancement during speech envelope encoding. At the same time, there appears to be a similar mechanism for visual entrainment in which cortical oscillations entrain to salient lip movements even when they are incongruous to the acoustic stream [101]. These studies of envelope responses to speech incongruence support an emerging model of correlated auditory and visual signals dynamically interacting in a discrete process of multisensory integration [88,102]. Prosody is a major factor in this integration, as it aligns a stable framework of temporal and acoustic–phonetic cues to be used in speech processing; however, the contribution of prosodic dimensions of the speech stimuli to neural entrainment in multisensory processing in these studies has not been explicitly considered.

Prosody shares a number of features with music, so an area for potential exploration is the connection between neural entrainment to speech and to music. Envelope entrainment is influenced by speech rhythm [103]. Because rhythm and temporal cues provide a common link between music and speech perception (e.g., [104,105]), several studies demonstrate associations between musical rhythm aptitude, speech perception, and literacy skills in children [106,107]. Some have hypothesized that entrainment to music leads to increased timing precision in the auditory system, which leads to increased perception of the timing of speech sounds [108,109]. Doelling and Poeppel [110] found that the accuracy of cortical entrainment to musical stimuli is contingent upon musical expertise, suggesting individual differences in cortical oscillations related to experience. However, musical expertise does not necessarily predict stronger entrainment to the speech envelope [111]. Additional work on individual differences between speech and music may help to target the neural mechanisms behind prosodic processing.

Table 1. List of papers investigating speech envelope tracking using various analysis approaches, data collection procedures, and topics of interest. Analysis abbreviations: CC—cross-correlation; PC—phase coherence; TRF—temporal response function; SR—stimulus reconstruction.

Author/Year	Data	Analysis	Relevant Amplitude Envelope Findings
Speech Intelligibility			
Ahissar et al., 2001 [63]	MEG	CC	Phase-locking predicts speech comprehension
Luo and Poeppel, 2007 [112]	MEG	PC	Phase-locking to speech is robust at 4–8 Hz
Abrams et al., 2008 [69]	EEG	CC	Right-hemisphere dominance for phase-locking
Hertrich et al., 2012 [64]	MEG	CC	Phase-locking with right-lateralized peak at 100 ms
Ding and Simon, 2013 [49]	MEG	TRF	Phase-locking at <4 Hz remains stable in noise
Peelle et al., 2013 [32]	MEG	PC	Phase-locking is strongest at 4–7 Hz in intelligible speech
Ding et al., 2014 [113]	MEG	TRF/PC	Phase-locking at 1–4 Hz predicts speech comprehension
Millman et al., 2015 [114]	MEG	CC	Phase-locking at 4–7 Hz regardless of intelligibility
Power et al., 2016 [115]	EEG	SR	Reconstruction of vocoded speech is strongest at 0–2 Hz
Cocktail Party			
Power et al., 2012 [61]	EEG	TRF	Attention elicits left-lateralized peak at 209 ms
Ding and Simon, 2012 [54]	MEG	SR	Attended speech phase-locks at <10 Hz around 100 ms lag
Zion Golumbic et al., 2013 [87]	ECoG	PC	Attended speech phase-locks at 1–7 Hz and 70–150 Hz
Horton et al., 2014 [50]	EEG	CC	Attended phase-locking improves with sample length
O'Sullivan et al., 2015 [83]	EEG	SR	Attended speech encodes maximally at 170–250 ms lag
O'Sullivan et al., 2017 [116]	ECoG	SR	Attention boosts reconstruction accuracy in dynamic switching
Audiovisual Speech			
Crosse et al., 2015 [97]	EEG	SR	AV speech encodes better than A + V at 2–6 Hz
Crosse et al., 2016 [98]	EEG	SR	AV speech improves reconstruction in noise at <3 Hz
Park et al., 2016 [101]	MEG	PC	Cortical activity entrains to lip movements at 1–7 Hz
Linguistic Information			
Di Liberto et al., 2015 [117]	EEG	SR	Cortical activity entrains to phonetic information
Ding et al., 2017 [118]	EEG	PC	Cortical activity entrains to multiple levels concurrently
Falk et al., 2017 [119]	EEG	PC	Phase-locking improves when rhythmic cue precedes speech
Broderick et al., 2018 [120]	EEG	TRF	Neural tracking depends on semantic congruency
Makov et al., 2017 [121]	EEG	PC	Phase-locking at 4 Hz during sleep, but not at higher levels

In summary, neural entrainment to the speech envelope likely reflects, at least in part, prosody perception. Prosodic fluctuations and prosody perception likely contribute to experimental findings linking envelope entrainment to intelligibility, selective attention, and audiovisual integration. Findings discussed in this section are highlighted in Table 1.

5. Developmental and Clinical Relevance of Envelope Entrainment

Children show a reliance on prosody processing from early infancy [122,123]; so, the envelope appears to be a critical tool for early language acquisition. The speech amplitude envelope contributes to the perception of linguistic stress, providing essential information for speech intelligibility and comprehension, e.g., [124]. Infant-directed speech is a manner of speaking that exaggerates prosodic cues, and infants show stronger cortical tracking of the infant-directed speech envelope compared to tracking of adult-directed speech [125]. Individuals who have difficulties with processing cues related to the speech amplitude envelope may demonstrate language-processing deficits [126].

Neuronal oscillatory activity in healthy adults entrains to adult-directed speech at various timescales, e.g., [33]. Frequencies in the delta band range (1–4 Hz) involve slower oscillations and track suprasegmental features of speech, such as phrase patterns, intonation, and stress [33,61]. Prosodic cues are particularly salient in the delta band and may be of particular relevance for envelope entrainment and language acquisition in children. Child-directed speech appears to bolster entrainment at the delta band specifically by amplifying these prosodic features [127]. The accuracy of delta band entrainment may also be indicative of higher-level linguistic abilities, as entrainment at the 0–2 Hz band is positively correlated with literacy [115,128]. The delta band may be crucially important because it provides the foundation for hierarchical linguistic structures of the incoming speech signal [33]. This could, in turn, affect cross-frequency neural synchronization, which may be particularly informative for the development of speech comprehension [129].

Autism spectrum disorders (ASD) are associated with atypical processing of various sensory modalities [130]. Individuals with ASD show less efficient neural integration of audio and visual information in non-speech [131] and speech input [132]. This is related to the temporal binding hypothesis in ASD, which suggests that these individuals have a deficit in synchronization across neural networks [133]. Jochaut et al. [134] showed deficient speech envelope tracking using fMRI and EEG when individuals with ASD perceive congruent audiovisual information. Possible impairment in coupling rhythms into oscillatory hierarchies could contribute to these results [135], and examining language deficits in ASD as oscillopathic traits may be a promising step forward in understanding these disorders [136,137].

Developmental dyslexia is a disorder of reading and spelling difficulties not associated with cognitive deficits or overt neurological conditions, and it is often considered a disorder of phonological processing skills [138]. Dyslexia is believed to affect the temporal coding in the auditory and visual modalities [139,140], and individuals with dyslexia often have difficulty identifying syllable structure or rhyme schemes, see [141]. The speech envelope is important to study in dyslexia because it carries syllable pattern information, and Abrams et al. [142] reported delayed phase-locking to the envelope in individuals with dyslexia. Specifically, the delta band in neuronal oscillations can reveal anomalies such as atypical phase of entrainment [43,143] and poor envelope reconstructions [115], which may ultimately have a downstream effect on establishing phonological representations [25]. Because the delta band reflects prosodic fluctuations, the atypical entrainment in this range suggests that individuals with dyslexia may have impaired encoding at the prosodic linguistic level [61].

Developmental language disorder (DLD) affects language abilities while leaving other cognitive skills intact, and it is sometimes studied in parallel with dyslexia due to similar deficits in phonological and auditory processing [144,145]. The prosodic phrasing hypothesis [146] suggests that children with DLD have difficulty detecting rhythmic patterns in speech, particularly related to impaired sensitivity to amplitude rise time [147] and sound duration [126], and difficulties in processing accelerated speech rate [148]. Given the growing behavioral evidence suggesting that children with DLD have deficits

in prosody perception (see [149]), it stands to reason that they would show poor speech envelope entrainment, particularly in the delta frequency band [33]. To our knowledge, there has not been an electrophysiological study looking at neural entrainment to the speech envelope in children with DLD, but this would be an illuminating endeavor.

6. Directions for Future Research

There have been many recent advances related to speech envelope entrainment, and we argue that prosody has had a substantial—though at times underrated—role in many studies. It is well accepted that prosodic cues facilitate speech processing, e.g., [3], and these suprasegmental features are represented in the amplitude envelope, e.g., [64]. Therefore, studies investigating speech envelope entrainment inherently capture a response to prosody to some degree, yet the underlying mechanisms of prosody perception, and their effect on speech processing, remain somewhat a mystery. We suggest that including experimental manipulations of the prosodic dimensions of speech in future studies may inform the findings of previous works, and it may shed light on the future interpretation of entrainment, particularly in the low-frequency range. Ding et al. [118] have shown that removing prosodic cues from speech weakens envelope entrainment, which suggests that synchrony between neural oscillations and the speech envelope reflects perception of the acoustic manifestations of prosody, and future work should continue testing this relationship. More broadly, we present a series of potential future directions in Table 2.

Table 2. Potential future directions including key points and methodological considerations.

Future Directions	
Key Point	**Potential Directions and Methodological Considerations**
Prosodic characteristics of stimuli should be controlled and well-described	• Is it feasible to equalize the prosodic dimension across experimental conditions that are not meant to isolate prosody? At the least, authors could describe the metrical structure of speech stimuli in studies that examine entrainment to envelope features. • What is the variability of neural entrainment to stimuli that differ in prosodic structure?
Role of repetition in establishing neural entrainment to prosodic cues	• What are the implications of hearing the same sentence/stimulus repeated many times versus hearing novel speech? Repetition affects semantic and syntactic expectancies, as well as expectations for the unfolding envelope of the signal. • What is the relationship between predictive neural processes, entrainment to the envelope, and intelligibility? How do these concepts relate to prosody?
Low-frequency envelope fluctuations correlate with the syntactic structure of speech and are relevant to language development	• Does entraining to envelope phrase boundary markers (such as pauses and phrase-final lengthening that correlate with important syntactical information) explain variance in syntactic processing? • How does detection of these cues evolve over the course of childhood language development? • Is neural entrainment to the envelope a potential signature of development of sensitivity to these cues?
Individuals vary in their sensitivity to prosody	• Does neural entrainment to the envelope reflect how individuals differ in their prosodic sensitivity (when measured as a separate behavioral trait)? • Can environmental and genetic factors such as musical training and music aptitude affect individual differences in neural entrainment to speech? • Do some individuals with developmental disabilities have impaired neural entrainment to the envelope? How does this differ among different neurodevelopmental disorders? Is there a causal impact of this impairment on their speech/language/reading development? • Can speech/language/reading therapy enhance sensitivity to prosody via increased neural entrainment to the envelope (as a mediating mechanism to improving speech/language/reading outcomes)?

Synchronization occurs when internal oscillators adjust their phase and period to rate changes of speech rhythm, e.g., [150]. According to the dynamic attending theory [151,152], attentional effort is not uniformly distributed over time, but rather, it occurs periodically with salient sensory input. Prosody offers meaningful information through stressed syllables, which gives attentional rhythms a structure for scaffolding speech processing mechanisms [108]. Suprasegmental elements are present in a wide array of stimuli that demonstrate neural entrainment to the speech envelope (e.g., intelligible and unintelligible speech; attended and unattended speech; audiovisual and audio only and visual only speech). The suprasegmental cues may be one reason why stimuli of varying salience continue to reveal entrainment. Future empirical investigations may consider how prosody supports neural entrainment under these different experimental conditions.

Of course, prosodic fluctuations alone cannot fully explain neural entrainment to speech or speech comprehension [34]. When Ding and colleagues [118,153] removed prosodic cues from connected speech stimuli, they did find some low-frequency entrainment (<10 Hz), which they attribute to syntactic processing. However, they pointed out that neural tracking would likely be more prominent in natural speech with the addition of rich prosodic information. In spoken English, syntax can exist without prosody, but the inclusion of prosody certainly facilitates syntactic processing (with phrase segmentation, pitch inflection, etc.). Therefore, further study of prosodic versus syntactical manipulations will shed light on their respective contributions—and their interaction—to neural entrainment to speech, including when examined together with behavioral measures of speech comprehension. Studies have shown that phonetic [117] and semantic [120] levels of processing also contribute to neural activity at different hierarchical timescales. It may be informative to consider how prosodic cues organize and facilitate processing at these different levels. Future work should attempt to isolate prosodic cues from phonetic and semantic details to specify the contributions of prosody to these other structures in continuous speech. This could be accomplished by restricting prosodic cues (using monotone pitch and constant word durations, as in [118]) or by creating stimuli with a prosodic mismatch (using unpredictable changes in amplitude, pitch, and duration). These manipulations would allow researchers to more directly target the role of prosody in entrainment.

Examining the links between prosody and neural encoding of the speech envelope may also have relevance for additional topics and clinical populations. For example, it has been shown that features of prosody are directly linked to emotional expressiveness in speech [6], and one novel area of research would be to connect patterns of envelope entrainment with perception of emotional states. This would likely have implications for the clinical populations discussed above, as well as typical emotional development. Other recent work has investigated rhythmic cueing and temporal dynamics of speech in patients with Parkinson's disease [154], aphasia [155], and even blindness [156]. Because these populations show difficulty with prosodic cues in speech, a next step could be to examine speech envelope entrainment in these individuals to examine if there is a neural deficit in prosody encoding.

As conveyed in this review, low-frequency neural oscillations likely reflect in part a response to prosodic cues in speech. Future research can investigate how prosody impacts neural envelope entrainment and scaffolds higher-level speech processing, as well as examine individual differences in prosody perception and neural entrainment. Future research in speech entrainment ought to search for connections to prosody perception and determine what it takes to get the speech envelope signed, sealed, and delivered to the cortex.

Funding: The authors were supported by the National Institute on Deafness and Other Communication Disorders and the Office of Behavioral and Social Sciences Research of the National Institutes of Health under Award Numbers R03DC014802, 1R21DC016710-01, and K18DC017383. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was additionally supported by the Program for Music, Mind and Society at Vanderbilt (with funding from the Trans-Institutional Programs Initiative), the VUMC Faculty Research Scholars Program, and the Department of Otolaryngology at Vanderbilt University Medical Center.

Acknowledgments: The authors would like to thank Duane Watson, Cyrille Magne, Stephen Camarata, and anonymous reviewers for invaluable theoretical and conceptual insight to this review, as well as Edmund Lalor, Giovanni Di Liberto, Fleur Bouwer, and Andrew Lotto for thoughtful methodological considerations herein.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kunert, R.; Jongman, S.R. Entrainment to an auditory signal: Is attention involved? *J. Exp. Psychol. Gen.* **2017**, *146*, 77–88. [CrossRef] [PubMed]
2. Dahan, D.; Tanenhaus, M.K.; Chambers, C.G. Accent and reference resolution in spoken-language comprehension. *J. Mem. Lang.* **2002**, *47*, 292–314. [CrossRef]
3. Pitt, M.A.; Samuel, A.G. The use of rhythm in attending to speech. *J. Exp. Psychol. Hum. Percept. Perform.* **1990**, *16*, 564–573. [CrossRef] [PubMed]
4. Scherer, K.R. Vocal affect expression: A review and a model for future research. *Psychol. Bull.* **1986**, *99*, 143–165. [CrossRef] [PubMed]
5. Zentner, M.; Grandjean, D.; Scherer, K.R. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* **2008**, *8*, 494–521. [CrossRef] [PubMed]
6. Coutinho, E.; Dibben, N. Psychoacoustic cues to emotion in speech prosody and music. *Cogn. Emot.* **2013**, *27*, 658–684. [CrossRef] [PubMed]
7. Scherer, K.R.; Banse, R.; Wallbott, H.G. Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *J. Cross-Cult. Psychol.* **2001**, *32*, 76–92. [CrossRef]
8. Thompson, W.F.; Balkwill, L.-L. Decoding speech prosody in five languages. *Semiotica* **2006**, *158*, 407–424. [CrossRef]
9. Watson, D.; Gibson, E. Intonational phrasing and constituency in language production and comprehension*. *Stud. Linguist.* **2005**, *59*, 279–300. [CrossRef]
10. Liu, F.; Jiang, C.; Wang, B.; Xu, Y.; Patel, A.D. A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia* **2015**, *66*, 111–118. [CrossRef]
11. Magne, C.; Astesano, C.; Aramaki, M.; Ystad, S.; Kronland-Martinet, R.; Besson, M. Influence of Syllabic Lengthening on Semantic Processing in Spoken French: Behavioral and Electrophysiological Evidence. *Cereb. Cortex* **2007**, *17*, 2659–2668. [CrossRef] [PubMed]
12. Gervain, J.; Werker, J.F. Prosody cues word order in 7-month-old bilingual infants. *Nat. Commun.* **2013**, *4*, 1490. [CrossRef]
13. Nazzi, T.; Ramus, F. Perception and acquisition of linguistic rhythm by infants. *Speech Commun.* **2003**, *41*, 233–243. [CrossRef]
14. Soderstrom, M. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *J. Mem. Lang.* **2003**, *49*, 249–267. [CrossRef]
15. Shintel, H.; Anderson, N.L.; Fenn, K.M. Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *J. Exp. Psychol. Gen.* **2014**, *143*, 1437–1442. [CrossRef] [PubMed]
16. Tzeng, C.Y.; Duan, J.; Namy, L.L.; Nygaard, L.C. Prosody in speech as a source of referential information. *Lang. Cogn. Neurosci.* **2018**, *33*, 512–526. [CrossRef]
17. Gordon, R.L.; Shivers, C.M.; Wieland, E.A.; Kotz, S.A.; Yoder, P.J.; Devin McAuley, J. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* **2015**, *18*, 635–644. [CrossRef] [PubMed]
18. Holt, C.M.; Yuen, I.; Demuth, K. Discourse Strategies and the Production of Prosody by Prelingually Deaf Adolescent Cochlear Implant Users. *Ear Hear.* **2017**, *38*, e101–e108. [CrossRef]
19. Goswami, U.; Gerson, D.; Astruc, L. Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Read. Writ.* **2010**, *23*, 995–1019. [CrossRef]
20. Grossman, R.B.; Bemis, R.H.; Plesa Skwerer, D.; Tager-Flusberg, H. Lexical and Affective Prosody in Children With High-Functioning Autism. *J. Speech Lang. Hear. Res.* **2010**, *53*, 778. [CrossRef]
21. Fletcher, J. The Prosody of Speech: Timing and Rhythm. In *The Handbook of Phonetic Sciences*; Hardcastle, W.J., Laver, J., Gibbon, F.E., Eds.; Blackwell Publishing Ltd.: Oxford, UK, 2010; pp. 521–602. ISBN 978-1-4443-1725-1.
22. Lehiste, I. *Suprasegmentals*; M.I.T. Press: Cambridge, MA, USA, 1970; ISBN 978-0-262-12023-4.

23. Fernald, A.; Simon, T. Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* **1984**, *20*, 104–113. [CrossRef]
24. Greenberg, S.; Carvey, H.; Hitchcock, L.; Chang, S. Temporal properties of spontaneous speech—A syllable-centric perspective. *J. Phon.* **2003**, *31*, 465–485. [CrossRef]
25. Leong, V.; Goswami, U. Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLoS ONE* **2015**, *10*, e0144411. [CrossRef]
26. Jusczyk, P.W.; Hirsh-Pasek, K.; Kemler Nelson, D.G.; Kennedy, L.J.; Woodward, A.; Piwoz, J. Perception of acoustic correlates of major phrasal units by young infants. *Cognit. Psychol.* **1992**, *24*, 252–293. [CrossRef]
27. Cooper, R.P.; Abraham, J.; Berman, S.; Staska, M. The development of infants' preference for motherese. *Infant Behav. Dev.* **1997**, *20*, 477–488. [CrossRef]
28. Houston, D.M.; Jusczyk, P.W.; Kuijpers, C.; Coolen, R.; Cutler, A. Cross-language word segmentation by 9-month-olds. *Psychon. Bull. Rev.* **2000**, *7*, 504–509. [CrossRef]
29. de Carvalho, A.; Dautriche, I.; Lin, I.; Christophe, A. Phrasal prosody constrains syntactic analysis in toddlers. *Cognition* **2017**, *163*, 67–79. [CrossRef]
30. Sharpe, V.; Fogerty, D.; den Ouden, D.-B. The Role of Fundamental Frequency and Temporal Envelope in Processing Sentences with Temporary Syntactic Ambiguities. *Lang. Speech* **2017**, *60*, 399–426. [CrossRef]
31. Rosen, S. Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1992**, *336*, 367–373.
32. Peelle, J.E.; Gross, J.; Davis, M.H. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cereb. Cortex* **2013**, *23*, 1378–1387. [CrossRef]
33. Giraud, A.-L.; Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **2012**, *15*, 511–517. [CrossRef]
34. Obleser, J.; Herrmann, B.; Henry, M.J. Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front. Hum. Neurosci.* **2012**, *6*, 250. [CrossRef]
35. Joris, P.X.; Schreiner, C.E.; Rees, A. Neural Processing of Amplitude-Modulated Sounds. *Physiol. Rev.* **2004**, *84*, 541–577. [CrossRef]
36. Fritz, J.; Shamma, S.; Elhilali, M.; Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **2003**, *6*, 1216–1223. [CrossRef]
37. Frazier, L.; Carlson, K.; Cliftonjr, C. Prosodic phrasing is central to language comprehension. *Trends Cogn. Sci.* **2006**, *10*, 244–249. [CrossRef]
38. Paraouty, N.; Ewert, S.D.; Wallaert, N.; Lorenzi, C. Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss. *J. Acoust. Soc. Am.* **2016**, *140*, 121–131. [CrossRef]
39. Shannon, R.V.; Zeng, F.G.; Kamath, V.; Wygonski, J.; Ekelid, M. Speech recognition with primarily temporal cues. *Science* **1995**, *270*, 303–304. [CrossRef]
40. Lehiste, I.; Olive, J.P.; Streeter, L.A. Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.* **1976**, *60*, 1199–1202. [CrossRef]
41. Adank, P.; Janse, E. Perceptual learning of time-compressed and natural fast speech. *J. Acoust. Soc. Am.* **2009**, *126*, 2649–2659. [CrossRef]
42. Aubanel, V.; Davis, C.; Kim, J. Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech. *Front. Hum. Neurosci.* **2016**, *10*, 430. [CrossRef]
43. Leong, V.; Goswami, U. Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hear. Res.* **2014**, *308*, 141–161. [CrossRef]
44. Dauer, R.M. Stress-timing and syllable-timing reanalyzed. *J. Phon.* **1983**, *11*, 51–62.
45. Liss, J.M.; LeGendre, S.; Lotto, A.J. Discriminating Dysarthria Type From Envelope Modulation Spectra. *J. Speech Lang. Hear. Res.* **2010**, *53*, 1246. [CrossRef]
46. Drullman, R.; Festen, J.M.; Plomp, R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* **1994**, *95*, 1053–1064. [CrossRef] [PubMed]
47. Ding, N.; Simon, J.Z. Cortical entrainment to continuous speech: Functional roles and interpretations. *Front. Hum. Neurosci.* **2014**, *8*, 311. [CrossRef]
48. Keitel, A.; Gross, J.; Kayser, C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biol.* **2018**, *16*, e2004473. [CrossRef] [PubMed]
49. Ding, N.; Simon, J.Z. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *J. Neurosci.* **2013**, *33*, 5728–5735. [CrossRef] [PubMed]

50. Horton, C.; Srinivasan, R.; D'Zmura, M. Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party'. *J. Neural Eng.* **2014**, *11*, 046015. [CrossRef] [PubMed]
51. Shamma, S.A.; Elhilali, M.; Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **2011**, *34*, 114–123. [CrossRef]
52. Arnal, L.H.; Morillon, B.; Kell, C.A.; Giraud, A.-L. Dual Neural Routing of Visual Facilitation in Speech Processing. *J. Neurosci.* **2009**, *29*, 13445–13453. [CrossRef]
53. Mesgarani, N.; Chang, E.F. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **2012**, *485*, 233–236. [CrossRef] [PubMed]
54. Ding, N.; Simon, J.Z. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11854–11859. [CrossRef] [PubMed]
55. Leong, V.; Stone, M.A.; Turner, R.E.; Goswami, U. A role for amplitude modulation phase relationships in speech rhythm perception. *J. Acoust. Soc. Am.* **2014**, *136*, 366–381. [CrossRef]
56. Mai, G.; Minett, J.W.; Wang, W.S.-Y. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage* **2016**, *133*, 516–528. [CrossRef]
57. Picton, T.W.; Hillyard, S.A.; Krausz, H.I.; Galambos, R. Human auditory evoked potentials. I: Evaluation of components. *Electroencephalogr. Clin. Neurophysiol.* **1974**, *36*, 179–190. [CrossRef]
58. Schmidt-Kassow, M.; Kotz, S.A. Event-related Brain Potentials Suggest a Late Interaction of Meter and Syntax in the P600. *J. Cogn. Neurosci.* **2009**, *21*, 1693–1708. [CrossRef]
59. Marie, C.; Magne, C.; Besson, M. Musicians and the Metric Structure of Words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef]
60. Astésano, C.; Besson, M.; Alter, K. Brain potentials during semantic and prosodic processing in French. *Cogn. Brain Res.* **2004**, *18*, 172–184. [CrossRef]
61. Power, A.J.; Foxe, J.J.; Forde, E.-J.; Reilly, R.B.; Lalor, E.C. At what time is the cocktail party? A late locus of selective attention to natural speech: A late locus of attention to natural speech. *Eur. J. Neurosci.* **2012**, *35*, 1497–1503. [CrossRef] [PubMed]
62. Horton, C.; D'Zmura, M.; Srinivasan, R. Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* **2013**, *109*, 3082–3093. [CrossRef] [PubMed]
63. Ahissar, E.; Nagarajan, S.; Ahissar, M.; Protopapas, A.; Mahncke, H.; Merzenich, M.M. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13367–13372. [CrossRef]
64. Hertrich, I.; Dietrich, S.; Trouvain, J.; Moos, A.; Ackermann, H. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal: MEG activity phase-locked to speech. *Psychophysiology* **2012**, *49*, 322–334. [CrossRef] [PubMed]
65. Hickok, G.; Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **2007**, *8*, 393–402. [CrossRef] [PubMed]
66. Jung, T.-P.; Makeig, S.; Westerfield, M.; Townsend, J.; Courchesne, E.; Sejnowski, T.J. Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* **2001**, *14*, 166–185. [CrossRef] [PubMed]
67. Poeppel, D. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **2003**, *41*, 245–255. [CrossRef]
68. Ding, N.; Simon, J.Z. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* **2012**, *107*, 78–89. [CrossRef] [PubMed]
69. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. *J. Neurosci.* **2008**, *28*, 3958–3965. [CrossRef] [PubMed]
70. Aiken, S.J.; Picton, T.W. Human Cortical Responses to the Speech Envelope. *Ear Hear.* **2008**, *29*, 139–157. [CrossRef] [PubMed]
71. Crosse, M.J.; Di Liberto, G.M.; Bednar, A.; Lalor, E.C. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* **2016**, *10*. [CrossRef] [PubMed]
72. Marmarelis, V.Z. *Nonlinear Dynamic Modeling of Physiological Systems*; IEEE Press Series in Biomedical Engineering; Wiley-Interscience: Hoboken, NJ, USA, 2004; ISBN 978-0-471-46960-5.
73. Boynton, G.M.; Demb, J.B.; Heeger, D.J. fMRI responses in human V1 correlate with perceived stimulus contrast. *NeuroImage* **1996**, *3*, S265. [CrossRef]

74. Ringach, D.; Shapley, R. Reverse correlation in neurophysiology. *Cogn. Sci.* **2004**, *28*, 147–166. [CrossRef]
75. Lalor, E.C.; Foxe, J.J. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.* **2010**, *31*, 189–193. [CrossRef] [PubMed]
76. Lalor, E.C.; Power, A.J.; Reilly, R.B.; Foxe, J.J. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *J. Neurophysiol.* **2009**, *102*, 349–359. [CrossRef] [PubMed]
77. Haufe, S.; Meinecke, F.; Görgen, K.; Dähne, S.; Haynes, J.-D.; Blankertz, B.; Bießmann, F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **2014**, *87*, 96–110. [CrossRef] [PubMed]
78. Yang, X.; Wang, K.; Shamma, S.A. Auditory representations of acoustic signals. *IEEE Trans. Inf. Theory* **1992**, *38*, 824–839. [CrossRef]
79. Theunissen, F.E.; David, S.V.; Singh, N.C.; Hsu, A.; Vinje, W.E.; Gallant, J.L. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Netw. Comput. Neural Syst.* **2001**, *12*, 289–316. [CrossRef]
80. Mesgarani, N.; David, S.V.; Fritz, J.B.; Shamma, S.A. Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. *J. Neurophysiol.* **2009**, *102*, 3329–3339. [CrossRef]
81. Nourski, K.V.; Reale, R.A.; Oya, H.; Kawasaki, H.; Kovach, C.K.; Chen, H.; Howard, M.A.; Brugge, J.F. Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. *J. Neurosci.* **2009**, *29*, 15564–15574. [CrossRef]
82. Howard, M.F.; Poeppel, D. Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *J. Neurophysiol.* **2010**, *104*, 2500–2511. [CrossRef]
83. O'Sullivan, J.A.; Power, A.J.; Mesgarani, N.; Rajaram, S.; Foxe, J.J.; Shinn-Cunningham, B.G.; Slaney, M.; Shamma, S.A.; Lalor, E.C. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* **2015**, *25*, 1697–1706. [CrossRef]
84. Cherry, E.C. Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.* **1953**, *25*, 975–979. [CrossRef]
85. Biesmans, W.; Das, N.; Francart, T.; Bertrand, A. Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 402–412. [CrossRef]
86. McDermott, J.H. The cocktail party problem. *Curr. Biol.* **2009**, *19*, R1024–R1027. [CrossRef]
87. Zion Golumbic, E.M.; Ding, N.; Bickel, S.; Lakatos, P.; Schevon, C.A.; McKhann, G.M.; Goodman, R.R.; Emerson, R.; Mehta, A.D.; Simon, J.Z.; et al. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron* **2013**, *77*, 980–991. [CrossRef]
88. Peelle, J.E.; Sommers, M.S. Prediction and constraint in audiovisual speech perception. *Cortex* **2015**, *68*, 169–181. [CrossRef] [PubMed]
89. Erber, N.P. Auditory-Visual Perception of Speech. *J. Speech Hear. Disord.* **1975**, *40*, 481. [CrossRef]
90. Peelle, J.E.; Davis, M.H. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* **2012**, *3*. [CrossRef]
91. Chandrasekaran, C.; Ghazanfar, A.A. Different Neural Frequency Bands Integrate Faces and Voices Differently in the Superior Temporal Sulcus. *J. Neurophysiol.* **2009**, *101*, 773–788. [CrossRef]
92. Carlyon, R.P.; Cusack, R.; Foxton, J.M.; Robertson, I.H. Effects of attention and unilateral neglect on auditory stream segregation. *J. Exp. Psychol. Hum. Percept. Perform.* **2001**, *27*, 115–127. [CrossRef]
93. Wagner, P.; Malisz, Z.; Kopp, S. Gesture and speech in interaction: An overview. *Speech Commun.* **2014**, *57*, 209–232. [CrossRef]
94. Loehr, D. Aspects of rhythm in gesture and speech. *Gesture* **2007**, *7*, 179–214. [CrossRef]
95. Krahmer, E.; Swerts, M. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* **2007**, *57*, 396–414. [CrossRef]
96. McGurk, H.; MacDonald, J. Hearing lips and seeing voices. *Nature* **1976**, *264*, 746–748. [CrossRef] [PubMed]
97. Crosse, M.J.; Butler, J.S.; Lalor, E.C. Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *J. Neurosci.* **2015**, *35*, 14195–14204. [CrossRef] [PubMed]
98. Crosse, M.J.; Di Liberto, G.M.; Lalor, E.C. Eye Can Hear Clearly Now: Inverse Effectiveness in Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal Integration. *J. Neurosci.* **2016**, *36*, 9888–9895. [CrossRef]

99. Sumby, W.H.; Pollack, I. Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am.* **1954**, *26*, 212–215. [CrossRef]
100. Ross, L.A.; Saint-Amour, D.; Leavitt, V.M.; Molholm, S.; Javitt, D.C.; Foxe, J.J. Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* **2007**, *97*, 173–183. [CrossRef]
101. Park, H.; Kayser, C.; Thut, G.; Gross, J. Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife* **2016**, *5*. [CrossRef] [PubMed]
102. Tye-Murray, N.; Sommers, M.; Spehar, B. Auditory and Visual Lexical Neighborhoods in Audiovisual Speech Perception. *Trends Amplif.* **2007**, *11*, 233–241. [CrossRef]
103. Kösem, A.; Bosker, H.R.; Takashima, A.; Meyer, A.; Jensen, O.; Hagoort, P. Neural Entrainment Determines the Words We Hear. *Curr. Biol.* **2018**, *28*, 2867–2875. [CrossRef]
104. Jäncke, L. The Relationship between Music and Language. *Front. Psychol.* **2012**, *3*. [CrossRef]
105. Hausen, M.; Torppa, R.; Salmela, V.R.; Vainio, M.; Särkämö, T. Music and speech prosody: A common rhythm. *Front. Psychol.* **2013**, *4*. [CrossRef]
106. Bonacina, S.; Krizman, J.; White-Schwoch, T.; Kraus, N. Clapping in time parallels literacy and calls upon overlapping neural mechanisms in early readers: Clapping in time parallels literacy. *Ann. N. Y. Acad. Sci.* **2018**, *1423*, 338–348. [CrossRef]
107. Ozernov-Palchik, O.; Wolf, M.; Patel, A.D. Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *J. Exp. Child Psychol.* **2018**, *167*, 354–368. [CrossRef]
108. Tierney, A.; Kraus, N. Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Front. Hum. Neurosci.* **2014**, *8*, 949. [CrossRef]
109. Woodruff Carr, K.; White-Schwoch, T.; Tierney, A.T.; Strait, D.L.; Kraus, N. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14559–14564. [CrossRef]
110. Doelling, K.B.; Poeppel, D. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6233–E6242. [CrossRef]
111. Harding, E.E.; Sammler, D.; Henry, M.J.; Large, E.W.; Kotz, S.A. Cortical tracking of rhythm in music and speech. *NeuroImage* **2019**, *185*, 96–101. [CrossRef]
112. Luo, H.; Poeppel, D. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron* **2007**, *54*, 1001–1010. [CrossRef] [PubMed]
113. Ding, N.; Chatterjee, M.; Simon, J.Z. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* **2014**, *88*, 41–46. [CrossRef]
114. Millman, R.E.; Johnson, S.R.; Prendergast, G. The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. *J. Cogn. Neurosci.* **2015**, *27*, 533–545. [CrossRef] [PubMed]
115. Power, A.J.; Colling, L.J.; Mead, N.; Barnes, L.; Goswami, U. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain Lang.* **2016**, *160*, 1–10. [CrossRef] [PubMed]
116. O'Sullivan, J.; Chen, Z.; Herrero, J.; McKhann, G.M.; Sheth, S.A.; Mehta, A.D.; Mesgarani, N. Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *J. Neural Eng.* **2017**, *14*, 056001. [CrossRef]
117. Di Liberto, G.M.; O'Sullivan, J.A.; Lalor, E.C. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Curr. Biol.* **2015**, *25*, 2457–2465. [CrossRef] [PubMed]
118. Ding, N.; Melloni, L.; Yang, A.; Wang, Y.; Zhang, W.; Poeppel, D. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Front. Hum. Neurosci.* **2017**, *11*, 481. [CrossRef] [PubMed]
119. Falk, S.; Lanzilotti, C.; Schön, D. Tuning Neural Phase Entrainment to Speech. *J. Cogn. Neurosci.* **2017**, *29*, 1378–1389. [CrossRef]
120. Broderick, M.P.; Anderson, A.J.; Di Liberto, G.M.; Crosse, M.J.; Lalor, E.C. Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Curr. Biol.* **2018**, *28*, 803–809. [CrossRef] [PubMed]
121. Makov, S.; Sharon, O.; Ding, N.; Ben-Shachar, M.; Nir, Y.; Zion Golumbic, E. Sleep Disrupts High-Level Speech Parsing Despite Significant Basic Auditory Processing. *J. Neurosci.* **2017**, *37*, 7772–7781. [CrossRef]
122. Curtin, S.; Mintz, T.H.; Christiansen, M.H. Stress changes the representational landscape: Evidence from word segmentation. *Cognition* **2005**, *96*, 233–262. [CrossRef]

123. Mehler, J.; Jusczyk, P.; Lambertz, G.; Halsted, N.; Bertoncini, J.; Amiel-Tison, C. A precursor of language acquisition in young infants. *Cognition* **1988**, *29*, 143–178. [CrossRef]
124. Ghitza, O. On the Role of Theta-Driven Syllabic Parsing in Decoding Speech: Intelligibility of Speech with a Manipulated Modulation Spectrum. *Front. Psychol.* **2012**, *3*, 238. [CrossRef] [PubMed]
125. Kalashnikova, M.; Peter, V.; Di Liberto, G.M.; Lalor, E.C.; Burnham, D. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* **2018**, *8*, 13745. [CrossRef] [PubMed]
126. Richards, S.; Goswami, U. Auditory Processing in Specific Language Impairment (SLI): Relations With the Perception of Lexical and Phrasal Stress. *J. Speech Lang. Hear. Res.* **2015**, *58*, 1292. [CrossRef] [PubMed]
127. Leong, V.; Kalashnikova, M.; Burnham, D.; Goswami, U. The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind* **2017**, *1*, 78–90. [CrossRef]
128. Molinaro, N.; Lizarazu, M.; Lallier, M.; Bourguignon, M.; Carreiras, M. Out-of-synchrony speech entrainment in developmental dyslexia: Altered Cortical Speech Tracking in Dyslexia. *Hum. Brain Mapp.* **2016**, *37*, 2767–2783. [CrossRef] [PubMed]
129. Leong, V.; Goswami, U. Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Front. Hum. Neurosci.* **2014**, *8*, 96. [CrossRef]
130. Marco, E.J.; Hinkley, L.B.N.; Hill, S.S.; Nagarajan, S.S. Sensory Processing in Autism: A Review of Neurophysiologic Findings. *Pediatr. Res.* **2011**, *69*, 48R–54R. [CrossRef]
131. Brandwein, A.B.; Foxe, J.J.; Butler, J.S.; Russo, N.N.; Altschuler, T.S.; Gomes, H.; Molholm, S. The Development of Multisensory Integration in High-Functioning Autism: High-Density Electrical Mapping and Psychophysical Measures Reveal Impairments in the Processing of Audiovisual Inputs. *Cereb. Cortex* **2013**, *23*, 1329–1341. [CrossRef] [PubMed]
132. Stevenson, R.A.; Siemann, J.K.; Schneider, B.C.; Eberly, H.E.; Woynaroski, T.G.; Camarata, S.M.; Wallace, M.T. Multisensory Temporal Integration in Autism Spectrum Disorders. *J. Neurosci.* **2014**, *34*, 691–697. [CrossRef]
133. Brock, J.; Brown, C.C.; Boucher, J.; Rippon, G. The temporal binding deficit hypothesis of autism. *Dev. Psychopathol.* **2002**, *14*, 209–224. [CrossRef]
134. Jochaut, D.; Lehongre, K.; Saitovitch, A.; Devauchelle, A.-D.; Olasagasti, I.; Chabane, N.; Zilbovicius, M.; Giraud, A.-L. Atypical coordination of cortical oscillations in response to speech in autism. *Front. Hum. Neurosci.* **2015**, *9*, 171. [CrossRef] [PubMed]
135. Kikuchi, M.; Yoshimura, Y.; Hiraishi, H.; Munesue, T.; Hashimoto, T.; Tsubokawa, T.; Takahashi, T.; Suzuki, M.; Higashida, H.; Minabe, Y. Reduced long-range functional connectivity in young children with autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* **2015**, *10*, 248–254. [CrossRef] [PubMed]
136. Benítez-Burraco, A.; Murphy, E. The Oscillopathic Nature of Language Deficits in Autism: From Genes to Language Evolution. *Front. Hum. Neurosci.* **2016**, *10*, 120. [CrossRef] [PubMed]
137. Simon, D.M.; Wallace, M.T. Dysfunction of sensory oscillations in Autism Spectrum Disorder. *Neurosci. Biobehav. Rev.* **2016**, *68*, 848–861. [CrossRef] [PubMed]
138. Stanovich, K.E. Refining the Phonological Core Deficit Model. *Child Psychol. Psychiatry Rev.* **1998**, *3*, 17–21. [CrossRef]
139. Lallier, M.; Thierry, G.; Tainturier, M.-J.; Donnadieu, S.; Peyrin, C.; Billard, C.; Valdois, S. Auditory and visual stream segregation in children and adults: An assessment of the amodality assumption of the 'sluggish attentional shifting' theory of dyslexia. *Brain Res.* **2009**, *1302*, 132–147. [CrossRef] [PubMed]
140. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
141. Ziegler, J.C.; Goswami, U. Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychol. Bull.* **2005**, *131*, 3–29. [CrossRef] [PubMed]
142. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Abnormal Cortical Processing of the Syllable Rate of Speech in Poor Readers. *J. Neurosci.* **2009**, *29*, 7686–7693. [CrossRef]
143. Power, A.J.; Mead, N.; Barnes, L.; Goswami, U. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* **2013**, *7*, 777. [CrossRef]
144. Goswami, U.; Cumming, R.; Chait, M.; Huss, M.; Mead, N.; Wilson, A.M.; Barnes, L.; Fosker, T. Perception of Filtered Speech by Children with Developmental Dyslexia and Children with Specific Language Impairments. *Front. Psychol.* **2016**, *7*, 791. [CrossRef]

145. Przybylski, L.; Bedoin, N.; Krifi-Papoz, S.; Herbillon, V.; Roch, D.; Léculier, L.; Kotz, S.A.; Tillmann, B. Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* **2013**, *27*, 121–131. [CrossRef]
146. Cumming, R.; Wilson, A.; Leong, V.; Colling, L.J.; Goswami, U. Awareness of Rhythm Patterns in Speech and Music in Children with Specific Language Impairments. *Front. Hum. Neurosci.* **2015**, *9*, 672. [CrossRef]
147. Beattie, R.L.; Manis, F.R. Rise Time Perception in Children With Reading and Combined Reading and Language Difficulties. *J. Learn. Disabil.* **2013**, *46*, 200–209. [CrossRef]
148. Guiraud, H.; Bedoin, N.; Krifi-Papoz, S.; Herbillon, V.; Caillot-Bascoul, A.; Gonzalez-Monge, S.; Boulenger, V. Don't speak too fast! Processing of fast rate speech in children with specific language impairment. *PLoS ONE* **2018**, *13*, e0191808. [CrossRef]
149. Cumming, R.; Wilson, A.; Goswami, U. Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: A new look at a perceptual hypothesis. *Front. Psychol.* **2015**, *6*, 972. [CrossRef]
150. Kotz, S.A.; Schwartze, M.; Schmidt-Kassow, M. Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex* **2009**, *45*, 982–990. [CrossRef]
151. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
152. Iversen, J.R.; Repp, B.H.; Patel, A.D. Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 58–73. [CrossRef]
153. Ding, N.; Melloni, L.; Zhang, H.; Tian, X.; Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **2015**, *19*, 158. [CrossRef]
154. Kotz, S.A.; Gunter, T.C. Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease?: Rhythmic auditory cuing and language. *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 62–68. [CrossRef]
155. Knilans, J.; DeDe, G. Online Sentence Reading in People With Aphasia: Evidence From Eye Tracking. *Am. J. Speech Lang. Pathol.* **2015**, *24*, S961. [CrossRef]
156. Van Ackeren, M.J.; Barbero, F.M.; Mattioni, S.; Bottini, R.; Collignon, O. Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. *eLife* **2018**, *7*, e31640. [CrossRef]

brain sciences

MDPI

Article

Infants Segment Words from Songs—An EEG Study

Tineke M. Snijders [1,2,*,†], Titia Benders [3,*,†] and Paula Fikkert [2,4]

1 Language Development Department, Max Planck Institute for Psycholinguistics, 6500 Nijmegen, The Netherlands
2 Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 Nijmegen, The Netherlands; p.fikkert@let.ru.nl
3 Department of Linguistics, Macquarie University, North Ryde 2109, Australia
4 Centre for Language Studies, Radboud University, 6500 Nijmegen, The Netherlands
* Correspondence: tineke.snijders@mpi.nl (T.M.S.); titia.benders@mq.edu.au (T.B.)
† These authors contributed equally to this work.

Received: 25 November 2019; Accepted: 6 January 2020; Published: 9 January 2020

Abstract: Children's songs are omnipresent and highly attractive stimuli in infants' input. Previous work suggests that infants process linguistic–phonetic information from simplified sung melodies. The present study investigated whether infants learn words from ecologically valid children's songs. Testing 40 Dutch-learning 10-month-olds in a familiarization-then-test electroencephalography (EEG) paradigm, this study asked whether infants can segment repeated target words embedded in songs during familiarization and subsequently recognize those words in continuous speech in the test phase. To replicate previous speech work and compare segmentation across modalities, infants participated in both song and speech sessions. Results showed a positive event-related potential (ERP) familiarity effect to the final compared to the first target occurrences during both song and speech familiarization. No evidence was found for word recognition in the test phase following either song or speech. Comparisons across the stimuli of the present and a comparable previous study suggested that acoustic prominence and speech rate may have contributed to the polarity of the ERP familiarity effect and its absence in the test phase. Overall, the present study provides evidence that 10-month-old infants can segment words embedded in songs, and it raises questions about the acoustic and other factors that enable or hinder infant word segmentation from songs and speech.

Keywords: word segmentation; infant; speech; song; EEG; ERP; familiarity; recognition; polarity

1. Introduction

Parents across cultures sing songs, words sung to a tune, for their infants. They sing lullabies to soothe and comfort, and they often sing play songs to try and make their babies laugh [1]. While parents initially sing for affect regulation and social engagement, they add didactic reasons around their infant's 10th month [2]. In fact, vocabulary acquisition is one of the primary areas in which mothers expect to see progress when they participate with their one-year-old in a musical education program [3]. Could songs indeed be beneficial for vocabulary learning?

There is evidence that infants preferentially process linguistic–phonetic information from songs compared to speech: Infants of 7 and 11 months old detect changes to syllable sequences when the syllables are sung rather than spoken [4,5], and neonates can already detect syllable co-occurrences in a continuous stream if these syllables are sung rather than produced in a flat speech register [6]. However, research so far has not yet convincingly shown that infants can use actual children's songs to learn actual language. Firstly, the songs used in previous experiments had lyrics of only four or five words and consistently paired syllables with a single pitch pattern throughout the song, thus not reflecting the lyrical and musical complexity of actual children's songs. Secondly, the skills assessed

were only partially relevant to language acquisition: While the detection of syllable co-occurrences, as tested by [6], is important to linguistic word segmentation and was associated with the participants' vocabulary size at 18 months of age, the ability to detect changes to syllable sequences, as assessed by [4,5], may be less critical to infants' concurrent or later language acquisition. Finally, only [4] provided the critical evidence that children could transfer the material learned from song to recognize words in speech, which is ultimately the primary modality of spoken language communication.

Therefore, the present study aims to test whether infants are able to learn linguistically relevant units from ecologically valid children's songs, and then also transfer these units to recognition in the spoken register. Specifically, we will assess infant word segmentation from children's songs with full lyrical and musical complexity, asking whether infants can segment word forms within songs, and subsequently recognize those word forms in speech. Moreover, we directly compare infants' segmentation across songs and the same materials presented in speech to assess whether songs present an advantage compared to speech.

As most research on the role of input in infant language acquisition has focused on the role of speech (for reviews: [7,8]), we will first contextualize the present study by discussing the potentially beneficial and hindering effects of songs for general language acquisition in adults, children, and infants. Then, we will review the literature on infant word segmentation, the fundamental ability to extract word forms from continuous speech input, which infants acquire in their first year of life. The present study, which assesses infant word segmentation from songs, is detailed in the final section of this introduction.

Songs can be expected to provide a good source for infant language learning, considering the beneficial effects of songs as well as music more generally on later language acquisition and processing. Songs directly aid memory for verbal material in both adults [9,10] and children [11]. Those findings have inspired research into the efficacy of songs for foreign- or second-language vocabulary acquisition, with the benefits, in particular for vocabulary acquisition, extending to children in the foreign language classroom (for reviews: [12,13]). Musical training can also enhance general auditory encoding, which indirectly improves a range of language skills (for a review: [14]), including children's speech segmentation [15], phonological abilities [16], as well as the perception of speech prosody [17] and durational speech cues [18].

The beneficial effects of songs and music for language acquisition are generally understood in terms of both emotional–attentional and cognitive mechanisms. Musical expertise fine-tunes and enhances sensitivity to the acoustic features shared by music and speech, and it also enhances auditory attention and working memory [19–29]. These explanations can be extended to hypothesize that songs also provide useful linguistic input for infants. Firstly, songs grab infants' attention at least as effectively as infant-directed speech [30–34], and they are more effective than speech in delaying and ameliorating distress [35,36]. Secondly, songs employ many features that infants are sensitive to in their early language acquisition, including phrasing [37,38] and rhythm [39,40]. Finally, it has been proposed that some of the beneficial effects of song on infants' well-being are a direct result of internal rhythmic entrainment [35], which is a mechanism that has also been hypothesized to be responsible for the improved encoding of linguistic material [27–29]. These three effects of song on infants render it likely that infants can effectively engage their speech encoding networks to learn from songs.

Nevertheless, it is not trivial that infants pick up linguistic information from songs. Firstly, infants' speech-honed language-learning skills may not be successful when applied to the acoustic signal of songs: lyrics sung to a melody are produced with different acoustic features than regular speech [41], including a more compressed and less consistently produced acoustic vowel space [42], cf. [43]. Secondly, even adults at times mishear words in songs, both in their non-native and native language [44–46]. Finally, even if infants learn words in songs, they may not be able to recognize these words in speech, the modality that is overwhelmingly used for spoken language communication. For example, the developmental literature on word segmentation shows that infants' ability to transfer the recognition of a learned word to a new type of acoustic stimulus slowly emerges in the second half of infants' first year. The ability to generalize across speakers, genders, or emotions emerges around 10.5 months

of age [47,48], with evidence of infants' ability to generalize across accents emerging around their first birthday [49,50]. Related work on infant word recognition suggests that infants between 8 and 10 months old might be particularly negatively impacted by speech variation [51]. Thus, it is conceivable that the infants up to one year of age are not yet able to transfer words learned from song to speech. This would pose clear boundaries to the effectiveness of songs for language acquisition in the first year of life.

The present study assesses the efficacy of songs for infant language learning through a word segmentation paradigm, testing word segmentation within songs and speech as well as subsequent generalization to speech. Segmentation, i.e., extracting individual word forms from the continuous speech stream, in which word boundaries are not typically marked by pauses, presents a sensible starting point for this research agenda, as adults find songs easier to segment than speech [52]; musical expertise is associated with better and faster segmentation in adults [53–57]; and musical training facilitates word segmentation in children [15]. Moreover, segmentation is critical to successful language acquisition, as the vast majority of words spoken to infants appear in continuous speech [58,59], even if parents are instructed to teach their infant a word [60,61]. Once infants have extracted word forms from the speech stream, they can more easily associate these with their meaning and thus start building a lexicon [62–65]. Word segmentation is also important in developing the language-ready brain, with word segmentation skills in infancy predicting language ability in the toddler years [6,66–70], although possibly not beyond [70]. Although the group-level effects of word segmentation are not always replicated [71,72], which is a topic that we will return to in the discussion, infants' ability to segment words from continuous speech is well established (see [73] for a meta-analysis). In the present study, we ask whether songs provide one source of information in infants' input from which they could segment words and start building their lexicon.

Infants rely heavily on language-specific rhythmic cues for word segmentation, with English-learning 7.5-month-olds relying on the strong–weak trochaic word stress, a rhythmic property, to segment words [74–78]. This metrical segmentation strategy is also developed by infants learning Dutch, another language with trochaic lexical stress [79], albeit at a slower rate compared to their English-learning peers [76], but not by infants learning French, a language without lexical stress [80–82].

Infant word segmentation is facilitated by the exaggeration of prosodic cues on the target word and across the entire speech stream. Prosodic accentuation on the target word is essential for segmentation by 6-month-olds and facilitates segmentation for 9-month-olds, although it may become less important when infants are 12 months of age [83]. Moreover, exact alignment of the accentuated pitch peak with the stressed syllable of the word appears to be critical [84]. General prosodic exaggeration across the speech stream, as observed in infant-directed speech (IDS), facilitates segmentation on the basis of transitional probabilities in 8-month-olds [85] and possibly even newborns [86]. In addition, word segmentation is easier from natural speech than from prosodically exaggerated speech [72], although the extent of the beneficial effect of IDS is still under investigation [87].

Considering that infants strongly rely on rhythmic cues and that prosodic exaggeration facilitates segmentation, it is conceivable that the clear musical rhythm and melodic cues of songs will enable and possibly facilitate infants' speech segmentation. The aforementioned study by François and colleagues [6] has provided the first support for this hypothesis by showing that newborns are only able to extract words from an artificial speech stream that is musically enriched. However, every syllable in these "songs" was paired with a consistent tone, resulting in a song that presented each of the four tri-syllabic words with its own unique tune throughout. Therefore, it is still an open question whether infants can segment words from songs with the full melodic and lyrical complexity of actual children's songs. Moreover, infants' aforementioned difficulties generalizing segmented words across speakers, accents, and emotions raise the question of whether they will be able to recognize words segmented from song in speech. The present study aims to address these issues by testing whether infants can segment words from realistic children's songs and subsequently recognize those words in

continuous speech. In addition, it will assess how infants' segmentation from song compares to their segmentation from continuous speech.

The present study employed an electroencephalography (EEG) familiarization paradigm for word segmentation [88,89]. This procedure is adapted from the behavioral two-step familiarization-then-test procedure, which first familiarizes infants with words and then tests their word recognition by comparing the (head turn) preference for speech with the familiarized target versus a novel control [90]. The EEG version of this paradigm exposes infants to a series of familiarization-then-test blocks and assesses word recognition on each block by comparing event-related potentials (ERPs) to familiarized targets and matched novel control words in the test phase. The EEG paradigm was preferred, as previous research has found it to be more sensitive than the behavioral method to (emerging) segmentation abilities [69,80,91,92]. Moreover, the EEG paradigm can uniquely reveal the time-course of the developing word recognition by comparing ERPs to the first and last target occurrences within the familiarization phase [89]. For example, tracking this temporal development of recognition in the EEG has revealed faster segmentation by newborns from a musically enriched compared to a monotonous speech stream [6].

The setup of the present study, as illustrated in Figure 1, was adapted from Junge and colleagues [89], who presented continuous speech in both the familiarization and test phase. Each block in the current study familiarized infants with a word embedded eight times within a sung or spoken fragment. Within this familiarization phase, the comparison of ERPs to the first two and last two occurrences of the target word enabled us to assess infants' ability to segment words from songs and contrast it with their ability to segment words from speech. After each sung or spoken familiarization phase, infants were presented with a spoken test phase consisting of two spoken phrases with the familiarized target word, and two others with a matched novel control word. The difference in ERP response to familiarized target and novel control words after song familiarization would index infants' ability to transfer words that are segmented from song to recognition in speech.

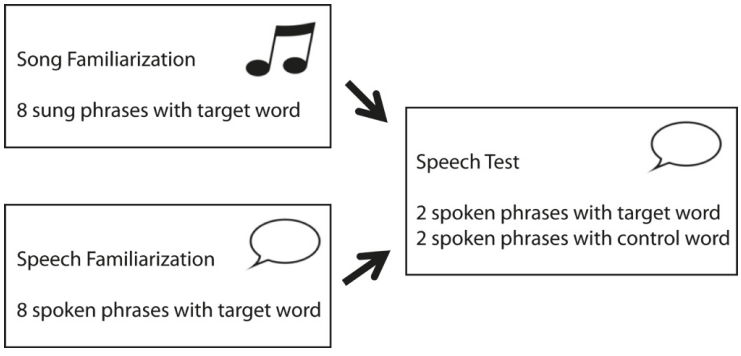


Figure 1. Setup of one experimental block of the study design.

ERP responses to familiarized compared to novel words are generally largest over left-frontal electrode sites, and can be either positive- or negative-going, with a negativity being considered a more mature response (see review and discussion by [71,83]). The polarity of infants' ERP familiarity response partly depends on stimulus difficulty, with 7-month-olds displaying positive-going responses to words embedded in speech after having negative-going responses to those same words presented in isolation [69]. The polarity of the response to stimuli of the same difficulty also changes developmentally, shifting from a positive-going response typically displayed by 6-month-olds to-7-month-olds [69,83] to a negative-going response after 8 months of age [67,79,83,88,92,93]. However, group effects in polarity are not consistently observed across studies due to large individual differences within age bands [71]. This variation between infants appears to be significant to early language development,

as the negative responders have more robust neural segmentation responses across various stages of the procedure [70,93], better concurrent vocabulary size [71,93], better and faster vocabulary development into toddlerhood [67,69,93], and better general language skills [69].

The general developmental shift from an initial positivity to a later negativity for word recognition responses has been ascribed to cortex maturation [83] (see also [94] for a similar reasoning on the polarity of the infant MMN for auditory discrimination), as the auditory cortex undergoes tremendous changes, specifically around 6 months of age [95,96], which can influence the polarity of ERP components [97]. However, the stimulus-dependent polarity shift within a single group of infants [69] reveals that a more functional explanation is required. The polarity of an ERP depends on the location and orientation of the underlying brain activation [98]. Männel and Friederici have proposed slightly different origins of the positive and negative ERP familiarity effects (secondary auditory cortices versus superior temporal cortex, respectively), with more lexical processing due to the infants' advancing linguistic experience resulting in the shift to superior temporal cortex activation [83]. In a similar vein, Kidd and colleagues have proposed that the negativity reflects the emergence of a lexicon [71]. We would argue similarly that the negative ERP familiarity effect reflects active lexical learning, and we interpret the negative familiarity effect as a 'repetition enhancement' effect (see below). When adults hear repetitions of words within sentences, a positive ERP repetition effect is elicited [99]. This can be interpreted as 'repetition suppression', which is a reduced neural response when a stimulus is repeated [100,101]. The positive infant ERP familiarity effect might reflect a similar repetition suppression response but now for low-level acoustic properties of the stimulus. However, the negative ERP familiarity effect might reflect 'repetition enhancement'-enhanced processing when a stimulus is repeated [102]. Repetition enhancement effects are thought to reflect a neural learning mechanism for building or strengthening novel neural representations [103,104]. This reasoning would support the notion of the negative infant ERP familiarity effect reflecting the active building of a lexicon, which is in accordance with the proposals of [71,83].

The present study tested word segmentation in 10-month-old Dutch infants, for whom a negative-going ERP familiarity response can generally be expected in speech [67,79,88,89]. For the speech sessions, we expect to replicate the negative ERP familiarity response seen in the work by Junge and colleagues [89]. Within the song familiarization, a left frontal negative-going response to the last two compared to the first two target occurrences would be taken as evidence that word segmentation from a song is unproblematic for infants. Both negative and positive ERP familiarity responses indicate that the repetition of the word form has been identified within the continuous speech stream. However, given the previous literature, a positive response would be interpreted as indicating difficulties with song segmentation. Within the subsequent spoken test phase, a negative-going response would similarly be interpreted as automatic generalization from song to speech, with a positivity indicating a more challenging transfer process.

2. Materials and Methods

2.1. Participants

Forty Dutch 10-month-old infants participated in two experimental sessions, resulting in eighty datasets (40 song, 40 speech). The number of participants tested was based on [89]. All infants were born in term (37–42 weeks gestational age), normally developing, without a history of neurological or language impairments in the immediate family. Twenty-one datasets were excluded from analysis because of too few artefact-free EEG trials (see below). One participant was excluded because he was raised bilingually. The remaining 57 included datasets came from 32 subjects, with 25 of them contributing good data in both the speech and the song session. The 32 included subjects (16 female) were all monolingual Dutch infants (session 1: mean age 299 days, range 288–313 days; session 2: mean age 306 days, range 293–321 days). Infants were recruited from the Nijmegen Baby and Child Research Center Database. The study was approved by the local ethics committee, and parent(s) gave

written informed consent for their infants prior to the experiment, in accordance with the Declaration of Helsinki.

2.2. Materials

The familiarization materials were 20 verses of eight phrases. Each verse contained one repeating target word in every phrase (see Table 1 and Figure 2 for an example). The verses were recorded in a sung and a spoken version, for the "song familiarization" and "speech familiarization", respectively. The "song" and "speech" stimuli used identical verses/lyrics, but only the "song" versions were recorded with the designated melodies. Each song and speech version of a verse was recorded with two different target words, for a total of four recordings per verse. The reader is referred to Supplementary Table S3 for the full set of materials.

Table 1. Example of a familiarization +test block with target word pair bellers-piefen. Target words are underlined. Materials were in Dutch. English word-for-word and semantic translations are given. The first two (first/second) target occurrences are indicated in blue, the last two (seventh/eighth) target occurrences of the familiarization phase are indicated in red. Familiarized target words in the test phase are indicated in purple, and novel control words are indicated in green.

	Familiarization Phase (Sung or Spoken)	**Word-for-Word Translation**	**Semantic Translation**
	Luister eens!	Listen once!	Listen to this!
1.	Met bellers kun je lachen	With callers can you laugh	One can have a laugh with callers
2.	De vrouw vindt bellers stom	The woman regards callers stupid	The woman thinks callers are stupid
3.	We spraken met de woeste bellers	We spoke to the wild callers	We spoke to the wild callers
4.	Dan praten bellers graag	Then speak callers preferably	Callers prefer to speak then
5.	Daar achterin zijn bellers	There behind are callers	Callers are there in the back
6.	Wat lopen bellers snel	What walk callers fast	Callers walk that fast
7.	Jouw bellers kletsen makkelijk	Your callers chat easily	Your callers are at ease chatting
8.	Ik zag de bellers niet	I saw the callers not	I did not see the callers
	Test phase (spoken)		
	Luister eens!	Listen once!	Listen to this!
9.	Aan die piefen gaf hij koffie	To those hotshots gave he coffee	He served those hotshots coffee
10.	Vaak gaan bellers op reis	Often go callers to travel	Callers travel often
11.	Alle bellers stappen laat uit	All callers get late off	All callers get off late
12.	Zij zijn goede piefen geworden	They have good hotshots become	They have become good hotshots

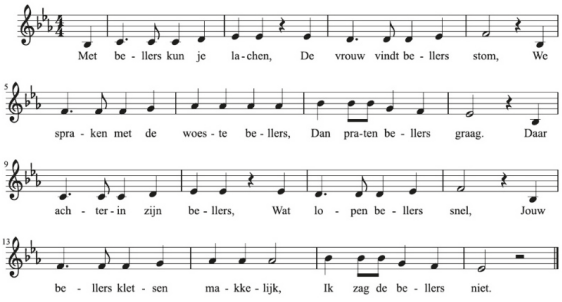


Figure 2. Example of score for the song that was used for target word 'bellers'.

The 20 melodies for the "song" versions all consisted of eight phrases, or of four melodic phrases that were repeated twice (with different lyrics). The melodies were (variations on) melodies of German, English, French, Norwegian, and Dutch children's songs and unknown to a sample of 22 native Dutch parents with a 10-month-old infant (see Supplementary Table S2). The 20 original target words in [89] were supplemented with three further target words from [79,88] and 17 new target words. New target words were added to avoid the repetition of target words across blocks. All 40 target words were low-frequency trochees (see Table 2), each with a CELEX frequency lower than 19 per million [105]. The 40 target words were combined into 20 word pairs that shared a semantic category (e.g., "emoe" and "hinde"; English: emu and doe; see Table 2). Yoked pairs were created between the 20 melodies and the 20 target word pairs (Table 2). The verses that were written to each melody were made with both target words of the word pair.

The verses had a mean phrase length of 5.71 words (range: 3–10) and 7.82 syllables (range: 4–14). All eight phrases of a verse contained the target word. The target word was never the first word of a phrase, and the target word occurred maximally twice per verse in final phrase position. With one exception due to experimenter error, target words were never the last word of the first, second, seventh, or eighth sentence. The word preceding the target word was unique across the eight phrases. The main word stress of the target word consistently matched the meter of the melody in the phrase, and the text setting of the phrases to the melodies was correct, which is a condition that is considered critical for learning from songs [24]. We also created four test sentences per target word pair. The position of the target words was never the first or last of a test sentence and was otherwise variable.

Stimuli were recorded in a sound-attenuated booth, using Adobe Audition. The stimuli were annotated and further processed in Praat [106]. All stimuli were recorded by a trained singer (mezzo-soprano), and sung and spoken in a child-directed manner. The verses for the familiarization phase were typically recorded in one take per verse. Those original recordings were kept intact in terms of the order of the phrases, the duration of the phrase intervals, and the speaker's breathing in those intervals. Three speech and one song stimuli were created by combining multiple takes to obtain stimuli without disturbing noises. Recording in one take was required to render naturally sounding song versions. In this respect, our stimulus creation is different from that used in [89], who recorded their sentences in a randomized order and combined them after the fact.

The spoken test sentences for the test phase were recorded in one take per target word pair, extracted individually from the original recordings, and played back in a randomized order in the experiment.

Finally, the attentional phrase "Luister eens!" ("Listen to this!"), which was used as a precursor to each training and test stimulus in the experiment, was recorded in both song and speech versions.

Supplementary information about the materials is given in Supplementary Table S1, with acoustic properties (duration, pitch, and loudness measures) given in Supplementary Table S1 (for phrases) and Supplementary Table S2 (for target words). The mean 'focus' is also reported in Supplementary Table S2, which approaches 1 if the target word is always the highest or loudest in the phrase, thus measuring

acoustic prominence. Acoustic properties of the target words in phrases one and two versus those in phrases seven and eight of the familiarization stimuli were matched (see Supplementary Table S3).

Table 2. The 20 target word pairs (English translation in parentheses), and songs that were the base for the melody of that specific word pair (language of lyrics of original song in parentheses).

	Word 1		Word 2		Song that Melody Was Based on
1	bellers	(callers)	piefen	(hotshots)	If all the world were paper (English)
2	hinde	(doe)	emoe	(emu)	Sing a song of sixpence (English)
3	gondels	(gondolas)	schuiten	(barques)	See-saw Margery Daw (English)
4	drummer	(drummer)	cantor	(cantor)	Georgie Porgie (English)
5	gieter	(watering cans)	silo's	(silos)	There was a crooked man (English)
6	hommels	(bumblebees)	kevers	(beetle)	Pat-a-cake (English)
7	fakirs	(fakirs)	dansers	(dancers)	Little Tommy Tucker (English)
8	krekels	(crickets)	hoenders	(fowl)	En elefant kom marsjerende (Norwegian)
9	krokus	(crocus)	anjer	(carnation)	Smil og vær glad (Norwegian)
10	lener	(borrower)	preses	(president)	Ute På Den Grønne Eng (Norwegian)
11	mammoet	(mammoth)	orka	(orca)	Jeg snører min sekk (Norwegian)
12	monnik	(monk)	frater	(friar)	Auf de Swäb'sche Eisenbahne (German)
13	otters	(otter)	lama's	(llamas)	Suse, liebe Suse (German)
14	mosterd	(mustard)	soja	(soya)	Schneeflöckchen Weißröckchen (German)
15	pelgrims	(pilgrim)	lopers	(runners)	Wem Gott will rechte Gunst erweisen (Ger.)
16	pudding	(pudding)	sorbet	(sorbet)	Wiesje (Dutch)
17	ronde	(round)	kuier	(saunter)	A l'intérieur d'une citrouille (French)
18	sitar	(sitar)	banjo	(banjo)	La bonne avonture o gué (French)
19	sultan	(sultan)	viking	(Viking)	Neige neige blanche (French)
20	zwaluw	(swallow)	kievit	(lapwing)	Entre le boeuf e l'âne gris (French)

2.3. Procedure

Infants participated in a separate song and speech session. The session order was counterbalanced across infants, with on average 7.6 days between sessions (range 5–14 days). Before the experiment started, the child could play on a play mat to get accustomed to the lab environment while the experimental procedure was explained to the parent. The EEG cap was pregelled to minimize the setup time. Then, the cap was fitted, electrode impedances were checked, and some extra gel was added where necessary. Next, the infants were seated on their parent's lap in a sound-attenuated booth with Faraday cage, and data collection was initiated. Sung and spoken sentences were presented to the infant over two loudspeakers at 65 dB. While listening to the sentences, the infant watched a silent screen-saver (not linked to auditory input) or played with silent toys. One experimenter sat next to the screen to maintain the engagement of the infant with silent toys or soap bubbles if necessary. Both parent and experimenter listened to masking music over closed headphones. A second experimenter ran the EEG acquisition from outside the experimental booth and monitored the infant through a closed-circuit video. The experiment was stopped if the infant became distressed. One full experimental session (including preparations and breaks) took about one hour, with the experiment proper taking about 20 min.

Stimuli were presented using Presentation software [107]. In each session, the infants listened to 20 blocks of familiarization-and-test trials, with a different melody and target word pair in every block. Each block consisted of a familiarization phase immediately followed by the corresponding test phase (see Figure 1; see Figure 2 and Table 1 for an example). The eight phrases of the verse in the familiarization phase, all containing the target word, were spoken (speech session) or sung (song session). The four sentences in the test phase were always spoken: two sentences contained the 'familiarized' word and two contained the second 'control' word of the word pair, presented in a randomized order. To reduce the effects of modality switching, the attentional phrase "Luister eens!" (English: Listen to this!) was played to the infants in the modality of the session before each familiarization block. The words "Luister eens!" were always presented in the spoken modality before each test phase.

The order of the blocks was counterbalanced across subjects. Within each session, every target word was the 'familiarized' word for half of the infants and the 'control' word for the other half, and

this assignment of words was counterbalanced across subjects. For each infant, the familiarized words in the spoken session were the control words in the sung session and vice versa. Note that in [89], this repetition of critical words already occurred in the second half of the experiment, which was why we accepted a repetition in the second session (after 5–14 days). The order of the blocks was such that the target words never started more than twice in a row with vowels or with the same consonantal manner or place of articulation.

2.4. EEG Recordings

EEG was recorded from 32 electrodes placed according to the International 10–20 system, using active Ag/AgCl electrodes (ActiCAP), Brain Amp DC, and Brain Vision Recorder software (Brain Products GmbH, Germany). The FCz electrode was used as the on-line reference. Electro-oculogram (EOG) was recorded from electrodes above (Fp1) and below the eye, and at the outer canthi of the eyes (F9, F10). The recorded EEG electrodes were F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP6, TP10, P7, P3, Pz, P4, P8, PO9, and Oz. The data were recorded with a sampling rate of 500 Hz and were filtered on-line with a time constant of 10 s and a high cutoff at 1000 Hz. Electrode impedances were typically kept below 25 kΩ.

2.5. Data Processing

EEG data were analyzed using Fieldtrip, an open source MATLAB toolbox (The MathWorks, Natick, MA, USA) for EEG and MEG analyses [108].

First, eye movement components and noise components in the EEG data were identified using independent component analysis (ICA, [109]). In order to identify components based on as much data as possible, prior to ICA analysis, all the EEG data of the whole session were filtered from 0.1 to 30 Hz and cut in 1 s segments. Bad channels were removed, as were data segments with flat channels or large artifacts (>150 μV for EEG channels, >250 μV for EOG channels). Then, we applied infomax ICA [110] as implemented in EEGlab [111]. A trained observer (T.M.S.) identified components that revealed eye components or noise on individual electrodes. Subsequently, time-locked data were made from the original EEG data by cutting the raw data in trials from 200 ms before to 900 ms after the onset of the critical words. Again, these data were filtered from 0.1 to 30 Hz, and the bad channels were removed, as well as trials with flat channels. Then, the identified eye and noise components were removed from the time-locked data. For the included datasets (see the end of this subsection), the mean number of removed eye and noise components was 2.8 and 2.7, respectively (range: 1–5 for eye components, 0–6 for noise components). After ICA component rejection, EEG channels were re-referenced to the linked mastoids. Electrodes PO10, Oz, and PO9 were discarded, because these were bad channels for too many infants. A baseline correction was applied in which the waveforms were normalized relative to the 200 ms epoch preceding the onset of the critical word, and trials containing EEG exceeding ±150 μV were removed. Six datasets were discarded because of too many (>4) bad channels, and 15 datasets were discarded because fewer than 10 trials per condition (<25%) remained after artifact rejection. For the remaining datasets (32 subjects, 57 datasets; 31 speech, 26 song; 29 first session, 28 second session; 25 subjects with good data in both sessions), bad channels were repaired using spherical spline interpolation ([112]; mean of 0.9 channels repaired, range 0–3). Finally, ERPs were made by averaging over relevant trials. For the analyses on the combined datasets of the speech and song sessions (see below), the trials were concatenated across sessions before averaging. The combined datasets had an average of 48 included trials per condition (range 16–69), while for the single sessions, this was 26 (range 12–36) for song and 28 (range 13–36) for speech.

2.6. Planned ERP Analyses

For the familiarization phase, the ERP familiarity effect was assessed by comparing the ERP in response to the last two (seventh/eighth) versus the first two (first/second) target occurrences. For the

test phase, the ERP familiarity effect was assessed by comparing the ERP in response to familiarized target words versus novel control words.

Analyses were performed first for the combined song and speech sessions (32 subjects). In a second step, differences between song and speech sessions were assessed, this time only including the 25 subjects that had >10 trials per condition in both sessions.

Analyses to assess the ERP familiarity effect were performed both on predefined time windows and left-frontal electrodes (based on previous literature), as well as on all time points or electrodes in a single test (to assess possible deviations from previous literature and considering the novel inclusion of the song modality).

Time windows of interest (250–500 ms and 600–800 ms) were defined based on previous literature reporting the infant ERP familiarity effect (see [71]). The left-frontal region of interest was defined as electrodes F7, F3, and F5, which are consistently included in the calculation of average amplitudes in previous literature [67,70,71,83,88,89].

The average ERPs of the left frontal region for the two time windows were assessed using SPSS, with a 2 × 2 repeated measures ANOVA on mean left-frontal ERP amplitude, with familiarity (familiar, novel) and time window (250–500, 600–800) as within-subject factors. For the modality comparison (on the 25 subjects who had good data in both sessions), modality (song, speech) was added as a within-subjects factor. For the ANOVAs, we used the Huynh–Feldt epsilon correction and reported original degrees of freedom, adjusted p-values, and adjusted effect sizes (partial eta-squared: ηp^2).

Additionally, to explore the possible effects outside the regions and time windows of interest, ERP familiarity effects were assessed using cluster randomization tests [113]. Cluster randomization tests use the clustering of neighboring significant electrodes or time points to effectively control for multiple comparisons, while taking the electrophysiological properties of EEG into account. In these tests, first, all the electrodes or all time points are identified that exceed some prior threshold (in this case, a dependent samples t-test was conducted and the p-value was compared to a threshold alpha level of 0.05, uncorrected for multiple comparisons). In a second step, clusters are made in which neighboring electrodes or time points that exceed the threshold are grouped. For every identified cluster, a cluster-level statistic is calculated, summing the t-statistics of all the included electrodes/time points. A reference randomization null distribution of the maximum cluster-level statistic is obtained by randomly pairing data with conditions (e.g., familiarized and novel conditions) within every participant. This reference distribution is created from 1000 random draws, and the observed cluster p-value can then be estimated as the proportion from this randomization null distribution with a maximum cluster-level test statistic exceeding the observed cluster-level test statistic (Monte Carlo p-value). Thus, the cluster randomization p-value denotes the chance that such a large summed cluster-level statistic will be observed in the absence of an effect (see [113]). Note that although the observed clusters provide some information about latency and location, the cluster statistic does not define the actual spatial and temporal extent of the effect, as time points and electrodes are included based on uncorrected statistics (see [114]). Thus, precise cluster onset and extent should not be over-interpreted.

For our purposes, three exploratory cluster randomization tests were performed: two assessing all the electrodes in the two time windows (for both time windows comparing the condition averages of the time window for all electrodes and then clustering over electrodes), and one assessing all the time points from 100 to 900 ms in the left-frontal region of interest (comparing the condition averages of the three left-frontal electrodes for all time points, and then clustering over time points, starting from 100 ms after stimulus onset, as no effects were expected to occur before this time).

3. Results

3.1. Planned Analyses—Segmentation from Song and Speech

3.1.1. ERP Familiarity Effect in the Familiarization Phase, Song and Speech Combined

At the left-frontal region of interest, the repeated measure ANOVA on the 32 combined datasets showed an ERP familiarity effect ($M(SD)_{\text{unfamiliar12}}$ = −1.44(0.67), $M(SD)_{\text{familiarized78}}$ = 0.97(0.64); $F_{\text{familiarity}}(1,31) = 11.8$, $p = 0.002$, $\eta p^2 = 0.28$) with no detected difference across time windows (250–500 ms: $M(SD)_{\text{unfamiliar12}}$ = −1.41(0.69), $M(SD)_{\text{familiarized78}}$ = 0.93(0.65); 600–800 ms: $M(SD)_{\text{unfamiliar12}}$ = −1.47(0.74), $M(SD)_{\text{familiarized78}}$ = 1.00(0.76); $F_{\text{familiarity x time window}}(1,31) = 0.05$, $p = 0.83$, $\eta p^2 = 0.002$). As can be seen in Figures 3 and 4, the ERP was more positive for the last two (seventh/eighth) than for the first two (first/second) target occurrences.

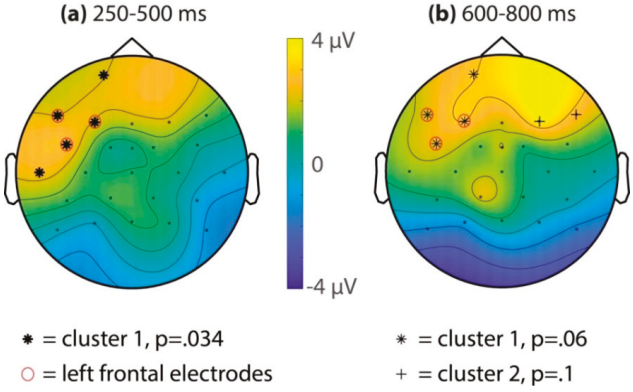


Figure 3. Familiarity effect in the familiarization phase for the combined sessions (32 subjects). Topographic isovoltage maps of the difference between the last two target occurrences (7th/8th) and the first two target occurrences (1st/2nd), in the (**a**) 250–500 ms and (**b**) 600–800 ms latency ranges. Electrodes that are part of output clusters in the cluster randomization test are shown with stars. The red circles indicate electrodes that are part of the left-frontal region of interest.

Subsequently, all the electrodes were assessed using a cluster-randomization test, comparing the average ERP amplitudes for the last two (seventh/eighth) and first two (first/second) target occurrences for both time windows of interest (see Figure 3). For the 250–500 ms time window, this resulted in a significant cluster ($p = 0.034$) of left-frontal electrodes (Fp1, F7, F3, FC5, and T7). A second cluster over the right hemisphere (F4) did not survive multiple comparison correction (cluster $p = 0.52$). For the 600–800 ms time window, a positive cluster over the left-frontal electrodes was marginally significant (cluster $p = 0.06$, electrodes Fp1, F7, F3, and FC5). Two other clusters did not survive multiple comparisons (cluster 2: F4 and F8, $p = 0.10$; cluster 3: CP1, $p = 0.48$).

To assess all time points between 100 and 900 ms after target onset, a cluster randomization was performed on the mean of the left-frontal region (electrodes F7, F3, and F5, see Figure 4). This resulted in two significant clusters, the first ranging from 268 to 594 ms (cluster $p = 0.004$) and the second ranging from 612 to 792 ms (cluster $p = 0.028$).

The clusters identified in the cluster randomization tests match the timing and topography of the previously reported infant ERP familiarity effect. However, note that the present ERP familiarity effect is a positive shift, while previous literature on segmentation by infants of the same age has most often reported a negative shift [79,88,89]. This discrepancy will be discussed in detail in the Discussion section.

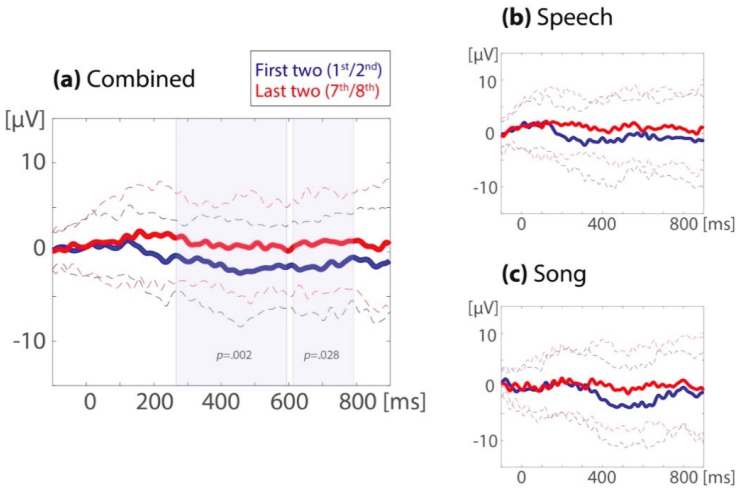


Figure 4. Familiarity effect in the familiarization phase. Event-related potentials (ERP) averaged over left-frontal electrodes for (**a**) the combined sessions (32 subjects), (**b**) the speech sessions (31 subjects), and (**c**) the song sessions (26 subjects). The solid lines are the ERPs from the first two target occurrences (1 and 2, in blue) and the last two target occurrences (7 and 8, in red). The means ±1 SD are given as dotted lines. The shaded areas indicate the clusters identified in the cluster randomization test.

3.1.2. ERP Familiarity Effect in the Familiarization Phase, Comparing Song to Speech

As can be seen in Figure 4, the ERP familiarity effects in the familiarization phase were similar across the song and speech modalities, although this effect occurred possibly somewhat later in the songs. For the 25 subjects that contributed enough data for both the speech and song sessions, the left-frontal ERP familiarity effect during the familiarization phase was compared between modalities. The repeated measures ANOVA showed a main effect of familiarity ($M(SD)_{\text{unfamiliar12}} = -1.23(0.76)$, $M(SD)_{\text{familiarized78}} = 0.32(0.71)$; $F_{\text{familiarity}}$ (1,24) = 4.14, $p = 0.053$, $\eta p^2 = 0.15$). There was no significant difference in the ERP familiarity effect between modalities (speech: $M(SD)_{\text{unfamiliar12}} = -0.74(1.19)$, $M(SD)_{\text{familiarized78}} = 0.42(0.80)$; song: $M(SD)_{\text{unfamiliar12}} = -1.72(1.18)$, $M(SD)_{\text{familiarized78}} = 0.23(1.10)$; $F_{\text{familiarity x modality}}$ (1,24) = 0.13, $p = 0.72$, $\eta p^2 = 0.005$), and no significant interaction between familiarity, modality, and time window ($F_{\text{familiarity x modality x time window}}$ (1,24) = 0.14, $p = 0.71$, $\eta p^2 = 0.006$).

Cluster-randomization tests were performed to explore the possible modality differences in ERP familiarity effects outside the regions and time windows of interest. First, all electrodes were assessed in the 250–500 ms as well as the 600–800 ms time window, comparing the difference between the first two (first/second) and last two (seventh/eighth) target occurrences across the speech and the song modality. No clusters were identified, meaning that not one electrode showed a significant interaction between modality and familiarity, even prior to corrections for multiple comparisons. Then, all time points were assessed in the left-frontal region of interest, comparing the familiarity effect across modalities. Again, no clusters were identified, meaning that not one time point showed a significant interaction between modality and familiarity, even when uncorrected for multiple comparisons. In sum, no differences could be identified in the ERP familiarity effect between the song and the speech modality, and there was no evidence for an earlier start of the ERP familiarity effect in speech.

3.2. Planned Analyses Effects in Test Phase (Transfer to Speech)

3.2.1. ERP Familiarity Effect in the Test Phase, Song and Speech Combined

Figure 5 displays the ERPs for familiarized target words and novel control words in the test phase, as well as the topography of their difference. For the left-frontal region of interest, the repeated measures ANOVA showed no difference between the ERP to the familiarized and the novel test items ($M(SD)_{\text{novelTest}} = -1.32(0.73)$, $M(SD)_{\text{familiarizedTest}} = -0.51(0.80)$; $F_{\text{familiarity}}$ (1,31) = 0.80, $p = 0.38$, $\eta p^2 = 0.025$), with no difference in the ERP familiarity effect between time windows (250–500 ms: $M(SD)_{\text{novelTest}} = -1.49(0.68)$, $M(SD)_{\text{familiarizedTest}} = -0.07(0.84)$; 600–800 ms: $M(SD)_{\text{novelTest}} = -1.16(0.89)$, $M(SD)_{\text{familiarizedTest}} = -0.95(0.92)$; $F_{\text{familiarity x time window}}$ (1,31) = 1.69, $p = 0.20$, $\eta p^2 = 0.05$).

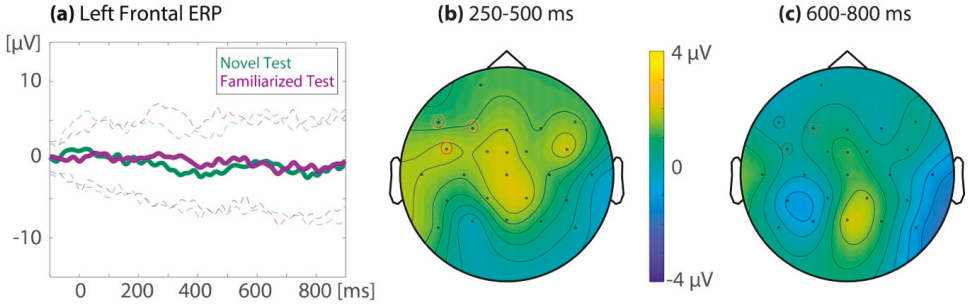


Figure 5. Familiarity effect in the test phase for the combined sessions (32 subjects). (**a**) Event-related potentials (ERP) averaged over left-frontal electrodes (red circles in (**b**) and (**c**)). The solid lines are the ERPs from the novel control words in the test phase (in green) and the familiarized target words in the test phase (in purple). The means ±1 SD are given as dotted lines. Right: Topographic isovoltage maps of the difference between the familiarized target words and novel control words during test, in the (**b**) 250–500 ms and (**c**) 600–800 ms latency ranges. The red circles indicate electrodes that are part of the left-frontal region of interest.

When assessing the ERP familiarity effect during test on all electrodes using the cluster randomization test, one effect was identified at T7 in the 250–500 ms time window, which did not survive multiple comparisons correction (cluster $p = 0.32$). No clusters were identified in the 600–800 ms time window. When assessing all time points between 100 and 900 ms for the left-frontal region of interest, one cluster was identified from 330 to 350 ms, which did not survive multiple comparisons (cluster $p = 0.26$). Thus, no significant ERP familiarity effect could be identified in the test phase.

3.2.2. ERP Familiarity Effect in the Test Phase, Comparing Song to Speech

To assess the possible differences in ERP familiarity effect in the test phase across modalities, a repeated measures ANOVA was performed on the 25 subjects that contributed enough data in both speech and song sessions. There was no significant difference between modalities in the left-frontal ERP familiarity effect (speech: $M(SD)_{\text{novelTest}} = -2.28(1.17)$, $M(SD)_{\text{familiarizedTest}} = -1.15(1.26)$; song: $M(SD)_{\text{novelTest}} = -0.86(1.20)$, $M(SD)_{\text{familiarizedTest}} = 0.75(1.28)$; $F_{\text{familiarity x modality}}$(1,24) = 0.045, $p = 0.83$, $\eta p^2 = 0.002$), with also no significant interaction between familiarity, modality, and time window ($F_{\text{familiarity x modality x time window}}$(1,24) = 0.42, $p = 0.52$, $\eta p^2 = 0.017$).

When comparing the ERP familiarity effect across modalities for all electrodes using cluster randomization, one cluster was identified (FC2, CP2) in the 250–500 ms time window, which did not survive multiple comparisons correction (cluster $p = 0.17$). An effect at CP2 was also identified in the 600–800 ms time window, again not surviving multiple comparisons correction (cluster $p = 0.31$). When comparing the ERP familiarity effect across modalities on all time points for the left frontal

region of interest, no clusters were identified. Thus, no differences in the ERP familiarity effect were identified for song compared to speech in the test phase.

To summarize, in the familiarization phase, we identified a positive ERP familiarity effect over the left-frontal electrodes in both the 250–500 and 600–800 ms time windows (see Figures 3 and 4). This effect did not differ significantly between the song and speech modality (see Figure 3b,c). In the test phase, there was neither an ERP familiarity effect in song nor in speech.

3.3. Follow-Up Analyses—Motivation and Methods

The planned analyses did not render all predicted effects, most notably a positive-going instead of a negative-going familiarity response in the familiarization phase and the absence of any group-level effects in the test phase. As reviewed in the Introduction, the polarity of infants' responses to familiar words has been associated with stimulus difficulty as well as developmental maturity [71,83], and the absence of group-level effects in previous work has been ascribed to large individual variation in development even within narrow age bands [71]. Therefore, we conducted a series of follow-up analyses to explore individual differences and to assess whether the ERP polarity to the target word shifted with target occurrence during the familiarization phase.

The first follow-up analysis considered the familiarization phase in more detail. In order to expand the comparison of the first two (first/second) versus last two (seventh/eighth) target occurrences from the planned analyses, it assessed how the familiarity responses developed across all eight occurrences of the target word therein. Additional follow-up analyses are reported in Supplementary Table S4, asking whether the lack of a group-level effect in the test phase might be due to a mix of positive and negative responders and associating the responder type to the responses across the familiarization phase.

3.3.1. Follow-Up Analysis #1—Development Over Eight Familiarization Occurrences

The first set of follow-up analyses was conducted to scrutinize the identified positive-going response in the last two (seventh/eighth) compared to the first two (first/second) occurrences of the target word in the familiarization passage. Building on the work by Junge and colleagues [89], this analysis targeted the development of the word recognition response across all eight target word occurrences. Figure 6 displays this development, averaged over participants, separated by time window (250–500 ms versus 600–800 ms) and modality (speech versus song). As can be seen in Figure 6 as well as from the model results described below, the word recognition response increased in positivity over the first four occurrences of the target word for both speech and song passages. Then, the response became more negative on occurrences five and six when children listened to speech, but it remained stable when they listened to song. On the final occurrences seven and eight, the word recognition response became more positive again when children listened to speech, whereas it now became more negative when children listened to song.

The statistical analyses were conducted using mixed effects regression models as implemented in the *lmer* function in the *lme4* package [115] in the *R* statistical programming environment [116]. The dependent variable was the per-occurrence average EEG amplitude (in μV) over the left-frontal electrodes (F7, F3, and F5) in the two time windows of interest (250–500 ms and 600–800 ms after stimulus onset), thus adhering to the same electrodes and time windows of interest as in the planned analyses. The random-effects structures were selected to be parsimonious, i.e., containing only those random-effects parameters required to account for the variance in the data, and were determined following an iterative procedure adapted from [117]. The resulting *p*-values were interpreted against the conventional alpha level of 0.05, while *p*-values between 0.05 and 0.07 were interpreted as marginally significant and also worthy of discussion.

The first model assessed the word recognition effect over the eight familiarization occurrences (captured by the first to third-order polynomials, to assess linear, cubic, and quadratic trends over occurrences), comparing across the two modalities (speech = −1; song = 1), the two sessions (first session = −1; second session = 1), and the two time windows (250–500 ms = −1; 600–800 ms = 1).

The three fixed factors were fully crossed (i.e., all the possible two-way and three-way interactions were included), and each interacted with the three polynomials. In the first stage of determining the random-effects structure, the dimensionality of the by-subject random effects was reduced to include the (non-correlated) random intercept and slope for modality. In the second stage, the by-word random effects nested within word pairs were added and systematically reduced to by-word random intercepts and by-word slopes for modality and session.

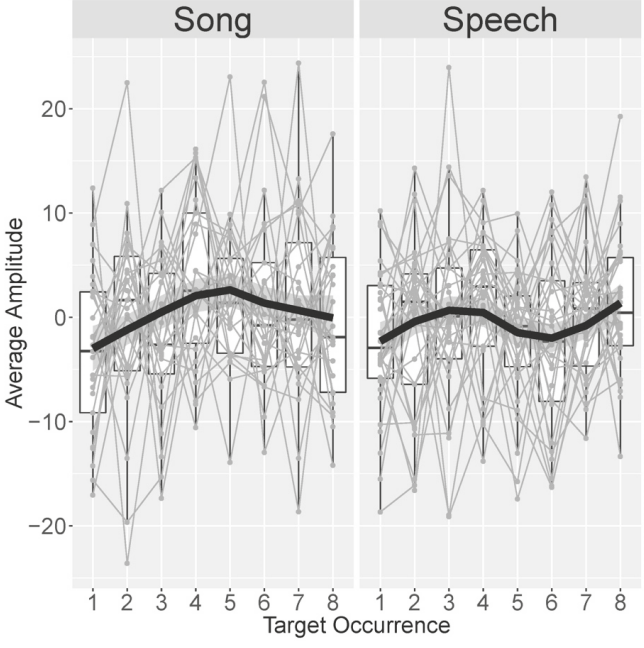


Figure 6. Average electroencephalography (EEG) amplitude in μV (averaged over blocks) in the early (250–500 ms) time window across the eight target occurrences in the familiarization phase in song (left panel) and speech (right panel). Gray lines connect individual participants' averages, indicated by gray points. The black lines provide a locally estimated scatterplot smoothing (LOESS)-smoothed development averaged over participants.

The follow-up models assessed the development of the word recognition effect separately in speech and in song, with as predictors the three polynomials for the eight occurrences, the contrasts for session and time window, the fully crossed fixed effects, and their interactions with the polynomials. To maintain consistency across analyses, we aimed for the random-effects structure of these analyses to contain a by-subject random intercept as well as the by-word random intercepts and slope for session. A model with this random-effects structure converged successfully for the song data, but it had to be reduced to by-subject and by-word random intercepts for the speech data.

The first model, which included both song and speech, revealed a marginally significant positive linear trend over occurrences ($\beta = 51.31$, $t = 1.862$, $p = 0.063$), a significant quadratic trend ($\beta = -78.66$, $t = -2.858$, $p = 0.004$), and a significant cubic trend ($\beta = 81.74$, $t = 2.968$, $p = 0.003$). The linear trend was modulated by a significant positive interaction with session ($\beta = 5.634$, $t = 2.044$, $p = 0.041$). The quadratic trend was modulated by a significant interaction with modality ($\beta = -82.44$, $t = -2.996$, $p = 0.003$). No other effects were statistically significant or marginally significant. Due to the interactions with session and, especially, modality, the interpretation of these effects requires analyzing subsets of

the data. Since the effect of modality was of primary interest, these subset analyses separately analyzed the speech and the song data.

The model on only the speech data revealed a significant cubic trend over occurrences ($\beta = 124.16$, $t = 3.367$, $p = 0.0007$) and a significant interaction between session and the linear effect of occurrence ($\beta = 96.52$, $t = 2.619$, $p = 0.009$). No other effects were statistically significant or marginally significant. These findings support the observations regarding the speech sessions as made from Figure 6: the EEG amplitude developed non-linearly over the course of a speech block, with two changes in the direction of the developing response (after the initial increase followed a decrease and another increase).

The model on only the song data revealed a significant linear trend over occurrences ($\beta = -94.47$, $t = 2.295$, $p = 0.022$) as well as a quadratic trend over occurrences ($\beta = -161.61$, $t = -3.952$, $p < 0.001$). There was a marginally significant effect of session ($\beta = -1.393$, $t = -1.904$, $p = 0.065$). No other effects were statistically significant or marginally significant. These findings support the observations regarding the song sessions as made from Figure 6: the EEG amplitude developed non-linearly over the course of a song block, with one change in direction of the developing response (the initial increase was followed by a decrease).

These effects were further teased apart in by-occurrence analyses reported in the Supplementary Table S4 (D1: "By-Occurrence analyses"), which confirmed that the difference between modalities was only apparent on occurrences five and six. A final series of models reported in the Supplementary Table S4 (D2: "Occurrence–session interaction") explored the interaction between the linear trend over occurrences and session, which was apparent in the analyses comparing song and speech as well as the speech-only data. Across the separate analyses of Sessions 1 and 2, the linear effect was only statistically significant in Session 2 of both song and speech data, but it was not statistically significant in Session 1. These results had no implications for the conclusions regarding the development of the familiarity response in speech and song as outlined above.

3.3.2. Additional Follow-up Analyses: Responder Types in the Test Trials

Supplementary Table S4 (D3–5) report additional exploratory follow-up analyses, which suggest that the group-level null effect in the test phase may mask a split between negative and positive responders and thus a robust segmentation effect (D3: "Formally establishing responder types"). Using these groups in further analyses seemed warranted, as there was no strong indication that the positive and negative responders were unequally distributed across the versions (D4: "Experimental control and responder types"). Comparisons of negative and positive responders suggest a more pronounced sinusoid-shaped development across speech familiarization for negative than positive responders (D5: "Comparing development in the familiarization phase across responder types in the test phase").

4. Discussion

The present study set out to test whether infants are able to segment words from ecologically valid children's songs and then transfer these units to recognition in the spoken register. This was done through an EEG familiarization paradigm for word segmentation [88,89], which presented infants with a series of familiarization-then-test blocks in separate song and speech sessions. Each block commenced with a familiarization phase of one word embedded eight times within an (unique) eight-phrase song or spoken fragment with the same words. The comparison between ERPs to the first two and last two target word occurrences was taken as an index of recognition in the familiarization phase, and speech and song sessions were compared to assess the potential beneficial or hindering effects of song compared to speech. Then, blocks continued with a spoken test phase consisting of two spoken phrases with the familiarized target word and two other spoken phrases containing a matched novel control word to which infants had not been familiarized. The difference in ERP response to the familiar target and control words would index word recognition in the test phase. For the speech sessions, we sought to replicate segmentation from speech as found in the study by Junge and colleagues [89]; for the song

sessions, this familiarity effect in the test phase could indicate infants' ability to transfer words that are segmented from song to recognition in speech.

The planned analyses of the familiarization phase revealed that infants are able to segment words from songs as well as from speech. Specifically, infants' ERPs showed an increased positivity to the last two compared to the first two target word tokens in both song and speech. This finding shows that infants identify the repetition of word forms within songs with the lyrical and musical complexity of actual children's songs, thus extending previous results on infants' abilities to process syllables in segmentally and musically much simpler songs than those employed here [4–6]. However, in contrast to this previous work, the present findings provide no clear indication that the segmentation response is stronger in song than speech. Factors that may have contributed to the lack of an observed song benefit are discussed below.

The planned analysis of the test phase provided no evidence of segmentation of either song or speech. This means that the status of the elements that children segment from songs is currently unclear. The interpretation of the absence of song-to-speech transfer is complicated by the lack of evidence for segmentation in the speech test phase following the speech familiarization (thus not replicating [89]). This leaves us unable to provide a definitive answer as to whether infants transfer words that are segmented from song to recognition in speech. Infants' difficulties with the test phase will be addressed in more detail below.

In addition to these planned analyses, we also conducted a set of follow-up analyses to better understand the unexpected aspects of the results: the positive-going instead of negative-going response in the familiarization phase and the absence of a group-level effect in the test phase. The follow-up analyses of the familiarization data suggested that the response to target words develops non-linearly over a passage and develops differently for song compared to speech. The response to target words in songs showed an inverted U-shape: the positivity increased incrementally over the first four occurrences; then, it remained relatively stable in occurrences five and six and reduced slightly in amplitude in occurrences seven and eight. The response to target words in speech developed with a sinusoid-like shape: regarding song, the positivity increased over the first four occurrences. However, the amplitude then attenuated in occurrences five and six to increase again in the last two occurrences. This was the only difference observed between infants' responses to song and speech. One could very tentatively consider the inverted U in the development of song responses as a slower version of the first half of the sinusoid development of responses to speech. However, as this analysis was exploratory and the modulation over target occurrences differed from the previously found linear trajectory followed by a plateau at maximum negative amplitude [89], these patterns and thus the potential difference between infants' processing of speech and songs should be cautiously interpreted until replication.

The exploratory follow-up analyses of the test phase data indicate tentative support for a binary split between infants that displayed a positive versus a negative response to the familiar target versus novel control words. Moreover, infants with a more mature negative response in the test phase showed a stronger modulation over the familiarization phase than the infants with a less mature positive response, which is reminiscent of previous suggestions that infants with negative-going responses have more robust neural segmentation responses across various stages of the procedure [70,93]. While this binary split may explain the absence of a group-level effect in the test phase and the possibility of interpreting the modulation in the familiarization phase, these results await replication, considering the exploratory nature of the binary split between infants with negative-going and positive-going responses (however, see precedents in [69–71,93]), as well as the unexpected shape of the modulation of the response over the familiarization phase.

Although the timing and topography of the ERP familiarity effect was comparable to previous reports, the present results differ from previous studies in two critical respects. Firstly, the familiarity response was positive going for both song and speech, instead of the negative-going response that was predicted for the segmentation of these 10-month-old infants [67,79,83,88,92,93]. A positive-going

response as observed in the present study is generally considered less mature (see reviews by [71,83]). In the context of speech segmentation, a positivity is associated with segmentation by younger infants [69,83], with infants' individual poorer concurrent and later language outcomes [67,69,71,93] as well as with stimulus materials that are more difficult to process [69].

A second discrepancy from most of these previous studies is that we found no evidence of segmentation of either speech or song in the test phase. Thus, this (null) finding may qualify claims about the sensitivity of the ERP familiarity response compared to behavioral familiarization-then-test methods [69,80,91,92]. However, the apparent lack of a group-level segmentation response in the present study is in line with the aforementioned recently published results from over 100 9-month-old participants [71] as well as a large-scale failure to replicate speech segmentation in a behavioral task from 8 to 10.5 months [72]. While Kidd and colleagues explain the lack of a group-level response with reference to the large individual differences between infants of roughly the same age and find that the response polarity relates to vocabulary size [71], Floccia and colleagues refer to the acoustic properties of the stimuli [72]. This latter explanation is in keeping with the literature suggesting that the prosodic exaggeration of infant-directed speech facilitates segmentation by infants [85,86]; cf. [87].

The two discrepancies from the literature, a positive-going response and a lack of a segmentation effect in the test phase, could thus be explained with reference to one of two factors: the (linguistic) maturity of the infants, or at least a subgroup of them, and the properties of the stimuli. As we did not collect data about the participants' concurrent or later language outcomes, the relationship between the EEG responses and (linguistic) maturity cannot be further addressed. However, the role of the stimulus properties can be addressed by directly comparing the acoustic features of the present stimulus set to the stimuli from Junge and colleagues [89], who found a negative-going response in the familiarization as well as the test phase. For a valid comparison between the present stimuli and those of [89], we re-analyzed the average pitch and pitch range of the stimuli of Junge and colleagues and added the acoustic focus measure, using the same procedure as reported in the present Methods section. The duration values to quantify speaking rate are taken directly from [89].

This comparison between the stimulus sets will focus on three features that have been suggested as facilitating infant speech processing in the literature. First, an overall higher pitch with more expanded range is often cited as the prime difference between IDS and adult-directed speech (ADS) (for reviews: [7,118], facilitating segmentation on the basis of transitional probabilities [85,86] as well as possibly segmentation from natural speech [72]; c.f., [87]. However, exaggerated pitch properties have not been found to enhance infant word recognition [119]. Second, pitch changes may also provide a specific acoustic focus on the target word. This focus appears essential for segmentation by 6-month-olds, and still aids segmentation by 9-month-olds [83]. Third, a slower speaking rate is often cited as a key difference between IDS and ADS (cf. [120]), which benefits infants' word-processing efficiency [119]. A fast speaking rate is disruptive to infants' segmentation abilities, eliminating a group-level response for 11-month-olds and shifting the behavioral response to a less mature familiarity preference in 14-month-olds [121].

Of the three features scrutinized, overall, pitch characteristics probably do not affect the maturity of the speech segmentation response observed in the present study: The average pitch was in fact higher in the present speech stimuli compared to those used by [89] (234 Hz versus 201 Hz), and the pitch range was highly comparable between the two datasets (10.90 semitones versus 10.28 semitones). Unless a lower-pitched voice is easier to segment, the present data support the conclusion from [119] that overall, pitch characteristics probably do not facilitate infant word processing, extending the conclusion to segmentation. The specific acoustic emphasis on the word could be a contributing factor to the relatively immature speech segmentation response in the present study, as the acoustic focus on the target word was somewhat smaller in the present compared to the [89] stimuli (0.896 versus 0.930, respectively). This suggests that prosodic focus may still aid segmentation at 10 months, extending the age range for which prosodic focus is found to be beneficial from 9 to 10 months [83]. Finally, the speaking rate of the stimuli could (partially) account for the positive-going response.

The speaking rate in our speech stimuli is faster than in [89], as evidenced by shorter target words and phrases (target words: 515 ms versus 694 ms, respectively; phrases: 2068 ms versus 2719, respectively) despite both studies using disyllabic trochees as target words and having a very similar number of words per phrase (5.71 versus 5.75). This comparison agrees with the possible beneficial effect of a slow speaking rate for word recognition [119] and the hindering effects of a fast speaking rate for word segmentation [121].

The suggestion that acoustic prominence and speaking rate facilitate segmentation is also tentatively supported by the properties of the test phase stimuli. Recall that no group-level segmentation effect was observed in the test phase, but that about half of the participants displayed a (mature) negative-going response. Interestingly, the test stimuli were spoken with more acoustic focus on the target words than the familiarization stimuli (0.942 versus 0.896, respectively), which is comparable to the focus in the stimuli from Junge and colleagues ([89]; 0.930). Moreover, the test stimuli were somewhat slower than the familiarization stimuli, although not as slow as the [89] stimuli (target word duration: 538 ms; phrase duration: 2216 ms). In other words, the test stimuli might have been somewhat easier to segment than the familiarization stimuli. This might have facilitated a negative-going response in a larger subset of participants (or trials), possibly resulting in the disappearance of the group-level positive-going effect from the familiarization phase.

The present stimuli may have emphasized the target words occurring in the familiarization phase less than previous studies, because our stimuli were recorded in short stories rather than in individual sentences. Target words may be emphasized less in narratives, as the emphasis on repeated words diminishes over subsequent utterances, even in infant-directed speech [122]. If the recording of narratives and associated reduced prosodic emphasis is indeed responsible for the present study's less mature segmentation response, this raises questions about infants' ability to segment words from the speech that they are presented with during their daily interactions.

However, the acoustic properties of the stimuli do not explain all aspects of the present data, which becomes apparent when the song stimuli are considered. The song stimuli provided more prosodic focus on the target words (0.959) and were lower in speaking rate (target word duration: 794; phrase duration: 3181 ms), even compared to the stimuli of Junge and colleagues [89]. If acoustic focus and/or a slow speaking rate were necessary and sufficient for a mature segmentation response, the song familiarization should have elicited a group-level negative-going response or enabled more children to display this response. Thus, the observed positivity in the song familiarization could reflect additional challenges segmenting words from songs, such as the atypical acoustic signal of sung compared to spoken language [41,42]; cf. [43], or the increased risk of mis-segmentations from songs [44,45,123]. Alternatively, the acoustic focus and speaking rate might not explain the positive-going response in the present speech familiarization either, in which case we need to look to other factors to explain the different result patterns across studies.

One possible deviation in procedure compared to some previous studies (e.g., [88,89]) is that to maintain experimental engagement, besides showing desynchronized visual stimuli on a computer screen, in the present study, not only the parent but also an experimenter sat with the infant and showed silent toys when necessary. Note that other experimenters (e.g., [71], who did not find a group level effect in 9-month-olds) have also used such an entertainer. The additional visual distraction might have resulted in less selective attention to the auditory stimuli for at least some of the children or trials. Note that the allocation and maintenance of attentional focus undergo major developmental changes in infancy, with individual differences herein possibly being related to lexical development [124]. The precise effect of visual distraction (life or on screen) on the ERP word familiarity effect should be established in future studies. If life visual distraction were found to reduce or eliminate a mature negative-going response at the group level, this would raise questions about the effect of selective attention to auditory stimuli on the ERP familiarity effect in general. Is a negative ERP word familiarity effect possibly a reflection of more focused attention to the auditory stimuli, and do children who are better at selectively attending to auditory stimuli then develop language quicker (as in [71])?

As described in the Introduction, the developmental shift from an initial positivity to a later negativity for word recognition responses has been ascribed to cortical maturation as well as to the development of the lexicon [71,83], with various stimulus and task characteristics giving rise to a 'lexical processing mode', resulting in the negative ERP familiarity effect. This repetition enhancement effect might reflect the strengthening of novel neural representations during active learning ([103]; see Introduction). The positive ERP familiarity effect for both speech and song stimuli for the 10-month-olds in the current study suggests that acoustic prominence, speaking rate, and selective auditory attention might be important to get infants into such an active lexical processing mode that is necessary for building a lexicon. Future research is needed to further clarify the development and precise characteristics of the infant ERP familiarity effect and to specify the exact role of lexical processing in its polarity—for example, by actively manipulating the lexical status of stimuli or the processing mode of the infant.

The absence of evidence for better segmentation from song compared to speech, although not the primary focus of the present study, could be considered surprising in light of the previous research finding benefits of songs for language learning in adults and infants [4–6]. Two factors might have contributed to the lack of a song advantage in the present study. First, the present speech stimuli might have been more attractive than those used in the previous infant studies comparing language learning from song and speech. The speech stimuli of previous work were spoken in a completely flat contour [6], adult-directed speech [4], or at a pitch intermediate between infant-directed and adult-directed speech [5]. In contrast, the present stimuli were elicited in an infant-directed register and had an overall pitch that was highly comparable to the average pitch in a sample of Dutch-speaking mothers interacting with their own 15-month-old infant (236/237 Hz and 234 Hz, respectively; [125]). Thus, the lack of a song advantage in the present study could reflect the beneficial effects of infant-directed speech for infant speech segmentation [85,86]; cf. [87]. A second factor contributing to the lack of an observed song advantage might have been the relative complexity of the songs in the present study. The stimuli of previous work all presented consistent syllable–tone pairings, which were combined into one [5], two [4], or four [6] four-syllable melodies that were repeatedly presented to the infants during familiarization. In contrast, the present stimuli consisted of 20 novel melodies, some of which included a melodic repetition in the second four-phrase verse. As melodic repetition facilitates word recall from songs [24], the children's songs in the present study may have been too novel and complex to enhance segmentation. We are looking forward to future research investigating whether song complexity and, in particular, experience with the songs increases the benefit of songs for word learning.

In sum, this study provided electrophysiological evidence that 10-month-old Dutch children can segment words from songs as well as from speech, although the segmentation response was less mature than that observed in previous studies and did not persist into the test phase. Close inspection of the stimuli suggested that a limited prosodic focus on the target words and a relatively fast speaking rate may have inhibited more mature responses to the speech stimuli. However, these same cues were strongly present in the song stimuli, suggesting that other factors may suppress mature segmentation from songs. Future research will need to establish which aspects of songs, including children's familiarity with them, contribute to children's segmentation from speech and song and to what extent children can recognize words segmented from songs in speech.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3425/10/1/39/s1, Table S1: Acoustic Properties of the Experimental Stimuli, Table S2: Pre-Test on the Melodies, Table S3: Full Set of Stimulus Materials, Table S4: Additional Analyses to the Follow-up Analyses.

Author Contributions: Conceptualization, T.M.S., T.B., and P.F.; methodology, T.M.S., T.B., and P.F.; software, T.M.S. and T.B.; validation, T.M.S. and T.B.; formal analysis, T.M.S. and T.B.; investigation, T.M.S., and T.B.; resources, T.M.S., T.B., and P.F.; data curation, T.M.S. and T.B.; writing—original draft preparation, T.B. and T.M.S.; writing—review and editing, T.B., T.M.S., and P.F.; visualization, T.M.S. and T.B.; supervision, T.B., T.M.S., and P.F.; project administration, T.M.S.; funding acquisition, P.F. and T.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NWO (Nederlandse Organisatie for Wetenschappelijk Onderzoek), grant number 275-89-023 (VENI) to T.M.S.

Acknowledgments: We thank all the lab assistants for their help with the pretest and with stimulus creation, particularly Catheleyne Creusen, Esther Kroese, Melissa van Wijk, and Lisa Rommers. We thank Annelies van Wijngaarden for lending her voice to the speech and song stimuli. Many thanks to Renske van der Cruijsen for scheduling and testing most of the infant EEG sessions, with the assistance of Laura Hahn and several lab rotation students. We are greatly indebted to all infants and parents who participated in the experiment. Thanks to Laura Hahn and other members of the FLA-group of Radboud University for commenting on earlier drafts of the paper, and to George Rowland for proofreading. We thank Caroline Junge for generously sharing her stimuli for the further acoustic analyses reported in the discussion of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Trehub, S.E.; Trainor, L. Singing to infants: Lullabies and play songs. In *Advances in Infancy Research*; Rovee-Collier, D., Lipsitt, L., Hayne, H., Eds.; Ablex: Stamford, CT, USA, 1998; Volume 12, pp. 43–77.
2. Custodero, L.A.; Johnson-Green, E.A. Caregiving in counterpoint: Reciprocal influences in the musical parenting of younger and older infants. *Early Child Dev. Care* **2008**, *178*, 15–39. [CrossRef]
3. Adachi, M.; Trehub, S.E. Musical lives of infants. In *Music Learning and Teaching in Infancy, Childhood, and Adolescence. An Oxford Handbook of Music Education*; McPherson, G., Welch, G., Eds.; Oxford University Press: New York, NY, USA, 2012; Volume 2, pp. 229–247.
4. Thiessen, E.D.; Saffran, J.R. How the melody facilitates the message and vice versa in infant learning and memory. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 225–233. [CrossRef] [PubMed]
5. Lebedeva, G.C.; Kuhl, P.K. Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behav. Dev.* **2010**, *33*, 419–430. [CrossRef] [PubMed]
6. François, C.; Teixidó, M.; Takerkart, S.; Agut, T.; Bosch, L.; Rodriguez-Fornells, A. Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. *Sci. Rep.* **2017**, *7*, 12451. [CrossRef] [PubMed]
7. Cristia, A. Input to language: The phonetics and perception of infant-directed speech. *Lang. Linguist. Compass* **2013**, *7*, 157–170. [CrossRef]
8. Golinkoff, R.M.; Can, D.D.; Soderstrom, M.; Hirsh-Pasek, K. (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Curr. Dir. Psychol. Sci.* **2015**, *24*, 339–344. [CrossRef]
9. Klahr, D.; Chase, W.G.; Lovelace, E.A. Structure and process in alphabetic retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* **1983**, *9*, 462–477. [CrossRef]
10. Ludke, K.M.; Ferreira, F.; Overy, K. Singing can facilitate foreign language learning. *Mem. Cogn.* **2014**, *42*, 41–52. [CrossRef]
11. Calvert, S.L.; Billingsley, R.L. Young children's recitation and comprehension of information presented by songs. *J. Appl. Dev. Psychol.* **1998**, *19*, 97–108. [CrossRef]
12. Engh, D. Why use music in English language learning? A survey of the literature. *Engl. Lang. Teach.* **2013**, *6*, 113–127. [CrossRef]
13. Davis, G.M. Songs in the young learner classroom: A critical review of evidence. *Engl. Lang. Teach.* **2017**, *71*, 445–455. [CrossRef]
14. Benz, S.; Sellaro, R.; Hommel, B.; Colzato, L.S. Music makes the world go round: The impact of musical training on non-musical cognitive functions—A review. *Front. Psychol.* **2016**, *6*, 2023. [CrossRef]
15. Francois, C.; Chobert, J.; Besson, M.; Schon, D. Music training for the development of speech segmentation. *Cereb. Cortex* **2013**, *23*, 2038–2043. [CrossRef] [PubMed]
16. Bhide, A.; Power, A.; Goswami, U. A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind Brain Educ.* **2013**, *7*, 113–123. [CrossRef]
17. Thompson, W.F.; Schellenberg, E.G.; Husain, G. Decoding speech prosody: Do music lessons help? *Emotion* **2004**, *4*, 46–64. [CrossRef] [PubMed]
18. Chobert, J.; Francois, C.; Velay, J.-L.; Besson, M. Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* **2014**, *24*, 956–967. [CrossRef] [PubMed]

19. Besson, M.; Chobert, J.; Marie, C. Transfer of training between music and speech: Common processing, attention, and memory. *Front. Psychol.* **2011**, *2*, 94. [CrossRef]
20. Strait, D.; Kraus, N. Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music Percept. Interdiscip. J.* **2011**, *29*, 133–146. [CrossRef]
21. Patel, A.D. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* **2011**, *2*, 142. [CrossRef]
22. Patel, A.D. The OPERA hypothesis: Assumptions and clarifications. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 124–128. [CrossRef]
23. Patel, A.D. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* **2014**, *308*, 98–108. [CrossRef] [PubMed]
24. Wallace, W.T. Memory for music: Effect of melody on recall of text. *J. Exp. Psychol. Learn. Mem. Cogn.* **1994**, *20*, 1471. [CrossRef]
25. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
26. Schroeder, C.E.; Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **2009**, *32*, 9–18. [CrossRef] [PubMed]
27. Thaut, M.H. Temporal entrainment of cognitive functions: Musical mnemonics induce brain plasticity and oscillatory synchrony in neural networks underlying memory. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 243–254. [CrossRef] [PubMed]
28. Doelling, K.B.; Poeppel, D. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6233–E6242. [CrossRef]
29. Myers, B.R.; Lense, M.D.; Gordon, R.L. Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain Sci.* **2019**, *9*, 70. [CrossRef]
30. Corbeil, M.; Trehub, S.; Peretz, I. Speech vs. singing: Infants choose happier sounds. *Front. Psychol.* **2013**, *4*, 372. [CrossRef]
31. Costa-Giomi, E.; Ilari, B. Infants' preferential attention to sung and spoken stimuli. *J. Res. Music Educ.* **2014**, *62*, 188–194. [CrossRef]
32. Trehub, S.E.; Plantinga, J.; Russo, F.A. Maternal vocal interactions with infants: Reciprocal visual influences. *Soc. Dev.* **2016**, *25*, 665–683. [CrossRef]
33. Nakata, T.; Trehub, S.E. Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* **2004**, *27*, 455–464. [CrossRef]
34. Tsang, C.D.; Falk, S.; Hessel, A. Infants prefer infant-directed song over speech. *Child. Dev.* **2017**, *88*, 1207–1215. [CrossRef] [PubMed]
35. Corbeil, M.; Trehub, S.E.; Peretz, I. Singing delays the onset of infant distress. *Infancy* **2016**, *21*, 373–391. [CrossRef]
36. Trehub, S.E.; Ghazban, N.; Corbeil, M. Musical affect regulation in infancy: Musical affect regulation in infancy. *Ann. N. Y. Acad. Sci.* **2015**, *1337*, 186–192. [CrossRef]
37. Nazzi, T.; Nelson, D.G.K.; Jusczyk, P.W.; Jusczyk, A.M. Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy* **2000**, *1*, 123–147. [CrossRef]
38. Johnson, E.K.; Seidl, A. Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy* **2008**, *13*, 440–455. [CrossRef]
39. Nazzi, T.; Bertoncini, J.; Mehler, J. Language discrimination by newborns: Toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* **1998**, *24*, 756. [CrossRef]
40. Ramus, F. Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annu. Rev. Lang. Acquis.* **2002**, *2*, 85–115. [CrossRef]
41. Falk, S.; Maslow, E.; Thum, G.; Hoole, P. Temporal variability in sung productions of adolescents who stutter. *J. Commun. Disord.* **2016**, *62*, 101–114. [CrossRef]
42. Evan, D. Bradley A comparison of the acoustic vowel spaces of speech and song. *Linguist. Res.* **2018**, *35*, 381–394. [CrossRef]
43. Audibert, N.; Falk, S. Vowel space and f0 characteristics of infant-directed singing and speech. In Proceedings of the 9th International Conference on Speech Prosody, Poznań, Poland, 13–16 June 2018; pp. 153–157.
44. Wright, S. The Death of Lady Mondegreen. *Harper's Mag.* **1954**, *209*, 48–51.

45. Otake, T. Interlingual near homophonic words and phrases in L2 listening: Evidence from misheard song lyrics. In Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007), Saarbrücken, Germany, 6–10 August 2007; pp. 777–780.
46. Kentner, G. Rhythmic segmentation in auditory illusions—Evidence from cross-linguistic mondegreens. In Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK, 10–14 August 2015.
47. Houston, D.M.; Jusczyk, P.W. The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percept. Perform.* **2000**, *26*, 1570–1582. [CrossRef] [PubMed]
48. Singh, L.; Morgan, J.L.; White, K.S. Preference and processing: The role of speech affect in early spoken word recognition. *J. Mem. Lang.* **2004**, *51*, 173–189. [CrossRef]
49. Schmale, R.; Seidl, A. Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Dev. Sci.* **2009**, *12*, 583–601. [CrossRef]
50. Schmale, R.; Cristià, A.; Seidl, A.; Johnson, E.K. Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy* **2010**, *15*, 650–662. [CrossRef]
51. Bergelson, E.; Swingley, D. Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Dev.* **2018**, *89*, 1567–1576. [CrossRef]
52. Schön, D.; Boyer, M.; Moreno, S.; Besson, M.; Peretz, I.; Kolinsky, R. Songs as an aid for language acquisition. *Cognition* **2008**, *106*, 975–983. [CrossRef]
53. Francois, C.; Schön, D. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* **2011**, *21*, 2357–2365. [CrossRef]
54. François, C.; Tillmann, B.; Schön, D. Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 108–115. [CrossRef]
55. François, C.; Jaillet, F.; Takerkar, S.; Schön, D. Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE* **2014**, *9*, e101340. [CrossRef]
56. Shook, A.; Marian, V.; Bartolotti, J.; Schroeder, S.R. Musical experience influences statistical learning of a novel language. *Am. J. Psychol.* **2013**, *126*, 95. [CrossRef] [PubMed]
57. Larrouy-Maestri, P.; Leybaert, J.; Kolinsky, R. The benefit of musical and linguistic expertise on language acquisition in sung material. *Musicae Sci.* **2013**, *17*, 217–228. [CrossRef]
58. Morgan, J.L. Prosody and the roots of parsing. *Lang. Cogn. Process.* **1996**, *11*, 69–106. [CrossRef]
59. Brent, M.R.; Siskind, J.M. The role of exposure to isolated words in early vocabulary development. *Cognition* **2001**, *81*, 33–44. [CrossRef]
60. Aslin, R.N.; Woodward, J.Z.; LaMendola, N.P. Models of word segmentation in fluent maternal speech to infants. In *Signal to Syntax*; Psychology Press: New York, NY, USA; London, UK, 1996; pp. 117–134.
61. Johnson, E.K.; Lahey, M.; Ernestus, M.; Cutler, A. A multimodal corpus of speech to infant and adult listeners. *J. Acoust. Soc. Am.* **2013**, *134*, EL534–EL540. [CrossRef] [PubMed]
62. Graf Estes, K.; Evans, J.L.; Alibali, M.W.; Saffran, J.R. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* **2007**, *18*, 254–260. [CrossRef]
63. Swingley, D. Lexical exposure and word-form encoding in 1.5-year-olds. *Dev. Psychol.* **2007**, *43*, 454–464. [CrossRef]
64. Lany, J.; Saffran, J.R. From statistics to meaning: Infants' acquisition of lexical categories. *Psychol. Sci.* **2010**, *21*, 284–291. [CrossRef]
65. Hay, J.F.; Pelucchi, B.; Estes, K.G.; Saffran, J.R. Linking sounds to meanings: Infant statistical learning in a natural language. *Cogn. Psychol.* **2011**, *63*, 93–106. [CrossRef]
66. Newman, R.; Bersntein Ratner, N.; Jusczyk, A.M.; Jusczyk, P.W.; Dow, K.A. Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Dev. Psychol.* **2006**, *42*, 643–655. [CrossRef]
67. Junge, C.; Kooijman, V.; Hagoort, P.; Cutler, A. Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* **2012**, *15*, 463–473. [CrossRef] [PubMed]
68. Singh, L.; Steven Reznick, J.; Xuehua, L. Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Dev. Sci.* **2012**, *15*, 482–495. [CrossRef] [PubMed]
69. Kooijman, V.; Junge, C.; Johnson, E.K.; Hagoort, P.; Cutler, A. Predictive brain signals of linguistic development. *Front. Psychol.* **2013**, *4*, 25. [CrossRef] [PubMed]
70. Junge, C.; Cutler, A. Early word recognition and later language skills. *Brain Sci.* **2014**, *4*, 532–559. [CrossRef] [PubMed]

71. Kidd, E.; Junge, C.; Spokes, T.; Morrison, L.; Cutler, A. Individual differences in infant speech segmentation: Achieving the lexical shift. *Infancy* **2018**, *23*, 770–794. [CrossRef]
72. Floccia, C.; Keren-Portnoy, T.; DePaolis, R.; Duffy, H.; Delle Luche, C.; Durrant, S.; White, L.; Goslin, J.; Vihman, M. British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition* **2016**, *148*, 1–9. [CrossRef]
73. Bergmann, C.; Cristia, A. Development of infants' segmentation of words from native speech: A meta-analytic approach. *Dev. Sci.* **2016**, *19*, 901–917. [CrossRef]
74. Cutler, A. Segmentation problems, rhythmic solutions. *Lingua* **1994**, *92*, 81–104. [CrossRef]
75. Jusczyk, P.W.; Houston, D.M.; Newsome, M. The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* **1999**, *39*, 159–207. [CrossRef]
76. Houston, D.M.; Jusczyk, P.W.; Kuijpers, C.; Coolen, R.; Cutler, A. Cross-language word segmentation by 9-month-olds. *Psychon. Bull. Rev.* **2000**, *7*, 504–509. [CrossRef]
77. Johnson, E.K.; Jusczyk, P.W. Word segmentation by 8-month-olds: When speech cues count more than statistics. *J. Mem. Lang.* **2001**, *44*, 548–567. [CrossRef]
78. Curtin, S.; Mintz, T.H.; Christiansen, M.H. Stress changes the representational landscape: Evidence from word segmentation. *Cognition* **2005**, *96*, 233–262. [CrossRef] [PubMed]
79. Kooijman, V.K.; Hagoort, P.; Cutler, A. Prosodic structure in early word segmentation: ERP evidence from Dutch 10-month-olds. *Infancy* **2009**, *14*, 591–612. [CrossRef]
80. Nazzi, T.; Iakimova, G.; Bertoncini, J.; Frédonie, S.; Alcantara, C. Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *J. Mem. Lang.* **2006**, *54*, 283–299. [CrossRef]
81. Nazzi, T.; Mersad, K.; Sundara, M.; Iakimova, G.; Polka, L. Early word segmentation in infants acquiring Parisian French: Task-dependent and dialect-specific aspects. *J. Child Lang.* **2014**, *41*, 600–633. [CrossRef]
82. Polka, L.; Sundara, M. Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy* **2012**, *17*, 198–232. [CrossRef]
83. Männel, C.; Friederici, A.D. Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex* **2013**, *49*, 2788–2798. [CrossRef]
84. Zahner, K.; Schönhuber, M.; Braun, B. The limits of metrical segmentation: Intonation modulates infants' extraction of embedded trochees. *J. Child Lang.* **2016**, *43*, 1338–1364. [CrossRef]
85. Thiessen, E.D.; Hill, E.A.; Saffran, J.R. Infant-directed speech facilitates word segmentation. *Infancy* **2005**, *7*, 53–71. [CrossRef]
86. Bosseler, A.N.; Teinonen, T.; Tervaniemi, M.; Huotilainen, M. Infant directed speech enhances statistical learning in newborn infants: An ERP study. *PLoS ONE* **2016**, *11*, e0162177. [CrossRef]
87. Schreiner, M.S.; Altvater-Mackensen, N.; Mani, N. Early word segmentation in naturalistic environments: Limited effects of speech register. *Infancy* **2016**, *21*, 625–647. [CrossRef]
88. Kooijman, V.; Hagoort, P.; Cutler, A. Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cogn. Brain Res.* **2005**, *24*, 109–116. [CrossRef] [PubMed]
89. Junge, C.; Cutler, A.; Hagoort, P. Successful word recognition by 10-month-olds given continuous speech both at initial exposure and test. *Infancy* **2014**, *19*, 179–193. [CrossRef]
90. Jusczyk, P.W.; Hohne, E.A. Infants' memory for spoken words. *Science* **1997**, *277*, 1984–1986. [CrossRef]
91. Kuijpers, C.T.L.; Coolen, R.; Houston, D.; Cutler, A. Using the head-turning technique to explore cross-linguistic performance differences. In *Advances in Infancy Research*; Rovee-Collier, D., Lipsitt, L., Hayne, H., Eds.; Ablex: Stamford, CT, USA, 1998; Volume 12, pp. 205–220.
92. Goyet, L.; de Schonen, S.; Nazzi, T. Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Res.* **2010**, *1332*, 75–89. [CrossRef]
93. Von Holzen, K.; Nishibayashi, L.-L.; Nazzi, T. Consonant and vowel processing in word form segmentation: An infant ERP study. *Brain Sci.* **2018**, *8*, 24. [CrossRef]
94. Trainor, L.; McFadden, M.; Hodgson, L.; Darragh, L.; Barlow, J.; Matsos, L.; Sonnadara, R. Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *Int. J. Psychophysiol.* **2003**, *51*, 5–15. [CrossRef]
95. Moore, J.K.; Guan, Y.-L. Cytoarchitectural and axonal maturation in human auditory cortex. *J. Assoc. Res. Otolaryngol.* **2001**, *2*, 297–311. [CrossRef]

96. Moore, J.K.; Linthicum, F.H. The human auditory system: A timeline of development. *Int. J. Audiol.* **2007**, *46*, 460–478. [CrossRef]
97. Eggermont, J.J.; Moore, J.K. Morphological and functional development of the auditory nervous system. In *Human Auditory Development*; Werner, L., Fay, R.R., Popper, A.N., Eds.; Springer Handbook of Auditory Research; Springer: New York, NY, USA, 2012; pp. 61–105.
98. Luck, S.J. Ten simple rules for designing and interpreting ERP experiments. In *Event-related Potentials: A Methods Handbook*; Handy, T.C., Ed.; MIT Press: Cambridge, MA, USA, 2005; pp. 17–32.
99. Snijders, T.M.; Kooijman, V.; Cutler, A.; Hagoort, P. Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Res.* **2007**, *1178*, 106–113. [CrossRef]
100. Grill-Spector, K.; Henson, R.; Martin, A. Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn. Sci.* **2006**, *10*, 14–23. [CrossRef] [PubMed]
101. Henson, R.N.A.; Rugg, M.D. Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* **2003**, *41*, 263–270. [CrossRef]
102. Henson, R.; Shallice, T.; Dolan, R. Neuroimaging evidence for dissociable forms of repetition priming. *Science* **2000**, *287*, 1269–1272. [CrossRef] [PubMed]
103. Weber, K.; Christiansen, M.H.; Petersson, K.M.; Indefrey, P.; Hagoort, P. fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *J. Neurosci.* **2016**, *36*, 6872–6880. [CrossRef]
104. Segaert, K.; Weber, K.; de Lange, F.P.; Petersson, K.M.; Hagoort, P. The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia* **2013**, *51*, 59–66. [CrossRef]
105. Baayen, R.H.; Piepenbrock, R.; Gulikers, L. *The CELEX Lexical Database (CD-ROM)*; Linguistic Data Consortium, University of Pennsylvania: Philadelphia, PA, USA, 1993.
106. Boersma, P.; Weenink, D. *Praat: Doing Phonetics by Computer*, Version 5.3.45; Available online: http://www.fon.hum.uva.nl/praat/ (accessed on 22 April 2013).
107. Neurobehavioral Systems. Available online: https://www.neurobs.com (accessed on 2 February 2015).
108. Oostenveld, R.; Fries, P.; Maris, E.; Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 156869. [CrossRef]
109. Makeig, S.; Bell, A.J.; Jung, T.-P.; Sejnowski, T.J. Independent component analysis of electroencephalographic data. In *Advances in Neural Information Processing Systems 8*; Touretzky, D., Mozer, M., Hasselmo, M., Eds.; MIT Press: Cambridge, MA, USA, 1996; pp. 145–151.
110. Bell, A.J.; Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **1995**, *7*, 1129–1159. [CrossRef]
111. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef]
112. Perrin, F.; Pernier, J.; Bertrand, O.; Echallier, J.F. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* **1989**, *72*, 184–187. [CrossRef]
113. Maris, E.; Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **2007**, *164*, 177–190. [CrossRef]
114. Sassenhagen, J.; Draschkow, D. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* **2019**, *56*, e13335. [CrossRef] [PubMed]
115. Bates, D.; Maechler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
116. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
117. Bates, D.; Kliegl, R.; Vasishth, S.; Baayen, H. Parsimonious mixed models. *arXiv* **2015**, arXiv:1506.04967.
118. Soderstrom, M. Beyond Babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* **2007**, *27*, 501–532. [CrossRef]
119. Song, J.Y.; Demuth, K.; Morgan, J. Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* **2010**, *128*, 389–400. [CrossRef]
120. Martin, A.; Igarashi, Y.; Jincho, N.; Mazuka, R. Utterances in infant-directed speech are shorter, not slower. *Cognition* **2016**, *156*, 52–59. [CrossRef]
121. Wang, Y.; Llanos, F.; Seidl, A. Infants adapt to speaking rate differences in word segmentation. *J. Acoust. Soc. Am.* **2017**, *141*, 2569–2578. [CrossRef]

122. Fernald, A.; Mazzie, C. Prosody and focus in speech to infants and adults. *Dev. Psychol.* **1991**, *27*, 209–221. [CrossRef]
123. Beck, C.; Kardatzki, B.; Ethofer, T. Mondegreens and soramimi as a method to induce misperceptions of speech content - Influence of familiarity, wittiness, and language competence. *PLoS ONE* **2014**, *9*, e84667. [CrossRef]
124. Gomes, H. The development of auditory attention in children. *Front. Biosci.* **2000**, *5*, d108. [CrossRef]
125. Benders, T. Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behav. Dev.* **2013**, *36*, 847–862. [CrossRef]

brain sciences

MDPI

Article

How the Brain Understands Spoken and Sung Sentences

Sonja Rossi [1,*], Manfred F. Gugler [2], Markus Rungger [3], Oliver Galvan [3], Patrick G. Zorowka [3] and Josef Seebacher [3]

[1] ICONE-Innsbruck Cognitive Neuroscience, Department for Hearing, Speech, and Voice Disorders, Medical University of Innsbruck, 6020 Innsbruck, Austria
[2] Department for Medical Psychology, Medical University of Innsbruck, 6020 Innsbruck, Austria; manfred.gugler@tirol-kliniken.at
[3] Department for Hearing, Speech, and Voice Disorders, Medical University of Innsbruck, 6020 Innsbruck, Austria; markus.rungger@i-med.ac.at (M.R.); oliver.galvan@tirol-kliniken.at (O.G.); patrick.zorowka@i-med.ac.at (P.G.Z.); josef.seebacher@i-med.ac.at (J.S.)
* Correspondence: sonja.rossi@i-med.ac.at; Tel.: +43-512-504-26129

Received: 29 November 2019; Accepted: 6 January 2020; Published: 8 January 2020

Abstract: The present study investigates whether meaning is similarly extracted from spoken and sung sentences. For this purpose, subjects listened to semantically correct and incorrect sentences while performing a correctness judgement task. In order to examine underlying neural mechanisms, a multi-methodological approach was chosen combining two neuroscientific methods with behavioral data. In particular, fast dynamic changes reflected in the semantically associated N400 component of the electroencephalography (EEG) were simultaneously assessed with the topographically more fine-grained vascular signals acquired by the functional near-infrared spectroscopy (fNIRS). EEG results revealed a larger N400 for incorrect compared to correct sentences in both spoken and sung sentences. However, the N400 was delayed for sung sentences, potentially due to the longer sentence duration. fNIRS results revealed larger activations for spoken compared to sung sentences irrespective of semantic correctness at predominantly left-hemispheric areas, potentially suggesting a greater familiarity with spoken material. Furthermore, the fNIRS revealed a widespread activation for correct compared to incorrect sentences irrespective of modality, potentially indicating a successful processing of sentence meaning. The combined results indicate similar semantic processing in speech and song.

Keywords: semantics; speech comprehension; singing; N400; event-related brain potentials (ERPs); functional near-infrared spectroscopy (fNIRS)

1. Introduction

Speech communication is a unique human ability. However, also listening to and playing music is only present in human people. Both speech and music include productive and perceptual aspects. In the present study, we will focus on perceptual abilities. Language as well as music processing can be partitioned into several sub-abilities such as the identification of single sounds, syntactic-combinatorial rule extraction, melodic perception, and meaning extraction [1]. We will put the emphasis on the perception of linguistic meaning in speech and song. Singing is a form of music which also carries direct semantic meaning as in spoken language but with additional melody. Speech and song differ with respect to several aspects: songs display a more precise articulation and longer vowel duration than speech [2,3]. Furthermore, pitch is altered in song exhibiting a more discrete F0 contour and a fine-grained accurate pitch processing compared to speech [4]. Singing is an important evolutionary phenomenon, as early humans already used a protolanguage similar to singing [5], and still nowadays

the parent–children interaction is characterized by singing which supports bonding [6,7]. We opted for investigating this kind of music, as hearing impaired patients show more difficulties in extracting meaning from sung sentences (e.g., [8]). In a currently ongoing study in our lab, we aim at better understanding the neural processing of speech comprehension in different groups of hearing-impaired patients. We are particularly interested in whether they show similar or altered neural mechanisms to semantic processing. Before being able to interpret pathological data, it is however important to clearly understand processing mechanisms in healthy subjects. An overall musical training and/or singing in particular were found to positively impact semantic processing [9], foreign language learning in general [10,11], and perception of speech in noise [12–14]. Furthermore, music is beneficial for language production abilities during rehabilitation of language disorders such as aphasia [15–19]. Furthermore, deaf children supplied with a cochlear implant were found to benefit from musical training as they improve auditory abilities as well as language production and perception [20–22]. We are primarily interested in neural mechanisms as a direct measure of semantic processing and will compare these to behavioral performance during a correctness judgement task. Because different neuroscientific methods assess different neural signals of the brain, results may lead to different modulations and conclusions. Hence, we opt for a multi-methodological approach in which we simultaneously apply the electroencephalography (EEG) and the functional near-infrared spectroscopy (fNIRS). EEG, and—in particular—the investigation of event-related brain potentials (ERPs), assesses electrical signals from the scalp and bears the potential to assess fast dynamic processing mechanisms in the range of tens of milliseconds. The topographical resolution is only rough in EEG but in order to assess this information the fNIRS method represents an ideal candidate. fNIRS is an optical method assessing the vascular response by means of near-infrared light (for a review on fNIRS see [23]). Even though this response proceeds on a much larger timescale than EEG it allows for reliably identifying involved brain areas. It can measure brain responses from about 3 cm depth from the scalp. Through this, only cortical regions can be reached in adult participants. The combination of these two methods is perfectly suitable for investigation of auditory stimuli as (1) they are both soundless in contrast to the application of functional magnetic resonance imaging (fMRI), which produces loud noise during data acquisition; (2) they do not interfere with each other compared to EEG-fMRI; (3) they allow a comfortable measuring setting while subjects are seated in a chair instead of lying in an MRI scanner.

1.1. Electrophysiological Correlates of Semantic Processing in Speech and Song

In language comprehension research, semantic processing was investigated through several experimental designs. One of these is the priming design. During this paradigm, a prime followed by a target stimulus is presented. Usually, the target stimulus is semantically related or unrelated to the preceding prime. An electrophysiological correlate of semantics elicited in priming paradigms is the N400 component. The amplitude of the N400 reduces with repetition of stimuli and was thus found to be enhanced for unrelated targets (e.g., [24–27]). This centro-parietal ERP component reflects the degree of semantic relatedness between prime and target, and is thus an index of semantic processing (for a review see [28]). A similar paradigm was also adopted for investigating meaning in music. Some studies addressed the question whether instrumental music without lyrics can convey extra-musical meaning such as iconic, indexical/emotional, or symbolic/cultural meaning or intra-musical meaning (i.e., structure of musical elements) (please refer to reviews by [29,30]). As primes, musical excerpts without lyrics [31], single chords [32], or single tones [33] were used followed by a semantically related or unrelated target word. A similar N400 modulation (i.e., larger amplitude for unrelated targets) as for speech was also found in this musical context suggesting shared mechanisms of semantic priming in speech and music. Electrophysiological studies specifically investigating sung material are scarce. However, semantic processing in songs also elicited larger N400s for target words unrelated to the final word of familiar and unfamiliar pop song excerpts compared to related words [34] or when sung target words differed from sung prime words compared to the repetition of the same words [35].

Another important paradigm suitable for investigating perception of semantic processing is to integrate selection restriction errors in sentences and compare them to semantically correct sentences [36,37]. Such a design can be adopted in both spoken and sung sentences or phrases. Similar to the semantic priming paradigm, also in this experimental context, the N400 component in the EEG reliably indexes the integration of semantic information. The amplitude was found to be larger for semantically incorrect compared to correct sentences [36,38,39]. To our knowledge, there is only one electrophysiological study integrating semantic errors in sung musical excerpts [40]. In this study, familiar excerpts from French operas which were sung a cappella were presented to professional opera musicians. Songs were manipulated in such a way that the original version was contrasted with a version containing either semantically incorrect final words or melodic incongruities. Semantically incorrect words elicited a larger N400 component compared to correct words in these sung excerpts. This finding is in line with a magnetoencephalographic study in professional singers and actors using spoken and sung excerpts from Franz Schubert in which, above all, the final word was either semantically correct or incorrect [41].

Electrophysiological findings seem to suggest that semantic processing in speech and song is supported by similar processing mechanisms (for a review please refer to [42]). However, there is no study so far which directly compares electrophysiological processes in relation to selection restriction errors in spoken and sung sentences.

1.2. Brain Regions Supporting Semantic Processing in Speech and Song

Several models tried to assign different aspects of speech and music to the two hemispheres of the brain. Zatorre and colleagues postulated that auditory cortices of both hemispheres are specialized for different auditory analyses whereby the left temporal area is more sensitive for fine-grained temporal analyses and the right temporal area reacts more to spectral variations [43,44]. This difference led to the conclusion that speech is processed predominantly by left and music by right temporal areas [43]. The multi-time resolution hypothesis proposed by Poeppel and colleagues postulates a dichotomy of the left and right auditory cortices based on temporal variations contained in speech and music [45]. Fast auditory transitions are assumed to be processed bilaterally while slow transitions predominantly recruit right temporal areas. Such a hemispheric specialization is already visible in newborn infants when confronted with auditory stimuli with varying temporal modulations [46]. These models predominantly focus on the auditory cortex. The Dynamic Dual Pathway model [47], in contrast, differentiates between different linguistic functions and allocates them to cortical regions of the two hemispheres. The model postulates that segmental information such as phonology, syntax, and semantics are predominantly processed by a fronto-temporal network in the left hemisphere while prosody is located primarily in homologous right-hemispheric areas.

When focusing on semantic processing in particular, a ventral stream including the superior and middle portions of the temporal lobe was proposed [48]. This stream seems to be bilaterally distributed with a weak left-hemispheric dominance. A more or less dominant lateralization usually depends on the linguistic or musical aspects contrasted with each other.

Brain regions activated by priming paradigms—and thus supporting semantic relatedness, access to the lexical storage, and semantic selection—were found to be located predominantly in temporal (particularly the middle temporal gyrus (MTG) and the superior temporal sulcus (STS)), and frontal regions (especially the inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC)) in speech (e.g., [49–51]) but also in music [32]. Using unfamiliar songs which were repeated either with the same or different lyrics or with the same or different tunes, non-musicians showed a larger left-hemispheric activation in anterior STS for lyrics in contrast to tunes suggesting a greater autonomy of linguistic meaning probably because participants could rely more on their linguistic than musical expertise [52]. It should be noted that this study did not introduce any experimental task, thus subjects simply passively listened to the same/different repetition. In contrast, Schön and colleagues [53] presented pairs of spoken, vocalized (i.e., sung without words), or sung words and asked subjects

to explicitly judge whether word pairs were the same or different. The authors found similar brain regions being activated in spoken, vocalized, and sung processing compared to a noise stimulus. Differences arose at a quantitative rather that qualitative level. A larger activation in the left IFG was found for sung compared to vocalized words as they contained real words, thus more linguistically relevant features. Temporal areas (MTG and the superior temporal gyrus (STG)), on the contrary, were found for both linguistic and non-linguistic features, leading to the conclusion that these regions are recruited domain-independently.

Some neuroimaging studies compared different degrees of melodic information together with speech/lyrics and found differences especially in hemispheric lateralization. Merrill and colleagues [54] presented six different sentence categories: spoken, sung, with hummed (i.e., including prosodic pitch) or song melody (i.e., including melodic pitch), and with speech or musical rhythm. While activations were similar in bilateral temporal lobes, differences were present with respect to the inferior frontal areas. Sung sentences elicited increased activations in the right IFG similar to melodic pitch. Prosodic pitch, on the contrary, gave rise to activations predominantly in the left IFG. These findings fit with results obtained in a study [55] contrasting linguistic prosody (i.e., whether the phrase was spoken as a statement or question) to speech recognition (i.e., identifying whether a word was the same or different relative to the previous word). In this contrast, a larger recruitment of a right-hemispheric temporo-frontal network was found for linguistic prosody because of a stronger reliance on prosodic, thus melodic, aspects. The degree of linguistic or melodic features contained in the presented acoustic material seems relevant for a correct interpretation of found activations. In this vein of reasoning, a direct comparison between the listening to/production of spoken and sung material (e.g., familiar songs, words, phrases) showed an increased right-hemispheric dominance of the middle STG, the planum temporale (PT), and the OFC for sung compared to spoken songs [56–58]. The authors interpret the PT to be involved in the transformation of auditory input into motor representation relevant for speech production. The OFC is assumed to process pleasant and unpleasant emotional aspects during music perception. Interestingly, when comparing sung (i.e., linguistic and melodic information) as well as instrumental music (i.e., only melodic information) to spoken songs (i.e., only linguistic information), an increased activation was found not only in the right planum temporale but also in bilateral anterior planum polare, suggesting that these regions encode music/timbre in both instrumental and sung music [57]. Furthermore, spoken and sung material activated the STS bilaterally, indicating that this area is sensitive to human nonlinguistic vocalizations.

Brain regions subserving semantic processing at the sentential level in speech similarly include activations in left or bilateral temporal (particularly in STG and MTG), left or right frontal areas, and sometimes left parietal areas (i.e., angular gyrus) [59–62]. Bilateral temporal areas are assumed to reflect the semantic integration or semantic evaluation of the sentence, however some fMRI studies found increased activations in this region for correct compared to incorrect sentences [59] while others revealed a reversed activation pattern [62]. Frontal regions were found to be associated with semantic selection processes [61,62] whereas left temporal and temporo-parietal areas were also discussed to be involved in the integration of different syntactic and semantic information in a sentence [59,63]. While such a paradigm which integrates semantic errors in sentences was successfully applied in language research, to date no neuroimaging study used this paradigm for investigating semantic processing in songs. In the present study we opted for integrating semantic anomalies in sentences which were either spoken or sung in order to directly compare the underlying neural processing mechanisms.

1.3. The Present Study

The focus of the present study lies on semantic processing in speech and song. Even though several studies examined semantics by means of a priming design, it is not well understood how melodic aspects contribute to the extraction of meaning from semantically correct and incorrect sentences. Thus, we created a set of semantically correct and incorrect sentences which were either spoken or

sung. While subjects listened to these sentences and performed a correctness judgement task, neural processing was assessed via the simultaneous application of the electroencephalography (EEG) and the functional near-infrared spectroscopy (fNIRS). This multi-methodological approach was chosen for several reasons: (1) only one electrophysiological study [40] so far investigated semantic errors in sung sentences in professional musicians, but no direct comparison with spoken sentences in the same non-musically trained subjects was performed until now, (2) no neuroimaging study so far directly investigated semantic errors in sung sentences, and (3) the use of fNIRS in contrast to fMRI, especially, is very advantageous as this method is completely silent without any scanner noise, and is thus suitable for measuring acoustic stimuli. In the EEG we will focus on the well-established ERP component of the N400, while the fNIRS is capable of identifying underlying brain areas. In particular, the involvement of same or different neural networks in sung and spoken sentences as well as the degree of lateralization will provide important insights into the neural underpinnings of semantic processing in speech and song and potentially be relevant for therapeutic interventions in hearing impaired patients in future.

2. Materials and Methods

2.1. Participants

Twenty German native speakers (10 female) participated in the study (mean age: 38.65 years; range: 28–53 years). All participants grew up monolingually with German, and had learned foreign languages mostly at school or through friends, but not intensively in their family surroundings like bilingual subjects. All subjects learned English as their first foreign language at a mean age of 10.67 years (range: 3–11 years; 1 missing data). Other foreign languages were also learned (1 subject had learned 4 additional foreign languages, 2 subjects had learned 3 additional foreign languages, 4 subjects had learned 2 additional foreign languages, 7 subjects had learned 1 additional foreign language). All subjects were right-handed according to the Oldfield Handedness Inventory [64] (mean % right-handedness: 73.68; range: 0–100), had no neurological disorders, were not born prematurely, took no medication affecting cognitive functioning, had normal or corrected-to-normal vision and a normal hearing ability at both ears (assessed by an otorhinolaryngologist and an audiologist of the Department of Hearing, Speech, and Voice Disorders of the Medical University of Innsbruck by means of a pure tone audiogram (PTA) with the following criteria: thresholds <30 dB HL at audiometric test frequencies 500, 1000, 2000, and 4000 Hz-PTA4 average). No subject was a professional musician. 18 subjects entered EEG analyses (2 were excluded due to technical problems) and 18 subjects entered fNIRS analyses (2 were excluded due to technical problems). The excluded subjects differed between EEG and fNIRS as technical problems only referred to one single method. The respective other method remained unaffected.

2.2. Materials

The language material consisted of 88 German sentences (44 semantically correct, 44 semantically incorrect) of the following structure: definite article-subject-auxiliary-definite article-object-past participle). All sentences were constructed in past perfect tense. All nouns (subject and object) were bisyllabic, past participle verbs were trisyllabic. Example of a correct sentence: "Der Forscher hat die Firma gegründet" (engl. translation with German word order: "The researcher has the company founded"). Example of an incorrect sentence: "Der Forscher hat die Birke gegründet" (engl. translation with German word order: "The researcher has the birch founded"). Semantic incorrectness was achieved by a selection restriction error.

All correct and incorrect sentences were naturally spoken and sung by a male speaker who was working as a speech therapist and was trained as a professional singer. Sung sentences were assigned to four different melodies (2 rising, 2 descending) whereas rhythm was kept constant in order to provide a greater melodic variety to subjects (please refer to Supplementary Materials for an auditory example

of a correct/incorrect spoken/sung sentence). Acoustic stimuli were digitally recorded in an anechoic chamber at a sampling rate of 44 kHz and 16 bits. Afterwards, acoustic stimuli were edited using the editing program Audacity (www.audacityteam.org). This included inserting 30 ms of silence at the onset and offset of each sentence as well as loudness normalizing. Furthermore, each individual word of the sentences was marked, and the individual onset times of each word were extracted. This was necessary in order to insert the exact timing of each word into the EEG and fNIRS marker files for neuroscientific analyses. Duration of the critical verb was as following: correct spoken: 1198 ms, incorrect spoken: 1190 ms, correct sung: 1744 ms, and incorrect sung: 1722 ms. An ANOVA with the within-subject factors *condition* (correct vs. incorrect) and *modality* (spoken vs. sung) revealed a significant main effect of *modality* [F (1,43) = 449.293, $p < 0.0001$] suggesting longer verbs in sung compared to spoken sentences. We also tested the whole duration of sentences and found a similar main effect of *modality* [F (1,43) = 130.862, $p < 0.0001$]. Again, sung sentences were longer than spoken ones (correct spoken: 4492 ms, incorrect spoken: 4362 ms, correct sung: 5193 ms, and incorrect sung: 5069 ms).

2.3. Experimental Procedure

The present study was approved by the Ethics Committee of the Medical University of Innsbruck (approval code: 1041/2017). Prior to participating in the experiment, subjects were informed in detail about the aims of the study, the sequence of the experiment, the methods, the exact application procedures, the risks, and the actions to minimize these risks. After having the possibility to clarify any questions, subjects gave written informed consent to take part in the study. Subjects did not receive any compensation for participation.

The experiment was controlled by means of the software Presentation (www.neurobs.com). The presentation sequence started with a fixation cross for 500 ms on a 24'' monitor positioned 1 m in front of the subject. Afterwards the acoustic presentation of the sentence started via stereo loudspeakers positioned below the monitor. Sentences were presented at a sound level of approximately 70 dB. The maximum duration of a slot to present a sentence was 6 s. During this time the fixation cross remained on the screen in order to mitigate effects of eye movements on the EEG signal. After the sentence the fixation cross was again presented for 500 ms. This was followed by the visual presentation of a sad and a happy smiley initiating the correctness judgement task. During this task, subjects had to press either the left or right mouse button indicating whether the previously heard sentence was semantically correct (indicated by a happy smiley) or not (indicated by a sad smiley). The position of the smileys on the monitor as well as the required button presses was counter-balanced across participants. Subjects had to respond within 3 s and the presentation sequence continued as soon as they pressed the button. Afterwards a variable inter-stimulus-interval (ISI) of 6 s on average (range: 4–8 s) followed. This long ISI had to be introduced because of the assessment of functional near-infrared spectroscopy which measures the sluggish hemodynamic response (HRF) peaking around 5 s and returning to baseline after 15–20 s [65]. Because the HRF for each sentence would overlap in time, the introduction of a variable ISI prevents a systematic overlap and allows disentangling brain activation for each experimental condition.

Eight different pseudo-randomization versions were created based on the following rules: (1) not more than 3 correct or incorrect sentences in succession, (2) not more than 3 spoken or sung sentences in succession, (3) at least 10 items between sentences of the same sentence pair, (4) in each experimental half an equal amount of correct and incorrect sentences, and (5) in each experimental half the same amount of spoken and sung sentences.

Completing the experiment took about 45 min on average for all participants. In order to prevent subjects' fatigue, two standardized pauses were introduced after each 15 min.

2.4. Neuroscientific Recording

2.4.1. EEG Recording

The electroencephalogram (EEG) was recorded from 13 AgAgCl active electrodes (Brain Products GmbH, Gilching, Germany). Nine electrodes were placed on the scalp (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4; see Figure 1), while the ground electrode was positioned at AFz and the reference electrode at the nasal bone. One electrode above the right eye (at position FP2) measured the vertical electro-oculogram while one electrode at the outer canthus of the right eye (at position F10) assessed the horizontal electro-oculogram. Electrode impedance was controlled using actiCap Control software (Brain Products GmbH, Gilching, Germany) and kept below 10 kΩ. The EEG signal was recorded by means of the software Brain Vision Recorder (Brain Products GmbH, Gilching, Germany) at a sampling rate of 1000 Hz and amplified between 0.016 and 450 Hz. An anti-aliasing filter with a cut-off at 450 Hz (slope: 24 dB/oct) was applied prior to analogue to digital conversion.

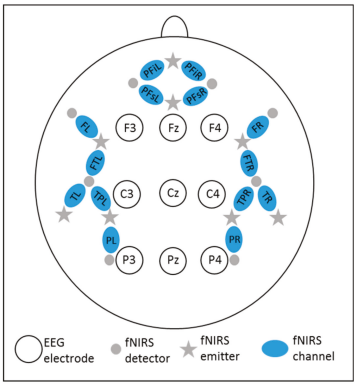


Figure 1. EEG-fNIRS positioning. PFi = prefrontal inferior, PFs = prefrontal superior, F = frontal, FT = fronto-temporal, T = temporal, TP = temporo-parietal, P = parietal, L = left, R = right.

2.4.2. fNIRS Recording

Functional near-infrared spectroscopy (fNIRS) was recorded by means of the NIRScout device (NIRx Medical Technologies, LLC, USA), using a dual wavelength (850 and 760 nm) continuous-wave system. Signals were recorded from 14 channels, resulting from a combination of 8 light emitters and 8 light detectors positioned over bilateral prefrontal, frontal, temporal, temporo-parietal, and parietal areas (see Figure 1). The emitter–detector distance was 3.5 cm. Sampling rate was 7.81 Hz.

2.5. Data Analyses

2.5.1. Behavioral Data Analyses

Based on the correctness judgement task subjects had to indicate whether the heard sentence was semantically correct or incorrect. Percentage of correct responses as well as associated reaction times were extracted and analyzed by means of an ANOVA with the within-subject factors *condition* (correct vs. incorrect) and *modality* (spoken vs. sung). Significance level was set at $p < 0.05$. In case of a significant interaction, post-hoc *t*-tests were performed adjusted by the False Discovery Rate [66].

2.5.2. EEG Data Analyses

EEG data analyses were performed with the software Brain Vision Analyzer 2 (Brain Products GmbH, Gilching, Germany). EEG data were first low-pass filtered with a cut-off of 30 Hz (slope: 12 dB/oct, Butterworth zero-phase filter). Afterwards, a segmentation based on the critical verb in

the sentence was performed from 200 ms before verb onset until 1500 ms after verb onset. An ocular correction based on the Gratton and Coles algorithm [67] was applied in order to correct vertical eye blinks. Other artifacts were manually rejected. A baseline correction (−200–0 ms) was applied. Event-related brain potentials (ERPs) were extracted for each subject and each experimental condition (correct spoken, incorrect spoken, correct sung, incorrect sung) which was followed by the calculation of grand averages in a time window from −200 ms until 1500 ms time-locked to verb onset. After artifact rejection 75.6% (range: 50%–96.2%) of correct spoken, 75% (range: 39.4%–97%) of incorrect spoken, 76.6% (range: 51.3%–94.4%) of correct sung, and 77.8% (range: 51%–95.2%) of incorrect sung sentences entered final statistical analyses.

Statistical analyses were conducted on mean amplitudes. Because the difference between correct and incorrect sentences for spoken and sung sentences was delayed in time, two time windows, 500–900 ms and 800–1200 ms, were chosen based on visual inspection of the grand averages. The first time window characterized the N400 differences between correct and incorrect sentences for spoken sentences, while the second time window indicated the difference for sung sentences. For these analyses, a repeated-measures ANOVA with the within-subject factors *condition* (correct vs. incorrect), *modality* (spoken vs. sung), *region* (anterior vs. central vs. posterior), and *hemisphere* (left vs. right) was performed for lateral electrodes. Midline electrodes underwent an ANOVA with the factors *condition*, *modality*, and *electrodes*. With respect to modality, the mean amplitudes of the two time windows were used in the above-mentioned statistical analysis. Significance level was set at $p < 0.05$. Posthoc *t*-tests were performed and the False Discovery Rate [66] was applied for correcting for multiple comparisons. Whenever Mauchly's test of sphericity became significant, the Greenhouse–Geisser correction [68] was applied.

2.5.3. fNIRS Data Analyses

fNIRS data were first separated into artifact-free segments by eliminating potential artifact-contaminated segments at the beginning and end of experiment as well as additionally introduced pauses in between the experiment in which no markers were presented. Further artifacts during the experiment were visually selected and corrected by a linear interpolation approach (e.g., [69]). A low-pass Butterworth filter of 0.4 Hz (filter order: 3) was applied. Stimulus duration was set at 3 s and used afterwards for applying the general linear model (GLM). Light attenuation was converted into concentration changes of oxygenated hemoglobin [oxy-Hb] and deoxygenated hemoglobin [deoxy-Hb] by means of the modified Beer–Lambert law [70]. For statistical analyses, a GLM-approach was used—in which a box-car-predictor of the stimulus duration was convolved with a canonical hemodynamic response function [71]—peaking at 5 s and fitted to the measured data. This procedure resulted in Beta-values corresponding to μmolar changes which were used for statistical analyses. These comprised repeated-measure ANOVAs with the within-subject factors *condition* (correct vs. incorrect), *modality* (spoken vs. sung), *region* (each of the 7 channels), and *hemisphere* (left vs. right), performed for [oxy-Hb] and [deoxy-Hb], separately. The significance level was set at $p < 0.05$. Posthoc *t*-Tests were performed and the False Discovery Rate [66] was applied for correcting multiple comparisons. Whenever Mauchly's test of sphericity became significant Greenhouse–Geisser correction [68] was applied. Increases in [oxy-Hb] as well as decreases in [deoxy-Hb] are both signs of increased brain activation and were thus analyzed separately.

3. Results

3.1. Behavioral Results

The ANOVA with respect to percentage of correctly answered trials during the correctness judgement task yielded no significant main effect or interaction indicating an equally high percentage: correct spoken (95%), incorrect spoken (97%), correct sung (94%), and incorrect sung (97%) (please refer to Figure 2).

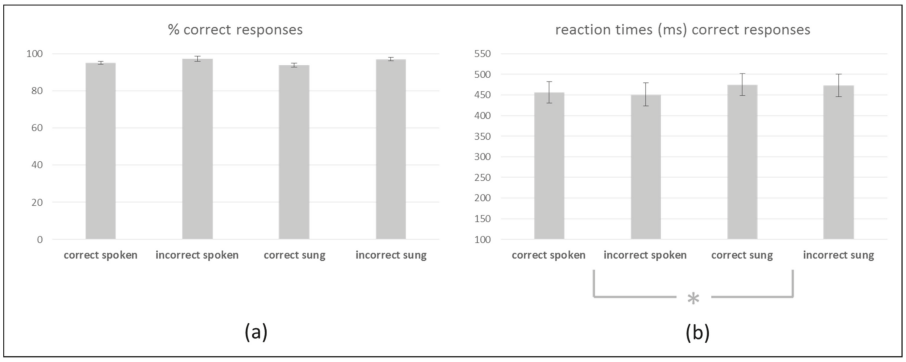


Figure 2. Behavioral data from the correctness judgement task. (**a**) Percentage of correctly answered trials per experimental condition including SEMs. (**b**) Reaction times (in ms) of correctly answered trials per experimental conditions including SEMs. * indicates the significant main effect of *modality* reflecting longer reaction times for sung compared to spoken sentences.

The ANOVA for reaction times of correctly answered trials during the correctness judgement task yielded a significant main effect of *modality* [F (1,19) = 4.602, p = 0.045]. Posthoc t-tests revealed longer reaction times for sung (474 ms) compared to spoken sentences (454 ms) (see Figure 2).

3.2. EEG Results

The ANOVA for lateral electrodes revealed significant main effects of *condition* and *modality* as well as significant interactions *condition* × *region* and *modality* × *region* (Table 1). Subsequent posthoc t-Tests resolving the interaction *condition* × *region* revealed a larger negative amplitude for incorrect compared to correct sentences at central [C3 and C4: t (17) = 3.326, p = 0.004] and posterior regions [P3 and P4: t (17) = 3.223, p = 0.005] (see Figures 3 and 4). Posthoc t-tests resolving the interaction *modality* × *region* revealed a larger negativity for spoken compared to sung sentences at central [C3 and C4: t (17) = −3.064, p = 0.007] and posterior regions [P3 and P4: t (17) = −3.076, p = 0.007].

Table 1. Statistical results of the ANOVA *condition* × *modality* × *region* × *hemisphere* for event-related brain potentials (ERP) data on lateral electrodes. Data from the time window 500–900 ms was considered for spoken sentences, while data from the time window 800–1200 ms entered analyses for sung sentences. Significant effects ($p < 0.050$) are marked in bold.

Effect Lateral Electrodes	*df*	*F*	*p*
condition	1,17	8.498	**0.010**
modality	1,17	8.572	**0.009**
condition × *modality*	1,17	0.221	0.644
condition* × *region	2,34	6.316	**0.010**
modality* × *region	2,34	6.929	**0.010**
condition × *modality* × *region*	2,34	0.159	0.761
condition × *hemisphere*	1,17	0.258	0.618
modality × *hemisphere*	1,17	0.008	0.932
condition × *modality* × *hemisphere*	1,17	0.041	0.842
condition × *region* × *hemisphere*	2,34	0.956	0.372
modality × *region* × *hemisphere*	2,34	0.288	0.751
condition × *modality* × *region* × *hemisphere*	2,34	1.553	0.226

Findings for midline electrodes revealed the following significant effects (Table 2): main effect of *condition*, main effect of *modality*, and interaction *condition* × *electrodes*. The main effect of modality revealed a more negative shift for spoken compared to sung sentences. Subsequent posthoc t-tests

resolving the interaction *condition* × *electrodes* revealed a larger negativity for incorrect compared to correct sentences at Fz [t (16) = 3.199, p = 0.006], Cz [t (17) = 3.340, p = 0.004], and Pz [t (17) = 3.264, p = 0.005] (see Figures 3 and 4).

Table 2. Statistical results of the ANOVA *condition* × *modality* × *region* × *hemisphere* for ERP data on midline electrodes. Data from the time window 500–900 ms was considered for spoken sentences, while data from the time window 800–1200 ms entered analyses for sung sentences. Significant effects ($p < 0.050$) are marked in bold.

Effect Midline Electrodes	*df*	*F*	*p*
condition	1,16	14.749	**0.001**
modality	1,16	8.033	**0.012**
condition × *modality*	1,16	0.191	0.668
condition* × *electrodes	2,32	5.428	**0.009**
modality × *electrodes*	2,32	3.142	0.057
condition × *modality* × *electrodes*	2,32	0.159	0.745

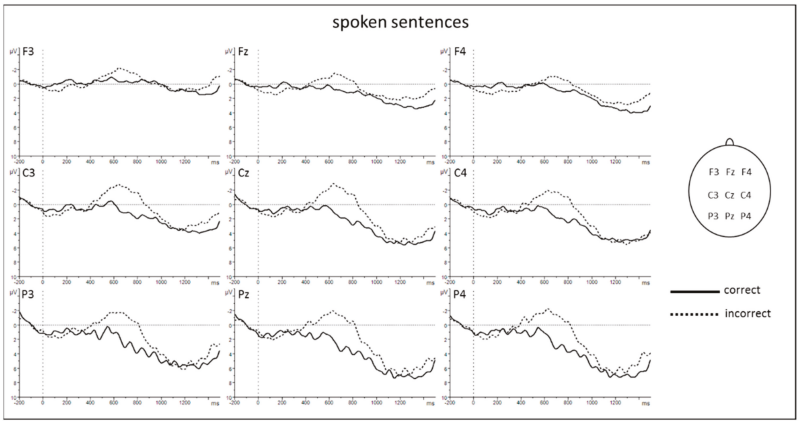


Figure 3. ERP results for spoken sentences. Grand averages from −200 ms to 1500 ms after verb onset. Negativity is plotted upwards. An 8 Hz low-pass filter was applied for presentation purposes only.

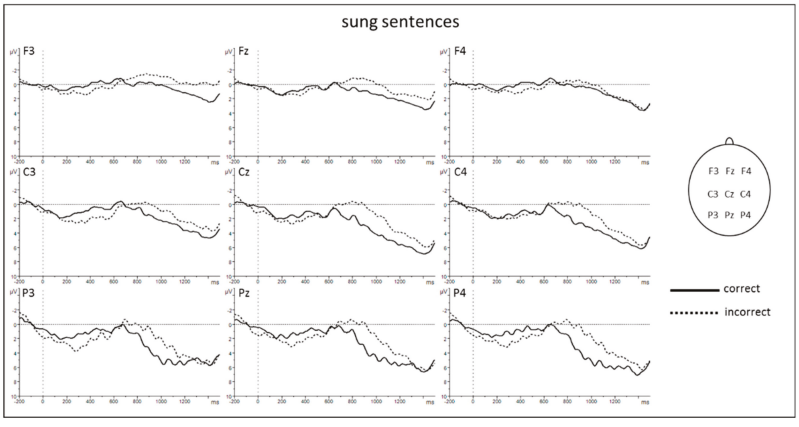


Figure 4. ERP results for sung sentences. Grand averages from −200 ms to 1500 ms after verb onset. Negativity is plotted upwards. An 8 Hz low-pass filter was applied for presentation purposes only.

3.3. fNIRS Results

3.3.1. Results for [oxy-Hb]

The ANOVA revealed a significant main effect of *modality* as well as significant interactions *modality* × *region* and *modality* × *region* × *hemisphere* (Table 3). Subsequent posthoc *t*-tests resolving the three-way interaction revealed a stronger activation for spoken compared to sung sentences at the following channels: left prefrontal inferior [PFiL: t (17) = 2.974, p = 0.009], left and right prefrontal superior [PFsL: t (17) = 2.615, p = 0.018; PFsR: t (17) = 2.814, p = 0.012], left temporal [TL: t (17) = 2.140, p = 0.047], left temporo-parietal [TPL: t (17) = 2.902, p = 0.010], as well as left and right parietal [PL: t (17) = 2.242, p = 0.039; PR: t (17) = 3.041, p = 0.007] (see Figure 5).

Table 3. Statistical results of the ANOVA *condition* × *modality* × *region* × *hemisphere* for [oxy-Hb] of functional near-infrared spectroscopy (fNIRS) data. Significant effects ($p < 0.050$) are marked in bold.

Effect [oxy-Hb]	*df*	*F*	*p*
condition	1,17	2.100	0.166
modality	1,17	4.897	**0.041**
condition × *modality*	1,17	0.041	0.841
condition × *region*	6,102	0.392	0.592
modality* × *region	6,102	3.211	**0.046**
condition × *modality* × *region*	6,102	0.614	0.624
condition × *hemisphere*	1,17	1.249	0.279
modality × *hemisphere*	1,17	1.012	0.329
condition × *modality* × *hemisphere*	1,17	0.018	0.896
condition × *region* × *hemisphere*	6,102	1.025	0.391
modality* × *region* × *hemisphere	6,102	4.165	**0.008**
condition × *modality* × *region* × *hemisphere*	6,102	1.615	0.198

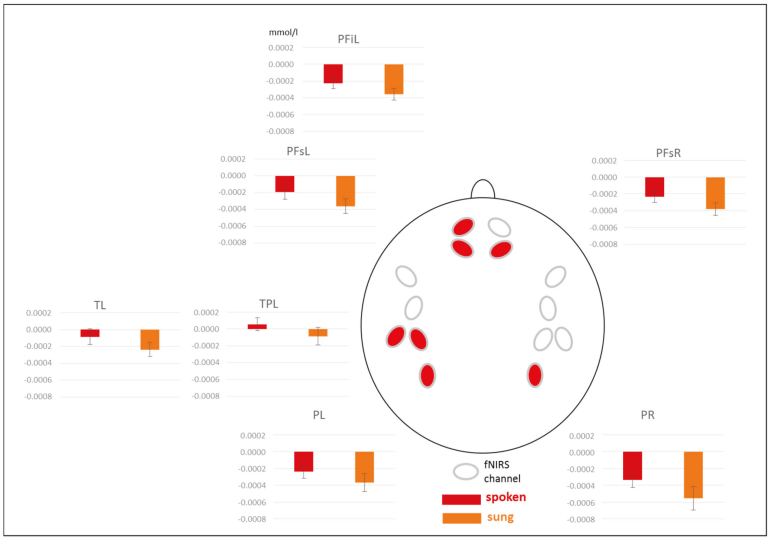


Figure 5. fNIRS results for [oxy-Hb]. Beta-values for spoken and sung sentences merged across correct and incorrect sentences. Red channels indicate significant differences. PFi = prefrontal inferior, PFs = prefrontal superior, T = temporal, TP = temporo-parietal, P = parietal, L = left, R = right. Please note that a more positive value indicates an increased activation.

3.3.2. Results for [deoxy-Hb]

The ANOVA revealed a significant main effect of *condition* (Table 4) indicating a stronger activation for correct compared to incorrect sentences (see Figure 6).

Table 4. Statistical results of the ANOVA *condition* × *modality* × *region* × *hemisphere* for [deoxy-Hb] of fNIRS data. Significant effects ($p < 0.50$) are marked in bold.

Effect [deoxy-Hb]	*df*	*F*	*p*
condition	1,17	12.530	**0.003**
modality	1,17	1.153	0.298
condition × *modality*	1,17	0.020	0.889
condition × *region*	6,102	1.183	0.316
modality × *region*	6,102	0.936	0.416
condition × *modality* × *region*	6,102	1.555	0.210
condition × *hemisphere*	1,17	0.330	0.572
modality × *hemisphere*	1,17	0.675	0.423
condition × *modality* × *hemisphere*	1,17	1.573	0.227
condition × *region* × *hemisphere*	6,102	0.849	0.405
modality × *region* × *hemisphere*	6,102	0.669	0.520
condition × *modality* × *region* × *hemisphere*	6,102	0.532	0.648

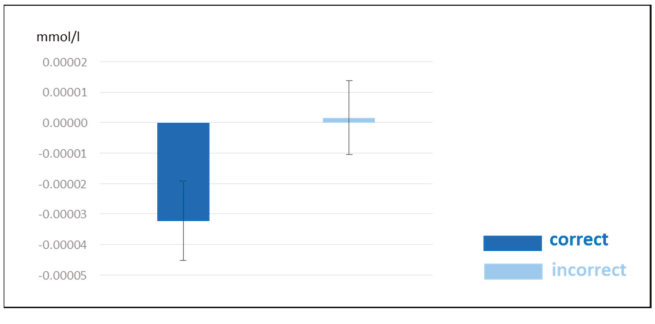


Figure 6. fNIRS results for [deoxy-Hb]. Beta-values for correct and incorrect sentences merged across spoken and sung sentences and across all channels including SEMs. Please note that a more negative value indicates an increased activation.

4. Discussion

The present study investigated neural mechanisms of semantic processing in speech and song. Semantic processing was operationalized by acoustically presenting semantically correct and incorrect sentences which were either spoken or sung. Singing is a form of music including both melodic as well as linguistic aspects. However, is meaning extracted similarly or differently from singing compared to pure spoken information? This research question guided the present study. In order to assess neural foundations of semantic processing, two neuroscientific methods were applied simultaneously, namely the EEG and the fNIRS.

4.1. The N400 Differentiates between Correct and Incorrect Sentences

EEG results for spoken and sung sentences showed a clear difference between semantically correct and incorrect sentences indexed by a classical N400 component. The N400 is usually found in several semantic contexts and reflects lexical access and semantic integration [27,28,37]. It shows larger amplitudes when semantic processing is difficult. Such a modulation was also found in our study, revealing larger N400 amplitudes for incorrect compared to correct sentences. This N400 effect was equally present in both modalities. However, an important difference was nevertheless observable.

The N400 for spoken and sung sentences was generally delayed compared to previous studies, and the N400 for sung sentences was even more delayed (500–900 ms for spoken and 800–1200 ms for sung sentences). A first consideration for this general delay of the N400 was that the critical verb is a past participle containing a clear syntactic marker "ge" in German. Only after this prefix an identification of semantic correctness is possible. Thus, we averaged ERPs aligned after this prefix. However, the N400 for sung sentences was still delayed in time compared to spoken sentences (cf. Supplementary Figure S5). Thus, we opted for carrying out the standard analysis procedure aligning ERPs to critical word onsets. Another explanation for the delayed N400 might concern the subjects' age range (mean age of 39 years). Studies investigating the N400 in differential semantic paradigms in younger (usually in the mid 20s) and older subjects show ambiguous results. Some studies report some delays of the N400 in older subjects [72–74] while others do not find any delayed processing [40,75,76]. Our delay might rather be driven by the longer duration of spoken but especially sung sentences as well as final words. A classical N400 to spoken sentences was usually reported between 300 and 500 ms [28]. In our study, the N400 to spoken sentences was found between 500 and 900 ms, thus delayed. However, giving a closer look to the grand averages of N400s in spoken sentences in previous studies shows that even though smaller time windows were analyzed (400–700 ms in [38] and 250–700 ms in [39]), the differences between semantically correct and incorrect sentences lasted longer (~until 1000 ms). This was the case for young (around 25 years [38,39]) but also middle age (around 43 years [77]) and older participants (around 60 years [78,79]). It should be noted that sentence duration in these studies [38,39] was about 1700 ms while in our study spoken sentences lasted much longer (around 4400 ms). This longer duration resulted from a slow presentation rate in order to approximate sentence length of spoken to sung sentences. Furthermore, this slow presentation rate was introduced because the study is currently also performed in hearing impaired patients supplied with cochlear implants and/or hearing aids with difficulties in language comprehension. In order to give these patients a chance to understand these sentences they were spoken very slowly. In fact, normal-hearing participants noticed this slow presentation rate, indicating that they experienced the experiment as effortful. Patients, on the other hand, did not complain about this slow presentation rate. Unfortunately, Besson and colleagues [40] do not report the exact duration of their sung final word. Gordon and colleagues [35], however, report the duration of their word stimuli used in a priming study. Their sung stimuli were 913 ms long while our critical words lasted around 1700 ms, thus much longer. While in Gordon et al. the N400 occurred between 300 and 500 ms, the longer duration of the sung stimuli in our study could explain the delayed N400 effect. Further support for this assumption is provided by the reaction times during the correctness judgement task in the present experiment also showing longer reaction times for sung compared to spoken stimuli. Finally, EEG results seem to show qualitatively similar semantic processing in spoken and sung sentences, with a quantitative difference displayed in a delayed N400 component. These EEG findings might be important with respect to hearing impaired patients who clearly show more behavioral difficulties in extracting meaning from sung sentences as from spoken speech [8] but also benefits from a musical training [21,22]. These findings are moreover interesting in the light of therapeutic interventions such as melodic intonation therapy (MIT) postulating a beneficial effect on language processing in aphasic patients through singing [15,16]. It should, however, be considered that MIT predominantly reveals its favorable effects with respect to speech production and not necessarily speech comprehension which was studied in the present study.

4.2. Brain Areas Recruited for Semantic Processing in Spoken and Sung Sentences

fNIRS results showed a twofold pattern: (1) an increased activation for spoken compared to sung sentences, irrespective of semantic correctness in bilateral prefrontal, left temporal and temporo-parietal, and bilateral parietal areas, and (2) an increased activation for correct compared to incorrect sentences—irrespective of modality widespread over the whole cortex.

The larger activation for spoken compared to sung sentences in the fNIRS goes in line with the larger negativity for spoken versus sung sentences in the EEG. However, in the EEG this difference

can hardly be interpreted due to the different time windows analyzed for spoken and sung sentences. This increased activation for spoken compared to sung sentences in the fNIRS shows a stronger left-hemispheric lateralization, which might potentially be driven by the fact that our participants were non-musicians. Thus, they are more familiar with understanding spoken compared to sung language in everyday life. Furthermore, the correctness judgement task directed attention to the linguistic content and not to the melodic features of sentences. Similar findings were also shown by Sammler and colleagues [52] in a repetition priming study with fMRI contrasting lyrics and tunes in unfamiliar songs. The authors also found larger activations in the left superior temporal sulcus for lyrics than tunes in musically untrained subjects suggesting a link between subjects' expertise with music and language and a predominant processing of linguistic meaning.

The second important fNIRS finding was the widespread increased activation for correct compared to incorrect sentences, irrespective of modality. This result is in line with previous studies which also contrast semantically correct to incorrect sentences [59–61]. In particular, the direction of effects conforms to the fMRI findings of Humphries and colleagues [59]. They contrasted semantically correct with random sentences (i.e., words were scrambled resulting in a meaningless sentence). The authors also found increased activations for correct sentences in similar regions as in our study. Especially temporo-parietal areas were proposed to be related to combinatory semantic processes at the sentence level relevant for the formation of a more complex meaning. Such an interpretation would also fit with our activation pattern. The fact that a differentiation between correct and incorrect sentences was equally present for spoken and sung material might be attributed to the task in our experiment which primarily directed attention to the semantic content of sentences.

In general, however, topographic aspects of fNIRS results should be considered with caution as spatial resolution is limited compared to fMRI due to the possibility to assess neural activation from maximally 3 cm depth from scalp. Thus, only cortical areas can be reached. Due to the simultaneous assessment of EEG and fNIRS, only a limited number of light emitters and detectors can be positioned in between EEG electrodes. Consequently, specific tomographic analyses with multi-distance emitter-detector-pairs potentially leading to a better spatial resolution are not possible.

5. Conclusions

Findings from our multi-methodological approach indicate that the extraction of meaning from sentences is equally processed in spoken compared to sung sentences. A predominant processing of spoken compared to sung sentences could furthermore be attested. This effect seems to be at least partially influenced by a stronger familiarity with spoken material as well as with the correctness judgement task directing subjects' attention to the linguistic content of sentences. It would be interesting to conduct the same experiment without any experimental task; for example, simply during passive listening to spoken and sung sentences. Importantly, these fine-grained mechanisms appear only in the neural response but not in behavioral data, showing an equally high percentage of identification of correct and incorrect sentences in both spoken and sung modality. Interestingly, both neuroscientific methods show concordant results with respect to the direction of effects. However, the EEG—with its high temporal resolution—showed quantitative differences between spoken and sung sentences, as semantic processing in sung sentences was delayed in time. Based on these findings, we pursue the next step to investigate semantics in spoken and sung sentences in hearing-impaired listeners who are supplied with either hearing aids or cochlear implants, as these patients experience language comprehension problems. This would provide insights into the neural processing mechanisms which are present at the beginning and during the course of the rehabilitation process.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3425/10/1/36/s1, Audio file S1: An example of a correct spoken sentence. Audio file S2: An example of a correct sung sentence. Audio file S3: An example of an incorrect spoken sentence. Audio file S4: An example of an incorrect sung sentence. Figure S5: Grand averages at the electrode Cz for semantically correct versus incorrect sentences for spoken and sung sentences aligned after the prefix "ge" of the critical past participle.

Author Contributions: Conceptualization, S.R. and J.S.; methodology, S.R. and J.S.; formal analysis, S.R., M.F.G. and J.S.; investigation, S.R. and M.R.; data curation, S.R.; writing—original draft preparation, S.R.; writing—review and editing, J.S., M.F.G., P.G.Z., M.R. and O.G.; visualization, S.R.; supervision, S.R.; project administration, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Thomas Lungenschmid for speaking and singing the experimental material as well as all participating subjects. A special thank goes to all collaborators helping during neuroscientific data acquisition and analyses.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kraus, N.; Slater, J. Chapter 12-Music and language: Relations and disconnections. In *Handbook of Clinical Neurology*; The Human Auditory System; Aminoff, M.J., Boller, F., Swaab, D.F., Eds.; Elsevier: Amsterdam, The Netherlands, 2015; pp. 207–222.
2. Seidner, W.; Wendler, J. *Die Sängerstimme*; Henschel Verlag: Berlin, Germany, 1978.
3. Sundberg, J. Formant Structure and Articulation of Spoken and Sung Vowels. *FPL (Folia Phoniatrica et Logopaedica)* **1970**, *22*, 28–48. [CrossRef]
4. Zatorre, R.J.; Baum, S.R. Musical Melody and Speech Intonation: Singing a Different Tune. *PLoS Biol.* **2012**, *10*, e1001372. [CrossRef]
5. Mithen, S.; Morley, I.; Wray, A.; Tallerman, M.; Gamble, C. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body, by Steven Mithen*; Weidenfeld & Nicholson: London, UK, 2005; pp. 97–112, ISBN 0-297-64317-7 hardback £20 & US$25.2; ix+374. *Camb. Archaeol. J.* **2006**, *16*, 97–112.
6. De l'Etoile, S.K. Infant behavioral responses to infant-directed singing and other maternal interactions. *Infant Behav. Dev.* **2006**, *29*, 456–470. [CrossRef] [PubMed]
7. Nakata, T.; Trehub, S.E. Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* **2004**, *27*, 455–464. [CrossRef]
8. Crew, J.D.; Galvin, J.J.; Fu, Q.-J. Perception of Sung Speech in Bimodal Cochlear Implant Users. *Trends Hear.* **2016**, *20*. [CrossRef] [PubMed]
9. Yu, M.; Xu, M.; Li, X.; Chen, Z.; Song, Y.; Liu, J. The shared neural basis of music and language. *Neuroscience* **2017**, *357*, 208–219. [CrossRef] [PubMed]
10. Ludke, K.M.; Ferreira, F.; Overy, K. Singing can facilitate foreign language learning. *Mem. Cogn.* **2014**, *42*, 41–52. [CrossRef]
11. Dittinger, E.; Barbaroux, M.; D'Imperio, M.; Jäncke, L.; Elmer, S.; Besson, M. Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations. *J. Cogn. Neurosci.* **2016**, *28*, 1584–1602. [CrossRef]
12. Kraus, N.; Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **2010**, *11*, 599–605. [CrossRef]
13. Anderson, S.; White-Schwoch, T.; Parbery-Clark, A.; Kraus, N. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear. Res.* **2013**, *300*, 18–32. [CrossRef] [PubMed]
14. Strait, D.L.; Kraus, N. Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hear. Res.* **2014**, *308*, 109–121. [CrossRef] [PubMed]
15. Schlaug, G.; Marchina, S.; Norton, A. From Singing to Speaking: Why Singing May Lead to Recovery of Expressive Language Function in Patients with Broca's Aphasia. *Music Percept. Interdiscip. J.* **2008**, *25*, 315–323. [CrossRef] [PubMed]
16. Merrett, D.L.; Peretz, I.; Wilson, S.J. Neurobiological, Cognitive, and Emotional Mechanisms in Melodic Intonation Therapy. *Front. Hum. Neurosci.* **2014**, *8*, 401. [CrossRef] [PubMed]
17. Sihvonen, A.J.; Särkämö, T.; Leo, V.; Tervaniemi, M.; Altenmüller, E.; Soinila, S. Music-based interventions in neurological rehabilitation. *Lancet Neurol.* **2017**, *16*, 648–660. [CrossRef]
18. Orellana, C.P.; van de Sandt-Koenderman, M.E.; Saliasi, E.; van der Meulen, I.; Klip, S.; van der Lugt, A.; Smits, M. Insight into the neurophysiological processes of melodically intoned language with functional MRI. *Brain Behav.* **2014**, *4*, 615–625. [CrossRef]

19. Akanuma, K.; Meguro, K.; Satoh, M.; Tashiro, M.; Itoh, M. Singing can improve speech function in aphasics associated with intact right basal ganglia and preserve right temporal glucose metabolism: Implications for singing therapy indication. *Int. J. Neurosci.* **2016**, *126*, 39–45. [CrossRef]
20. Good, A.; Gordon, K.A.; Papsin, B.C.; Nespoli, G.; Hopyan, T.; Peretz, I.; Russo, F.A. Benefits of Music Training for Perception of Emotional Speech Prosody in Deaf Children With Cochlear Implants. *Ear Hear.* **2017**, *38*, 455–464. [CrossRef]
21. Torppa, R.; Faulkner, A.; Laasonen, M.; Lipsanen, J.; Sammler, D. Links of Prosodic Stress Perception and Musical Activities to Language Skills of Children With Cochlear Implants and Normal Hearing. *Ear Hear.* **2019**. [CrossRef]
22. Torppa, R.; Huotilainen, M. Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hear. Res.* **2019**, *380*, 108–122. [CrossRef]
23. Rossi, S.; Telkemeyer, S.; Wartenburger, I.; Obrig, H. Shedding light on words and sentences: Near-infrared spectroscopy in language research. *Brain Lang.* **2012**, *121*, 152–163. [CrossRef] [PubMed]
24. Holcomb, P.J. Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* **1993**, *30*, 47–61. [CrossRef] [PubMed]
25. Rugg, M.D. The Effects of Semantic Priming and Word Repetition on Event-Related Potentials. *Psychophysiology* **1985**, *22*, 642–647. [CrossRef] [PubMed]
26. Matsumoto, A.; Iidaka, T.; Haneda, K.; Okada, T.; Sadato, N. Linking semantic priming effect in functional MRI and event-related potentials. *NeuroImage* **2005**, *24*, 624–634. [CrossRef] [PubMed]
27. Lau, E.F.; Phillips, C.; Poeppel, D. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* **2008**, *9*, 920–933. [CrossRef]
28. Kutas, M.; Federmeier, K.D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* **2011**, *62*, 621–647. [CrossRef]
29. Koelsch, S. Towards a neural basis of processing musical semantics. *Phys. Life Rev.* **2011**, *8*, 89–105. [CrossRef]
30. Koelsch, S. Toward a Neural Basis of Music Perception-A Review and Updated Model. *Front. Psychol.* **2011**, *2*, 110. [CrossRef]
31. Koelsch, S.; Kasper, E.; Sammler, D.; Schulze, K.; Gunter, T.; Friederici, A.D. Music, language and meaning: Brain signatures of semantic processing. *Nat. Neurosci.* **2004**, *7*, 302–307. [CrossRef]
32. Steinbeis, N.; Koelsch, S. Comparing the Processing of Music and Language Meaning Using EEG and fMRI Provides Evidence for Similar and Distinct Neural Representations. *PLoS ONE* **2008**, *3*, e2226. [CrossRef]
33. Painter, J.G.; Koelsch, S. Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music: Processing of meaning in music. *Psychophysiology* **2011**, *48*, 645–655. [CrossRef] [PubMed]
34. Chien, P.-J.; Chan, S. Old songs can be as fresh as new: An ERP study on lyrics processing. *J. Neurolinguist.* **2015**, *35*, 55–67. [CrossRef]
35. Gordon, R.L.; Schön, D.; Magne, C.; Astésano, C.; Besson, M. Words and Melody Are Intertwined in Perception of Sung Words: EEG and Behavioral Evidence. *PLoS ONE* **2010**, *5*, e9889. [CrossRef] [PubMed]
36. Kutas, M.; Hillyard, S.A. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* **1980**, *207*, 203–205. [CrossRef] [PubMed]
37. Friederici, A.D. Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci. (Regul. Ed.)* **2002**, *6*, 78–84. [CrossRef]
38. Hahne, A.; Friederici, A.D. Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cogn. Brain Res.* **2002**, *13*, 339–356. [CrossRef]
39. Friederici, A.D.; Pfeifer, E.; Hahne, A. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* **1993**, *1*, 183–192. [CrossRef]
40. Besson, M.; Faïta, F.; Peretz, I.; Bonnel, A.-M.; Requin, J. Singing in the Brain: Independence of Lyrics and Tunes. *Psychol. Sci.* **1998**, *9*, 494–498. [CrossRef]
41. Rosslau, K.; Herholz, S.C.; Knief, A.; Ortmann, M.; Deuster, D.; Schmidt, C.-M.; Zehnhoff-Dinnesen, A.; Pantev, C.; Dobel, C. Song Perception by Professional Singers and Actors: An MEG Study. *PLoS ONE* **2016**, *11*, e0147986. [CrossRef]
42. Besson, M.; Schön, D. Comparison between Language and Music. *Ann. N. Y. Acad. Sci.* **2001**, *930*, 232–258. [CrossRef]

43. Zatorre, R.J.; Belin, P.; Penhune, V.B. Structure and function of auditory cortex: Music and speech. *Trends Cogn. Sci.* **2002**, *6*, 37–46. [CrossRef]
44. Zatorre, R.J.; Belin, P. Spectral and Temporal Processing in Human Auditory Cortex. *Cereb. Cortex* **2001**, *11*, 946–953. [CrossRef]
45. Poeppel, D.; Idsardi, W.J.; van Wassenhove, V. Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 1071–1086. [CrossRef] [PubMed]
46. Telkemeyer, S.; Rossi, S.; Koch, S.P.; Nierhaus, T.; Steinbrink, J.; Poeppel, D.; Obrig, H.; Wartenburger, I. Sensitivity of Newborn Auditory Cortex to the Temporal Structure of Sounds. *J. Neurosci.* **2009**, *29*, 14726–14733. [CrossRef] [PubMed]
47. Friederici, A.D.; Alter, K. Lateralization of auditory language functions: A dynamic dual pathway model. *Brain Lang.* **2004**, *89*, 267–276. [CrossRef]
48. Hickok, G.; Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **2007**, *8*, 393–402. [CrossRef]
49. Mummery, C.J.; Shallice, T.; Price, C.J. Dual-Process Model in Semantic Priming: A Functional Imaging Perspective. *NeuroImage* **1999**, *9*, 516–525. [CrossRef] [PubMed]
50. Fang, Y.; Han, Z.; Zhong, S.; Gong, G.; Song, L.; Liu, F.; Huang, R.; Du, X.; Sun, R.; Wang, Q.; et al. The Semantic Anatomical Network: Evidence from Healthy and Brain-Damaged Patient Populations. Available online: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.22858 (accessed on 6 August 2019).
51. Rissman, J.; Eliassen, J.C.; Blumstein, S.E. An event-related FMRI investigation of implicit semantic priming. *J. Cogn. Neurosci.* **2003**, *15*, 1160–1175. [CrossRef] [PubMed]
52. Sammler, D.; Baird, A.; Valabrègue, R.; Clément, S.; Dupont, S.; Belin, P.; Samson, S. The Relationship of Lyrics and Tunes in the Processing of Unfamiliar Songs: A Functional Magnetic Resonance Adaptation Study. *J. Neurosci.* **2010**, *30*, 3572–3578. [CrossRef]
53. Schön, D.; Gordon, R.; Campagne, A.; Magne, C.; Astésano, C.; Anton, J.-L.; Besson, M. Similar cerebral networks in language, music and song perception. *NeuroImage* **2010**, *51*, 450–461. [CrossRef]
54. Merrill, J.; Sammler, D.; Bangert, M.; Goldhahn, D.; Lohmann, G.; Turner, R.; Friederici, A.D. Perception of Words and Pitch Patterns in Song and Speech. *Front. Psychol.* **2012**, *3*, 76. [CrossRef] [PubMed]
55. Kreitewolf, J.; Friederici, A.D.; von Kriegstein, K. Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage* **2014**, *102*, 332–344. [CrossRef]
56. Callan, D.E.; Tsytsarev, V.; Hanakawa, T.; Callan, A.M.; Katsuhara, M.; Fukuyama, H.; Turner, R. Song and speech: Brain regions involved with perception and covert production. *NeuroImage* **2006**, *31*, 1327–1342. [CrossRef] [PubMed]
57. Whitehead, J.C.; Armony, J.L. Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Hum. Brain Mapp.* **2018**, *39*, 4913–4924. [CrossRef] [PubMed]
58. Özdemir, E.; Norton, A.; Schlaug, G. Shared and distinct neural correlates of singing and speaking. *NeuroImage* **2006**, *33*, 628–635. [CrossRef]
59. Humphries, C.; Binder, J.R.; Medler, D.A.; Liebenthal, E. Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *J. Cogn. Neurosci.* **2006**, *18*, 665–679. [CrossRef]
60. Kuperberg, G.R.; McGuire, P.K.; Bullmore, E.T.; Brammer, M.J.; Rabe-Hesketh, S.; Wright, I.C.; Lythgoe, D.J.; Williams, S.C.R.; David, A.S. Common and Distinct Neural Substrates for Pragmatic, Semantic, and Syntactic Processing of Spoken Sentences: An fMRI Study. *J. Cogn. Neurosci.* **2000**, *12*, 321–341. [CrossRef]
61. Rüschemeyer, S.-A.; Fiebach, C.J.; Kempe, V.; Friederici, A.D. Processing Lexical Semantic and Syntactic Information in First and Second Language: fMRI Evidence from German and Russian. Available online: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.20098 (accessed on 6 August 2019).
62. Friederici, A.D.; Rüschemeyer, S.-A.; Hahne, A.; Fiebach, C.J. The Role of Left Inferior Frontal and Superior Temporal Cortex in Sentence Comprehension: Localizing Syntactic and Semantic Processes. *Cereb. Cortex* **2003**, *13*, 170–177. [CrossRef]
63. Rogalsky, C.; Hickok, G. Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex. *Cereb. Cortex* **2009**, *19*, 786–796. [CrossRef]
64. Oldfield, R.C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **1971**, *9*, 97–113. [CrossRef]
65. Huettel, S.A.; Song, A.W.; McCarthy, G. *Functional Magnetic Resonance Imaging*, 2nd ed.; Sinauer Associates, Inc.: Sunderland, MA, USA, 2008.

66. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* **1995**, *57*, 289–300. [CrossRef]
67. Gratton, G.; Coles, M.G.; Donchin, E. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* **1983**, *55*, 468–484. [CrossRef]
68. Greenhouse, S.W.; Geisser, S. On methods in the analysis of profile data. *Psychometrika* **1959**, *24*, 95–112. [CrossRef]
69. Scholkmann, F.; Spichtig, S.; Muehlemann, T.; Wolf, M. How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol. Meas.* **2010**, *31*, 649–662. [CrossRef]
70. Cope, M.; Delpy, D.T.; Wray, S.; Wyatt, J.S.; Reynolds, E.O.R. A CCD Spectrophotometer to Quantitate the Concentration of Chromophores in Living Tissue Utilising the Absorption Peak of Water at 975 nm. In *Oxygen Transport to Tissue XI*; Advances in Experimental Medicine and Biology; Springer: Boston, MA, USA, 1989; pp. 33–40. ISBN 978-1-4684-5645-5.
71. Boynton, G.M.; Engel, S.A.; Heeger, D.J. Linear systems analysis of the fMRI signal. *NeuroImage* **2012**, *62*, 975–984. [CrossRef]
72. Cheimariou, S.; Farmer, T.A.; Gordon, J.K. Lexical prediction in the aging brain: The effects of predictiveness and congruency on the N400 ERP component. *Aging Neuropsychol. Cogn.* **2019**, *26*, 781–806. [CrossRef]
73. Kutas, M.; Iragui, V. The N400 in a semantic categorization task across 6 decades. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials Sect.* **1998**, *108*, 456–471. [CrossRef]
74. Hunter, C.R. Is the time course of lexical activation and competition in spoken word recognition affected by adult aging? An event-related potential (ERP) study. *Neuropsychologia* **2016**, *91*, 451–464. [CrossRef]
75. Mohan, R.; Weber, C. Neural activity reveals effects of aging on inhibitory processes during word retrieval. *Aging Neuropsychol. Cogn.* **2019**, *26*, 660–687. [CrossRef]
76. Federmeier, K.D.; Van Petten, C.; Schwartz, T.J.; Kutas, M. Sounds, Words, Sentences: Age-Related Changes Across Levels of Language Processing. *Psychol. Aging* **2003**, *18*, 858–872. [CrossRef]
77. Friederici, A.D.; von Cramon, D.Y.; Kotz, S.A. Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain* **1999**, *122*, 1033–1047. [CrossRef] [PubMed]
78. Hagoort, P.; Brown, C.M.; Swaab, T.Y. Lexical-semantic event–related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. *Brain* **1996**, *119*, 627–649. [CrossRef] [PubMed]
79. Swaab, T.; Brown, C.; Hagoort, P. Spoken Sentence Comprehension in Aphasia: Event-related Potential Evidence for a Lexical Integration Deficit. *J. Cogn. Neurosci.* **1997**, *9*, 39–66. [CrossRef] [PubMed]

brain sciences

MDPI

Article

Music Training Positively Influences the Preattentive Perception of Voice Onset Time in Children with Dyslexia: A Longitudinal Study

Aline Frey [1,*], Clément François [2,3], Julie Chobert [4], Jean-Luc Velay [4], Michel Habib [5] and Mireille Besson [4,6]

1 ESPE de l'académie de Créteil, Université Paris-Est Créteil, Laboratoire CHArt, 94380 Bonneuil-sur-Marne, France
2 Laboratoire Parole et Langage, CNRS et Aix Marseille Université, 13640 Aix-en-Provence, France; fclement24@hotmail.com
3 Cognition and Brain Plasticity Group, IDIBELL, University of Barcelona, 08193 Barcelona, Spain
4 Laboratoire de Neurosciences Cognitives, CNRS et Aix-Marseille Université, 13331 Marseille, France; julie@chobert.fr (J.C.); jean-luc.velay@univ-amu.fr (J.-L.V.); mireille.besson@univ-amu.fr (M.B.)
5 Département de Neurologie Pédiatrique, CHU Timone, 13005 Marseille, France; michel.habib@resodys.org
6 Cuban Neuroscience Center, La Havane 4850, Cuba
* Correspondence: aline.frey@u-pec.fr; Tel.: +01-49-56-35-27

Received: 1 April 2019; Accepted: 13 April 2019; Published: 21 April 2019

Abstract: Previous results showed a positive influence of music training on linguistic abilities at both attentive and preattentive levels. Here, we investigate whether six months of active music training is more efficient than painting training to improve the preattentive processing of phonological parameters based on durations that are often impaired in children with developmental dyslexia (DD). Results were also compared to a control group of Typically Developing (TD) children matched on reading age. We used a Test–Training–Retest procedure and analysed the Mismatch Negativity (MMN) and the N1 and N250 components of the Event-Related Potentials to syllables that differed in Voice Onset Time (VOT), vowel duration, and vowel frequency. Results were clear-cut in showing a normalization of the preattentive processing of VOT in children with DD after music training but not after painting training. They also revealed increased N250 amplitude to duration deviant stimuli in children with DD after music but not painting training, and no training effect on the preattentive processing of frequency. These findings are discussed in view of recent theories of dyslexia pointing to deficits in processing the temporal structure of speech. They clearly encourage the use of active music training for the rehabilitation of children with language impairments.

Keywords: Music training; longitudinal study; children with dyslexia; Mismatch Negativity (MMN); syllables

1. Introduction

Developmental Dyslexia (DD) is a neurodevelopmental disorder that impairs the acquisition of reading despite conventional instruction and sociocultural opportunities, normal intelligence, and motivation [1–3]. This disorder affects ~5% of children in primary school [1,4–6]. While recent computational models provide evidence that the causes of DD are likely to be multifactorial [7], deficits in phonological processing have long been considered as one of the hallmarks of DD in a large majority of dyslexic children [8–10]. In fact, deficient phonological processing may impact reading, writing and the acquisition of novel phonological forms [11–13] even with intact phonemic representations [14,15]. Specifically, the poor development of reading skills in children at risk for or with dyslexia [16–18] possibly reflects processing deficits of temporal and spectral acoustic cues [19,20]. In a previous

experiment comparing children with dyslexia and Typically Developing (TD) children [21], we tested for this hypothesis at the preattentive level, that is, when children are not asked to focus their attention on the stimuli of interest. Therefore, these stimuli are processed implicitly rather than explicitly. We used the Mismatch Negativity (MMN), which is considered as a good index of preattentive processing ([22]; for review, see [23]). The MMN is elicited by infrequent changes in an auditory stimulus sequence of standard repeated stimuli, even when participants are watching a silent movie and not paying attention to the stimuli. The MMN, measured as the difference between the Event-Related Potentials (ERPs) to the deviant and the ERPs to the standard stimuli, can be recorded from adults as well as from children, infants, and patients which make the MMN an invaluable tool to study perceptive and cognitive processes from populations that are often difficult to test [24]. Chobert et al. [21] used a multifeature MMN design that allowed them to test, within the same auditory sequence, the preattentive processing of several types of deviant syllables that differed from the standard stimulus (/ba/) in one specific feature [25]. The multifeature MMN design is particularly useful with clinical populations that cannot be tested in long duration sessions. Syllables differed in Voice Onset Time (VOT), defined as the time interval between noise bursts at consonant release and the onset of vocal cord vibrations [26], that allows perceiving stop consonant as voiced (e.g., /b/, negative VOT ~−100 ms in French) or voiceless (e.g., /p/, VOT values higher than 0 ms in French, [27]). They also differed in vowel duration and in vowel frequency. In line with several results in the literature, we found abnormal processing of VOT and duration, but not of frequency changes in children with DD compared to typically developing (TD) children ([16,28–30], see [31], for abnormal processing of frequency deviant stimuli in adults with DD, and [32] for pitch discrimination deficits in sentence contexts).

Several findings point to atypical processing of the temporal structure of speech in children with DD, particularly rise time and slow amplitude modulations of the speech envelope [19,20,33,34]. This led Goswami and collaborators to propose the temporal sampling theory of dyslexia [35,36]. They assumed that these deficits are possibly linked to abnormal synchronization of brain oscillations, at frequencies involved in syllabic and prosodic perception (theta—from 4 to 10 Hz—and delta—from 1 to 4 Hz—respectively; [35,37,38]. Very recently, Cantiania and collaborators [39] used a multifeature MMN design with nonverbal (complex tones) frequency and duration deviant stimuli. In line with the temporal sampling theory, they reported reduced left gamma power in Italian 6-month-old infants at risk for language and learning impairments compared to children with no known family deficits (see [40] for enhanced phase locking at 4–7 Hz in children with DD).

Despite evidence showing abnormal phonological processing of the structure of speech in children with DD ([21,35,41–43], see [29] for review), whether this deficit is causally linked to difficulties in the perception of acoustic-phonetic features still remains an open issue [21,29,35,44–48]. To address this causality issue, we used a longitudinal approach with children with dyslexia trained with music or with painting. The reasons for using music training in the experimental group are detailed below. Painting training was used in the control group because painting is an activity that can be as interesting and motivating for the children, thereby controlling for general factors known to influence learning such as attention and motivation.

Many results in the literature have shown that musicians are more sensitive than nonmusicians to acoustic cues that are common to music and speech sounds (i.e., duration, frequency, intensity, and timbre), possibly because processing these cues in speech and nonspeech sounds draw upon the same pool of neural resources and rely on common processing [49–56]. At the behavioral level, music training increases pitch and duration discrimination accuracy for pure and harmonic tones [57–60] and decreases discrimination thresholds for syllables that vary on temporal cues (i.e., vowel duration, VOT, and rise time; [61–63]). More generally, results of an increasing number of experiments demonstrated music to language transfer effects, so that music training and musical expertise influence several levels of language processing, including the processing of linguistic and emotional prosody, categorical perception, word segmentation as well as syntactic and semantic processing (for reviews see [51,64]).

Taken together, these results opened the interesting perspective to use music training as a rehabilitation tool for adults and children with language deficits.

At the electrophysiological level, previous results in adults demonstrated larger amplitude and/or shorter Mismatch Negativitys (MMNs) latency in musicians than in nonmusicians for frequency, duration and timber manipulations in pure or harmonic tones (e.g., [65,66]) and in speech or speech-like stimuli (e.g., [61]). Moreover, children with high musical aptitudes and pronunciation skills showed enhanced MMNs to speech duration deviant stimuli compared to children who lacked these skills [67]. Also, using a multifeature MMN design [25], we found larger MMNs to duration deviant stimuli in 9-year-old TD children with four years of music training than in TD children with no formal music training (cross-sectional study, [68]). Moreover, the deviance size effects (i.e., the difference between large and small deviant stimuli) for VOT deviant stimuli was also larger in TD children with music training than without (Group by Deviance size interaction). Finally, as in previous studies with children with DD [16,21,29,30], no group by deviant size interaction was found for frequency changes. Taken together, these results suggest that in both adults and in TD children, music training improves several aspects of speech perception, in particular VOT and duration, possibly because increased sensitivity to features that are common to music and speech allows musicians and musically trained children to construct more reliable phonological representations of speech sounds than nonmusicians [47,49,51,52,56]. This interpretation is also in line with results showing positive correlations between musical aptitude and phonological abilities (e.g., [47,69–71]).

Correlation is not causality, however [72,73]. Because the studies described above used cross-sectional designs, it is not possible to ascertain that music training is at the origin of the improvements in speech perception. To directly test for causality, Chobert and collaborators [74] implemented a longitudinal study over two school years, in which TD children without formal musical background, were pseudo-randomly assigned to music or to painting training programs in a controlled, randomized trial (CRT). In line with the predictions directly issued from the results of the cross-sectional study described above [68], results for TD children showed that the MMNs to duration and VOT deviant stimuli was enhanced after 12 months of music training compared to before training but not after 12 months of painting training. No training effect was found for frequency deviant stimuli in either group. By controlling for preexisting between-group differences before training and by using pseudo-random assignment to music or painting training in a CRT, the results of Chobert et al. [74] demonstrated that enhanced sensitivity to temporal (duration) and phonological (VOT) features of syllables in TD children trained with music did not result from predispositions for music but was causally linked to music training.

In the present study, we report results for children with DD who were involved in the same CRT as the TD children of Chobert et al. [74] and who were trained with music or painting. The multifeature MMN design [25] included the syllable "Ba" as the standard stimulus as well as large and small changes in VOT, vowel duration, and vowel frequency as deviant stimuli. The logic of the experiment is based on the results described above that were obtained with different groups of children. Compared to TD children, children with DD showed deficits in the preattentive processing of VOT and duration deviant stimuli [21]. By contrast, TD children with four years of music training (cross-sectional study, [68]) and nonmusician TD children trained with music for 12 months (longitudinal study, [74]) showed enhanced sensitivity to VOT and duration deviant stimuli compared to TD children with no formal music training or to TD nonmusician children trained with painting. Based on these results, the first aim of the present experiment was to determine whether children with DD would develop an enhanced sensitivity to VOT and to duration deviant stimuli after music training but not after painting training. Specifically, comparing children with DD involved in the two types of training, we expected the deviance size effect for VOT and duration deviant stimuli not to be significant before training (as in [21]), but to be significant after music training and not after painting training [74]. Finally, we also compared results for children with DD after music or painting training with results for TD children before training. We predicted that the deviance size effects for VOT and duration would be similar

for children with DD after music training and for TD children before training (normalization of the deviance effect, no significant Group × Deviance size effect). By contrast, the deviance size effect would still be significantly different for children with DD after painting training and for TD children before training (significant Group × Deviance size interaction).

The second aim of the present experiment was to better understand the relationship between the MMNs and the ERP components of interest. To this aim, we analysed the N100 component, an exogenous component and obligatory brain response elicited by the presentation of any stimulus, be it a sound, a light, a touch, etc. that is typically taken to reflect perceptual processing (e.g., [75]). The N100 component shows maximum amplitude (peak) at ~100 ms, and mean amplitude of the N100 is measured in a latency window surrounding the peak. Interestingly, in the studies mentioned above, results pointed to differences between children with dyslexia and TD children in the frontocentral N1 associated to stimuli with different rise times [19,20]. The N250 component probably belongs to the N200 family of components, taken to reflect stimulus categorization [76] with larger amplitude in explicit than in implicit categorization tasks [77]. The N250 is measured in the same way as the N100 component (mean amplitude in a latency window centered on the peak). In children, the N250 possibly reflects the building-up of sound representations in sensory memory [78]. We predicted enhanced amplitude of the N100 component to large, and possibly small, VOT deviant stimuli compared to standard stimuli after music training but not after painting training. Similarly, we expected the N250 component to duration deviant stimuli to be larger compared to standard stimuli after music but not after painting training.

2. Materials and Methods

2.1. Participants

A total of 57 children participated in the study with 33 children with DD and 24 TD children attending the 3rd grade in two schools in Aix-en-Provence and Marseille. Dyslexic children in each school had been formally diagnosed with dyslexia by an interdisciplinary team of neurologists, neuropsychologists and speech therapists and they were part of a specialized dyslexia class (called CLIS in French for "CLasse pour l'Inclusion Scolaire" or class for inclusive schooling). Children in the present study were tested before training using several cognitive and reading measures (see below) in order to compare cognitive functioning and reading abilities between children with DD and TD before and after training, but not for redoing a formal diagnosis of dyslexia. Eleven children were excluded from the dyslexic group either because they left during the school year (4 children) or because of too many artifacts in their electrophysiological recordings (7 children). Out of the 22 remaining children with DD, 11 were trained with music (3 girls; 8 right-handers) and 11 were trained with painting (4 girls; 9 right-handers). Finally, three children were excluded from the group of TD children because of too many artifacts in EEG recordings and the final group comprised 21 TD children (11 girls; 18 right-handers).

The mean chronological age at the start of the study was not significantly different in the two groups of children with DD (music group: 10.24-year-old (sd = 0.93) and painting group 10.75-year-old (sd = 0.73)), but TD children were significantly younger (8.26-year-old (sd = 0.15) than children with DD; see Table 1). Reading age was assessed with the Alouette reading test [79], which is the most commonly used standardized reading test in France, with the most reliable norms for calculating reading age [80]. Reading age of children with dyslexia in the music group was 8.14 years (sd = 0.66) and in the painting group 7.85 years (sd = 0.55), which corresponds approximately to a reading delay of 3 years. Thus, we can safely assume that children with DD who participated in the present experiment were still quite severely impaired at the time of the study. TD children were matched for reading age based on the Alouette standardized reading test. Their reading age was 8.07 years (sd = 0.33), which was not significantly different from that of the children with dyslexia (see Table 1).

Table 1. Before training. Results of children with dyslexia (DysMus and DysPaint) and of typically developing readers (TD) who were matched on reading age, on measures of memory, on verbal and nonverbal intelligence, on phonology, on reading regular words, irregular words and pseudo-words, and on visual and auditory attention. For each test, the number in brackets (e.g., /44) corresponds to maximum score.

Test	TD	DysM.	DysP.	F(2,40)	Post hoc comparisons
Chronological age (month)	99.14	122.91	129.00	60.25, $p < 0.001$	DysM. vs. TD: $p < 0.001$ DysP. vs. TD: $p < 0.001$ DysM. vs. DysP.: ns
Reading age [a] (month)	92.44	87.24	86.27	F < 1	
Memory [b] Digit Span (/32)	13.38	11.45	11.73	3.71, $p = 0.03$	DysM. vs. TD: $p < 0.05$ DysP.vs TD: ns DysM. vs. DysP.: ns
Verbal IQ Similarities (/44) [b]	15.67	17.73	18.36	F < 1	
Nonverbal IQ Symboles (/60) [b]	16.38	20.73	16.82	2.16, $p = 0.13$	
Nonverbal IQ Progressive Matrices (/36) [c]	26.81	28.91	27.64	F < 1	
Phonology[d] RAN (seconds)	28.58	27.18	29.45	F < 1	
Phonology [d] Phoneme Deletion (/10)	5.58	5.54	4.82	F < 1	
Phonology [d] Phoneme Fusion (/10)	5.92	5.64	5.18	F < 1	
Phonology [d] Nonword repetition (/20)	17.58	16.64	17.09	F < 1	
Reading irregular words (/20)	7.91	4.45	6.90	2.24, $p = 0.15$	
Reading regular words (/20)	14.27	9.27	12.36	3.07, $p = 0.08$	
Reading of pseudowords (/20)	13.60	8.82	10.08	7.84, $p < 0.03$	DysM. vs. TD: $p < 0.02$ DysP. vs. TD: $p < 0.05$ DysM. vs. DysP.: ns
Attention [e] Visual Attention Score (/45)	17.08	16.00	18.09	F < 1	
Attention [e] Auditory Attention (/132)	93.92	84.36	89.27	F < 1	
Attention [e] Orientation (/10)	6.56	7.45	7.18	F < 1	
Attention [e] Visuomotor precision (/52)	22.42	25.36	24.09	F < 1	
Attention [e] Arrows (/30)	19.33	20.73	20.09	F < 1	

[a] Alouette Standardized Reading Test. [b] Wechsler Intelligence Scale for Children WISCIV. [c] Progressive Matrices PM47. [d] ODEDYS. [e] NEPSY.

Children were tested using several tests from the Wechsler Intelligence Scale for Children (WISC-IV; Digit Span: direct and reverse, similarities, symbols; [81]), from the NEPSY battery (visual attention, auditory attention and associated responses, orientation, and visuomotor precision; [82]), from the ODEDYS battery (reading regular and irregular words as well as pseudo-words, Rapid Automatized Naming (RAN), and Phonological awareness: phoneme deletion, phoneme fusion, and nonword repetition), and the Progressive Matrices (PM47, nonverbal cognitive abilities, [83]). Results at these tests together with age, school level, gender, and socioeconomic background were used to pseudo-randomly assign children to the music or painting groups and to ensure that no significant differences existed between the two groups before training. These measures are presented in Table 1. Note that children

with DD in the Dys-Mus and Dys-Paint groups showed higher scores for verbal and nonverbal IQ compared to TD children, most likely because they were almost two years older on average.

All children were native French speakers and had normal or corrected-to-normal vision, normal audition, and no known neurological deficits as determined from a detailed questionnaire completed by the parents prior to the experiment. Children had similar socioeconomic backgrounds ranging from middle to low social class as determined from the parents' profession according to the criteria of the National Institute of Statistics and Economic Studies. Most children were involved in extracurricular activities (i.e., mainly sports), but none of the children or their parents had formal training in music or painting.

This study was conducted in accordance with norms and guidelines for the protection of human subjects. Informed consent from the inspector and school directors as well as from the children and their parents was granted before the start of the project that was approved by the National Ethics Committee for Biomedical Research (RCB: 2011-A00172-39). Parents were informed in detail on the procedure (see below *2.2 Longitudinal Study: Procedure*) and on music and painting training that were described as challenging, interesting and rewarding experiences for their children. At the end of each school year, children from the painting group displayed their artwork at a school exhibition and children from the music group performed a concert. Children were given gifts at the end of each testing session to thank them for their participation.

2.2. Longitudinal Study: Procedure

Children were tested before training and after six months of music or painting training (while children were enrolled in this program for two school years (12 months), too many children with DD had left the program after two years to obtain reliable results). In both cases, they were tested individually in a quiet room of their school in two separate sessions that included neuropsychological assessments (as described above in *2.1. Participants*) and electrophysiological tests (as described below in *2.3.MMN Experiment: Procedure*). Each session lasted for two hours (including many pauses) and was separated by four or five days.

Two teachers professionally trained in music or visual arts were specifically hired for this project. Training lasted for 6 months (20 weeks excluding holidays) twice a week for 45 min which amounts to a total of 300 h of training. Music training was based on a combination of Kodály and Orff methods (http://www.iks.hu/; http://www.orff.de/en.html). During the music training sessions, children progressively learned how to play musical pieces of increasing complexity on diverse musical instruments including drums, timbales, guitars, and xylophones. Each session started with relaxation and vocal exercises that were followed by vocal and body games focusing on pitch, musical intervals and rhythms (finding the beat, following the beat, counting the pulse, performing polyrhythmic pieces, singing together, in canon, etc.). These games encouraged the mapping of vocal pitch modulations to hand movements. Specific time slots focused on improvising melodies and rhythms in order to foster group listening and coordination (stopping all the instruments except one, playing crescendo and decrescendo, silencing all the instruments, starting all the instruments together, starting one by one, etc.). Children also learned to synchronize their walk on the pulse while tapping in their hands (i) on the beat, (ii) on the strong beats only (2 and 4 beats), and (iii) on $\frac{1}{4}$ notes. Each session involved the recordings of short live performances in such a way that children always listened to their own productions and could comment on their quality, in order to increase their conscious awareness of the vocal and instrumental performances.

Painting training was based on the approach developed by Arno Stern (http://www.arnostern.com/) and based on the idea that children' personal development is linked to social experiences. Autonomy and creativity are developed through painting considered as a game that is to be played together. Training sessions were built to progressively develop a better understanding of concepts such as lines and perspectives, static and dynamic figures, matter and texture, lights and colours. Children were sensitized to these different components in different training sessions using concrete approaches (e.g.,

combine colours that were available from cans of paint presented on a rack in the middle of the room; draw lines and perspectives; mix different textures; etc.) and different themes (e.g., favourite animals, houses and cities, nature, flowers, trees, etc.). As children were painting on large paper sheets fixed on mural panels, they learned to coordinate their movements to produce large as well as small motives in the paintings. Members of the research group coordinated the training activities and ensured that both groups of children were similarly motivated and stimulated.

2.3. MMN Experiment: Procedure

Children sat in a comfortable chair 1 meter from a computer screen and EEG was recorded before and after training while the children watched a silent subtitled movie displayed on a computer screen. Children were told to watch the movie without paying attention to the sounds that were presented through headphones. VOT, duration, and frequency deviant stimuli, each with two levels of deviance-size (large and small distance from the standard) were randomly presented within the auditory sequence with a sound onset asynchrony of 600 ms synchronized with vowel onset. A total of 1200 stimuli were presented with 432 deviant stimuli (72 stimuli (6% probability) for each of the 6 deviant types). All stimuli were presented within a single block that lasted for 12.2 min. At the end of the experiment, children were asked questions to ensure they had paid attention to the movie.

2.4. Stimuli

Stimuli were syllables with Consonant-Vowel (CV) structure (see Figure 1). The standard stimulus /Ba/ was naturally produced and had a VOT of −70 ms and vowel duration of 208 ms, for a total duration of the stimulus equal to 278 ms and a fundamental frequency (F0) of 103 Hz. For VOT deviant stimuli, F0 and vowel duration were the same as for the standard stimulus but VOT changed. Large and small deviant stimuli were selected on a "Ba-Pa" continuum that comprised 9 sounds. The large deviant stimulus was "$Ba_{0\ ms}$" (VOT = 0 ms; i.e., 70 ms shorter than the standard, 100% decrease) and the small deviant stimulus was "$Ba_{-40\ m}s$" (VOT = −40 ms; i.e., 30 ms shorter than the standard, 42% decrease). For duration deviant simuli, VOT and F0 were the same as for the standard stimulus but vowel duration was shortened using "Adobe Audition" software [84]. Vowel duration was 128 ms for the large deviant stimulus (i.e., 80 ms shorter than the standard, 38% decrease; total duration large deviant = 198 ms) and 158 ms for the small deviant stimulus (i.e., 50 ms shorter than the standard, 24% decrease; total duration small deviant = 228 ms). For frequency deviant stimuli, VOT and vowel duration were the same as for the standard stimulus but the F0 of the vowel was increased using the Praat software [85]. For the large deviant stimulus, the F0 was increased to 154 Hz (i.e., 51 Hz higher than standard, 49% increase) and for the small deviant stimulus to 117 Hz (i.e., 14 Hz higher than standard, 13% increase).

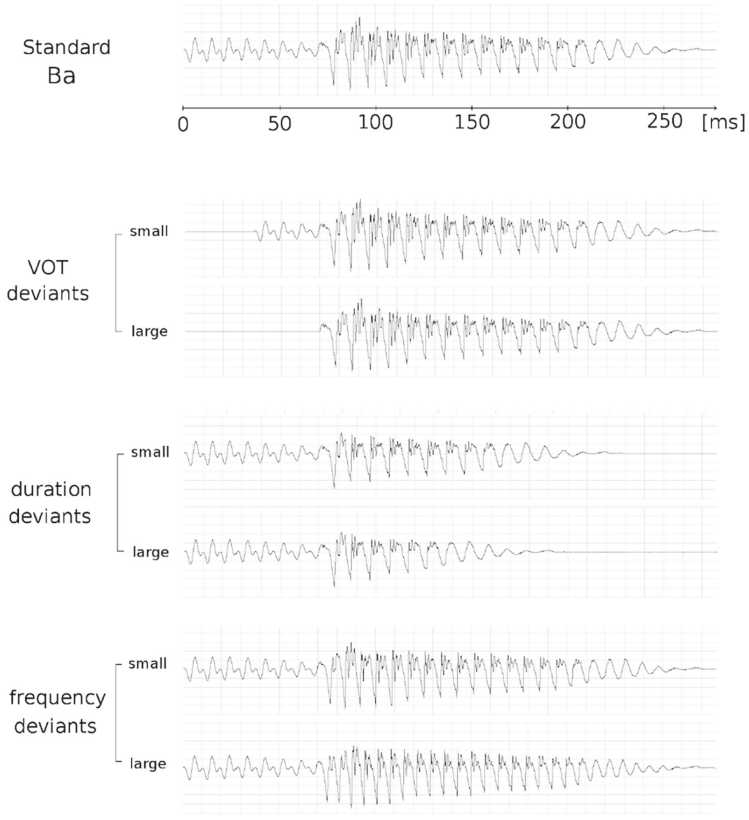


Figure 1. Illustration of the stimuli used in the experiment. Standard stimulus /Ba/ (Voice Onset Time (VOT): −70 ms, vowel duration: 208 ms, total duration: 278 ms and F0: 103 Hz); VOT deviant stimuli (vowel duration and F0: same as for /Ba/ but VOT for large deviant stimulus = 0 ms (/$Ba_{0\ ms}$/) and VOT for small deviant stimulus = −40 ms (was /$Ba_{-40\ ms}$/); duration deviant simuli (VOT and F0: same as for /Ba/ but vowel duration for the large deviant stimulus = 128 ms and for the small deviant stimulus = 158 ms; and frequency deviant stimuli (VOT and vowel duration: same as for /Ba/ but F0 for the large deviant stimulus = 154 Hz and for the small deviant stimulus = 117 Hz).

2.5. ERP Recording and Processing

The Electroencephalogram (EEG) was continuously recorded at a sampling rate of 512 Hz with a 0–102.4 Hz band-pass using a Biosemi amplifier system (Amsterdam, BioSemi Active 2) from 32 active Ag–Cl electrodes mounted on a child-sized elastic cap (Biosemi Pintype) at standard positions of the International 10/20 System [86]. Data were re-referenced offline both to a nose reference, to verify the typical MMN inversion between Fz/Cz and the mastoids electrodes [23] and to the averaged activity over the left and right mastoids, to quantify MMN amplitude since these averages typically show a better signal-to-noise ratio than the nose-referenced averages [24,87]. EEG data were filtered with a bandpass of 1 to 30 Hz (12 dB/oct; as recommended by [24]).

The electrooculogram (EOG) was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Three additional electrodes were placed on the left and right mastoids and on the nose. EEG data were analysed using the Brain Vision Analyser software (Version 01/04/2002; Brain Products, Gmbh). Recordings were segmented into 700 ms epochs (from −100 ms until 600 ms poststimulus onset). Epochs with electric

activity exceeding baseline activity by 60 μV were considered as artifacts and were automatically rejected from further processing (~10%).

2.6. Data Analysis

Data from the various psychometric tests were analysed using repeated measures Multivariate Analyses of Variance (MANOVAs) that included Group (DysMus vs. DysPaint vs. TD) as a between-subjects factor, Session (before vs. after training), and Tests as within-subject factors.

Electrophysiological data were analysed using BrainVision Analyzer v.2.0 software (Brain Products, Germany). ERPs to standard, large and small deviant stimuli were computed separately for each dimension (VOT, duration, and frequency). For VOT, the mean amplitude of the N1 component was measured in the 50 to 150 ms latency band and for duration, the mean amplitude of the N250 component was measured in the 200 to 350 ms latency band. MMNs were obtained for each deviant stimulus by subtracting ERPs to standard stimuli from ERPs to large or small deviant stimuli, separately for each participant and for each dimension (VOT, duration, and frequency) at each electrode. MMN mean amplitude was computed for each deviant over 50 ms windows (VOT: 80–130 ms, duration: 280–330 ms and frequency: 280–330 ms). Time windows were chosen based on visual inspection and on results of previous studies analysing MMN in children [21,74].

Repeated-measures of ANalysis Of VAriance (ANOVAs) were computed on MMN mean amplitude, as well as on N100 and N250 mean amplitude, for each dimension separately (VOT, duration, and frequency). Analyses typically included Group (DysMus vs. DysPaint) as a between-subject factor, and Session (before vs. after training), deviance size for MMNs analyses (large vs. small deviant stimuli), Anterior–Posterior Dimension (frontal, central, and parietal) and Laterality (left, central, and right) as within-subject factors, or only frontal sites (F3, Fz, and F4). Separate ANOVAs were conducted for each session separately when results of interactions including the Group and Session factors were significant. Finally, further analyses were also conducted to test for the hypothesis of a normalization of VOT and duration processing with music training that included the group of TD children (before training) and the groups of children with DD after music or painting training. Greenhouse–Geisser corrections were applied when appropriate and conservative post hoc Tukey tests (reducing the probability of Type I errors) were used to determine the source of significant interactions.

3. Results

3.1. Neuropsychological and Speech Assessments

No between-group differences were found before training ($F(1,20) = 2.04$; $p = 0.17$). Overall, the level of performance was higher after six months of training than before training (main effect of Session: ($F(1,20) = 11.6$; $p < 0.003$, see Table 2 for the main effect of Session in each specific test). This improvement was not significantly different in the music and painting training groups (main effect of Group: $F < 1$; Group by Session interaction $F < 1$).

Table 2. Results of children with dyslexia (DysMus and DysPaint) after 6 months of training, on measures of memory, verbal and nonverbal intelligence, phonology, reading regular and irregular words and pseudo-words, and visual and auditory attention. T0: before the start of the experiment and T6: six months after.

Test	DysM.	DysP.	Main Effect Session F(1,20)	Post hoc Comparisons
Reading age [a] (month)	90.18	90.45	18.50 $p < 0.001$	T0 = 82.64 < T6 = 90.32
Memory [b] Digit Span (/32)	11.73	11.18	F < 1	
Verbal IQ Similarities (/44) [b]	23.55	24.09	34.93 $p < 0.001$	T0 = 18.04 < T6 = 23.82

Table 2. *Cont.*

Test	DysM.	DysP.	Main Effect Session F(1,20)	Post hoc Comparisons
Nonverbal IQ Symboles (/60) [b]	22.64	19.00	F = 2.29 $p < 0.15$	
Nonverbal IQ Progressive Matrices (/36) [c]	30.36	27.64	F < 1	
Phonology [d] RAN (seconds)	23.27	23.64	5.62 $p < 0.03$	T0 = 32.44 > T6 = 26.40
Phonology [d] Phoneme Deletion (/10)	6.00	5.18	F < 1	
Phonology [d] Phoneme Fusion (/10)	7.27	7.09	11.20 $p < .003$	T0 = 5.41 < T6 = 7.18
Phonology [d] Nonword repetition (/20)	17.27	18.00	F < 1	
Reading irregular words (/20)	6.82	8.45	8.52 $p < 0.01$	T0 = 5.68 < T6 = 7.64
Reading regular words (/20)	11.18	12.00	F < 1	
Reading Pseudowords (/20)	8.64	11.73	F < 1	
Attention [e] Visual Attention Score (/45)	18.55	17.55	F < 1	
Attention [e] Auditory Attention (/132)	106.18	98.91	5.76 $p < 0.03$	T0 = 86.82 < T6 = 102.55
Attention [e] Orientation (/10)	8.18	8.18	10.87 $p < 0.005$	T0 = 7.32 < T6 = 8.18
Attention [e] Visuomotor precision (/52)	26.91	26.73	F = 2.13 $p < 0.16$	
Attention [e] Arrows (/30)	20.45	19.00	F < 1	

3.2. MMN Amplitude

As is typical in MMN paradigms, MMNs to deviant stimuli in VOT, duration, and frequency in children with dyslexia showed the typical polarity inversion between Fz and the mastoids electrodes when using the nose reference (see Figure 2).

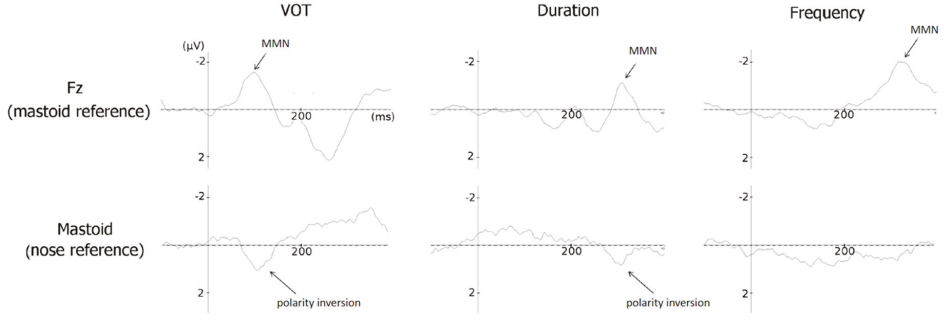


Figure 2. Mismatch Negativities (MMNs—i.e., Event-Related Potentials (ERPs) to the deviant stimuli minus ERPs to standard stimuli) averaged across large and small deviant stimuli and before and after training at Fz (with mastoid reference, top) and at mastoid electrode (mean of left and right mastoid electrodes, with nose reference, bottom) for Voice Onset Time (VOT, left), Duration (middle), and Frequency (right) deviant stimuli. Note the clear polarity inversion between Fz and mastoid for VOT and duration changes but not for frequency changes.

Moreover, independently of the type of deviant stimuli, MMNs were always larger over fronto-central regions than over parietal regions (main effect of Anteroposterior factor for VOT: $F(2,40) = 11.22$, $p < 0.001$; for duration: $F(2,40) = 7.81$, $p < 0.001$; and for frequency: $F(2,40) = 18.40$, $p < 0.001$; see Table 3). Based on these results, further analyses were conducted on the averaged responses for each deviant over frontal sites (F3, Fz, and F4). Analyses of the deviance size effect for each type of deviant stimuli (VOT, duration, and frequency) are reported below.

Table 3. Mean MMNs amplitude in children with dyslexia averaged across sessions and across large and small deviant stimuli on Voice Onset Time (VOT), duration, and frequency at Frontal, Central, and Parietal sites.

	VOT	Duration	Frequency
Frontal	−1.61 μV	−1.12 μV	−1.85 μV
Central	−1.27 μV	−1.16 μV	−1.24 μV
Parietal	−0.64 μV	−0.69 μV	−0.59 μV

3.2.1. VOT (MMN Amplitude)

Results of ANOVAs on MMN amplitude including Group (DysMus vs. DysPaint) as a between-subject factor as well as Session (before vs. after training), Deviance size (large vs. small VOT deviant stimuli) and Laterality (F3 vs. Fz vs. F4) as within-subjects factors showed that the main effects of Group and Session were not significant (both Fs < 1) but the Group × Session × Deviance size was significant ($F(1,20) = 3.39$, $p < 0.04$).

Results of separate ANOVAs before training revealed that the VOT deviance size effect on MMN amplitude was not significant either in DysMus (large = −1.21 μV and small = −1.27 μV, $p < 0.99$) or in DysPaint (large = −2.07 μV and small = −1.80 μV, $p < 0.99$; Group × Deviance size: F < 1). By contrast, after training, the deviance size effect was significant in DysMus ($p < 0.003$), with larger MMNs to large (−2.48 μV) than to small (−0.42 μV) VOT changes, but not in DysPaint (large = −1.99 μV and small = −1.69 μV, $p = 0.93$; Group × Deviance size: ($F(1,20) = 6.03$, $p < 0.02$; see Figure 3A). In other words, in DysMus, MMNs to large VOT deviant stimuli were larger after (−2.48 μV) than before training (−1.21 μV, $p < 0.001$) with no training-related differences in DysPaint (before training: −2.07 μV and after training: −1.99, μV, $p < 0.99$). For small VOT changes, the differences between after and before training did not reach significance either for DysMus or for DysPaint (see Figure 3C).

As mentioned in the introduction, we hypothesized that the deviance size effects for VOT and duration would be similar for children with DD after music training and for TD children before training. In other words, we expected a normalization of the deviance effect in the music group but not in the painting group. To specifically test for this hypothesis, we compared the group of TD children (before training) and DysMus and DysPaint groups after training. Results showed that the Group by Deviance size interaction was significant ($F(2,40) = 4.28$, $p < 0.02$). Separate comparisons showed that the deviance size effect was similar for TD children and DysMus (Group × Deviance size: $F(1,30) = 1.16$, $p < 0.30$), but it was larger for TD children than for DysPaint (Group × deviance size: $F(1,30) = 4.37$, $p < 0.04$). As can be seen on Figure 3A,B, the deviance size effect is significant for DysMus and for TD children, but not for DysPaint.

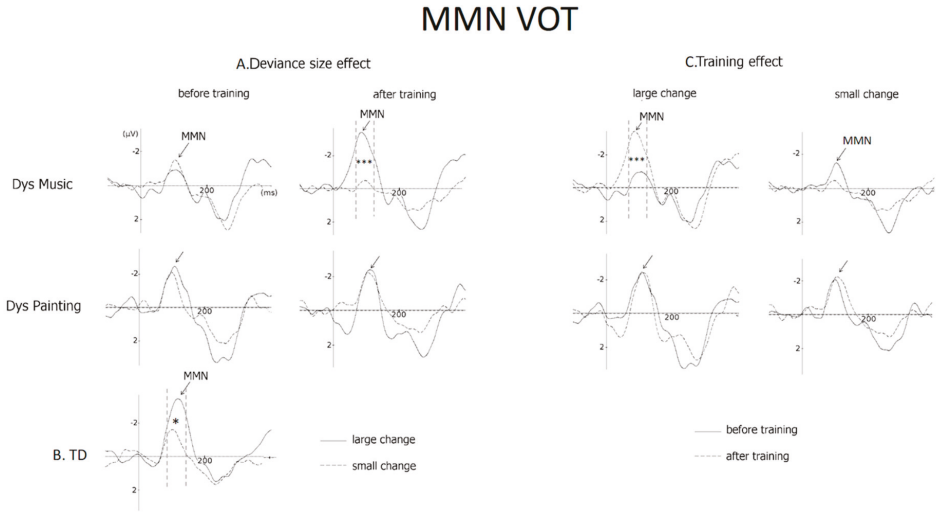


Figure 3. Mismatch Negativities (MMNs) to Voice Onset Time (VOT) deviant stimuli at Fz. (**A**) The deviance size effect (i.e., the difference in MMNs mean amplitude between large and small deviant stimuli) is illustrated before and after training for children with dyslexia trained with music (DysMus, top) or with painting (DysPaint), middle) and (**B**) for Typically Developing children (TD). (**C**) The training effect to large and small changes in VOT is illustrated before (solid lines) and after training (dashed line) for children with dyslexia in the music and in the painting groups.

3.2.2. VOT (N1 Amplitude)

Results of ANOVAs on N1 amplitude that included Group (DysMus vs. DysPaint) as a between-subject factor as well as Session (before vs. after training), Deviance size (large vs. small VOT deviant stimuli), Anterior-Posterior Dimension (frontal, central, and parietal), and Laterality (Left, Central, and Right) as within-subject factors were very similar to those reported above for MMN amplitude. The main effects of Group and Session were not significant ($F < 1$ and $F(1,20) = 3.48$, $p < 0.07$, respectively) but the Group × Session × Deviance size was significant ($F(1,20) = 5.69$, $p < 0.03$).

Before training, results of separate ANOVAs revealed that the VOT deviance size effect on N1 amplitude was not significant either in DysMus (large = 2.12 μV and small = 1.98 μV, $p = 0.99$) or in DysPaint group (large = 1.32 μV and small = 1.54 μV, $p = 0.96$; Group by Deviance size: $F< 1$). By contrast, after training, the deviance size effect was significant in DysMus ($p < 0.02$) with larger N1s to large (0.83 μV) than to small (1.93μV) VOT deviant stimuli, but not in DysPaint (large = 1.93 μV and small = 1.59 μV, $p = 0.85$; Group by Deviance size: ($F(1, 20) = 5.14$, $p < 0.03$); see Figure 4A).

In DysMus, the N1 to large VOT deviant stimuli was larger (i.e., less positive) after (0.83 μV) than before training (2.12 μV, $p < 0.03$) across all scalp sites, with larger differences over the midlines ($p < 0.001$) and the right hemisphere locations ($p < 0.001$) than over the left hemisphere ($p < 0.05$). No such differences were found in DysPaint (after training: 1.93 μV and before training: 1.32 μV, $p = 0.85$; Group by Session by Deviance size by Laterality: $F(2,40) = 2.75$, $p < 0.03$). For small VOT changes, the differences between after and before training were not significant in either group (DysMus = 1.93 μV vs. 1.98μV; DysPaint = 1.59 μV vs. 1.54μV; see Figure 4C).

To specifically test for the normalization of VOT processing with music training, we compared the group of TD children (before training) and the groups of DysMus and DysPaint after training. Results showed that the Group by Deviance size interaction was significant ($F(2,40) = 7.56$, $p < 0.001$). Separate comparisons showed that the deviance size effect on N1 amplitude was similar for TD children before training and DysMus after music training (Group × Deviance size: $F < 1$), but it was larger for TD

children than for DysPaint after painting training (Group × deviance size: F(1,30) = 14.72, $p < 0.001$; compare Figure 4A,B).

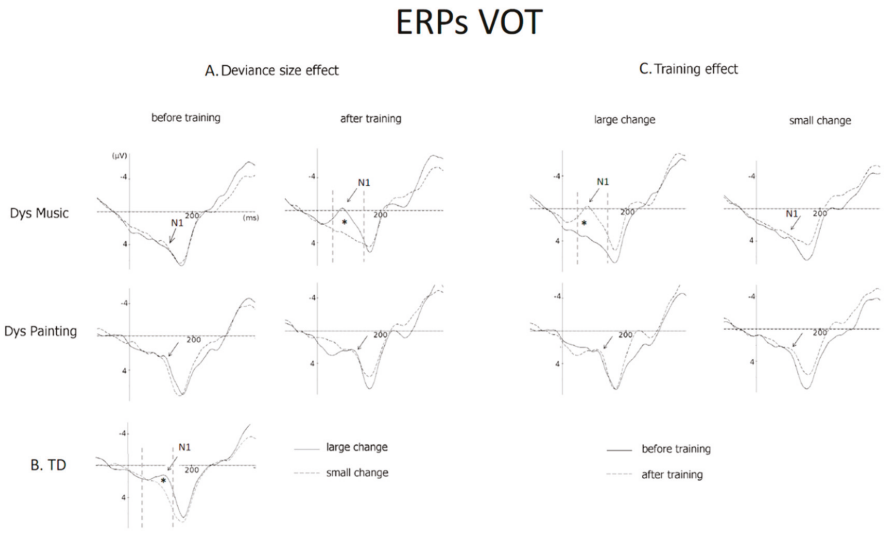


Figure 4. ERPs (original averages) to VOT deviant stimuli at Fz. (**A**) The deviance size effect (i.e., ERPs to large and small deviant stimuli) is illustrated before and after training for children with dyslexia trained with music (DysMus, top) or with painting (DysPaint), middle) and (**B**) for Typically Developing children (TD). (**C**) The training effect to large and small changes in VOT is illustrated before (solid lines) and after training (dashed line) for children with dyslexia in the music and in the painting groups.

3.3. Duration

3.3.1. Duration (MMN Amplitude)

Results of ANOVAs on MMN amplitude including Group (DysMus vs. DysPaint) as a between-subject factor as well as Session (before vs. after training), Deviance size (large vs. small duration deviant stimuli) and Laterality (F3 vs. Fz vs. F4) as within-subjects factors showed that neither the main effects of Group and Session (F(1,20) = 1.13, $p = 0.30$ and F(1,20) = 2.54, $p = 0.12$, respectively) nor the Group by Session by Deviance size and the Group by Session by Deviance size by Laterality interactions were significant (F < 1). Thus, in contrast to our hypothesis, the deviance size effect after training was not significantly different for DysMus or DysPaint (see Figure 5A).

3.3.2. Duration (N250 Amplitude)

Analyses of the N250 amplitude revealed that the main effects of Group and Session were not significant (F < 1 and F(1,20) = 1.37, $p < 0.30$), but there was a trend toward significance in the interaction Group by Session by Deviance size by Laterality (F(4, 80) = 2.32, $p < 0.06$).

Separate ANOVAs for DysMus showed that the N250 to both large and small duration deviant stimuli increased in amplitude from before to after training over frontal and central sites (Session by Anteroposterior interaction: F(2,20) = 5.39, $p < 0.01$; see Table 4). By contrast, for DysPaint, the N250 to duration deviant stimuli was not significantly different before and after training (main effect of Session: F < 1; see Figure 5B).

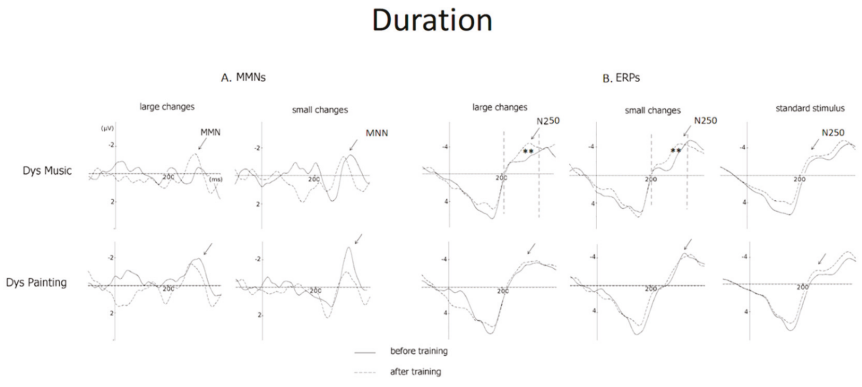


Figure 5. (**A**) Mismatch Negativities (MMNs) and (**B**) ERPs at Fz to large and small duration changes as well as for standard stimuli (for ERPs) before training (solid line) and after training (dashed line) for children with dyslexia in the music group (Dys Mus) and in the painting group (DysPaint).

Table 4. Mean N250 amplitude (in microvolts) at Frontal and Central sites for children with DD trained with music (DysMus) or with painting (DysPaint) before and after training, for large and small duration deviant stimuli.

		Frontal		Central	
		Before	**After**	**Before**	**After**
	Large Dev.	−1.96	−2.99	−1.90	−2.60
DysMus	Small Dev.	−1.88	−2.76	−1.54	−2.13
	Large Dev.	−1.44	−1.37	−1.66	−1.38
DysPaint	Small Dev.	−0.86	−1.19	−0.96	−1.51

3.4. Frequency (MMN and N250 Amplitude)

Results of ANOVAs showed that before training the main effect of Group was significant on MMN amplitude ($F(1,20) = 3.94$, $p < 0.05$, see Figure 6), with larger MMNs in DysPaint (−3.10 μV) than in DysMus (−1.54 μV), thereby precluding further analyses of training effects on the preattentive processing of frequency deviant stimuli.

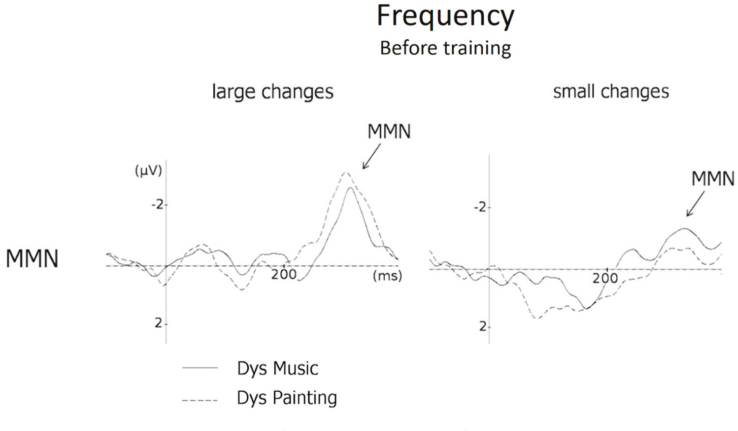


Figure 6. MMN at Fz to large and small frequency changes before training. MMNs are overlapped for children with dyslexia in the music group (solid line) and in the painting group (dashed line).

4. Discussion

The aim of this longitudinal study was to determine whether music training improves the preattentive processing of VOT and duration deviant stimuli in children with DD and normalizes the deviance size effects so that children with dyslexia after music training would not significantly differ from TD children before training. Results clearly support this hypothesis for VOT deviant stimuli but were not as clear-cut for duration deviant stimuli, as discussed below together with results for frequency deviant stimuli.

4.1. VOT Deviant Stimuli

The present study capitalized on previous results in the literature and on specific findings from our group using the same design and stimuli, that showed (1) deficits in the preattentive processing of VOT deviant stimuli in children with DD compared to TD children (cross-sectional study, [21]), (2) improved processing of VOT deviant stimuli in TD children with four years of music training, on average, compared to control nonmusician TD children (cross-sectional study, [68]), and (3) improved processing of VOT deviant stimuli in nonmusician TD children trained with music for 18 months compared to TD children training with painting (longitudinal study, [74]). The novelty of the present experiment is to use a longitudinal approach as in Chobert et al. [74] but with children with DD instead of TD children. Based on the findings summarized above, we hypothesized that six months of music training, but not of painting training, would enhance the preattentive processing of VOT in children with DD. In line with these hypotheses, the VOT deviance size effect was not significant in children with DD (either DysMus or DysPaint) before training but it was significant after music training and not after painting training. This was mainly due to an increase in MMN amplitude to large VOT deviant stimuli with no significant difference for small VOT deviant stimuli. Moreover, the VOT deviance effect for DysMus after music training was not significantly different from the VOT deviance size effect in the control group of TD children before training. By contrast, the VOT deviance size effect was still significantly smaller in DysPaint after training than in TD children before training. Thus, music training, but not painting training, helped to normalize the deviance size effect for VOT deviant stimuli in children with dyslexia. Because the VOT deviance size effect was not significant in either group of children with dyslexia before training and nonsignificant in DysPaint after training, the normalization found for DysMus is more likely to result from the positive influence of music training than from the influence of general factors such as maturation, attention and/or motivation [73].

Results of this longitudinal CRT in children with DD showed that music training improved sensitivity to VOT, a phonological parameter that is contrastive in French, and that in the present design allowed to discriminate the deviant (/pa/) from the standard syllables (/ba/). In other words, music training improved categorical perception, a cornerstone of speech perception that has been largely investigated in children with dyslexia [88–91]. Our findings at the preattentive level, as reflected by increased MMN and N1 amplitude to VOT deviant stimuli, are in line with previous results evidencing enhanced categorical perception with music training in children with DD. For instance, Habib et al. [92] tested for the influence of a Cognitive Musical Training (CMT) method in children with DD and in TD children matched on reading age. The CMT method was designed by speech therapists to include three main components: (1) an auditory component to mobilize the language-music similarity, with exercises based on pitch, duration, tempo, pulsation, and rhythm that aimed at developing both the perception and the production sides; (2) a motor component to engage the child into rhythm production and imitation (e.g., tapping in synchrony with sounds, tapping the written notation of a rhythm ...); and (3) a cross-modal component, to tap into simultaneous processing of information from different modalities including auditory, visual, sensory, and motor modalities as well as their combinations. The rhythmic aspect was always emphasised (for more details, see [93]).

CMT was used for 18 h either during three consecutive days (Experiment 1) or spread over six weeks (Experiment 2), and was based on rhythm production and imitation as well as on cross-modal integration of information from auditory, visual, sensory, and motor modalities. Results showed

that CMT improved the level of performance of children with DD in both the identification (/ba/ to /pa/ continuum) and the discrimination (different /ba/ - /pa/ pairs from the continuum) tasks used to investigate categorical perception. Moreover, differences between children with DD and control TD children after CMT were no longer significant. More generally, such findings are also in line with previous reports that music training improved categorical perception of speech sounds in both younger and older musicians compared to controls without formal music training [94]. Increased auditory sensitivity may thus be one of the driving forces behind enhanced categorical perception with music training.

As mentioned above, MMN amplitude to large VOT deviant stimuli significantly increased from pre to post training but results showed no effect of music or painting training on the preattentive processing of small VOT deviant stimuli. It could be that the difference between the standard and small VOT deviant stimuli was too small to be preattentively detected or that the duration of music training was not sufficiently long to increase auditory sensitivity to small changes in VOT. In line with this interpretation, results of the longitudinal study with TD children conducted by Chobert et al. [74] showed increased MMN to small VOT deviant stimuli after 12 months of music training, but not after six months of music training as done here. Thus, for both TD children and children with DD, more than six months seem necessary to find an effect of active music training on subtle changes in VOT. However, the results of Habib et al. [92] described above showed that only 18 h of music training influenced categorical perception. These different results are possibly linked with differences in the CMT method and the type of music training used here or with the small VOT deviant stimuli being less different from the standard than the stimuli used in the Habib et al. [92] experiment.

Results for the N1 component paralleled those found for the MMN: the VOT deviance size effect on N1 amplitude was not significant before training in either group of children with DD but it was significant after music training (i.e., larger N1 to large than to small VOT changes) and not after painting training. The normalization of the VOT deviance size effect was also reflected on N1 amplitude in DysMus after training (similar to the deviance effects in TD children) but not in DysPaint.

From an acoustic perspective, VOT deviant stimuli differed from standard stimuli right at stimulus onset. Thus, both the obligatory N1 response to stimulus onset and the MMN effects developed in the same latency band. This raises the possibility that the VOT deviance size effect reflects differences on N1 rather than on MMN amplitude [22,87,95,96]. However, two arguments lead us to consider that they are overlapping but different effects. First, the MMN and N1 deviance size effects showed a different scalp distribution. In line with previous studies (see [23,24], for reviews) the MMNs to large VOT deviant stimuli were clearly localized over frontal sites. By contrast, enhancement in N1 amplitude to large VOT deviant stimuli after music training was larger over midlines and right hemisphere than over left hemisphere sites. Thus, different generators seem at the origin of the N1 and MMN effects observed at the scalp. Second, while the N100 component is always time-locked to stimulus onset, the latency of the MMN varies as a function of when participants can perceive the deviance. Thus, in the present experiment, the MMN peaked earlier, ~100 ms, for VOT deviant stimuli than for duration and frequency deviant stimuli—~300 ms (see Figure 1, Figure 3, Figure 5, and Figure 6). Thus, in spite of the temporal overlap between the N1 and the MMN to VOT changes, these two effects may reflect the positive influence of music training on different processes: on the early perceptual processing of VOT (as reflected by the N1 component) and on the preattentive detection of a mismatch between standard and large VOT deviant stimuli (as reflected by the MMN).

Taken together, these results support the hypothesis that music training positively impacts the preattentive processing of VOT, mainly by influencing the perception of large VOT changes. Specifically, these results demonstrate that music training, but not painting training, can normalize the preattentive perception of VOT in children with dyslexia since, after music training, no significant differences were found between children with DD and TD children. These results nicely complement previous findings with TD children trained with music or painting [74]. Finally, they demonstrate transfer effects

from music training to the perception of an acoustic-phonological feature in speech—VOT—which is important for categorical perception.

4.2. Duration Deviant Stimuli

In contrast to our hypothesis, results on MMN amplitude revealed no advantage of music training over painting training on the preattentive processing of large and small duration deviant stimuli (no Group by Session by Deviance size interaction). This conclusion is also in contrast with results reported by [67] showing enhanced MMNs to speech duration deviant stimuli in 10-to-12-year-old children with high musical aptitudes, and with results of Chobert et al. [68], showing larger MMNs to duration deviant stimuli in musician children with an average of four years of music training compared to children with no formal music training. Several factors may explain these discrepancies: the children with DD involved in the present study may not have such high musical aptitudes as in previous studies, may not be as motivated as children who choose to be involved in formal music training or the teaching methods used here are possibly different than those used in classic music school. Moreover, results of the longitudinal study over two years conducted by Chobert et al. [74] with TD children showed that the MMN to duration deviant stimuli was larger after 12 months of music training, but not after six months of music training. It is therefore not surprising that children with DD showed no significant effect of six months of music training on the preattentive processing of duration deviant stimuli. As mentioned above, these children were involved in the same two-year project as the TD children of Chobert et al. [74] but the attrition rate at the end of the second year was too high to obtain reliable results (only six dyslexics were left in the music group and four dyslexics in the painting group).

In contrast with the MMN results, the amplitude of the N250 to both large and small duration deviant stimuli was larger after six months of music training than before training over fronto-central sites with no difference in the painting group. Thus, in line with results at the behavioral and electrophysiological levels showing that musical expertise improved duration discrimination accuracy in speech and perception of the metric structure of speech (e.g., [60,97]), music training seemed to enhance the preattentive perception of duration changes in children with dyslexia. However, this interpretation needs to be considered with caution, first because the Group by Session by Deviance size by Laterality interaction was only marginally significant ($p < 0.06$) and, second, because we found an increase in N250 amplitude that was as large for large than for small duration deviant stimuli. Thus, in contrast to our hypothesis, the deviance size effect was not significant.

4.3. Frequency Deviant Stimuli

As mentioned in the methods section, children with DD were pseudo-randomly assigned to music or to painting training to ensure that no between-group differences were found before training. The assignment was based on different factors (i.e., age, school level, sex, and socioeconomic background) as well as on the results at WISC-IV [81], NEPSY [82] and ODEDYS batteries [98]. However, it was not possible to control that no between-group differences were present before training on all the electrophysiological measures of interest. When results were analysed at the end of the experiment, they showed that before training, MMNs were larger in the painting group than in the music group. As a consequence, we did not conduct further analyses as it would be extremely difficult to disentangle the differences due to training from these significant pretraining differences.

Note that while it could have been interesting to determine whether 6 months of music training improves the preattentive processing of frequency deviant stimuli in children with DD, previous results with the same stimuli showed no significant improvements after 12 months of music training in TD children [74]. Based on these findings, we did not expect to find an effect of only six months of music training in children with DD. Moreover, as mentioned above, mixed results have been reported in cross-sectional studies comparing the processing of frequency changes in children with DD and TD children. For instance, Halliday and collaborators [40] reported no main effect of group on MMN amplitude but smaller Late Discriminative Negativity (LDN; 350–550 ms poststimulus onset)

to small frequency deviant stimuli in children with DD than in TD children. Similarly, Hämäläinen and collaborators [20] reported differences between children with reading disabilities and control children in the rapid processing of pitch changes on P3a but not on MMN amplitude. By contrast, Baldeweg et al. [31] found adult dyslexics to be impaired in auditory frequency discrimination, as reflected by the MMN and Maurer and collaborators reported that children at familial risk for dyslexia have more difficulties than controls to detect frequency deviant stimuli [99].

4.4. Psychometric Tests

Results of MANOVAs including the various psychometric tests as well as DysMus vs. DysPaint and before vs. after training as factors revealed that the level of performance was higher after six months of training than before training on several tests, including reading (Alouette and reading irregular words), verbal IQ (similarities), Rapid Automatized Naming (RAN), phoneme fusion, and Attention (auditory attention and orientation). This main effect of Session can be explained by repetition effects: the same tests were presented twice with typically higher level of performance on second that on first presentation. It may also reflect maturation effects since children were seven months older after compared to before training. However, and most importantly, we found no significant difference between DysMus and DysPaint, so that six months of music training did not improve the attentive use of phonological representations more than painting training. These results stand in contrast with previous results in the literature showing a positive impact of music training on speech perception and on different levels of language processing, including reading abilities. For instance, significant correlations between rise time perception and reading/spelling abilities have been reported in previous studies (e.g., [35,100]). Huss et al. [43] also found strong correlations between the perception of musical meter, sound rise time, phonological awareness, and reading abilities in children with DD and in TD children. In line with the temporal sampling theory of dyslexia proposed by Goswami and collaborators [35,36], rise time is possibly more important for syllabic discrimination than VOT. Another interesting interpretation of our discrepant results is based on the type of music training necessary to find improvements in phonological awareness and reading. For instance, Flaugnacco et al. [101] conducted a CRT with children with DD very similar to the CRT conducted by Chobert et al. [74] and here. They also used music training based on a combination of Kodály and Orff methods but with a strong focus on rhythm, temporal processing, and sensorimotor synchronization. Under these conditions, the level of performance in a metric perception task (i.e., perceiving changes in note duration within recurrent series) predicted reading speed and accuracy as well as phonological processing in Italian children with DD. Thus, focusing training on the rhythmic and motor components of music is possibly more efficient to normalize the attentive perception of the temporal structure of speech in children with DD (e.g., [19]). Finally, other factors may also account for these contrasting results, such the age of the children with DD, the severity and homogeneity of the dyslexia deficit and the impact of the speech therapy.

5. Conclusions

The present results reveal that six months of music training clearly improved the preattentive processing of VOT—a phonological cue determinant for categorical perception [92]—as reflected by strong changes in MMN and N1 amplitude. To a lesser extent, results also showed a larger influence of music compared to painting training on the processing of vowel duration in children with DD, as reflected by increased N250 amplitude to both large and small duration deviant stimuli. However, in contrast to previous results [43,92,100,101], six months of music training were not sufficient to improve attentive phonological processing or reading abilities, as revealed by results at the standardized psychometric tests. Thus, music training differentially influenced the preattentive processing of VOT and duration, as measured with the MMN to large and small deviant stimuli (and the deviance size effect), and the attentive processing of phonological cues, as measured in the psychometric tests. It is very possible that the effects of music training are seen in the brain waves before being seen in behavior,

and that they would manifest in the various behavioral tests after longer training. Moreover, based on the discussion above, focusing music training on the rhythmic and motor components of music is possibly the most efficient strategy to improve speech perception. There is already strong evidence in the literature that improving temporal processing has a strong impact on phonological and reading abilities as well as on semantic and syntactic processing. For instance, Przybylski et al. [102] reported that the level of performance in a syntactic task (decide whether a spoken sentence was syntactically correct or incorrect: e.g., "Laura has/have forgotten her violin") was increased by the prior presentation of rhythmic primes (a succession of notes played either regularly or irregularly). There is also recent evidence for preserved semantic processing in both children [103] and adults with dyslexia [104].

Taken together, these results provide new evidence for a positive influence of music training on preattentive speech perception in children with dyslexia. Importantly, because children were pseudo-randomly assigned to music or to painting training these results more likely reflect the impact of active music training than the influence of genetic predispositions for music. The direct implication of these findings is that rehabilitation methods of dyslexia should focus, at least in part, on restoring the ability to process temporal structures that sequentially unfold in time, such as speech and music. More generally, and based on recent results pointing to multi-deficits (orthographic, phonological and vocabulary) rather than single-deficit problems in children with dyslexia [7], rehabilitation methods should aim at increasing the integration of the different components that are important for reading and learning, using music training and possibly preserved semantic processing abilities in children with dyslexia to overcome their difficulties [103].

Author Contributions: Conceptualization, M.B., J.-L.V., and M.H.; Methodology, A.F., J.C., and C.F.; Formal Analysis, A.F. and J.C.; Resources, M.B., J.-L.V., and M.H.; Data Curation, J.C. and C.F.; Writing—Original Draft Preparation, A.F., J.C., and M.B.; Visualization, A.F. and J.C.; Supervision, M.B.; Project Administration, M.B. and J.-L.V.; Funding Acquisition, M.B.

Funding: This research was supported by a grant from the ANR-Neuro (#024-01) to Mireille Besson. At the time the study was conducted, Julie Chobert and Clément François were Ph.D. students supported by the ANR-Neuro (#024-01).

Acknowledgments: We thank Daniele Schön for helping us prepare the stimuli; Jitpanut Makchoay, Usanee Sotthiwat, Chizuru Deguchi, Julien Marec, Manon Castello, and Martin Chastenet for helping us run part of the experiment; and Carine Verse and Amandine Dettori-Campus for conducting neuropsychological testing. We also thank the directors of the two schools where the children were tested—Muriel Gaiarsa and Jean-Jacques Gaubert—the teachers of the schools as well as all the children who participated in this study and their parents.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Démonet, J.F.; Taylor, M.J.; Chaix, Y. Developmental dyslexia. *Lancet* **2004**, *363*, 1451–1460. [CrossRef]
2. Habib, M. The neurological basis of developmental dyslexia. *Brain* **2000**, *123*, 2373–2399. [CrossRef] [PubMed]
3. Collective Expertise INSERM, CNDRSDI. *Dyslexie, Dysorthographie, Dyscalculie: Bilan des Données Scientifiques*; INSERM: Paris, France, 2007.
4. Norton, E.S.; Beach, S.D.; Gabrieli, J.D. Neurobiology of dyslexia. *Curr. Opin. Neurobiol.* **2015**, *30*, 73–78. [CrossRef] [PubMed]
5. Snowling, M.J. *Dyslexia*; Blackwell: Oxford, UK, 2000.
6. Valdois, S.; Bosse, M.L.; Tainturier, M.J. The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attention disorder. *Dyslexia* **2004**, *10*, 1–25. [CrossRef] [PubMed]
7. Perry, C.; Zorzi, M.; Ziegler, J.C. Understanding Dyslexia Through Personalized Large-Scale Computational Models. *Psychol. Sci.* **2019**, *30*, 386–395. [CrossRef]
8. Ramus, F. Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Curr. Opin. Neurobiol.* **2003**, *13*, 212–218. [CrossRef]

9. Saksida, A.; Iannuzzi, S.; Bogliotti, C.; Chaix, Y.; Demonet, J.F.; Bricout, L.; Billard, C.; Nguyen-Morel, M.A.; Le Heuzey, M.F.; Soares-Boucaud, I.; et al. Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Dev. Psychol.* **2016**, *52*, 1503–1516. [CrossRef]
10. White, S.; Milne, E.; Rosen, S.; Hansen, P.; Swettenham, J.; Frith, U.; Ramus, F. The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Dev. Sci.* **2006**, *9*, 237–255. [CrossRef]
11. Ahissar, M.; Lubin, Y.; PutterKatz, H.; Banai, K. Dyslexia and the failure to form a perceptual anchor. *Nat. Neurosci.* **2006**, *9*, 1558–1564. [CrossRef] [PubMed]
12. Kimppa, L.; Shtyrov, Y.; Partanen, E.; Kujala, T. Impaired neural mechanism for online novel word acquisition in dyslexic children. *Sci. Rep.* **2018**, *24*, 12779. [CrossRef]
13. Thomson, J.M.; Goswami, U. Learning novel phonological representations in developmental dyslexia: Associations with basic auditory processing of rise time and phonological awareness. *Read. Writ.* **2010**, *23*, 453–473. [CrossRef]
14. Ramus, F.; Marshall, C.R.; Rosen, S.; van der Lely, H.K. Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain* **2013**, *136*, 630–645. [CrossRef] [PubMed]
15. Boets, B.; de Beeck, H.; Vandermosten, M.; Scott, S.K.; Gillebert, C.R.; Mantini, D.; Bulthé, J.; Sunaert, S.; Wouters, J.; Ghesquière, P. Intact but less accessible phonetic representations in adults with dyslexia. *Science* **2013**, *342*, 1251–1254. [CrossRef]
16. Lovio, R.; Näätänen, R.; Kujala, T. Abnormal pattern of cortical speech feature discrimination in 6-year-old children at risk for dyslexia. *Brain Res.* **2010**, *1335*, 53–62. [CrossRef] [PubMed]
17. Nagarajan, S.; Mahncke, H.; Salz, T.; Tallal, P.; Roberts, T.; Merzinech, M. Cortical auditory signal processing in poor readers. *Proc. Natl. Acad. Sci. USA* **1999**, *25*, 6483–6488. [CrossRef]
18. Ziegler, J.C.; Ferrand, L. Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychon. Bull. Rev.* **1998**, *5*, 683–689. [CrossRef]
19. Hämäläinen, J.A.; Leppänen, P.H.T.; Guttorm, T.K.; Lyytinen, H. N1 and P2 components of auditory event-related potentials in children with and without reading disabilities. *Clin. Neurophysiol.* **2007**, *118*, 2263–2275. [CrossRef] [PubMed]
20. Hämäläinen, J.A.; Leppänen, P.H.T.; Guttorm, T.K.; Lyytinen, H. Event-related potentials to pitch and rise time change in children with reading disabilities and typically reading children. *Clin. Neurophysiol.* **2008**, *119*, 100–115. [CrossRef] [PubMed]
21. Chobert, J.; François, C.; Habib, M.; Besson, M. Deficit in the preattentive processing of syllables in children with dyslexia. *Neuropsychologia* **2012**, *50*, 2044–2055. [CrossRef]
22. Näätänen, R.; Gaillard, A.W.K.; Mäntysalo, S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* **1978**, *42*, 313–329. [CrossRef]
23. Näätänen, R.; Paavilainen, P.; Rinne, T.; Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin. Neurophysiol.* **2007**, *118*, 2544–2590. [CrossRef]
24. Kujala, T.; Tervaniemi, M.; Schröger, E. The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biol. Psychol.* **2007**, *74*, 1–19. [CrossRef]
25. Näätänen, R.; Pakarinen, S.; Rinne, T.; Takegata, R. The mismatch negativity (MMN): Towards the optimal paradigm. *Clin. Neurophysiol.* **2004**, *115*, 140–144. [CrossRef]
26. Lisker, L.; Abramson, A.S. Some effects of context on voice onset time in English stops. *Lang. Speech* **1967**, *10*, 1–28. [CrossRef] [PubMed]
27. Serniclaes, W. Etude Expérimentale de la Perception du Trait de Voisement des Occlusives du Français. Unpublished Doctoral's Thesis, Université Libre de Bruxelles, Bruxelles, Belgium, 1987.
28. Banai, K.; Hornickel, J.; Skoe, E.; Nicol, T.; Zecker, S.G.; Kraus, N. Reading and subcortical auditory function. *Cereb. Cortex* **2009**, *19*, 2699–2707. [CrossRef] [PubMed]
29. Hämäläinen, J.; Salminen, H.; Leppanen, P. Basic auditory processing deficits in dyslexia; A systematic review of the behavioural and event-related potential field evidence. *J. Learn. Disabil.* **2013**, *46*, 413–427. [CrossRef] [PubMed]
30. Lovio, R.; Pakarinen, S.; Huotilainen, M.; Alku, P.; Silvennoinen, S.; Näätänen, R.; Kujala, T. Auditory discrimination profiles of speech sound changes in 6-year-old children as determined with the multi-feature MMN paradigm. *Clin. Neurophysiol.* **2009**, *120*, 916–921. [CrossRef]

31. Baldeweg, T.; Richardson, A.; Watkins, S.; Foale, C.; Gruzelier, J. Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Ann. Neurol.* **1999**, *45*, 495–503. [CrossRef]
32. Santos, A.; Joly-Pottuz, B.; Moreno, S.; Habib, M.; Besson, M. Behavioral and event-related potentials evidence for pitch discrimination deficits in dyslexic children: Improvement after intensive phonic intervention. *Neuropsychologia* **2007**, *45*, 1080–1090. [CrossRef]
33. Cutini, S.; Szucs, D.; Mead, N.; Huss, M.; Goswami, U. Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia. *Neuroimage* **2016**, *143*, 40–49. [CrossRef] [PubMed]
34. Power, A.J.; Colling, L.J.; Mead, N.; Barnes, L.; Goswami, U. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain Lang.* **2016**, *160*, 1–10. [CrossRef]
35. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
36. Goswami, U.; Power, A.J.; Lallier, M.; Facoetti, A. Oscillatory "temporal sampling" and developmental dyslexia: Toward an over-arching theoretical framework. *Front. Hum. Neurosci.* **2014**, *8*, 904. [CrossRef] [PubMed]
37. Hämäläinen, J.; Rupp, A.; Soltész, F.; Szücs, D.; Goswami, U. Reduced phase locking to slow amplitude modulation in adults with dyslexia: An MEG study. *NeuroImage* **2012**, *59*, 2952–2961. [CrossRef] [PubMed]
38. Lehongre, K.; Ramus, F.; Villiermet, N.; Schwartz, D.; Giraud, A.L. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* **2011**, *72*, 1080–1090. [CrossRef]
39. Cantiania, C.; Ortiz-Mantillab, S.; Rivaa, V.; Piazzac, C.; Bettonia, R.; Musacchia, G.; Moltenia, M.; Marino, C.; Benasich, A.A. Reduced left-lateralized pattern of event-related EEG oscillations in infants at familial risk for language and learning impairment. *Neuroimage Clin.* **2019**, *22*, 101778. [CrossRef]
40. Halliday, L.F.; Barry, J.G.; Hardiman, M.J.; Bishop, D.V.M. Late, not early mismatch responses to changes in frequency are reduced or deviant in children with dyslexia: An event-related potential study. *J. Neurodev. Disord.* **2014**, *6*, 1–15. [CrossRef] [PubMed]
41. Bishop, D.V.M. Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychol. Bull.* **2007**, *133*, 651–672. [CrossRef] [PubMed]
42. Goswami, U.; Huss, M.; Mead, N.; Fosker, T.; Verney, J.P. Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex* **2013**, *49*, 1363–1376. [CrossRef]
43. Huss, M.; Verney, J.P.; Fosker, T.; Mead, N.; Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* **2011**, *47*, 674–689. [CrossRef]
44. Forgeard, M.; Schlaug, G.; Norton, A.; Rosam, C.; Iyengar, U.; Winner, E. The relation between music and phonological processing innormal-reading children and children with dyslexia. *Music Percept. Interdiscip. J.* **2008**, *25*, 383–390. [CrossRef]
45. Frey, A.; François, C.; Chobert, J.; Besson, M.; Ziegler, J. Behavioral and electrophysiological investigation of speech perception deficits in silence, noise and envelope conditions in developmental dyslexia. *Neuropsychologia* **2018**. [CrossRef] [PubMed]
46. Liberman, I.Y.; Shankweiler, D. Phonology and the problems of learning to read and write. *Remedial Spec. Educ.* **1985**, *6*, 8–17. [CrossRef]
47. Overy, K. Dyslexia and music: From timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* **2003**, *999*, 497–505. [CrossRef] [PubMed]
48. Ziegler, J.C.; Goswami, U. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychol. Bull.* **2005**, *131*, 3–29. [CrossRef]
49. Abrams, D.A.; Bhatara, A.; Ryali, S.; Balaban, E.; Levitin, D.J.; Menon, V. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* **2011**, *21*, 1507–1518. [CrossRef] [PubMed]
50. Besson, M. Meaning, structure and time in language and music. Neural substrates of cognitive processes. Special issue in homage to Jean Requin. *Curr. Psychol. Cogn.* **1998**, *17*, 921–951.
51. Besson, M.; Chobert, J.; Marie, C. Transfer of training between music and speech: Common processing, attention, and memory. *Front. Psychol.* **2011**, *2*, 94. [CrossRef] [PubMed]

52. Kraus, N.; Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **2010**, *11*, 599–605. [CrossRef]
53. Maess, B.; Koelsch, S.; Gunter, T.C.; Friederici, A.D. Musical syntax is processed in Broca's area: An MEG study. *Nat. Neurosci.* **2001**, *4*, 540–545. [CrossRef] [PubMed]
54. Patel, A.D. Language, music, syntax and the brain. *Nat. Neurosci.* **2003**, *6*, 674–681. [CrossRef] [PubMed]
55. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: Oxford, UK, 2008.
56. Zatorre, R.J.; Gandour, J.T. Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos. Trans. R. Soc. B Biol. Sci.* **2008**, *363*, 1087–1104. [CrossRef] [PubMed]
57. Kishon-Rabin, L.; Amir, O.; Vexler, Y.; Zaltz, Y. Pitch discrimination: Are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* **2001**, *12*, 125–143. [CrossRef] [PubMed]
58. Marie, C.; Kujala, T.; Besson, M. Musical and linguistic expertise influence preattentive and attentive processing of non-speech sounds. *Cortex* **2012**, *10*, 1016.
59. Spiegel, M.F.; Watson, C.S. Performance on frequency discrimination tasks by musicians and nonmusicians. *J. Acoust. Soc. Am.* **1984**, *76*, 1690–1695. [CrossRef]
60. Tervaniemi, M.; Just, V.; Koelsch, S.; Widmann, A.; Schröger, E. Pitch discrimination accuracy in musicians vs. nonmusicians: An event-related potential and behavioral study. *Exp. Brain Res.* **2005**, *161*, 1–10. [CrossRef]
61. Kühnis, J.; Elmer, S.; Meyer, M.; Jäncke, L. The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia* **2013**, *5l*, 1608–1618. [CrossRef]
62. Zuk, J.; Ozernov-Palchik, O.; Kim, H.; Lakshminarayanan, K.; Gabrieli, J.D.E.; Tallal, P.; Gaab, N. Enhanced Syllable Discrimination Thresholds in Musicians. *PLoS ONE* **2013**, *8*, e80546. [CrossRef]
63. Parbery-Clark, A.; Tierney, A.; Strait, D.L.; Kraus, N. Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* **2012**, *219*, 111–119. [CrossRef]
64. Besson, M.; Dittinger, E.; Barbaroux, M. How music training influences language processing: Evidence against informational encapsulation. *L'année Psychol.* **2018**, *118*, 273–288.
65. Nikjeh, D.A.; Lister, J.J.; Frisch, S.A. The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *J. Acoust. Soc. Am.* **2009**, *125*, 328–338. [CrossRef]
66. Tervaniemi, M.; Rytkönen, M.; Schröger, E.; Ilmoniemi, R.J.; Näätänen, R. Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* **2001**, *8*, 295–300. [CrossRef]
67. Milovanov, R.; Huotilainen, M.; Esquef, P.A.A.; Välimäki, V.; Alku, P.; Tervaniemi, M. The role of musical aptitude and language skills in preattentive duration determination in school-aged children. *Neurosci. Lett.* **2009**, *460*, 161–165. [CrossRef]
68. Chobert, J.; Marie, C.; François, C.; Schön, D.; Besson, M. Enhanced passive and active processing of syllables in musician children. *J. Cogn. Neurosci.* **2011**, *23*, 3874–3887. [CrossRef] [PubMed]
69. Anvari, S.H.; Trainor, L.J.; Woodside, J.; Levy, B.A. Relation among musical skills, phonological processing and early reading ability in preschool children. *J. Exp. Psychol.* **2002**, *83*, 111–130. [CrossRef]
70. Degé, F.; Schwarzer, G. The effect of a music program on phonological awareness in preschoolers. *Front. Psychol.* **2011**, *2*, 124. [CrossRef]
71. Slevc, L.R.; Miyake, A. Individual differences in second language proficiency: Does musical ability matter? *Psychol. Sci.* **2006**, *17*, 675–681. [CrossRef] [PubMed]
72. Moreno, S.; Marques, C.; Santos, A.; Santos, M.; Castro, S.L.; Besson, M. Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb. Cortex* **2009**, *19*, 712–723. [CrossRef] [PubMed]
73. Schellenberg, E.G. Music lessons enhance IQ. *Psychol. Sci.* **2004**, *15*, 511–514. [CrossRef] [PubMed]
74. Chobert, J.; François, C.; Velay, J.-L.; Besson, M. Twelve months of active musical training in 8 to 10 year old children enhances the preattentive processing of syllabic duration and Voice Onset Time. *Cereb. Cortex* **2014**, *24*, 956–967. [CrossRef] [PubMed]
75. Rugg, M.D.; Coles, M.G.H. *Electrophysiology of Mind*; Oxford University Press: New York, NY, USA, 1995.
76. Donchin, E.; Ritter, W.; McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In *Event-Related Brain Potentials in Man*; Callaway, E., Tueting, P., Koslow, S.H., Eds.; Academic Press: New York, NY, USA, 1978; pp. 349–411.
77. Ritter, W.; Simon, R.; Vaughan, H.G., Jr. Event-related potential correlates to two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology* **1983**, *20*, 168–179. [CrossRef]

78. Ceponiene, R.; Shestakova, A.; Balan, P.; Alku, P.; Yiaguchi, K.; Näätänen, R. Children's auditory event-related potentials index sound complexity and "speechness". *Int. J. Neurosci.* **2001**, *109*, 245–260. [CrossRef] [PubMed]
79. Lefavrais, J. *Test de l'Alouette, rev. version*; ECPA: Paris, France, 2005.
80. Bertrand, D.; Fluss, J.; Billard, C.; Ziegler, J.C. Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture [Efficiency, sensitivity, specificity: Comparison of different reading tests]. *Année Psychol.* **2010**, *110*, 299–320. [CrossRef]
81. Wechsler, D. *Wechsler Intelligence Scale for Children*, 4th ed.; WISC-IV; The Psychological Corporation: San Antonio, TX, USA, 2003.
82. Korkman, M.; Kemp, S.L.; Kirk, U. *NEPSY: Bilan Neuropsychologique de L'enfant*; ECPA (Editions du centre de psychologie appliquée): Paris, France, 2004.
83. Raven, J.C. *Standard Progressive Matrices: Sets A, B, C, D, E*; Psychologists Press: Oxford, UK, 1976.
84. Chavez, M.; Day, R.; Deyell, S.; Ellis, P.; Fazio, S.; Green, P.; Johnston, D.; Jonasson, B.; Levine, J.; Orler, T.; et al. Adobe Audition Software. 2003. Available online: http://www.adobe.com/products/audition.html (accessed on 25 April 2011).
85. Boersma, P.; Weenink, D. Praat [Computer Software], Version 4.0. Available online: http://www.fon.hum.uva.nl/praat/ (accessed on 18 April 2001).
86. Jasper, H.A. The ten—twenty system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* **1958**, *10*, 371–375.
87. Schröger, E.; Wolff, C. Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport* **1998**, *9*, 3355–3358. [CrossRef] [PubMed]
88. Dufor, O.; Serniclaes, W.; Sprenger-Charolles, L.; Démonet, J.F. Left premotor cortex and allophonic speech perception in dyslexia: A PET study. *Neuroimage* **2009**, *46*, 241–248. [CrossRef] [PubMed]
89. Hoonhorst, I.; Medina, V.; Colin, C.; Markessis, E.; Radeau, M.; Deltenre, P.; Serniclaes, W. Categorical perception of voicing, colors and facial expressions: A developmental study. *Speech Commun.* **2011**, *53*, 417–430. [CrossRef]
90. Noordenbos, M.W.; Segers, E.; Serniclaes, W.; Mitterer, H.; Verhoeven, L. Neural evidence of allophonic perception in children at risk for dyslexia. *Neuropsychologia* **2012**, *50*, 2010–2017. [CrossRef] [PubMed]
91. Serniclaes, W.; Heghe, S.V.; Mousty, P.; Carré, R.; Sprenger-Charolles, L. Allophonic mode of speech perception in dyslexia. *J. Exp. Child Psychol.* **2004**, *87*, 336–361. [CrossRef]
92. Habib, M.; Lardy, C.; Desiles, T.; Commeiras, C.; Chobert, J.; Besson, M. Music and dyslexia: A new musical training method to improve reading and related disorders. *Front. Psychol.* **2016**, *7*, 22. [CrossRef] [PubMed]
93. Habib, M.; Commeiras, C. *«Mélodys»: Remédiation Cognitivo-Musicale des Troubles de L'apprentissage*; De Boeck: Bruxelles, Belgium, 2014.
94. Bidelman, G.M.; Alain, C. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.* **2015**, *35*, 1240–1249. [CrossRef]
95. Horvath, J.; Czigler, I.; Jacobsen, T.; Maess, B.; Schröger, E.; Winkler, I. MMN or no MMN: No magnitude deviance effect on the MMN amplitude. *Psychophysiology* **2008**, *45*, 60–69. [CrossRef] [PubMed]
96. Näätänen, R.; Alho, K. Mismatch negativity-the measure for central sound representation accuracy. *Audiol. Neurotol.* **1997**, *2*, 341–353. [CrossRef]
97. Marie, C.; Magne, C.; Besson, M. Musicians and the metric structure of words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef] [PubMed]
98. Jacquier-Roux, M.; Valdois, S.; Zorman, M.O. *Outil de Dépistage des Dyslexies*; Cogni-Sciences: Grenoble, France, 2005.
99. Maurer, U.; Bucher, K.; Brem, S.; Brandeis, D. Altered responses to tone and phoneme mismatch in kindergartners at familial dyslexia risk. *Neuroreport* **2003**, *14*, 2245–2250. [CrossRef]
100. Hämäläinen, J.; Leppänen, P.H.T.; Torppa, M.; Müller, K.; Lyytinen, H. Detection of sound rise time by adults with dyslexia. *Brain Lang.* **2005**, *94*, 32–42. [CrossRef]
101. Flaugnacco, E.; Lopez, L.; Terribili, C.; Montico, M.; Zoia, S.; Schön, D. Music Training Increases Phonological Awareness and Reading Skills in Developmental Dyslexia: A Randomized Control Trial. *PLoS ONE* **2015**, *25*, e0138715. [CrossRef] [PubMed]

102. Przybylski, L.; Bedoin, N.; Krifi-Papoz, S.; Herbillon, V.; Roch, D.; Léculier, L.; Kotz, S.A.; Tillmann, B. Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* **2013**, *27*, 121–131. [CrossRef] [PubMed]
103. Van der Kleij, S.W.; Groen, M.A.; Segers, E.; Verhoeven, L. Enhanced semantic involvement during word recognition in children with dyslexia. *J. Exp. Child Psychol.* **2019**, *178*, 15–29. [CrossRef] [PubMed]
104. Silva, P.B.; Ueki, K.; Oliveira, D.G.; Boggio, P.S.; Macedo, E.C. Early Stages of Sensory Processing, but Not Semantic Integration, Are Altered in Dyslexic Adults. *Front. Psychol.* **2016**, *7*, 430. [CrossRef] [PubMed]

brain sciences

MDPI

Review

Neurophysiological Markers of Statistical Learning in Music and Language: Hierarchy, Entropy and Uncertainty

Tatsuya Daikoku

Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; daikoku@cbs.mpg.de; Tel.: +81-5052157012

Received: 10 May 2018; Accepted: 18 June 2018; Published: 19 June 2018

Abstract: Statistical learning (SL) is a method of learning based on the transitional probabilities embedded in sequential phenomena such as music and language. It has been considered an implicit and domain-general mechanism that is innate in the human brain and that functions independently of intention to learn and awareness of what has been learned. SL is an interdisciplinary notion that incorporates information technology, artificial intelligence, musicology, and linguistics, as well as psychology and neuroscience. A body of recent study has suggested that SL can be reflected in neurophysiological responses based on the framework of information theory. This paper reviews a range of work on SL in adults and children that suggests overlapping and independent neural correlations in music and language, and that indicates disability of SL. Furthermore, this article discusses the relationships between the order of transitional probabilities (TPs) (i.e., hierarchy of local statistics) and entropy (i.e., global statistics) regarding SL strategies in human's brains; claims importance of information-theoretical approaches to understand domain-general, higher-order, and global SL covering both real-world music and language; and proposes promising approaches for the application of therapy and pedagogy from various perspectives of psychology, neuroscience, computational studies, musicology, and linguistics.

Keywords: statistical learning; implicit learning; domain generality; information theory; entropy; uncertainty; order; *n*-gram; Markov model; word segmentation

1. Introduction

The brain is a learning system that adapts to multiple external phenomena existing in its living environment, including various types of input such as auditory, visual, and somatosensory stimuli, and various learning domains such as music and language. By means of this wide-ranging system, humans can comprehend structured information, express their own emotions, and communicate with other people [1]. According to linguistic [2,3] and musicological studies [4,5], music and language have domain-specific structures including universal grammar, tonal pitch spaces, and hierarchical tension. Neurophysiological studies likewise suggest that there are specific neural bases for language [6,7] and music comprehension [8,9]. Nevertheless, a body of research suggests that the brain also possesses a domain-general learning system, called statistical learning (SL), that is partially shared by music and language [10,11]. SL is a process by which the brain automatically calculates the transitional probabilities (TPs) of sequential phenomena such as music and language, grasps information dynamics without an intention to learn or awareness of what we know [12,13], and further continually updates the acquired statistical knowledge to adapt to the variable phenomena in our living environments [14]. Some researchers also indicate that the sensitivity to statistical regularities in sequences could be a by-product of chunking [15].

The SL phenomenon can partially be supported by a unified brain theory [16]. This theory tries to provide a unified account of action and perception, as well as learning under a free-energy principle [17,18], which views several keys of brain theories in the biological (e.g., neural Darwinism), physical (e.g., information theory), and neurophysiological (e.g., predictive coding) sciences. This suggests that several brain theories might be unified within a free-energy framework [19], although its capacity to unify different perspectives has yet to be established. This theory suggests that the brain models phenomena in its living environment as a hierarchy of dynamical systems that encode a causal chain structure in the sensorium to maintain low entropy [16], and predicts a future state based on the internalized model to minimize sensory reaction and optimize motor action. This prediction is in keeping with the theory of SL in the brain. That is, in SL theory, the brain models sequential phenomena based on TP distributions, grasps entropy in the whole sequences, and predicts a future state based on the internalized stochastic model in the framework of predictive coding [20] and information theory [21]. The SL also occurs in action sequences [22,23], suggesting that SL could contribute to optimization of motor action.

SL is considered an implicit and ubiquitous process that is innate in humans, yet not unique to humans, as it is also found in monkeys [24,25], songbirds [26,27], and rats [28]. The terms implicit learning and SL have been used interchangeably and are regarded as the same phenomenon [15]. A neurophysiological study [29] has suggested that conditional probabilities in the Western music corpus are reflected in the music-specific neural responses referred to as early right anterior negativity (ERAN) in event-related potential (ERP) [8,9]. The corpus study also found statistical universals in music structures across cultures [30,31]. These findings also suggest that musical knowledge may be at least partially acquired through SL. Our recent studies have also demonstrated that the brain codes the statistics of auditory sequences as relative information, such as relative distribution of pitch and formant frequencies, and that this information can be used in the comprehension of other sequential structures [10,32]. This suggests that the brain does not have to code and accumulate all received information, and thus saves some memory capacity [33]. Thus, from the perspective of information theory [21], the brain's SL is systematically efficient.

As a result of the implicit nature of SL, however, humans cannot verbalize exactly what they statistically learn. Nonetheless, a body of evidence indicates that neurophysiological and behavioural responses can unveil musical and linguistic SL effects [14,32,34–44] in the framework of predictive coding [20]. Furthermore, recent studies have detected the effects of musical training on linguistic SL of words [41,43,45–47] and the interactions between musical and linguistic SL [10] and between auditory and visual SL [44,48–50]. On the other hand, some studies have also suggested that SL is impaired in humans with domain-specific disorders such as dyslexia [51–53] and amusia [54,55], disorders that affect linguistic and music processing, respectively (though Omigie and Stewart (2011) [56] have suggested that SL is intact in congenital amusia). Thiessen et al. [57] suggested that a complete-understanding statistical learning must incorporate two interdependent processes: one is the extracting process that computes TPs (i.e., local statistics) and extracts each item, such as word segmentation, and the other one is the integration process that computes distributional information (i.e., summary statistics) and integrates information across the extracted items. The entropy and uncertainty (i.e., summary statistics), as well as TPs, are used to understand the general predictability of sequences in domain-general SL that could cover music and language in the interdisciplinary realms of neuroscience, behavioral science, modeling, mathematics, and artificial intelligence. Recent studies have suggested that SL strategies in the brain depend on the hierarchy, order [14,35,58,59], entropy, and uncertainty in statistical structures [60]. Hasson et al. [61] also indicated that certain regions or networks perform specific computations of global or summary statistics (i.e., entropy), which are independent of local statistics (i.e., TP). Furthermore, neurophysiological studies suggested that sequences with higher entropy were learned based on higher-order TP, whereas those with lower entropy were learned based on lower-order TP [59]. Thus, it is considered that information-theoretical and neurophysiological concepts on SL link each other [62,63]. The integrated approach of

neurophysiology and informatics based on the notion of order of TP and entropy can shed light on linking concepts of SL among a broad range of disciplines. Although there have been a number of studies on SL in music and language, few studies have examined the relationships between the "order" of TPs (i.e., the order of local statistics) and entropy (i.e., summary statistics) in SL. This article focuses on three themes in SL from the viewpoint of information theory, as well as neuroscience: (1) a mathematical interpretation of SL that can cover music and language and the experimental paradigms that have been used to verify SL; (2) the neural basis underlying SL in adults and children; and (3) the applicability of therapy and pedagogy for humans with learning disabilities and healthy humans.

2. Mathematical Interpretation of Brain SL Process Shared by Music and Language

2.1. Local Statistics: Nth-Order Transitional Probability

According to SL theory, the brain automatically computes TP distributions in sequential phenomena (local statistics) [35], grasps uncertainty/entropy in the whole sequences (global statistics) [61], and predicts a future state based on the internalized statistical model to minimize sensory reaction [16,20]. The TP is a conditional probability of an event B given that the latest event A has occurred, written as P(B | A). The TP distributions sampled from sequential information such as music and language are often expressed by nth-order Markov models [64] or n-gram models [21] (Figure 1). Although the terminology of n-gram models has frequently been used in natural language processing, it has also recently been used in music models [65,66]. They have often been applied to develop artificial intelligence that gives computers learning abilities similar to those of the human brain, thus generating systems for data mining, automatic music composition [67–69], and automatic text classification in natural language processing [70,71]. The mathematical model of SL including nth-order Markov and $(n + 1)$-gram models is the conditional probability of an event e_{n+1}, given the preceding n events based on Bayes' theorem:

$$P(e_{n+1} \mid e_n) = P(e_{n+1} \cap e_n) / P(e_n) \quad (1)$$

From the viewpoint of psychology, the formula can be interpreted as positing that the brain predicts a subsequent event e_{n+1} based on the preceding events e_n in a sequence. In other words, learners expect the event with the highest TP based on the latest n states, whereas they are likely to be surprised by an event with lower TP (Figure 2).

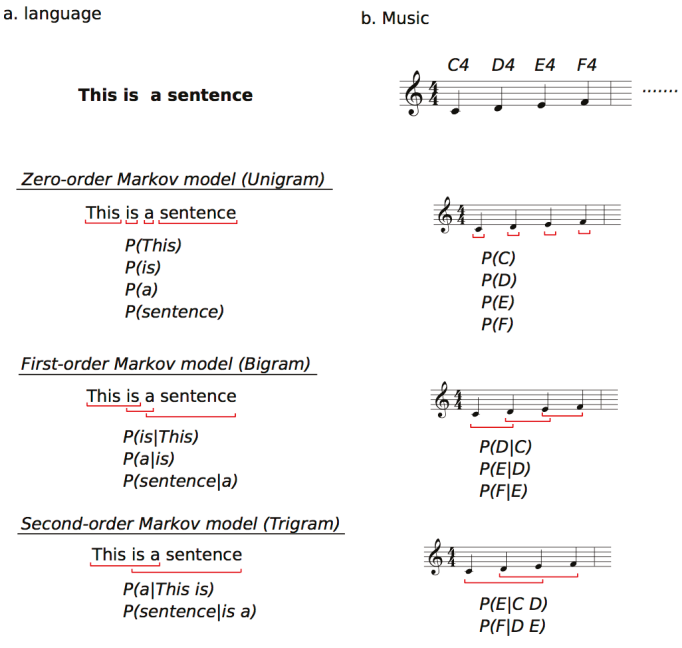


Figure 1. Example of n-gram and Markov models in statistical learning (SL) of language (**a**) and music (**b**) based on information theory. The top are examples of sequences, and the others explain how to calculate TPs ($P(e_{n+1} \mid e_n)$) based on zero- to second-order Markov models. They are based on the conditional probability of an event e_{n+1}, given the preceding n events based on Bayes' theorem. For instance, in language ((**a**), This is a sentence), the second-order Markov model represents that the "a" can be predicted based on the last subsequent two words of "This" and "is". In music ((**b**), C4, D4, E4, F4), second-order Markov model represents that the "E" can be predicted based on the last subsequent two tones of "C" and "D".

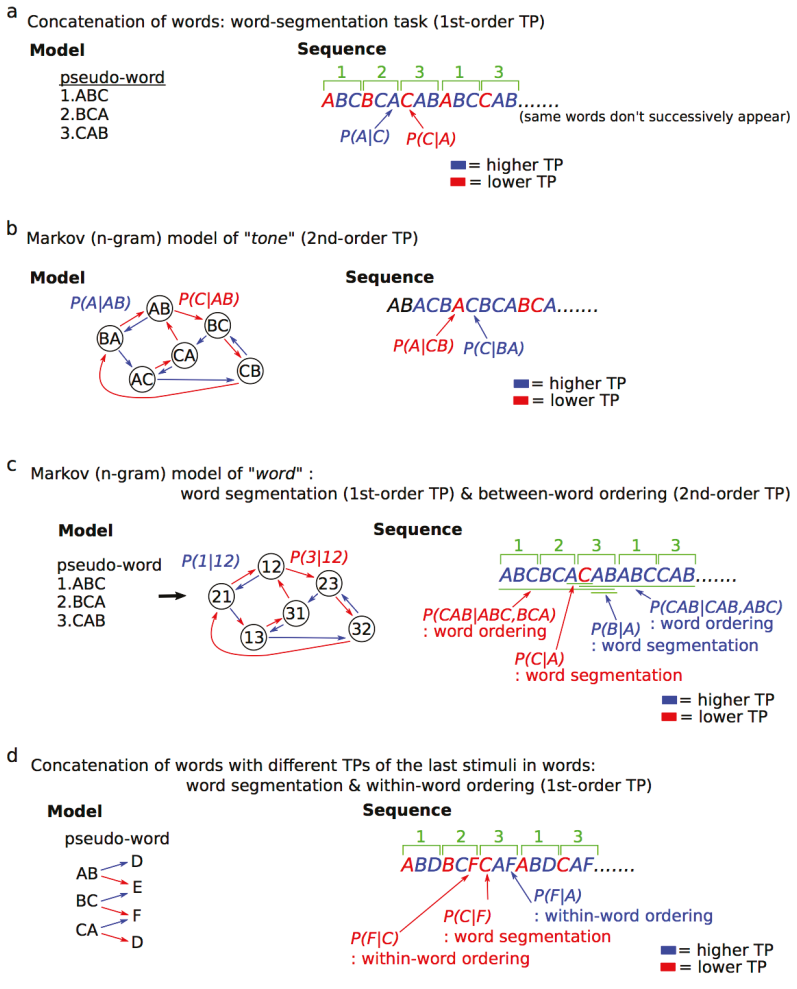


Figure 2. SL models and the sequences used in neural studies. All of the models and paradigms in sequences based on concatenation of words (**a**), Markov model of tone (**b**) and word (**c**), and concatenation of words with different TPs of the last stimuli in words (**d**) are simplified so that the characteristics of paradigms can be compared. In the example of word-segmentation paradigm (**a**), the same words do not successively appear. TP—transitional probability.

2.2. Global Statistics: Entropy and Uncertainty

SL models are sometimes evaluated in terms of entropy [72–75] in the framework of information theory, as done by Shannon [21]. Entropy can be calculated from probability distribution, interpreted as the average surprise (uncertainty) of outcomes [16,76], and used to evaluate the neurobiology of SL [60], as well as rule learning [77], decision making [78], anxiety, and curiosity [79,80] from the perspective of uncertainty. For instance, the conditional entropy (H(B | A)) in the *n*th order TP distribution (hereafter, Markov entropy) can be calculated from information contents:

$$H(X_{i+1} \mid X_i) = -\Sigma P(x_i)\Sigma P(x_{i+1} \mid x_i)\ log_2 P(x_{i+1} \mid x_i) \quad (2)$$

where $H(X_{i+1} \mid X_i)$ is the Markov entropy; $P(X_i)$ is the probability of event xi occurring; and $P(X_{i+1} \mid X_i)$ is the probability of X_{i+1}, given that X_i occurs previously. Previous articles have suggested that the degree of Markov entropy modulates human predictability in SL [61,81]. The uncertainty (i.e., global/summary statistics), as well as the TP (i.e., local statistics), of each event is applicable to and may be used to predict many types of sequential distributions, such as music and language, and to understand the predictability of a sequence (Figure 3). Indeed, entropy and uncertainty are often used to understand domain-general SL in the interdisciplinary realms of neuroscience, behavioural science, modeling, mathematics, and artificial intelligence.

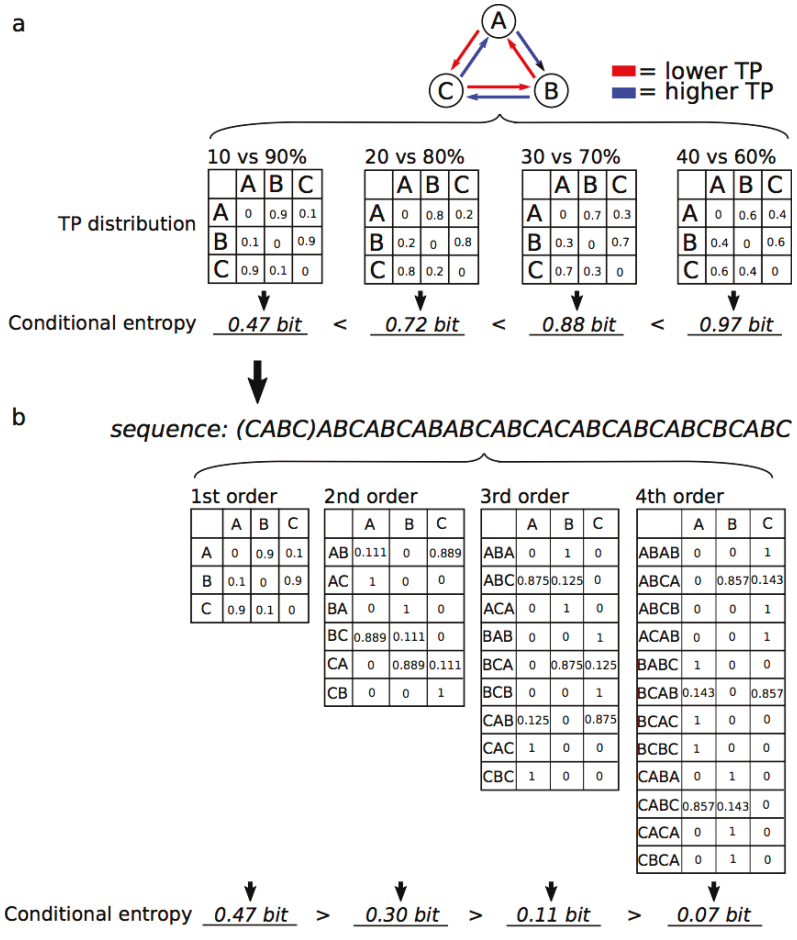


Figure 3. The entropy (uncertainty) of predictability in the framework of SL. The uncertainties depend on (**a**) TP ratios in a first-order Markov model (i.e., bigram model) and (**b**) orders of models in the TP ratio of 10% vs. 90%.

2.3. Experimental Designs of SL in Neurophysiological Studies

The word segmentation paradigm is frequently used to examine the neural basis underlying SL (e.g., [34,41,43,44,46,82–96]). This paradigm basically consists of a concatenation of pseudo-words (Figure 2a). In the pseudo-words sequence, the TP distributions based on a first-order Markov model represent lower TPs in the "first" stimulus of each word (Figure 2a: P(B | A), P(C | B), and

P(A | C)) than other stimuli of word (Figure 2a: P(C | A), P(A | A), P(A | B), P(B | B), P(B | C), and P(C | C)). When the brain statistically learns the sequences, it can identify the boundaries between words based on first-order TPs (Figure 2a) [97,98], and segment/extract each word. The SL of word segmentation based on first-order TPs has been considered as a mechanism for language acquisition in the early stages of language learning, even in infancy [12]. Recent studies have also demonstrated that SL can be performed based on within-word, as well as between-word, TPs ([40,98] for example, see Figure 2d). Although a number of studies have used a word segmentation paradigm consisting of words with a regular unit length (typically, three stimuli within a word), previous studies suggest that the unit length of words [99], the order of TPs [59], and the nonadjacent dependencies of TPs in sequences ([14,100–102] for example, see Figure 2c) can modulate the SL strategy used by the brain. Indeed, natural languages and music make use of higher-order statistics, including hierarchical, syntactical structures. To understand the brain's higher-order SL systems in a form closer to that used for natural language and music, sequential paradigms based on higher-order Markov models have also been used in neurophysiological studies ([32,35,103] for example, see Figure 2b). Furthermore, the *n*th-order Markov model has been applied to develop artificial intelligence that gives computers learning and decision-making abilities similar to those of the human brain, thus generating systems for automatic music composition [67–69] and natural language processing [70,71]. Information-theoretical approaches, including information content and entropy based on *n*th-order Markov models, may be useful in understanding the domain-general SL, as it functions in response to real-world learning phenomena in the interdisciplinary realms of brain and computational sciences.

3. Neural Basis of Statistical Learning

3.1. Event-Related Responses and Oscillatory Activity

The ERP and event-related magnetic fields (ERF) modalities directly measure brain activity during SL and represent a more sensitive method than the observation of behavioral effects [40,41,104]. Based on predictive coding [20], when the brain encodes the TP distributions of a stimulus sequence, it expects a probable future stimulus with a high TP and inhibits the neural response to predictable external stimuli for efficiency of neural processing. Finally, the effects of SL manifest as a difference in the ERP and ERF amplitudes between stimuli with lower and higher TPs (Figure 4). Although many studies of word segmentation detected SL effects on the N400 component [43,46,88,89,93,94,105], which is generally considered to reflect a semantic meaning in language and music [106–108], auditory brainstem response (ABR) [96], P50 [41], N100 [94], mismatch negativity (MMN) [40,44,98], P200 [46,89,105], N200–250 [44,47], and P300 [83] have also been reported to reflect SL effects (Table 1). In addition, other studies using Markov models also reported that SL is reflected in the P50 [14,36,37], N100 [10,14,32,35], and P200 components [35]. Compared with later auditory responses such as N400, the auditory responses that peak earlier than 10 ms after stimulus presentation (e.g., ABR) and at 20–80 ms, which is around P50 latency, have been attributed to parallel thalamo–cortical connections or cortico–cortical connections between the primary auditory cortex and the superior temporal gyrus [109]. Thus, the suppression of an early component of auditory responses to stimuli with a higher TP in lower cortical areas can be interpreted as the transient expression of prediction error that is suppressed by predictions from higher cortical areas in a top-down connection [96]. Thus, top-down, as well as bottom-up, processing in SL may be reflected in ERP/ERF. On the other hand, SL effects on N400 have been detected in word-segmentation tasks, but not in the Markov model. TPs of a word-segmentation task are calculated based on first-order models (Figure 2a). In other words, in terms of the "order" of TP, SL of word segmentation (i.e., sequence consisting of word concatenation) and first-order Markov model have same hierarchy of TP. Nevertheless, SL studies using the first-order Markov model did not detect learning effects of N400 (Table 1). The phenomenon of word segmentation itself has been considered as a mechanism of language acquisition in the early stages of language learning [12]. Several papers claim that the sensitivity to statistical regularities in sequences of word concatenation

could be a by-product of chunking [15]. Neurophysiological effects of word segmentation, such as N400, reflecting a semantic meaning in language [106–108] may be associated with the neural basis underlying linguistic functions, as well as statistical computation itself. On the other hand, our previous study using the first-order Markov model [36] struggled to detect N400 in terms of a stimulus onset asynchrony of sequences (i.e., 500 ms). A future study will be needed to verify SL effects of N400 using the Markov model.

Table 1. Overview of neurophysiological correlations with auditory statistical learning. TP—transitional probability; ABR—auditory brainstem response; MMN—mismatch negativity; STS—superior temporal sulcus; STG—superior temporal gyrus; IFG—inferior frontal gyrus; PMC—premotor cortex; PTC—posterior temporal cortex.

Paradigms	Order of TP	Neural Correlates	References
Word segmentation	First-order	ABR	Skoe et al., 2015 [96]
		P50	Paraskevopoulos et al., 2012 [41]
		N100	Sanders et al., 2002 [94]
		MMN	Koelsch et al., 2016 [40] Moldwin et al., 2017 [98] Francois et a., 2017 [44]
		P200	De Diego Balaguer et al., 2007 [89] Francois et al., 2011 [46] Cunillera et al., 2006 [105]
		N200–250	Mandikal Vasuki et al., 2017 [47] Francois et al., 2017 [44]
		P300	Batterink et al., 2015 [83]
		N400	Cunillera et al., 2009 [88], 2006 [105] De Diego Balaguer et al., 2007 [89] Sanders et al., 2002 [94] Francois et al., 2011 [46]; 2013 [43]; 2014 [93]
		STS, STG	Farthouat et al., 2017 [91] Tremblay et al., 2012 [110] Paraskevopoulos et al., 2017 [45]
		Left IFG	Abla and Okanoya, 2008 [111] McNealy et al., 2006 [112] Paraskevopoulos et al., 2017 [45]
		PMC	Cunillera et al., 2009 [88]
		Hippocampus	Schapiro et al., 2014 [113]
Markov model	First-order	P50	Daikoku et al., 2016 [36]
		Wernicke's area	Bischoff-Grethe et al., 2000 [114]
		Hippocampus	Harrison et al., 2006 [60]
	Higher-order	P50	Daikoku et al., 2017 [14]; 2017 [37]
		N100	Furl et al., 2011 [35] Daikoku et al., 2014 [32]; 2015 [10]; 2017 [14]
		P200	Furl et al., 2011 [35]
		Right PTC	Furl et al., 2011 [35]

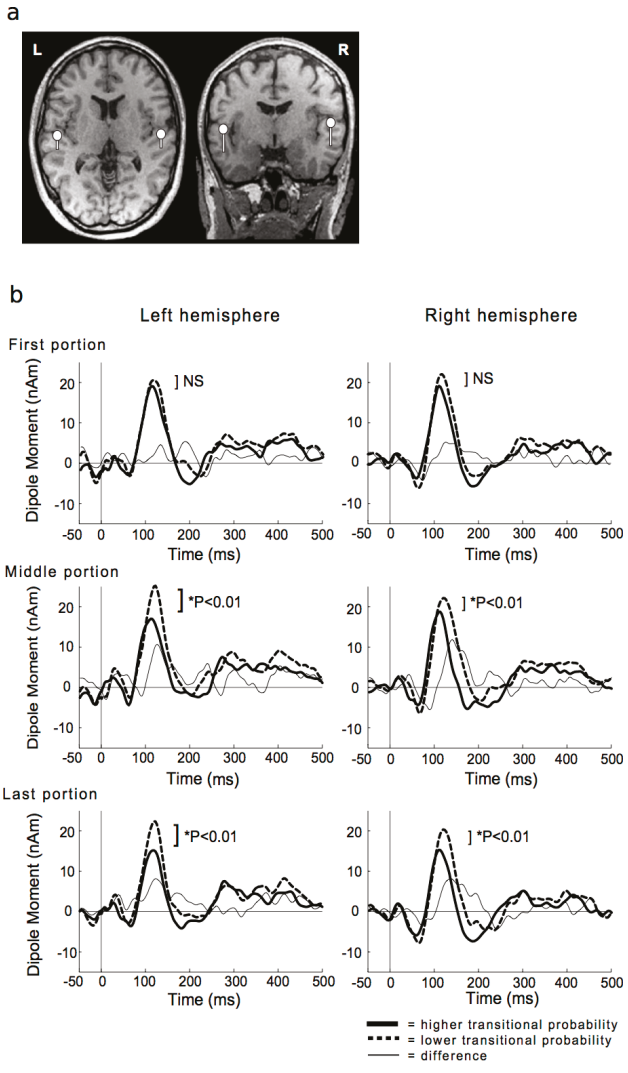


Figure 4. Representative equivalent current dipole (ECD) locations (dots) and orientations (bars) for the N100 m responses superimposed on the magnetic resonance images (**a**) (Daikoku et al., 2014 [32]; and the SL effects (**b**) (Daikoku et al., 2015 [10]) (NS = not significant). When the brain encodes the TP in a sequence, it expects a probable future stimulus with a high TP and inhibits the neural response to predictable stimuli. In the end, the SL effects manifest as a difference in amplitudes of neural responses to stimuli with lower and higher TPs (**b**).

It has been suggested that SL could also be reflected in oscillatory responses in the theta band [115,116]. Moreover, the human and monkey auditory cortices represent the neural marker of predictability based on SL in the form of modulations of transient theta oscillations coupling with gamma and concomitant effects [25], suggesting that SL processes are unlikely to have evolved convergently and are not unique to humans. According to previous studies, low-frequency oscillations may play an important role in speech segmentation associated with SL [73], and in tracking the envelope of the speech

signal, whereas high-frequency oscillations are fundamentally involved in tracking the fine structure of speech [117]. Furthermore, there is evidence of top-down effects in low-frequency oscillations during listening to speech (up to beta band: 15–30 Hz), whereas bottom-up processing dominates in higher frequency bands [118]. Studies on the auditory oddball paradigm have also demonstrated that the power and/or coherence of theta oscillations to low-probability sounds is increased relative to high-probability sounds. Thus, many studies suggest that the lower-frequency oscillations, including theta band, are related to the prediction error [119]. Top-down predictions also control the coupling between speech and low-frequency oscillations in the left frontal areas, most likely in the speech motor cortex [120]. Although low-frequency oscillations could cover ERP components that have been suggested to reflect SL effects, the studies on oscillation and prediction imply the importance of investigating SL effects on oscillatory responses, as well as ERP.

3.2. Anatomical Mechanisms

3.2.1. Local Statistics: Transitional Probability

Neuroimaging studies have indicated that both cortical and subcortical areas play an important role in SL. For instance, the auditory association cortex, including the superior temporal sulcus (STS) [91] and superior temporal gyrus (STG) [110], contributes to auditory SL of both speech and non-speech sounds. Previous studies have also reported the effects of laterality on SL. For instance, functional magnetic resonance imaging (fMRI) [121] and near-infrared spectroscopy (NIRS) [111] studies have suggested that SL is linked to the left auditory association cortex or the left inferior frontal gyrus (IFG) [112,122], which include Wernicke's and Broca's areas, respectively. Furthermore, one previous study has indicated that brain connectivity between bilateral superior temporal sources and the left IFG is important for auditory SL [45]. On the other hand, another study has shown that the right posterior temporal cortex (PTC), which represents the high levels of the peri-Sylvian auditory hierarchy, is related to higher-order auditory SL [35] (i.e., second-order TPs). Further study will be needed to examine the relationships between the order of TPs in sequences and the neural correlations that depend on the order of TPs and hierarchy of SL.

Some studies have suggested that the sensory type of each stimulus modulates the neural basis underlying SL. For instance, some previous studies have suggested that the right hemisphere contributes to visual SL [123]. Paraskevopoulos and colleagues [50] revealed that the cortical network underlying audiovisual SL was partly common with and partly distinct from the unimodal networks of visual and auditory SL, comprising the right temporal and left inferior frontal sources, respectively. fMRI studies have also reported that Heschl's gyrus and the medial temporal lobe [124] contribute to auditory and visual SL, respectively [113], and that motor cortex activity also contributes to visual SL of action words [22]. Furthermore, Cunillera et al. [88] have suggested that the superior part of the ventral premotor cortex (PMC), as well as the posterior STG, are responsible for SL of word segmentation, suggesting that linguistic SL is related to an auditory–motor interface. Another study has suggested that the abstraction of acquired statistical knowledge is associated with a gradual shift from memory systems in the medial temporal lobe, including the hippocampus, to those of the striatum, and that this may be mediated by slow wave sleep [125].

3.2.2. Global Statistics: Entropy

Perceptive mechanisms of summary structure (i.e., global statistics) are considered to be independent of the prediction of each stimulus with different TPs (local statistics) [57,61]. Recent studies have examined the brain systems that are responsible for encoding the uncertainty of global statistics in sequences by comparing brain activities while listening to Markov/word-concatenation and random sequences, which have lower and higher entropies, respectively. Regardless of whether music or language is assessed, the hippocampus and the lateral temporal region [88], including Wernicke's area [114], are considered to play important roles in encoding uncertainty and conditional entropy of statistical information [60].

Bischoff-Grethe et al. have also indicated that Wernicke's area may not be exclusively associated with uncertainty of language information [114]. Furthermore, uncertainty in auditory and visual statistics is coded by modality-general, as well as modality-specific, neural mechanisms [126,127], supporting the hypothesis that the neural basis underlying the brain's perception of global statistics (i.e., uncertainty), as well as local statistics (i.e., prediction of each stimulus with different TPs), is a domain-general system. Our previous neural study also suggested that reorganization of acquired statistical knowledge requires more time than the acquisition of new statistical knowledge, even if the new and previously acquired information sets have equivalent entropy levels [14]. Furthermore the results suggested that humans learn larger structures, such as phrases, first and subsequently extract smaller structures, such as words, from the learned phrases (global-to-local learning strategy). To the best of our knowledge, however, no study has yet demonstrated the differences and neural basis interactions between global and local statistics. Further study is needed to reveal how the coding of global statistics affects that of local statistics.

4. Clinical and Pedagogical Viewpoints

4.1. Disability

Although SL is a domain-general system, some studies have reported that SL is impaired in domain-specific disabilities such as dyslexia [51–53] and amusia [54,55], which are language- and music-related disabilities, respectively. Ayotte and colleagues [128] have suggested that individuals with congenital amusia fail to learn music SL but can learn linguistic SL, even if the sequences of both types have the same degree of statistical regularity [54]. Another study has suggested, in contrast, that SL is intact in amusia [56], and that individuals with amusia lack confidence in their SL ability, although they can engage in SL of music. Peretz et al. [54] stated that the input and output of the statistical computation might be domain-specific, whereas the learning mechanism might be domain-general. Furthermore, previous studies have indicated that SL ability is impaired in patients with damage to a specific area of the brain. For instance, SL is impaired in connection with hippocampal [129] and right-hemisphere damage [130]. Indeed, it has been suggested that the hippocampus plays an important role in SL [124]. One recent study indicated that auditory deprivation leads to disability of not only auditory SL [131] but also visual SL [132]. This implies that there may be specific neural mechanisms for SL that can be shared among distinct sensory modalities. Another study [133], however, suggested that a period of early deafness is not associated with SL disability. Further study is needed to clarify whether SL disability is related to temporary auditory deprivation.

4.2. Music-to-Language Transfer

4.2.1. Neural Underpinnings of SL That Overlap across Music and Language Processing

Because of the acoustic similarity [134], cortical overlap [135,136], and domain generality of SL across language and music, experienced listeners to particular spectrotemporal acoustic features, such as rhythm and pitch, in either speech or music have an advantage when perceiving similar features in the other domain [137]. According to neural studies, musical training leads to a different gray matter concentration in the auditory cortex [138] and a larger planum temporale (PT) [139–143]; the region where both language and music are processed. An ERP study has demonstrated that both the linguistic and the musical effects of SL on the N100–P200 response, which could originate in the belt and parabelt auditory regions [144,145], were larger in musicians than in non-musicians [46]. Thus, the increased PT volume associated with musical training may facilitate auditory processing in SL. A magnetoencephalographic (MEG) study also reported that the effect of SL on the P50 response was larger in musicians than in non-musicians [41], suggesting that musical training also boosts corticofugal projections in a top-down manner regarding predictive coding [96].

Musical training could also facilitate the effects of SL on N400 [46], which is considered to be associated with IFG and PMC [88]. According to the results of a neural study, musicians have an increased gray matter density of the left IFG (i.e., Broca's area) and PMC [146]. Other studies have suggested that, during SL of word segmentation, musicians exhibit increased left-hemispheric theta coherence in the dorsal stream projecting from the posterior superior temporal (pST) and inferior parietal (IP) brain regions toward the prefrontal cortex, whereas non-musicians show stronger functional connectivity in the right hemisphere [115]. An MRI study also demonstrated that SL of word segmentation leads to pronounced left-hemisphere activity of the supratemporal plane, IP lobe, and Broca's area [147]. Thus, the left dorsal stream is considered to play an important role in SL, as well as language [7] and music learning [148].

The SL of word segmentation plays an important role in various speech abilities. Recent studies have revealed a strong link between SL of word segmentation and more general linguistic proficiency such as expressive vocabulary [149] and foreign language [150]. An fMRI study [151] has suggested that, during SL of word segmentation, participants with strong SL effects of familiar language on which they had been pretrained had decreased recruitment of fronto-subcortical and posterior parietal regions, as well as a dissociation between downstream regions and early auditory cortex, whereas participants with strong SL effects of novel language that had never been exposed showed the opposite trend. Furthermore, children with language disorders perform poorly when compared with typical developing children in tasks involving musical metrical structures [152], and have more difficulty in SL of word segmentation [153] and perception of speech rhythms [154,155]. Thus, musical training, including rhythm perception and production, is important for the development of language skills in children. Together, a body of study indicates that musical expertise may transfer to language learning [104]. It is generally considered that the left auditory cortex is more sensitive to temporal information, such as musical beat and the voice-onset (VOT) time of consonant-vowel (CV) syllables, whereas the right auditory cortex plays a role in spectral perception, such as pitch and vowel discriminations. Recent studies have indicated relationships between rhythm perception and SL [156].

Recent neural studies have demonstrated that SL of speech, pitch, timbre, and chord sequences can be performed and reflected in ERP/ERF [10,36,37,40,46]. Furthermore, the brain codes statistics of auditory sequences as relative information, such as relative distribution of pitch and formant frequencies, which could be used for comprehension of another sequential structure [10,32], suggesting that SL is ubiquitous and domain-general. On the other hand, the relative importance of acoustic features such as rhythm, pitch, intensity, and timbre varies depending on the domain, that is, music or language [157]. For instance, unlike spoken language, music contains various pitch frequencies. Recent studies have suggested that, compared with speech sequences, sung sequences with various pitches facilitate auditory SL based on word segmentation [92] and the Markov model [10]. These results further support the advantage of musical training for language SL. In addition, Hansen and colleagues have suggested that musical training also facilitates the hippocampal perception of global statistics of entropy (i.e., uncertainty) [158], as well as local statistics of each TP. Thus, musical training contributes to the improvement of SL systems in various brain regions, including the auditory cortex. Together, the facilitation of SL may be related to enhancement of the left dorsal stream via the IFG and PMC, as well as PT, enhanced low-level auditory processing in a top-down manner, and enhanced hippocampal processing. Musical training including rhythm perception contributes to these enhancements and facilitates the involvement of SL in language skills, and thus could be an important clinical and pedagogical strategy in persons with any of a variety of language-related disorders such as dyslexia [159,160] and aphasia [161].

4.2.2. Children and Adults: Critical Periods and Plasticity in the Brain

Previous studies have demonstrated that auditory SL can be performed even by sleeping neonates [85,86,162]. SL is ubiquitously performed at birth, showing that the human brain is innately prepared for it. An infant's SL extends to rhythms [163], visual stimuli [164], objects [165],

social learning [23,166], and a general mechanism by which infants form meaningful representations of the environment [167]. Furthermore, infants can also learn non-adjacent statistics [101]. This suggests that SL plays an important role in an infant's syntactic learning, as well as the simple segmentation of words. These results may enable us to disentangle the respective contributions of nature and nurture in the acquisition of language and music. On the other hand, an MEG study has suggested that the strategies for language acquisition in infants could shift from domain-general SL to domain-specific processing of native language between 6 and 12 months [116], a "critical period" for language acquisition [168]. A comparable developmental change from domain-general to domain-specific learning strategies can also occur in music perception [169]. During the "critical period" of heightened plasticity, the brain is formed by sensory experience [170–172]. The development of primary cortical acoustic representations can be shaped by the higher-order TP of stimulus sequences [58]. An ERP study [173] suggested that sensitivity to speech stimuli in infants gradually shifts from accentuation to repetition during a critical period. These results may suggest that cortical reorganization depending on early experience interacts with SL [174], and that fluctuations in the degree of dependence on SL for the acquisition of language and music are part of the developmental process during critical periods. On the other hand, the SL system in the brain can be preserved even in adults (e.g., [32,35,40,41]). According to previous studies, neural plasticity can occur in adults through SL [175] and musical training [176]. In fact, there is no doubt that SL occurs in adults who are already beyond the critical periods, and that their SL ability can be modulated by auditory training. Recent studies have revealed that the process of reorganization of acquired statistical knowledge can be detected in neurophysiological responses [14]. Furthermore, a computational study on music suggested the possibility that the time-course variation of statistical knowledge over a composer's lifetime can be reflected in that composer's music from different life stages [177]. Thus, implicit updates of statistical knowledge could be enabled by the combined and interdisciplinary approach of brain, behavioral, and computational methodologies [178].

5. General Discussion

5.1. Information-Theoretical Notions for Domain-General SL: Order of TP and Entropy

SL is a domain-general and interdisciplinary notion in psychology, neuroscience, musicology, linguistics, information technology, and artificial intelligence. To generate SL models that are applicable to all of these various realms, the *n*th-order Markov and *n*-gram models based on information theory have frequently been used in natural language processing [70,71] and in the creation of automatic music composition systems [67–69]. Such models can verify hierarchies of SL based on various-order TPs. Natural languages and music include higher-order statistics, such as hierarchical syntactical structures and grammar. Thus, information-theoretical approaches, including information content and entropy based on *n*th-order Markov models [59,61,81], can express domain-general statistical structures closer to those of real-world language and music. The SL models are often evaluated in terms of entropy [72–75]. From a psychological viewpoint, entropy is interpreted as the average surprise (uncertainty) of outcomes [16,76]. Previous studies have demonstrated that the perception of entropy and uncertainty based on SL could be reflected in neurophysiological responses [59] and activity of the hippocampus [60]. Hasson et al. [61] indicated that certain regions or networks perform specific computations of global or summary statistics (i.e., entropy), which are independent of local statistics (i.e., TP). Furthermore, Thiessen and colleagues [57] proposed that a complete-understanding statistical learning must incorporate two interdependent processes: one is the extracting process that computes TPs and extracts each item, such as word segmentation, and the other one is the integration process that computes distributional information and integrates information across the extracted items. Our previous studies [59] investigated correlation among entropy, order of TP, and the SL effect. As a result, the SL effects of sequences with higher entropy were lower than those with lower entropy, even when TP itself is same between these two sequences. This suggests that an evaluation of computational model of sequential information by entropy in the field of informatics may partially be

able to predict learning effect in human's brain. Thus, the integrated methodology of neurophysiology and informatics based on the notion of entropy can shed light on linking the concept of SL among a broad range of disciplines. To understand the domain-general SL system that incorporates notions from both information theory and neuroscience, it is important to investigate both global and local SL.

5.2. Output of Statistical Knowledge: From Learning to Using

According to recent studies, acquired statistical knowledge contributes to the comprehension and production of complex structural information, such as music and language [179], intuitive decision-making [77,78,180–182], auditory-motor planning [183], and creativity involved in musical composition [62]. Several studies suggest that musical representation is mainly formed by a tacit knowledge [184–186]. Thus, statistical knowledge is closely tied to musical and speech expression such as composition, playing, and conversation. In addition, global statistical knowledge (i.e., entropy and uncertainty), as well as local statistical knowledge (each TP), is also supposed to contribute to decision-making [78], anxiety [80], and curiosity [79]. A number of studies have reported, however, that humans cannot verbalize exactly what they have learned statistically, even when an SL effect is detected in neurophysiological responses [14,32,34–44]. Nevertheless, our previous study suggested that statistical knowledge could alternatively be expressed via abstract medium such as musical melody [32]. In these studies, learners could behaviorally distinguish between sequences with more than eight tones with only higher TPs and those with only lower TPs, suggesting that humans can distinguish sequences with different TPs when they are provided longer sequences when compared with a conventional way in word-segmentation studies that present sequences with three tones. These studies may also suggest that that SL of auditory sequences partially interact with the Gestalt principle [5]. Furthermore, an fMRI study has suggested that the abstraction of statistical knowledge is associated with a gradual shift from the memory systems in the medial temporal lobe, including the hippocampus, to those of the striatum, and that this may be mediated by slow wave sleep [125]. Future study is needed to examine how/when statistical learning contributes to mental expression of music and language.

5.3. Applicability in Clinical and Pedagogy

Previous studies suggest that neurophysiological correlations of SL can disclose subtle individual differences that might be underestimated by behavioral levels [34,88,89,187], although recent studies showed individual differences in SL by behavioral tasks [188]. Some studies suggest that neurophysiological responses disclose SL effects, even when no SL effects cannot be detected in behavioral levels [40,41]. Neurophysiological markers of SL may at least be informative when studying less accessible populations such as infants, who are unable to deliver an obvious behavioral response [86,162]. For instance, ERP/ERF could be a useful method for the evaluation of the individual ability of SL, which is linked to individual skill in language and music learning [189,190], and which is impaired in humans with language- and music-based learning impairments such as dyslexia [51–53] and amusia [54,55]. Thus, neurophysiological markers of SL may be applicable for the evaluation of therapeutic and educational effects for patients and healthy humans [191] across any domain in which the conditional probabilities of sequential events vary systematically. Francois's findings [43] suggest the possibility of music-based remediation for children with language-based SL impairments. In addition, by using information theoretic approaches such as higher-order Markov models and entropy, SL ability can be evaluated in the form that is closest to that used in learning natural language and music [14,63]. The integration of neural, behavioral, and information-theoretical approaches may enhance our ability to evaluate SL ability in terms of both music and language.

5.4. Challenges and Future Prospects: SL in Real-World Music and Language

Although SL is generally considered domain-general, many studies also report that comprehension of language and music, which have domain-specific structures including universal grammar, tonal pitch spaces, and hierarchical tension [2–5], may rely on domain-specific neural

bases [6–9,192]. Furthermore, current SL paradigms are not sufficient to account for all levels of the music- and language-learning process. Some studies suggest two steps of the learning process [193,194]. The first is SL, which shares a common mechanism among all the domains (domain generality). The second is domain-specific learning, which has different mechanisms in each domain (domain specificity). This learning process implies that, at least in an earlier step of the learning process, SL plays an essential role that covers music and language learning abilities [195]. On the other hand, few studies investigated how statistically acquired knowledge was represented in real-world communication, conversation, action, and music expression. Future studies will be needed to investigate how neural systems underlying SL contribute to comprehension and production in real-world music and language. Information-theoretical approaches based on higher-order Markov models can be used to understand SL systems in a form closer to that used for natural language and music, from a perspective of linguistics, musicology, and a unified brain theory such as the free-energy principle [16], including optimisation of action, as well as perception and learning.

6. Conclusions

This paper reviews a body of recent neural studies on SL in music and language, and discusses the possibility of therapeutic and pedagogical application. Because of a certain degree of acoustic similarity, neural overlap, and domain generality of SL between speech and music, musical training positively affects language skills in SL. Recent studies also suggested that SL strategies in the brain depend on the hierarchy, order [14,35,58,59], entropy, and uncertainty in statistical structures [60], and that certain brain regions perform specific computations of entropy that are independent of those of TP [61]. Yet few studies have investigated the relationships between the order of TPs (i.e., order of local statistics) and entropy (i.e., global statistics) in terms of SL strategies of the human brain. Information-theoretical approaches based on higher-order Markov models that can express hierarchical information dynamics as they are expressed in real-world language and music represent a possible means of understanding domain-general, higher-order, and global SL in the interdisciplinary realms of psychology, neuroscience, computational studies, musicology, and linguistics.

Funding: This work was supported by Suntory Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Ackermann, H.; Hage, S.R.; Ziegler, W. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav. Brain Sci.* **2014**, *37*, 529–604. [CrossRef] [PubMed]
2. Chomsky, N. *Syntactic Structures*; Mouton: The Hague, The Netherlands, 1957.
3. Hauser, M.D.; Chomsky, N.; Fitch, W.T. The faculty of language: What is it, who has it, and how did it evolve? *Science* **2002**, *298*, 1569–1579. [CrossRef] [PubMed]
4. Lerdahl, F.; Jackendoff, R. *A Generative Theory of Tonal Music*; MIT Press: Cambridge, MA, USA, 1983.
5. Jackendoff, R.; Lerdahl, F. The capacity for music: What is it, and what's special about it? *Cognition* **2006**, *100*, 33–72. [CrossRef] [PubMed]
6. Friederici, A.D.; Pfeifer, E.; Hahne, A. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Brain Res. Cogn. Brain Res.* **1993**, *1*, 183–192. [CrossRef]
7. Friederici, A.D.; Chomsky, N.; Berwick, R.C.; Moro, A.; Bolhuis, J.J. Language, mind and brain. *Nat. Hum. Behav.* **2017**, *1*, 713–722. [CrossRef]
8. Koelsch, S.; Gunter, T.; Friederici, A.D.; Schroger, E. Brain indices of music processing: "Non-musicians" are musical. *J. Cogn. Neurosci.* **2000**, *12*, 520–541. [CrossRef] [PubMed]
9. Koelsch, S. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology* **2009**, *46*, 179–190. [CrossRef] [PubMed]

10. Daikoku, T.; Yatomi, Y.; Yumoto, M. Statistical learning of music- and language-like sequences and tolerance for spectral shifts. *Neurobiol. Learn. Mem.* **2015**, *118*, 8–19. [CrossRef] [PubMed]
11. Saffran, J.R.; Johnson, E.K.; Aslin, R.N.; Newport, E.L. Statistical learning of tone sequences by human infants and adults. *Cognition* **1999**, *70*, 27–52. [CrossRef]
12. Saffran, J.R.; Aslin, R.N.; Newport, E.L. Statistical learning by 8-month-old infants. *Science* **1996**, *274*, 1926–1928. [CrossRef] [PubMed]
13. Cleeremans, A.; Destrebecqz, A.; Boyer, M. Implicit learning: News from the front. *Trends Cogn. Sci.* **1998**, *2*, 406–416. [CrossRef]
14. Daikoku, T.; Yatomi, Y.; Yumoto, M. Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia* **2017**, *95*, 1–10. [CrossRef] [PubMed]
15. Perruchet, P.; Pacton, S. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends Cogn. Sci.* **2006**, *10*, 233–238. [CrossRef] [PubMed]
16. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [CrossRef] [PubMed]
17. Friston, K.; Kilner, J.; Harrison, L. A free energy principle for the brain. *J. Physiol. Paris* **2006**, *100*, 70–87. [CrossRef] [PubMed]
18. Friston, K.; Kiebel, S. Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2009**, *364*, 1211–1221. [CrossRef] [PubMed]
19. Von Helmholtz, H. *Treatise on Physiological Optics*, 3rd ed.; Courier Corporation: Hamburg, Germany, 1909.
20. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B* **2005**, *360*, 815–836. [CrossRef] [PubMed]
21. Shannon, C.E. Prediction and entropy of printed english. *Bell Syst. Tech. J.* **1951**, *30*, 50–64. [CrossRef]
22. De Zubicaray, G.; Arciuli, J.; McMahon, K. Putting an "end" to the motor cortex representations of action words. *J. Cogn. Neurosci.* **2013**, *25*, 1957–1974. [CrossRef] [PubMed]
23. Monroy, C.D.; Gerson, S.A.; Domínguez-Martínez, E.; Kaduk, K.; Hunnius, S.; Reid, V. Sensitivity to structure in action sequences: An infant event-related potential study. *Neuropsychologia* **2017**. [CrossRef] [PubMed]
24. Saffran, J.R.; Hauser, M.; Seibel, R.; Kapfhamer, J.; Tsao, F.; Cushman, F. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* **2008**, *107*, 479–500. [CrossRef] [PubMed]
25. Kikuchi, Y.; Attaheri, A.; Wilson, B.; Rhone, A.E.; Nourski, K.V.; Gander, P.E.; Kovach, C.K.; Kawasaki, H.; Griffiths, T.D.; Howard, M.A., 3rd; et al. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol.* **2017**, *15*, e2000219. [CrossRef] [PubMed]
26. Lu, K.; Vicario, D.S. Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14553–14558. [CrossRef] [PubMed]
27. Lu, K.; Vicario, D.S. Familiar but Unexpected: Effects of Sound Context Statistics on Auditory Responses in the Songbird Forebrain. *J. Neurosci.* **2017**, *37*, 12006–12017. [CrossRef] [PubMed]
28. Toro, J.M.; Trobalón, J.B. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* **2005**, *67*, 867–875. [CrossRef] [PubMed]
29. Kim, S.G.; Kim, J.S.; Chung, C.K. The effect of conditional probability of chord progression on brain response: An MEG study. *PLoS ONE* **2011**, *6*. [CrossRef] [PubMed]
30. Savage, P.E.; Brown, S.; Sakai, E.; Currie, T.E. Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8987–8992. [CrossRef] [PubMed]
31. Stevens, C.J. Music perception and cognition: A review of recent cross-cultural research. *Top. Cogn. Sci.* **2012**, *4*, 653–667. [CrossRef] [PubMed]
32. Daikoku, T.; Yatomi, Y.; Yumoto, M. Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia* **2014**, *63*, 194–204. [CrossRef] [PubMed]
33. Olshausen, B.A.; Field, D.J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **2004**, *14*, 481–487. [CrossRef] [PubMed]
34. Abla, D.; Katahira, K.; Okanoya, K. On-line assessment of statistical learning by event related potentials. *J. Cogn. Neurosci.* **2008**, *20*, 952–964. [CrossRef] [PubMed]
35. Furl, N.; Kumar, S.; Alter, K.; Durrant, S.; Shawe-Taylor, J.; Griffiths, T.D. Neural prediction of higher-order auditory sequence statistics. *Neuroimage* **2011**, *54*, 2267–2277. [CrossRef] [PubMed]
36. Daikoku, T.; Yatomi, Y.; Yumoto, M. Pitch-class distribution modulates the statistical learning of atonal chord sequences. *Brain Cogn.* **2016**, *108*, 1–10. [CrossRef] [PubMed]

37. Daikoku, T.; Yumoto, M. Single, but not dual, attention facilitates statistical learning of two concurrent auditory sequences. *Sci. Rep.* **2017**, *7*, 10108. [CrossRef] [PubMed]
38. Daikoku, T.; Takahashi, Y.; Futagami, H.; Tarumoto, N.; Yasuda, H. Physical fitness modulates incidental but not intentional statistical learning of simultaneous auditory sequences during concurrent physical exercise. *Neurol. Res.* **2017**, *39*, 107–116. [CrossRef] [PubMed]
39. Daikoku, T.; Takahashi, Y.; Tarumoto, N.; Yasuda, H. Auditory Statistical Learning during Concurrent Physical Exercise and the Tolerance for Pitch, Tempo, and Rhythm Changes. *Motor Control* **2017**, *5*, 1–24. [CrossRef] [PubMed]
40. Koelsch, S.; Busch, T.; Jentschke, S.; Rohrmeier, M. Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Sci. Rep.* **2016**, *6*, 19741. [CrossRef] [PubMed]
41. Paraskevopoulos, E.; Kuchenbuch, A.; Herholz, S.C.; Pantev, C. Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia* **2012**. [CrossRef] [PubMed]
42. François, C.; Tillmann, B.; Schön, D. Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 108–115. [CrossRef] [PubMed]
43. François, C.; Chobert, J.; Besson, M.; Schön, D. Music training for the development of speech segmentation. *Cereb. Cortex* **2013**, *23*, 2038–2043. [CrossRef] [PubMed]
44. François, C.; Cunillera, T.; Garcia, E.; Laine, M.; Rodriguez-Fornells, A. Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia* **2017**, *98*, 56–67. [CrossRef] [PubMed]
45. Paraskevopoulos, E.; Chalas, N.; Bamidis, P. Functional connectivity of the cortical network supporting statistical learning in musicians and non-musicians: An MEG study. *Sci. Rep.* **2017**, *7*, 16268. [CrossRef] [PubMed]
46. François, C.; Schön, D. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* **2011**, *21*, 2357–2365. [CrossRef] [PubMed]
47. Mandikal Vasuki, P.R.; Sharma, M.; Ibrahim, R.; Arciuli, J. Statistical learning and auditory processing in children with music training: An ERP study. *Clin. Neurophysiol.* **2017**, *128*, 1270–1281. [CrossRef] [PubMed]
48. Mitchel, A.D.; Christiansen, M.H.; Weiss, D.J. Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Front. Psychol.* **2014**, *5*, 407. [CrossRef] [PubMed]
49. Conway, C.M.; Christiansen, M.H. Modality-constrained statistical learning of tactile visual and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **2005**, *31*, 24–39. [CrossRef] [PubMed]
50. Paraskevopoulos, E.; Chalas, N.; Kartsidis, P.; Wollbrink, A.; Bamidis, P. Statistical learning of multisensory regularities is enhanced in musicians: An MEG study. *Neuroimage* **2018**, *175*, 150–160. [CrossRef] [PubMed]
51. Vicari, S.; Finzi, A.; Menghini, D.; Marotta, L.; Baldi, S.; Petrosini, L. Do children with developmental dyslexia have an implicit learning deficit? *J. Neurol. Neurosurg. Psychiatry* **2005**, *76*, 1392–1397. [CrossRef] [PubMed]
52. Howard, J.H., Jr.; Howard, D.V.; Japikse, K.C.; Eden, G.F. Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia* **2006**, *44*, 1131–1144. [CrossRef] [PubMed]
53. Menghini, D.; Hagberg, G.E.; Caltagirone, C.; Petrosini, L.; Vicari, S. Implicit learning deficits in dyslexic adults: An fMRI study. *Neuroimage* **2006**, *33*, 1218–1226. [CrossRef] [PubMed]
54. Peretz, I.; Saffran, J.; Schön, D.; Gosselin, N. Statistical learning of speech, not music, in congenital amusia. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 361–367. [CrossRef] [PubMed]
55. Loui, P.; Schlaug, G. Impaired learning of event frequencies in tone deafness. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 354–360. [CrossRef] [PubMed]
56. Omigie, D.; Stewart, L. Preserved statistical learning of tonal and linguistic material in congenital amusia. *Front. Psychol.* **2011**, *2*, 109. [CrossRef] [PubMed]
57. Thiessen, E.D.; Kronstein, A.T.; Hufnagle, D.G. The extraction and integration framework: A two-process account of statistical learning. *Psychol. Bull.* **2013**, *139*, 792–814. [CrossRef] [PubMed]
58. Köver, H.; Gill, K.; Tseng, Y.T.; Bao, S. Perceptual and neuronal boundary learned from higher-order stimulus probabilities. *J. Neurosci.* **2013**, *33*, 3699–3705. [CrossRef] [PubMed]
59. Daikoku, T.; Okano, T.; Yumoto, M. Relative difficulty of auditory statistical learning based on tone transition diversity modulates chunk length in the learning strategy. In Proceedings of the Biomagnetic, Sendai, Japan, 22–24 May 2017; p. 75.
60. Harrison, L.M.; Duggins, A.; Friston, K.J. Encoding uncertainty in the hippocampus. *Neural Netw.* **2006**, *19*, 535–546. [CrossRef] [PubMed]

61. Hasson, U. The neurobiology of uncertainty: Implications for statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 1711. [CrossRef] [PubMed]
62. Pearce, M.; Wiggins, G. Auditory expectation: The information dynamics of music perception and cognition. *Top. Cogn. Sci.* **2012**, *4*, 625–652. [CrossRef] [PubMed]
63. Pearce, M.T.; Ruiz, M.H.; Kapasi, S.; Wiggins, G.A.; Bhattacharya, J. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage* **2010**, *50*, 302–313. [CrossRef] [PubMed]
64. Markov, A.A. *Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain*; Markov Chains; John Wiley and Sons: Hoboken, NJ, USA, 1971; Volume 1.
65. Pearce, M.T.; Wiggins, G.A. Improved methods for statistical modelling of monophonic music. *J. New Music Res.* **2004**, *33*, 367–385. [CrossRef]
66. Rohrmeier, M.A.; Cross, I. Modelling unsupervised online-learning of artificial grammars: Linking implicit and statistical learning. *Conscious. Cogn.* **2014**, *27*, 155–167. [CrossRef] [PubMed]
67. Raphael, C.; Stoddard, J. Functional harmonic analysis using probabilistic models. *Comput. Music J.* **2004**, *28*, 45–52. [CrossRef]
68. Boenn, G.; Brain, M.; De Vos, M.; Ffitch, J. Automatic composition of melodic and harmonic music by answer set programming. In *International Conference on Logic Programming*, ICLP 2008, 5366 ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 160–174.
69. Eigenfeldt, A.; Pasquier, P. Realtime Generation of Harmonic Progressions Using Controlled Markov Selection. In Proceedings of the ICCC-X-Computational Creativity Conference, New York, NY, USA, 7–9 January 2010; pp. 16–25.
70. Brent, M.R. Speech segmentation and word discovery: A computational perspective. *Trends Cogn. Sci.* **1999**, *3*, 294–301. [CrossRef]
71. Manning, C.D.; Schutze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
72. Pearce, M.; Wiggins, G. Expectation in melody: The influence of context and learning. *Music Percept.* **2006**, *23*, 377–405. [CrossRef]
73. Manzara, L.C.; Witten, I.H.; James, M. On the entropy of music: An experiment with Bach chorale melodies. *Leonardo* **1992**, *2*, 81–88. [CrossRef]
74. Reis, B.Y. Simulating Music Learning with Autonomous Listening Agents: Entropy, Ambiguity and Context. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 1999.
75. Cox, G. On the relationship between entropy and meaning in music: An exploration with recurrent neural networks. In Proceedings of the Cognitive Science Society, Portland, OR, USA, 11–14 August 2010; Volume 32.
76. Applebaum, D. *Probability and Information: An Integrated Approach*; Cambridge Univ. Press: Cambridge, UK, 2008.
77. Bach, D.R.; Dolan, R.J. Knowing how much you don't know: A neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **2012**, *13*, 572–586. [CrossRef] [PubMed]
78. Summerfield, C.; de Lange, F.P. Expectation in perceptual decision making: Neural and computational mechanisms. *Nat. Rev. Neurosci.* **2014**, *15*, 745–756. [CrossRef] [PubMed]
79. Loewenstein, G. The psychology of curiosity: A review and reinterpretation. *Psychol. Bull.* **1994**, *116*, 75–98. [CrossRef]
80. Hirsh, J.B.; Mar, R.A.; Peterson, J.B. Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychol. Rev.* **2012**, *119*, 304–320. [CrossRef] [PubMed]
81. Agres, K.; Abdallah, S.; Pearce, M. Information-Theoretic Properties of Auditory Sequences Dynamically Influence Expectation and Memory. *Cogn. Sci.* **2018**, *42*, 43–76. [CrossRef] [PubMed]
82. Abla, D.; Okanoya, K. Visual statistical learning of shape sequences: An ERP study. *Neurosci. Res.* **2009**, *64*, 185–190. [CrossRef] [PubMed]
83. Batterink, L.J.; Reber, P.J.; Neville, H.J.; Paller, K.A. Implicit and explicit contributions to statistical learning. *J. Mem. Lang.* **2015**, *83*, 62–78. [CrossRef] [PubMed]
84. Batterink, L.J.; Paller, K.A. Online neural monitoring of statistical learning. *Cortex* **2017**, *90*, 31–45. [CrossRef] [PubMed]
85. Bosseler, A.N.; Teinonen, T.; Tervaniemi, M.; Huotilainen, M. Infant Directed Speech Enhances Statistical Learning in Newborn Infants: An ERP Study. *PLoS ONE* **2016**, *11*, e0162177. [CrossRef] [PubMed]

86. Teinonen, T.; Fellman, V.; Näätänen, R.; Alku, P.; Huotilainen, M. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* **2009**, *13*, 10–21. [CrossRef] [PubMed]
87. Teinonen, T.; Huotilainen, M. Implicit segmentation of a stream of syllables based on transitional probabilities: An MEG study. *J. Psycholinguist. Res.* **2012**, *41*, 71–82. [CrossRef] [PubMed]
88. Cunillera, T.; Càmara, E.; Toro, J.M.; Marco-Pallares, J.; Sebastián-Galles, N.; Ortiz, H.; Pujol, J.; Rodríguez-Fornells, A. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* **2009**, *48*, 541–553. [CrossRef] [PubMed]
89. De Diego Balaguer, R.; Toro, J.M.; Rodriguez-Fornells, A.; Bachoud-Lévi, A.C. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* **2007**, *2*, e1175. [CrossRef] [PubMed]
90. Buiatti, M.; Peña, M.; Dehaene-Lambertz, G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* **2009**, *44*, 509–519. [CrossRef] [PubMed]
91. Farthouat, J.; Franco, A.; Mary, A.; Delpouve, J.; Wens, V.; Op de Beeck, M.; De Tiège, X.; Peigneux, P. Auditory Magnetoencephalographic Frequency-Tagged Responses Mirror the Ongoing Segmentation Processes Underlying Statistical Learning. *Brain Topogr.* **2017**, *30*, 220–232. [CrossRef] [PubMed]
92. Francois, C.; Schön, D. Learning of musical and linguistic structures: Comparing event-related potentials and behavior. *Neuroreport* **2010**, *21*, 928–932. [CrossRef] [PubMed]
93. François, C.; Jaillet, F.; Takerkart, S.; Schön, D. Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE* **2014**, *9*, e101340. [CrossRef] [PubMed]
94. Sanders, L.D.; Newport, E.L.; Neville, H.J. Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* **2002**, *5*, 700–703. [CrossRef] [PubMed]
95. Sanders, L.D.; Ameral, V.; Sayles, K. Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia* **2009**, *47*, 1183–1186. [CrossRef] [PubMed]
96. Skoe, E.; Krizman, J.; Spitzer, E.; Kraus, N. Prior experience biases subcortical sensitivity to sound patterns. *J. Cogn. Neurosci.* **2015**, *27*, 124–140. [CrossRef] [PubMed]
97. François, C.; Schön, D. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hear. Res.* **2014**, *308*, 122–128. [CrossRef] [PubMed]
98. Moldwin, T.; Schwartz, O.; Sussman, E.S. Statistical Learning of Melodic Patterns Influences the Brain's Response to Wrong Notes. *J. Cogn. Neurosci.* **2017**, *29*, 2114–2122. [CrossRef] [PubMed]
99. Hoch, L.; Tyler, M.D.; Tillmann, B. Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychon. Bull. Rev.* **2013**, *20*, 142–147. [CrossRef] [PubMed]
100. Frost, R.L.; Monaghan, P. Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition* **2016**, *147*, 70–74. [CrossRef] [PubMed]
101. Kabdebon, C.; Pena, M.; Buiatti, M.; Dehaene-Lambertz, G. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain Lang.* **2015**, *148*, 25–36. [CrossRef] [PubMed]
102. Newport, E.L.; Aslin, R.N. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **2004**, *48*, 127–162. [CrossRef]
103. Yumoto, M.; Daikoku, T. IV Auditory system. 5 basic function. In *Clinical Applications of Magnetoencephalography*; Tobimatsu, S., Kakigi, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 97–112.
104. Schon, D.; Francois, C. Musical expertise and statistical learning of musical and linguistic structures. *Front. Psychol.* **2011**, *2*, 167. [CrossRef] [PubMed]
105. Cunillera, T.; Toro, J.M.; Sebastián-Gallés, N.; Rodríguez-Fornells, A. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Res.* **2006**, *1123*, 168–178. [CrossRef] [PubMed]
106. Koelsch, S.; Kasper, E.; Sammler, D.; Schulze, K.; Gunter, T.; Friederici, A.D. Music, language and meaning: Brain signatures of semantic processing. *Nat. Neurosci.* **2004**, *7*, 302–307. [CrossRef] [PubMed]
107. Tillmann, B.; Koelsch, S.; Escoffier, N.; Bigand, E.; Lalitte, P.; Friederici, A.D.; Von Cramon, D.Y. Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *Neuroimage* **2006**, *31*, 1771–1782. [CrossRef] [PubMed]
108. Kutas, M.; Federmeier, K.D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* **2011**, *62*, 621–647. [CrossRef] [PubMed]

109. Adler, L.E.; Pachtman, E.; Franks, R.D.; Pecevich, M.; Waldo, M.C.; Freedman, R. Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biol. Psychiatry* **1982**, *17*, 639–654. [PubMed]
110. Tremblay, P.; Baroni, M.; Hasson, U. Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* **2012**, *66*, 318–332. [CrossRef] [PubMed]
111. Abla, D.; Okanoya, K. Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia* **2008**, *46*, 2787–2795. [CrossRef] [PubMed]
112. McNealy, K.; Mazziotta, J.C.; Dapretto, M. Cracking the language code: Neural mechanisms underlying speech parsing. *J. Neurosci.* **2006**, *26*, 7629–7639. [CrossRef] [PubMed]
113. Schapiro, A.C.; Gregory, E.; Landau, B.; McCloskey, M.; Turk-Browne, N.B. The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* **2014**, *26*, 1736–1747. [CrossRef] [PubMed]
114. Bischoff-Grethe, A.; Proper, S.M.; Mao, H.; Daniels, K.A.; Berns, G.S. Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J. Neurosci.* **2000**, *20*, 1975–1981. [CrossRef] [PubMed]
115. Elmer, S.; Albrecht, J.; Valizadeh, S.A.; François, C.; Rodríguez-Fornells, A. Theta Coherence Asymmetry in the Dorsal Stream of Musicians Facilitates Word Learning. *Sci. Rep.* **2018**, *8*, 4565. [CrossRef] [PubMed]
116. Bosseler, A.N.; Taulu, S.; Pihko, E.; Mäkelä, J.P.; Imada, T.; Ahonen, A.; Kuhl, P.K. Theta brain rhythms index perceptual narrowing in infant speech perception. *Front. Psychol.* **2013**, *4*, 690. [CrossRef] [PubMed]
117. Giraud, A.L.; Poeppel, D. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat. Neurosci.* **2012**, *15*, 511–517. [CrossRef] [PubMed]
118. Fontolan, L.; Morillon, B.; Liegeois-Chauvel, C.; Giraud, A.L. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* **2014**, *5*, 4694. [CrossRef] [PubMed]
119. Makeig, S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* **1993**, *86*, 293. [CrossRef]
120. Park, H.; Ince, R.A.; Schyns, P.G.; Thut, G.; Gross, J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **2015**, *25*, 1649–1653. [CrossRef] [PubMed]
121. Asaridou, S.S.; Takashima, A.; Dediu, D.; Hagoort, P.; McQueen, J.M. Repetition Suppression in the Left Inferior Frontal Gyrus Predicts Tone Learning Performance. *Cereb. Cortex* **2016**, *26*, 2728–2742. [CrossRef] [PubMed]
122. Dehaene, S.; Meyniel, F.; Wacongne, C.; Wang, L.; Pallier, C. The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron* **2015**, *88*, 2–19. [CrossRef] [PubMed]
123. Roser, M.E.; Fiser, J.; Aslin, R.N.; Gazzaniga, M.S. Right hemisphere dominance in visual statistical learning. *J. Cogn. Neurosci.* **2011**, *23*, 1088–1099. [CrossRef] [PubMed]
124. Reddy, L.; Poncet, M.; Self, M.W.; Peters, J.C.; Douw, L.; Van Dellen, E.; Claus, S.; Reijneveld, J.C.; Baayen, J.C.; Roelfsema, P.R. Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nat. Commun.* **2015**, *6*, 8556. [CrossRef] [PubMed]
125. Durrant, S.J.; Cairney, S.A.; Lewis, P.A. Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex* **2013**, *23*, 2467–2478. [CrossRef] [PubMed]
126. Strange, B.A.; Duggins, A.; Penny, W.; Dolan, R.J.; Friston, K.J. Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Netw.* **2005**, *18*, 225–230. [CrossRef] [PubMed]
127. Nastase, S.; Iacovella, V.; Hasson, U. Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* **2014**, *35*, 1111–1128. [CrossRef] [PubMed]
128. Ayotte, J.; Peretz, I.; Hyde, K. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* **2002**, *125*, 238–251. [CrossRef] [PubMed]
129. Covington, N.V.; Brown-Schmidt, S.; Duff, M.C. The Necessity of the Hippocampus for Statistical Learning. *J. Cogn. Neurosci.* **2018**, *30*, 680–697. [CrossRef] [PubMed]
130. Shaqiri, A.; Anderson, B. Priming and statistical learning in right brain damaged patients. *Neuropsychologia* **2013**, *51*, 2526–2533. [CrossRef] [PubMed]
131. Studer-Eichenberger, E.; Studer-Eichenberger, F.; Koenig, T. Statistical Learning, Syllable Processing, and Speech Production in Healthy Hearing and Hearing-Impaired Preschool Children: A Mismatch Negativity Study. *Ear Hear.* **2016**, *37*, e57–e71. [CrossRef] [PubMed]

132. Conway, C.M.; Pisoni, D.B.; Anaya, E.M.; Karpicke, J.; Henning, S.C. Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci.* **2011**, *14*, 69–82. [CrossRef] [PubMed]
133. Torkildsen, J.V.K.; Arciuli, J.; Haukedal, C.L.; Wie, O.B. Does a lack of auditory experience affect sequential learning? *Cognition* **2018**, *170*, 123–129. [CrossRef] [PubMed]
134. Kraus, N.; Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **2010**, *11*, 599–605. [CrossRef] [PubMed]
135. Schon, D.; Gordon, R.; Campagne, A.; Magne, C.; Astésano, C.; Anton, J.L.; Besson, M. Similar cerebral networks in language, music and song perception. *Neuroimage* **2010**, *51*, 450–461. [CrossRef] [PubMed]
136. Peretz, I.; Vuvan, D.; Lagrois, M.E.; Armony, J.L. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*, 68–75. [CrossRef] [PubMed]
137. Ong, J.H.; Burnham, D.; Stevens, C.J.; Escudero, P. Naïve Learners Show Cross-Domain Transfer after Distributional Learning: The Case of Lexical and Musical Pitch. *Front. Psychol.* **2016**, *7*, 1189. [CrossRef] [PubMed]
138. Bermudez, P.; Lerch, J.P.; Evans, A.C.; Zatorre, R.J. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* **2009**, *19*, 1583–1596. [CrossRef] [PubMed]
139. Schlaug, G.; Jäncke, L.; Huang, Y.; Staiger, J.F.; Steinmetz, H. Increased corpus callosum size in musicians. *Neuropsychologia* **1995**, *33*, 1047–1055. [CrossRef]
140. Keenan, J.P.; Thangaraj, V.; Halpern, A.R.; Schlaug, G. Absolute pitch and planum temporale. *Neuroimage* **2001**, *14*, 1402–1408. [CrossRef] [PubMed]
141. Bermudez, P.; Zatorre, R.J. Differences in gray matter between musicians and nonmusicians. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 395–399. [CrossRef] [PubMed]
142. Elmer, S.; Meyer, M.; Jancke, L. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* **2012**, *22*, 650–658. [CrossRef] [PubMed]
143. Elmer, S.; Hanggi, J.; Meyer, M.; Jancke, L. Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* **2013**, *49*, 2812–2821. [CrossRef] [PubMed]
144. Liegeois-Chauvel, C.; Musolino, A.; Badier, J.M.; Marquis, P.; Chauvel, P. Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* **1994**, *92*, 204–214. [CrossRef]
145. Hackett, T.A.; Preuss, T.M.; Kaas, J.H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* **2001**, *441*, 197–222. [CrossRef] [PubMed]
146. Sluming, V.; Barrick, T.; Howard, M.; Cezayirli, E.; Mayes, A.; Roberts, N. Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage* **2002**, *17*, 1613–1622. [CrossRef] [PubMed]
147. Lopez-Barroso, D.; Catani, M.; Ripolles, P.; Dell'Acqua, F.; Rodríguez-Fornells, A.; de Diego-Balaguer, R. Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13168–13173. [CrossRef] [PubMed]
148. Oechslin, M.S.; Imfeld, A.; Loenneker, T.; Meyer, M.; Jancke, L. The plasticity of the superior longitudinal fasciculus as a function of musical expertise: A diffusion tensor imaging study. *Front. Hum. Neurosci.* **2010**, *3*, 76. [CrossRef] [PubMed]
149. Newman, R.; Ratner, N.B.; Jusczyk, A.M.; Jusczyk, P.W.; Dow, K.A. Infant's early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Dev. Psychol.* **2006**, *42*, 643–655. [CrossRef] [PubMed]
150. McNealy, K.; Mazziotta, J.C.; Dapretto, M. Age and experience shape developmental changes in the neural basis of language-related learning. *Dev. Sci.* **2011**, *14*, 1261–1282. [CrossRef] [PubMed]
151. Karuza, E.A.; Li, P.; Weiss, D.J.; Bulgarelli, F.; Zinszer, B.D.; Aslin, R.N. Sampling over Nonuniform Distributions: A Neural Efficiency Account of the Primacy Effect in Statistical Learning. *J. Cogn. Neurosci.* **2016**, *28*, 1484–1500. [CrossRef] [PubMed]
152. Huss, M.; Verney, J.P.; Fosker, T.; Mead, N.; Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* **2011**, *47*, 674–689. [CrossRef] [PubMed]
153. Evans, J.L.; Saffran, J.R.; Robe-Torres, K. Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* **2009**, *52*, 321–335. [CrossRef]

154. Abrams, D.A.; Nicol, T.; Zecker, S.; Kraus, N. Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* **2009**, *29*, 7686–7693. [CrossRef] [PubMed]
155. Goswami, U.; Wang, H.L.; Cruz, A.; Fosker, T.; Mead, N.; Huss, M. Language-universal sensory deficits in developmental dyslexia: English, Spanish and Chinese. *J. Cogn. Neurosci.* **2011**, *23*, 325–337. [CrossRef] [PubMed]
156. Bouwer, F.L.; Werner, C.M.; Knetemann, M.; Honing, H. Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia* **2016**, *85*, 80–90. [CrossRef] [PubMed]
157. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: Oxford, UK, 2008.
158. Hansen, N.C.; Pearce, M.T. Predictive uncertainty in auditory sequence processing. *Front. Psychol.* **2014**, *5*, 1052. [CrossRef] [PubMed]
159. Habib, M.; Lardy, C.; Desiles, T.; Commeiras, C.; Chobert, J.; Besson, M. Music and dyslexia: A new musical training method to improve reading and related disorders. *Front. Psychol.* **2016**, *7*, 26. [CrossRef] [PubMed]
160. Marie, C.; Magne, C.; Besson, M. Musicians and the metric structure of words. *J. Cogn. Neurosci.* **2011**, *23*, 294–305. [CrossRef] [PubMed]
161. Norton, A.; Zipse, L.; Marchina, S.; Schlaug, G. Melodic intonation therapy shared insights on how it is done and why it might help. *Neurosci. Music* **2009**, *1169*, 431–436.
162. Kudo, N.; Nonaka, Y.; Mizuno, N.; Mizuno, K.; Okanoya, K. On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* **2011**, *14*, 1100–1106. [CrossRef] [PubMed]
163. Hannon, E.E.; Johnson, S.P. Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cogn. Psychol.* **2005**, *50*, 354–377. [CrossRef] [PubMed]
164. Fiser, J.; Aslin, R.N. Statistical learning of higher-order temporal structure from visual shape-sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **2002**, *28*, 458–467. [CrossRef] [PubMed]
165. Wu, R.; Gopnik, A.; Richardson, D.C.; Kirham, N.Z. Infants learn about objects from statistics and people. *Dev. Psychobiol.* **2011**, *47*, 1220–1229. [CrossRef] [PubMed]
166. Kushnir, T.; Xu, F.; Wellman, H.M. Young children use statistical sampling to infer the preferences of other people. *Psychol. Sci.* **2010**, *21*, 1134–1140. [CrossRef] [PubMed]
167. Xu, F.; Griffiths, T.L. Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition* **2011**, *120*, 299–301. [CrossRef] [PubMed]
168. Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N.; Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **1992**, *255*, 606–608. [CrossRef] [PubMed]
169. Dawson, C.; Gerken, L. From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition* **2009**, *111*, 378–382. [CrossRef] [PubMed]
170. Zhang, L.I.; Bao, S.; Merzenich, M.M. Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nat. Neurosci.* **2001**, *4*, 1123–1130. [CrossRef] [PubMed]
171. Hensch, T.K. Critical period regulation. *Annu. Rev. Neurosci.* **2004**, *27*, 549–579. [CrossRef] [PubMed]
172. Sanes, D.H.; Bao, S. Tuning up the developing auditory CNS. *Curr. Opin. Neurobiol.* **2009**, *19*, 188–199. [CrossRef] [PubMed]
173. Männel, C.; Friederici, A.D. Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex* **2013**, *49*, 2788–2798. [CrossRef] [PubMed]
174. Arciuli, J.; Simpson, I.C. Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Dev. Sci.* **2011**, *14*, 464–473. [CrossRef] [PubMed]
175. Skoe, E.; Kraus, N. Hearing it again and again: On-line subcortical plasticity in humans. *PLoS ONE* **2010**, *5*, e13645. [CrossRef] [PubMed]
176. Munte, T.F.; Altenmuller, E.; Jancke, L. The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* **2002**, *3*, 473–478. [CrossRef] [PubMed]
177. Daikoku, T. Time-course variation of statistics embedded in music: Corpus study on implicit learning and knowledge. *PLoS ONE* **2018**, *13*, e0196493. [CrossRef] [PubMed]
178. Arciuli, J.; Monaghan, P.; Seva, N. Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *J. Mem. Lang.* **2010**, *63*, 180–196. [CrossRef]
179. Rohrmeier, M.; Rebuschat, P. Implicit learning and acquisition of music. *Top. Cogn. Sci.* **2012**, *4*, 525–553. [CrossRef] [PubMed]

180. Berry, D.C.; Dienes, Z. *Implicit Learning: Theoretical and Empirical Issues*; Lawrence Erlbaum: Hove, UK, 1993.
181. Reber, A.S. *Implicit Learning and Tacit Knowledge. An Essay on the Cognitive Unconscious*; Oxford University Press: New York, NY, USA, 1993.
182. Perkovic, S.; Orquin, J.L. Implicit Statistical Learning in Real-World Environments Leads to Ecologically Rational Decision Making. *Psychol. Sci.* **2017**, *29*, 34–44. [CrossRef] [PubMed]
183. Norgaard, M. How jazz musicians improvise: E central role of auditory and motor pa erns. *Music Percept.* **2014**, *31*, 271–287. [CrossRef]
184. Bigand, E.; Poulin-Charronnat, B. Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition* **2006**, *100*, 100–130. [CrossRef] [PubMed]
185. Ettlinger, M.; Margulis, E.H.; Wong, P.C.M. Implicit memory in music and language. *Front. Psychol.* **2011**, *211*. [CrossRef] [PubMed]
186. Huron, D. Two challenges in cognitive musicology. *Top. Cogn. Sci.* **2012**, *4*, 678–684. [CrossRef] [PubMed]
187. McLaughlin, J.; Osterhout, L.; Kim, A. Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nat. Neurosci.* **2004**, *7*, 703–704. [CrossRef] [PubMed]
188. Siegelman, N.; Bogaerts, L.; Christiansen, M.H.; Frost, R. Towards a theory of individual differences in statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**, *372*, 1711. [CrossRef] [PubMed]
189. Arciuli, J.; Simpson, I.C. Statistical learning is related to reading ability in children and adults. *Cogn. Sci.* **2012**, *36*, 286–304. [CrossRef] [PubMed]
190. Kidd, E.; Arciuli, J. Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Dev.* **2016**, *87*, 184–193. [CrossRef] [PubMed]
191. Shaqiri, A.; Anderson, B.; Danckert, J. Statistical learning as a tool for rehabilitation in spatial neglect. *Front. Hum. Neurosci.* **2013**, *7*, 224. [CrossRef] [PubMed]
192. Daikoku, T.; Ogura, H.; Watanabe, M. The variation of hemodynamics relative to listening to consonance or dissonance during chord progression. *Neurol. Res.* **2012**, *34*, 557–563. [CrossRef] [PubMed]
193. Ellis, R. Implicit and explicit learning, knowledge and instruction. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*; Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., Reinders, H., Eds.; Multilingual Matters: Bristol, UK, 2009; pp. 3–25.
194. Jusczyk, P.W. How infants begin to extract words from speech. *Trends Cogn. Sci.* **1999**, *3*, 323–328. [CrossRef]
195. Archibald, L.M.; Joanisse, M.F. Domain-specific and domain-general constraints on word and sequence learning. *Mem. Cogn.* **2013**, *41*, 268–280. [CrossRef] [PubMed]

brain sciences

MDPI

Article

Domain-Specific Expectations in Music Segmentation

Susana Silva *, Carolina Dias and São Luís Castro *

Center for Psychology at University of Porto (CPUP), Faculty of Psychology and Education Sciences, 4200-135 Porto, Portugal

* Correspondence: susanamsilva@fpce.up.pt (S.S.); slcastro@fpce.up.pt (S.L.C.); Tel.: +351-964-366-725 (S.S.); +351-22-607-97-56 (S.L.C.)

Received: 20 June 2019; Accepted: 16 July 2019; Published: 17 July 2019

Abstract: The acoustic cues that guide the assignment of phrase boundaries in music (pauses and pitch movements) overlap with those that are known for speech prosody. Based on this, researchers have focused on highlighting the similarities and neural resources shared between music and speech prosody segmentation. The possibility that music-specific expectations add to acoustic cues in driving the segmentation of music into phrases could weaken this bottom-up view, but it remains underexplored. We tested for domain-specific expectations in music segmentation by comparing the segmentation of the same set of ambiguous stimuli under two different instructions: stimuli were either presented as speech prosody or as music. We measured how segmentation differed, in each instruction group, from a common reference (natural speech); thus, focusing on how instruction affected delexicalization effects (natural speech vs. transformed versions with no phonetic content) on segmentation. We saw interactions between delexicalization and instruction on most segmentation indices, suggesting that there is a music mode, different from a speech prosody mode in segmentation. Our findings highlight the importance of top-down influences in segmentation, and they contribute to rethinking the analogy between music and speech prosody.

Keywords: Prosody; Phrasing; Perception; Melody

1. Introduction

Most speech listeners and music listeners segment the auditory input into phrase-like units [1–4]. In both domains, listeners detect phrase boundaries as the input unfolds. This leads to the possibility of building the segmentation map of an utterance or a music piece, defining how many phrases were heard, whether they were short, long, regular or irregular in length, and how they relate to each other. Language and music users have ways of emphasizing their intended segmentation maps (phrase boundary locations) using specific graphic signs in printed versions of language and music. In written language, the intended segmentation map of an utterance is sometimes achieved by printed punctuation marks [5,6]. Printed music does not have a mandatory analogue of punctuation marks to signal the presence of intended phrase boundaries. Slurs are perhaps the most obvious sign of intended segmentation maps, although other markers such as pause signs can also be used [7,8].

In speech research, the idea of segmentation map, a set of individual choices regarding segmentation, has been implemented mostly with pairs of syntactically ambiguous sentences [8–10], holding more than one meaning depending on how they are parsed. The way that participants judge and understand such speech materials is then taken as an index of their segmentation choices. While traditional behavioral approaches only allowed delayed (post-exposure) judgements, more recent techniques such as eye-tracking [11,12] or EEG [8,10] popularized the online monitoring of speech segmentation, often affording the tracking of participants' revisions of their initial segmentation choices [13]. Online monitoring techniques have also increased the interest in music segmentation (e.g., [14]). In the present study, we used a simple behavioral online monitoring approach to both

speech and music segmentation maps, which consisted of asking participants to press a key every time they heard a phrase ending.

The segmentation of speech into phrase-like units depends not only on linguistic content (lexico–syntactic structure), but also on the paralinguistic intonation patterns of speech prosody (e.g., [1]), which define intonational phrases. The perception of intonation patterns per se, regardless of interactions with linguistic content, is driven by low-level acoustic cues such as changes in pitch, duration, or the presence of silence. In this sense, it is possible to view the segmentation of intonation-related speech prosody (speech prosody hereafter) as a bottom-up process, i.e., as a process where extra-perceptual factors like previous expectations of what an intonational phrase should be do not play a major role (see [15] for a discussion on top-down vs. bottom-up). Pitch deflections and pauses are acoustic cues that play an important role in driving both the segmentation of music [3,14] and that of speech prosody [16–18]. Does it follow that music is segmented the same way as speech prosody? The answer depends on how we assume that music segmentation is driven: if it is driven by acoustic cues, as in speech prosody, the answer is yes. If it is driven both by acoustic cues plus music-specific expectations (i.e., an idea of what a musical phrase is), the answer is no. The literature is mixed on this matter, as we will see below.

The idea that boundary assignment in music is driven solely by acoustic cues, which we will refer to as a bottom-up view on auditory segmentation, is present in the literature on the Closure Positive Shift (CPS) event-related potential. The CPS is an electrophysiological marker of phrase boundary perception, which has been found for speech [8,10,19], delexicalized (hummed) speech [20] and music [14,21–23] with little morphological variation across the three [20,24]. The bottom line of the CPS approach is that segmentation shares neural resources across music and speech prosody, and a strong motivation for these studies has been the fact that the same type of segmentation cues (pitch deflections, pauses) can be detected in both domains [14]. CPS studies have focused on the acoustic features that characterize musical and prosodic segmentation points (music phrases vs. intonational phrases). These are expected to elicit a brain response corresponding to boundary detection, with little effects of prior knowledge or contextual aspects. An implication of this view is that the segmentation map of a music piece can be similar to the segmentation map of a sample of speech prosody, provided that both have the same acoustic boundary cues at the same time points.

The alternative view, which we will refer to as the *top-down view*, emphasizes the role of expectations in suppressing or counteracting acoustic boundary cues. For instance, it has been admitted that music-specific expectations can make the listener search for four-bar structures when judging whether a musical phrase has ended or not [22,23], possibly overriding pauses within the four-bar phrase. The top-down view also relates to the idea that music segmentation may rely on more global cues than the segmentation of speech prosody [25]; such cues extending in time beyond the limits of a local boundary mark such as a pause and requiring integration. In contrast to the bottom-up view, one should expect here that equivalent boundary cues in music and speech prosody would not lead to equivalent segmentation maps, since segmentation options would depend on additional music-specific top-down influences. To our knowledge, neither this top-down-based hypothesis nor its bottom-up alternative have been subject to testing.

In the present paper, we tested whether music segmentation into phrases is driven by music-specific expectations that add to the acoustic cues used to segment speech prosody into intonational phrases. Thus, we tested for a top-down view on music segmentation. To that end, we compared participants' segmentation maps of a single set of ambiguous auditory stimuli, which were either presented as music or as speech prosody. We manipulated only the instruction, inducing different processing modes on the very same acoustic materials: top-down expectations and bottom-up processing for music, against bottom-up processing (only) for speech prosody vs. no additional expectations for speech prosody.

The ambiguous stimuli were obtained by an audio-to-MIDI conversion of natural speech, resulting in pitch-and-rhythm auditory streams deprived of linguistic content. For convenience of expression, we will refer to these wordless auditory streams as delexicalized versions, even though the difference between them and natural speech lies, strictly speaking, at the phonetic level rather than just the lexical one. Due to the algorithms involved in the audio-to-MIDI conversion of natural speech (see

methods), two types of data (speech prosody) distortions were expected: first, continuous pitch would be converted into discontinuous pitch (octave divided into 12-semitone intervals), lending a music-like character to these ambivalent streams; second, timing-related information concerning speech syllables might not be integrally preserved, although an approximation was expected. The first type of speech prosody distortion (discontinuous pitch) was necessary to keep the credibility of the music instruction. The second type of distortion created additional differences between delexicalized versions and the original speech signal, such that the former were, strictly speaking, delexicalized and modified. Nevertheless, delexicalized versions contained the pitch-and-timing information listeners use for processing speech prosody, with pitch and timing value-ranges reflecting the ones that occur in natural language. In this sense, we considered our delexicalized versions to be representative of speech prosody, even though they were not an exact copy of the speech prosody patterns that generated them.

In order to minimize participants' awareness of our experimental manipulation, instruction was set as a between-subjects factor. To circumvent the risk of imperfect group matching inherent to a between-subjects approach, we sought for a common reference (baseline) in the two groups, against which we analyzed participants' segmentation maps of ambiguous, delexicalized stimuli. The common reference we used was natural speech. Therefore, we collected the segmentation maps of a single set of delexicalized (ambiguous) stimuli (i.e., speech without lexical content) of two groups of participants receiving different types of instruction (speech prosody – "This is prosody" vs. music – "This is music"), as well as the segmentation maps of their natural-speech counterparts, in which case the instruction for segmentation was common to both groups ("This is speech"). We then focused on determining whether delexicalization effects (natural speech vs. delexicalized versions, within-subjects factor) were equivalent under music vs. speech prosody instructions, thus probing between-subjects instruction effects with the benefit of a baseline. Similar deviations from the natural-speech baseline (similar delexicalization effects) across instruction conditions (delexicalized presented as music vs. delexicalized presented as speech prosody) would indicate that music participants adopted segmentation approaches to delexicalized versions similar to those of speech prosody participants. In this case, there would be no reason to admit that there are music-specific expectations in music segmentation. By contrast, different deviations from baseline (different delexicalization effects) would indicate that music participants adopted segmentation approaches to delexicalized versions differing from those of speech prosody participants. In this case, music-specific expectations could be considered real.

The existence of delexicalization effects was a precondition to the goal of comparing such effects across instruction conditions. Delexicalization effects were expected under the speech prosody instruction, at least for one reason: it is known that lexicality – the presence vs. absence of lexical content - affects speech segmentation, in the sense that lexical information may override prosodic boundary markers in phrase boundary assignment [26–28] and the so-called *linguistic bias* ([29], see also [30] for a similar phenomenon in word segmentation) emerges (cf. [31]). For instance, Buxó-Lugo and Watson [26] found that listeners consistently report hearing more boundaries at syntactically licensed locations than at syntactically unlicensed locations, even when the acoustic evidence for an intonational boundary was controlled. Cole, Mo and Baek [28] analyzed the predictors of phrase boundary assignment, and found syntactic structure to be the strongest one, winning over prosodic cues. Meyer and colleagues [29] found that 2-phrase prosodic sentences with 2-phrase lexical groups lead to segmentation in 2 phrases, but 1-phrase prosodic sentences do not necessarily lead to a single phrase when there are two lexical groups. In the latter case, an electrophysiological marker of the linguistic bias is visible. On the other hand, the existence of delexicalization effects was a precondition, but not a target of this study, and this is why we did not discuss delexicalization effects per se. Instead, our question was whether the delexicalization effect tested under the music instruction would, or would not, parallel the delexicalization effect under the speech prosody instruction - in other words, if delexicalization would interact with instruction in the generation of segmentation maps.

In our approach, we characterized segmentation maps from two different viewpoints: segment length (correlated with the number of segments), and the matching with predefined segmentation

models. Interactions between delexicalization and instruction on any of these measures would indicate music-specific expectations.

2. Materials and Methods

2.1. Participants

Seventy participants took part in the experiment. Half (n = 35) were assigned to the speech instruction (31 women), and the other half to the music instruction (27 women). There was no evidence of significant differences between the two groups concerning age ($M \pm SD$: 20.54 ± 2.85 for speech, 20.28 ± 1.52 for music; $t(68) = 0.47$, $p > 0.64$, $d = 0.12$) and musical training (11 participants in the speech condition had 3.27 ± 2.24 years of training, ten in the music condition with 3.90 ± 2.46; $t(68) = -0.17$, $p > 0.86$, $d = -0.04$). All participants had normal hearing. None reported psychiatric or neurological disorders. Participants signed informed consent, according to the Declaration of Helsinki.

2.2. Stimuli

Stimulus materials consisted of natural speech samples and delexicalized versions of these (see Supplementary materials). The latter were presented under two different instructions (speech prosody vs. music) but they were physically the same. We used five different samples of natural speech. In order to maximize prosodic spontaneity, we selected these samples from available personal and media recordings instead of laboratory recordings. Each sample contained an utterance, combining full sentences that were semantically related (see Appendix A for transcriptions and sentence structure). Four utterances were spoken by men, and one by a woman. Stimulus 1 contained the online description of a short movie that was being watched by the speaker; stimulus 2 was a fragment of an interview; stimulus 3 and 4 were poems recorded by a famous Portuguese diseur; stimulus 5 was an excerpt from a news broadcast. Stimuli were similar in length (~60 sec., see Table 1), and they were all normalized to 70 dB rms.

Table 1. Acoustic properties of the five stimuli used in the experiment.

	***M* ± *SD* Pitch in Hz Rel *SD*[a] Hz/Rel *SD* Mel[b] −1/2 *SD*, +1/3 in Mel**		**Pitch Change Rate (Pitches per Second)**	**Duration (sec)**	**Silence Proportion (%)**	
	Natural	**Delexicalized**	**Delexicalized [c]**		**Natural**	**Delexicalized**
1	185 ± 28 Hz 0.55/−0.53, +0.55 −18, +12	187 ± 26 Hz 0.51/−0.50, +0.50 −17, +11	3.47	63.4	44.7	41.7
2	112 ± 21 0.41/−0.43, +0.45 −14, +10	112±19 0.37/−0.38, +0.40 −13, +9	3.2	55.5	4.5	19.6
3	130 ± 35 0.69/−0.72, +0.73 −26, +16	128 ± 34 0.67/−0.69, +0.72 −12, +15	3.02	57.0	32.4	34.5
4	172 ± 51 1/−1, +1 −34, +21	170 ± 51 1/−1, +1 −34, +22	3.09	59.1	20.1	24.6
5	123 ± 20 0.39/−0.40, +0.42 −13, +9	123 ± 18 0.35/−0.35, +.39 −12, +9	4	40.2	23.7	30.8

[a] Relative SD = SD/highest SD (Stimulus 4). Note that the magnitude relation across stimuli is equivalent, whether it comes in Hz or in Mel; [b] Mel – Measure of pitch that accounts for different sensitivity levels across the frequency range; [c] in natural speech, pitch change is continuous.

To create delexicalized versions, natural speech samples were converted to MIDI with software Live 9 (www.ableton.com), using a bass timbre and settings for monophonic stimuli. This audio-to-MIDI conversion software detects stable-pitch fragments preceded by transients (an attack), disregarding intensity information. When dealing with music audio, the software searches for music notes. In speech-related audio, it should detect syllable-like events.

As shown in Table 1, pitch mean and standard deviation were preserved after audio-to-MIDI conversion (Wilcoxon signed rank tests: $Z = 0$, $p = 0.059$ for mean pitch; $Z = 2$, $p > 0.56$ for standard deviation of pitch). In delexicalized (discrete pitch) versions, the pitch change rate was close to the syllable rate of speech (3–4 syllables per second, see [32,33]), supporting the idea that the algorithm captured syllable-like units As for the proportion of silences, it was apparently higher in delexicalized versions, but statistical tests did not confirm this ($Z = 13$, $p > 0.13$).

2.3. Procedure

We started the experiment with auditory reaction time measurements. Participants heard a series of beeps, among which there was a human voice pronouncing a syllable. They were asked to press a key as soon as they heard the human voice. The purpose of these measurements was to provide a participant-specific correction for reaction times (time between perception and key press) for the task of detecting phrase endings that would be requested in the experiment.

All participants were first exposed to the five delexicalized stimuli. Those under the speech instruction were told that the stimuli were derived from real speech, thus containing "the melody of speech, without the words". Participants under the music instruction were told that stimuli were "excerpts of contemporary music". All participants were asked to press the space bar of the computer keyboard every time they perceived a phrase ending. Before the experimental trials, all were given a brief explanation of the concept of phrase ("a speech/music fragment, with a beginning and an end"), followed by a demonstration of a possible way of segmenting either speech prosody (speech instruction) or music (music instruction) into phrases. In these examples, we defined segments with similar length across instructions (6 sec. for speech prosody, 7 sec. for music). Given that the concept of music phrase is not trivial among non-experts, we told music-instruction participants that music phrases "were the equivalent of speech phrases, in that they defined unitary fragments". We stressed that there were no wrong answers. Participants were given one practice trial, either with a delexicalized utterance (speech instruction) or with a music excerpt (music instruction) and then they proceeded into the experimental trials. Each trial consisted of one stimulus to be segmented. Since segmentation was made online, they were unable to go back for corrections. Therefore, we gave participants a second chance: each stimulus was presented twice in succession, and participants did the segmentation on both (5 x 2 trials). Only the second presentation of each stimulus was considered in the analyses. We presented stimuli no more than twice in order to keep the experiment short enough to avoid fatigue.

After segmenting the five delexicalized stimuli, participants were asked to do the same on the 5×2 natural speech counterparts. They were informed that they would listen to "normal speech" and they should, again, press the space bar whenever they sensed the phrase had just ended. Participants were not informed that delexicalized and natural speech had the same source. We created three different versions of the experiment, in order to counterbalance the order of presentation of the five stimuli (1-2-3-4-5; 1-4-5-2-3; 4-1-5-3-2). In each version, stimulus order was common to delexicalized and lexicalized sets. Thus, in version 1, participants heard 1-2-3-4-5 delexicalized and then 1-2-3-4-5 lexicalized. We did so in order to keep delexicalized and lexicalized conditions as equivalent as possible.

At the end of the experiment, participants were given a questionnaire where they rated the level of confidence in their segmentation responses for each block (delexicalized vs natural speech) on a 5-point scale and made any comments they wished to. Stimulus delivery was made with Presentation software (www.neurobs.com, v. 20). The experiment lasted about 40 minutes.

2.4. Segmentation Models

Prior to the analysis, we defined virtual segmentation points in each stimulus according to four theoretical models, each model based on a different segmentation cue: Pause, Pitch break, Pitch rise and Pitch drop. The adopted models intended to explore the idea of pauses and pitch movements such as low-level acoustic cues subtending both speech prosody and music segmentation (see Introduction section). Considering the possibility that music segmentation may rely more on global cues (see Introduction section)

than local boundary marks, two models targeted local cues (Pauses and Pitch breaks), and two targeted global cues (Pitch rises and Pitch drops).

Pauses and Pitch breaks were considered as local cues, in the sense that they included a restricted number of events (silence onset/offset, sudden pitch change), which unfolded within a short time-window. Based on a preliminary inspection of our five natural speech stimuli, we defined Pauses as silent periods longer than 200 ms. The onset of the Pause was considered the segmentation point. Pitch breaks were marked if two consecutive pitch values that were separated by a silence (shorter than 200 ms, the threshold for pause) differed by more than one standard deviation of the stimulus mean pitch. The onset of the second pitch value was set as the segmentation point. Note that the perception of pitch breaks is necessarily context-dependent, since pitch is continuously changing, and we are focusing on salient pitch breaks, which depend on the overall pitch context. However, the break per se (two different pitch values, separated by a short pause) occurs in a short time window. This is the reason why we considered pitch break as a local cue.

Pitch rises and Pitch drops were viewed as global cues, since they require the integration of multiple (pitch) values across time, and they tend to occur within larger time windows. Pitch rises and Pitch drops were defined as unidirectional pitch movements. Since Pitch drops are more common in natural speech, given the F0 decline phenomenon ([34,35], a universal tendency for pitch to drop across sentences, we used more restrictive criteria for Pitch drops than for Pitch rises. Pitch rises and drops should be either wide in pitch range (at least one third of global pitch range) or long-lasting (minimum 500 ms for pitch rise, and 1000 ms for pitch drops). For Pitch drops, we set the additional criterion that pitch should reach a low-frequency range, namely half a standard deviation from the global mean pitch. Small pitch deflections up to 250/200 ms were allowed within Pitch rise/drop segments, as well as pauses up to 200 ms. The offset of pitch movements (rises or drops) corresponded to the segmentation point. Pitch drops or rises not complying with these criteria were not used as virtual segmentation points of any kind.

When Pauses coexisted with Pitch breaks, rises or drops, we considered these as different situations/models. Pauses combined with Pitch rises or drops were viewed as mixed cues (local plus global cues), and pauses combined with Pitch breaks (i.e., when the pause between contrasting pitch values was larger than 200 ms) were viewed as local cues. Thus, in total, we had seven models.

Virtual segmentation points (cues) were marked for delexicalized and natural speech versions separately, leading to version-specific segmentation models. There was not a complete overlap in the number of segmentation points across the two versions, which was due to the audio-to-MIDI conversion process (e.g., pause lengths became slightly different in some cases, making the number of pause points differ). However, such differences were irrelevant to our main research question, which concerned the influence of instruction on the delexicalization effect rather than the delexicalization effect itself.

2.5. Preprocessing and Statistical Analysis

We were interested in the interaction between delexicalization (delexicalized vs. natural speech, within-subjects) and instruction (speech vs. music, between-subjects) on segmentation maps. Such interactions would indicate music-specific expectations, non-overlapping with prosody-specific ones. We analyzed the effects of delexicalization and instruction; first on segment length, and then on the adherence to a number of segmentation models we created (model matching).

To compute participants' segment length, we calculated the interval between participants' key presses. Participants' metrics per stimulus (mean and standard deviation of segment length – the latter indexing segment length variability) were obtained.

To analyze the matching of participants' segmentations with the segmentation models, participant-specific reaction times (see procedure; $M + SD = 287 \pm 50$ ms) were first subtracted from the raw time of key presses in order to obtain corrected segmentation points for each stimulus (Figure 1B). Then, also for each stimulus, we merged the time stamps of the virtual segmentation points from all seven models into one global array of time values (Figure 1A).

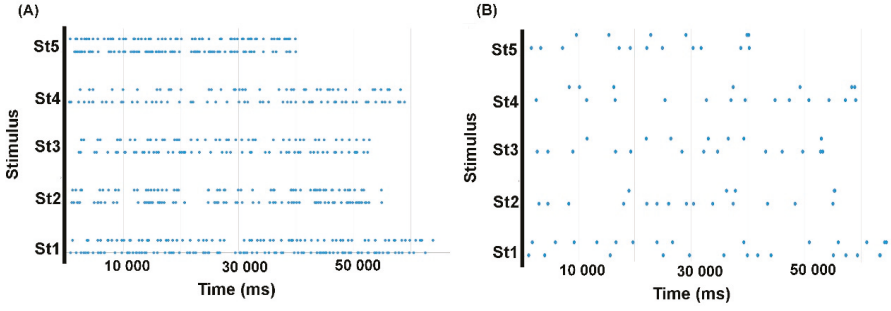


Figure 1. (A) Stimulus-specific global arrays, combining all segmentation models for each utterance (St1-5: Stimulus 1–5; lower line: delexicalized; upper line: natural). Dots represent virtual segmentation points; (B) Example of a segmentation maps (Participant 01) containing actual segmentation marks for each of the five stimuli in natural speech versions (above) and delexicalized ones (below, speech prosody instruction in this case).

With reference to each stimulus-specific global array of time values (lengths ranging from 45 to 96 virtual points depending on the stimulus, see Figure 1A), we derived separate logical arrays (1, true or present vs. 0, false or absent) for each model (1 marking the points of the model in question and 0 the points of other models), and one logical array per participant (1 marking participants' segmentation points and 0 absence of segmentation). When defining participants' logical arrays, the closest value of the global array of time values was always chosen. Maximum inter-point distances in global arrays of time values were 2690 and 3497 ms for stimulus 1 (delexicalized and natural), 4099 and 5116 ms for stimulus 2, 2397 and 3091 ms for stimulus 3, 2520 and 4886 ms for stimulus 4, 2075 and 1972 for stimulus 5. Therefore, this was the maximum error that could occur when fitting participants' marks to the available models. Finally, we computed the similarity between the logical array describing each participant's behavior and each of the seven logical arrays describing each model, using the Russell and Rao binary similarity coefficient [36]. The Russell and Rao coefficient evaluates the overlap of two data series concerning a binary attribute (present or absent). In our case, we measured how the distribution of participants' marks in time overlapped with the distribution of model-specific segmentation points; both filled with present vs. absent points in reference to the global array of time values. We referred to these coefficients as model matching scores, since they described participants' level of adherence to a given segmentation model.

For statistical analyses, we used mixed ANOVAs. We first analyzed the effects of delexicalization and instruction on the mean and standard deviation of segment length. We then considered the effects of delexicalization, instruction and model (within-subjects, seven levels/models: Pause, Pitch break, Pause plus pitch break, Pitch rise, Pitch drop, Pause plus pitch rise, Pause plus pitch drop) on model matching scores. In the presence of third-order interactions (delexicalization x instruction x model), delexicalization x instruction interactions were considered per model. Along the model matching analysis with seven models, we inspected whether the results fitted with the high-order classification of cues into local, global and mixed, to see whether it made sense to quantify the differences related to this triad. Mixed ANOVAs were also used to analyze questionnaire responses related to participants' confidence in their segmentation responses.

Even though participants heard delexicalized versions prior to natural speech, natural speech was the common reference against which the segmentation maps of the two delexicalized conditions (speech vs. music instruction) were evaluated. Therefore, we refer to the concept of delexicalization throughout the results section as a logical, rather than chronological process.

3. Results

3.1. Segment Length

The overall mean segment length was around 7000 ms (Figure 2), corresponding to an average of 8.6 segments per speech/music 60-sec sample (10.4/7.8 segments for delexicalized speech under speech/music instructions; 7.5/ 8.6 segments for natural speech). Mean segment length showed no main effects of delexicalization (p >0.17, η^2p = 0.027) or instruction (p >0.29, η^2p = 0.016), but there was an interaction between the two (F(1,68) = 10.12, p = 0.002, η^2p = 0.13, Figure 2): delexicalization led to decreased segment length under the speech instruction (t(34) = −4.66, p <0.001, d = −0.66), while it caused no significant changes under the music instruction (p >0.30, d = 0.21, Figure 2).

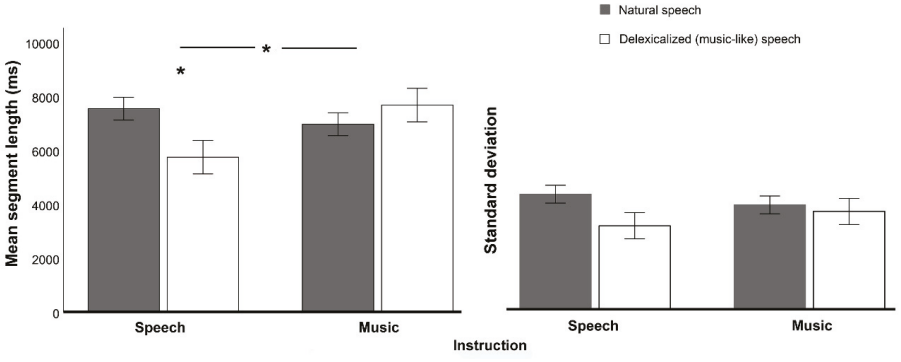


Figure 2. Delexicalization and instruction effects on the mean (left) and standard deviation (right) of segment length. Participants under the speech instruction decreased segment length in delexicalized versions, while those under the music instruction did not show any change. The standard deviation (variability) decreased in delexicalized versions for both instruction levels. Vertical bars represent the standard error of the mean.

Delexicalization decreased the standard deviation (variability) of segment length (main effect of delexicalization: F(1,68) = 4.60, p = 0.036, η^2p = 0.063, Figure 2), regardless of the instruction (non-significant delexicalization x instruction interaction: p >0.16, η^2p = 0.029).

3.2. Model Matching Scores

We found a significant interaction between delexicalization and instruction (F(1,68) = 19.32, p <0.001, η^2p = 0.22) on model matching scores. Both instruction conditions decreased general adherence to (all) models when given delexicalized versions (speech: F(1,34) = 99.84, p <0.001, η^2p = 0.75; music: F(1,34) = 10.48, p = 0.003, η^2p = 0.24), but the decrease was larger for the speech-prosody instruction. These effects came along with a significant (third-order) delexicalization x instruction x model interaction (Figure 3A, F(6,408) = 6.09, p <0.001, η^2p = 0.08), suggesting that delexicalization x instruction interactions differed across models.

When the three-way interaction was broken down into the seven models (Figure 3A,B), the pattern of effects and interactions (delexicalization x instruction) was indeed heterogeneous, and it did not overlap with the associated cue types (local, global, mixed). Pauses alone (local cue, p >0.45, η^2p = 0.008) and Pitch drops (global, p >0.95, η^2p = 0.000) showed non-significant interactions between delexicalization and instruction. For these, delexicalization increased model matching in both instruction levels (main effect of delexicalization on matching with Pauses: F(1,68) = 378.96, p <0.001, η^2p = 0.85; on matching with Pitch drop: F(1,68) = 19.16, p <0.001, η^2p = 0.22). Significant delexicalization x instruction interactions showed up for Pitch breaks (local cue, F(1,68) = 5.15, p = 0.026, η^2p = 0.07), Pitch rise (global, F(1,68) = 48.89, p <0.001, η^2p = 0.42), Pause + pitch rise (mixed, F(1,68) = 4.56, p = 0.036, η^2p = 0.06), and Pause + pitch

drop (mixed, $F(1,68) = 11.31$, $p = 0.001$, $\eta^2p = 0.14$). The interaction for Pause + pitch break was marginal (local cue, $F(1,68) = 3.07$, $p = 0.084$, $\eta^2p = 0.04$). All these interactions indicate different expectations for music compared to the speech prosody instruction.

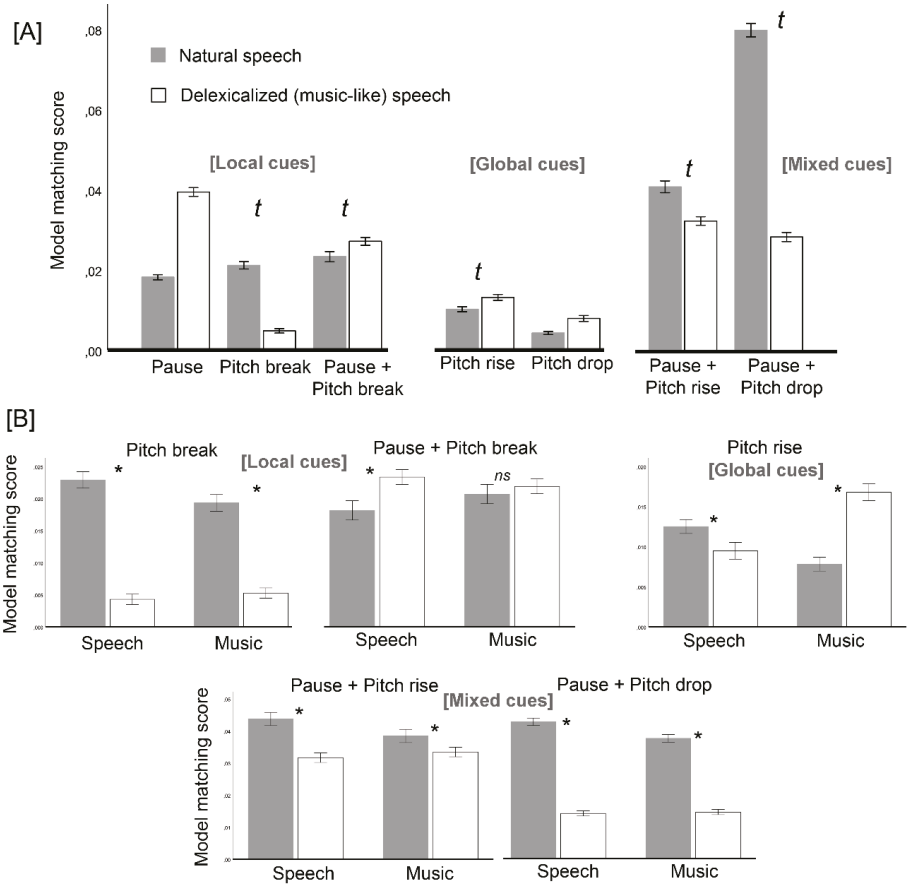


Figure 3. Effects of delexicalization (natural vs. delexicalized speech), model (7 models) and instruction (speech prosody vs. music) on model matching. A: Delexicalization effects per model, with five out of seven models (Pitch break, Pause + Pitch break, Pitch rise, Pause + Pitch rise, Pause + Pitch drop) showing delexicalization x instruction interactions (marked with *t*). B: Delexicalization x instruction interactions per model. Vertical bars represent the standard error of the mean.

The type of interaction was independent from cue type: we saw similar patterns for Pitch break (local cue), Pause + pitch rise and Pause + pitch drop (both mixed): for all, delexicalization had the effect of decreasing model matching scores for both speech and music, with a stronger effect in speech (Pitch break: $t(34) = 11.53$, $p < 0.001$, $d = 2.69$ speech, $t(34) = 11.82$, $p < 0.001$, $d = 2.56$ music; Pause + pitch rise: $t(34) = 5.41$, $p < 0.001$, $d = 1.05$ speech, $t(34) = 2.04$, $p = 0.049$, $d = 0.53$ music; Pause + pitch drop: $t(34) = 27.58$, $p < 0.001$, $d = 4.87$ speech, $t(34) = 17.81$, $p < 0.001$, $d = 3.97$ music). For Pause + pitch break—a local cue, just like Pitch break alone - delexicalization increased model matching for speech ($t(34) = -3.35$, $p = 0.002$, $d = -0.71$) while having no effect for music ($t(34) = -0.72$, $p > 0.47$, $d = -0.14$). Finally, for Pitch rise – a global cue, just like Pitch drop, which showed no interaction -

delexicalized versions decreased model matching scores under the speech instruction ($t(34) = 2.47$, $p = 0.019$, $d = 0.61$), while increasing it under the music one ($t(34) = -7.46$, $p < 0.001$, $d = -1.46$; Figure 3B).

3.3. Confidence in Segmentation

Participants' level of confidence in their segmentation responses was higher for natural speech ($M \pm SD$: 3.83 ± 0.55, 5-point scale) compared to delexicalized versions (2.89 ± 0.66; $F(1,64) = 89.61$, $p < 0.001$, $\eta^2p = 0.58$). Both speech prosody and music-instruction participants showed similar gains in confidence when going from delexicalized to natural speech ($p = 0.57$, $\eta^2p = 0.01$).

4. Discussion

Our goal was to determine whether the segmentation of music into phrases is driven by music-specific expectations, which would indicate that segmentation processes in music do not overlap with those that occur in speech prosody. To that end, we tested whether participants' segmentations of a single set of ambiguous stimuli without lexical content differed according only to the identity assigned to these stimuli (speech prosody vs. music), under a manipulation of the variable instruction. Since the effect of instruction was obtained from two different groups that could be imperfectly matched, we created a baseline-related measure of this effect: we focused on how instruction influenced the within-subjects difference between natural speech (a baseline or common reference) and ambiguous, delexicalized stimuli subject to manipulations of instruction. This within-subjects difference was named delexicalization effect. In our analysis, cross-group differences in the segmentation of the natural-speech baseline were indeed apparent (see Figures 2 and 3B), suggesting that participants' segmentation strategies differed a priori across groups and thus our baseline-related measure of instruction effects was prudent.

Supporting the hypothesis of music-specific expectations, the delexicalization effect changed according to instruction in several aspects. Instruction influenced delexicalization effects on segment length: mean segment length decreased with delexicalization for the speech instruction, but it did not change for the music one. In addition, instruction changed delexicalization effects on the matching of segmentation maps with five out of seven theoretical segmentation models: for instance, the matching with Pitch rise models decreased with delexicalization for speech instruction, but it increased for music instruction (Figure 3B).

Our primary goal was to determine whether expectations in music segmentation *differed* from those in the speech-prosody domain, and thus we were interested in *any* interactions between delexicalization and instruction. Based on previous literature, we admitted the possibility that music instructions would increase reliance on global cues but, beyond that, our approach was exploratory regarding the contents of music-specific expectations. Note that, in the context of our delexicalization-effect-based approach, expectations must be framed in relative terms, i.e., how participants in each level of instruction diverged from natural speech when confronted with delexicalized versions.

The hypothesis that music segmentation would favor global cues (Pitch drop and Pitch rise) did not get support. It was true that participants under the music instruction favored Pitch rise (global cue), while those under the speech instruction devalued Pitch rise. However, the same did not go for Pitch drop, which is also a global cue. Critically, music participants favored local cues (Pitch break) and mixed cues (Pause plus pitch drop, Pause plus pitch rise) more than speech participants. Therefore, the dichotomy global–local seems irrelevant to distinguish between music and speech prosody segmentation.

Having excluded the global-cue hypothesis on music segmentation, what were we left with? First, the music instruction seems to have preserved the mechanisms of natural speech segmentation more than the speech prosody instruction: unlike speech prosody participants, music participants did not decrease segment length with delexicalization. They also preserved Pitch breaks, Pauses plus pitch drops and Pauses plus pitch rises more than speech prosody participants. The possibility that natural speech expectations may be more similar to music than prosody-specific expectations

themselves is an intriguing finding that deserves further discussion. Explanations for this finding may generally relate to the disturbing potential of delexicalized speech prosody stimuli. One possibility may be that speech prosody requires phonetic content to be fully decoded, while the same does not apply to music. This might relate to the phonetic advantage effect, according to which it is easier to imitate prosody-related pitch when prosody is accompanied by phonetic content [37]. Although the authors also found a phonetic advantage for music, contradictory evidence is available [38]. In the context of our study, it is possible that dissociating speech prosody from its original phonetic content (delexicalized speech prosody versions) may have disturbed prosodic segmentation to such an extent that delexicalized music versions remained closer to natural speech. Specifically, it is possible that such disturbance was caused and/or amplified by the violation of expectations that takes place when a linguistic stimulus presents itself deprived of phonetic content. An alternative possibility may relate to the characteristics of our stimuli, namely the music-like characteristic of our delexicalized versions. We tested delexicalized stimuli using discontinuous (musical) pitch, resulting from the audio-to-MIDI transformation. Although this was necessary to maximize the credibility of the music instruction while keeping the stimuli unchanged across instruction levels, this may have created a sense of strangeness in participants from the speech group. As a result, it is possible that speech participants did not activate the bottom-up approach that we expected for speech prosody, nor the music-like set of expectations. So, although our results make it clear that speech prosody was approached differently from music, it is possible that speech prosody may have been perceived as an undetermined, unfamiliar type of auditory stream, eliciting hybrid, atypical and/or unstable expectations. In order to rule out any limitations brought by the music-like character of delexicalized stimuli - it may be helpful to add control conditions in future studies, wherein participants in each instruction condition are also presented with continuous pitch versions as delexicalized stimuli. Although these two possibilities, delexicalized speech prosody is generally disturbing, or/and particularly disturbing with discontinuous pitch, make sense, we should bear in mind that participants from the two instruction conditions did not differ in their confidence level regarding segmentation responses. From this viewpoint, one might think participants in the speech prosody instruction condition were, at least consciously, not more disturbed than those in the music condition. Still, it is possible that confidence may go along with changes in processing modes, and this may have occurred with speech prosody participants as they went from natural speech to delexicalized versions.

A second manifestation of music-specific segmentation was the increased adherence to pitch rise cues, with the opposite trend observed for speech prosody segmentation. Pitch drop is a universal, default feature of human speech [39], possibly because it is a natural outcome of decreased air flow as one vocalizes continuously without a specific pitch plan. Differently, pitch rises require planning and resources to be executed. This type of vocal attitude is characteristic of music, and it is not surprising that we saw increased expectations for pitch rise under the music instruction. Finally, music participants were unreactive to Pauses plus pitch breaks, unlike speech prosody participants, who relied more heavily on these after delexicalization. One possibility may be that the coexistence of pauses and pitch breaks tend to be interpreted more as the ending of a musical section rather than as the ending of a phrase, driving music participants to ignore this type of boundary cues. Further studies could test this possibility, by eliciting both types of segmentation; sections vs. phrases.

Concerning general aspects of our study that may deserve investigation in future studies, one relates to the order of presentation of natural vs. delexicalized versions, which may raise concerns over priming effects. In our study, participants were first exposed to delexicalized versions, and then to the natural speech counterparts. We did this because we were concerned that music-instruction participants might raise hypotheses on the origin of the delexicalized stimuli in case we had done the reverse and started with natural speech, and we wanted to avoid the risk of having to eliminate participants due to such awareness. Our choice for the present study may have introduced priming effects, but the reverse option would have done the same. Critically, potential priming effects of delexicalized over natural speech versions were common to both instruction levels, and this was all that we had to control for in

face of our research question (does instruction influence the delexicalization effect?). Although the order of presentation was not likely responsible for the differences between instruction levels, it may have affected the type of expectations that were observed. This is the reason why it might be useful to counterbalance the order of block (natural vs. delexicalized stimuli) presentation in future studies. Another aspect concerns the variety of segmentation models we used, which is not exhaustive and may be expanded. Specifically, future studies may benefit from considering pre-boundary lengthening phenomena [27], which are known to guide segmentation in music and speech prosody, but which we did not consider here.

Our main finding was that there are music-specific expectations, or top-down influences, in music segmentation. Our results suggest that there is a "music-segmentation mode", different from the processing mode engaged in speech prosody segmentation, which we assumed to be a bottom-up, data-driven approach. Although we found support for different modes, our findings do not inform us on whether speech prosody engages any expectations at all: it may be the case that music segmentation recruits expectations, or top-down processing, but speech prosody does not (our working assumption), but it may also be true that speech prosody also engages expectations, even though different from those engaged in music. A third scenario could be that speech prosody engages expectations while music segmentation is purely bottom-up, but this would go against evidence that listeners rely on metric, structural cues, such as 4-bar phrases, to perform segmentation (See Introduction Section). The best way to address these questions is to better specify the type of expectations in each domain and find cross-studies replicable patterns.

Our main finding arose from an experimental paradigm that approached music and speech prosody in ways that may not be considered fully representative of these phenomena. To probe music segmentation, we used a series of rhythmically organized discontinuous pitches without tonal organization (pitches did not organize according to tonal harmony [40]) and conveyed by a musical timbre (bass). While this approach captures basic elements of what is considered "music" (discontinuous pitch, rhythm, musical timbre), it misses important elements of common-practice music, namely tonal harmony (implicit in tonal melodies) and metric regularity. From this viewpoint, we should admit that we did not probe music segmentation in a broad sense but, rather, the segmentation of a particular music style, likely similar to contemporary jazz music (as we told our participants). So, what would happen if we used mainstream music, in case it would be possible with our paradigm? Our guess is that music-specific expectations would be more salient, since both tonal harmony and metric regularity, both absent in speech prosody, work as music-related segmentation cues [41]. In this sense, the limitations of the stimuli we presented as music concerning the ecological validity of our findings may not be significant. As for the ways we probed speech prosody, these may have limited the generality of our conclusions, as we already discussed.

5. Conclusions

In sum, our study was novel in testing for music-specific expectations in music segmentation, and we found evidence for these within the frame of our paradigm and assumptions. The existence of music-specific expectations in segmentation remained underacknowledged in the field of music-language comparative studies on segmentation [14,20,24]. Our findings contribute to challenge the analogy between speech prosody and music that has remained implicit in the field, setting the stage for a "music mode" and a "speech prosody mode" in segmentation.

Supplementary Materials: Stimulus materials can be downloaded from: https://drive.google.com/file/d/1XGyxhRsByrcTmymHiCNnPRcCegwYYIoo/view?usp=sharing.

Author Contributions: Conceptualization, S.S. and S.L.C.; methodology, S.S. and S.L.C.; formal analysis, S.S. and C.D.; investigation, S.S. and C.D.; data curation, C.D.; writing—original draft preparation, S.S. and C.D.; writing—review and editing, S.S. and S.L.C.; supervision, S.S. and S.L.C.; project administration, S.S. and S.L.C.; funding acquisition, S.L.C.

Funding: This research was supported by Fundação para a Ciência e a Tecnologia under grant UID/PSI/00050/2013. and COMPETE 2020 program.

Acknowledgments: We are grateful to Filipa Salomé and José Batista for their help with the analysis, and to our participants.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Pitch Structure and Linguistic Content of the Five Stimuli

In each figure, pitch structure is plotted above for natural speech stimuli, and below for delexicalized versions. Boundaries under natural speech versions indicate segmentation based on linguistic context, specifically on sentences as defined in transcription texts.

In transcriptions, // indicates clause boundaries, **bold and underlined** words indicate the main verb, and underlined words the verb (and conjunctions) in subordinate and/or coordinate clauses. Self-corrections and hums are included.

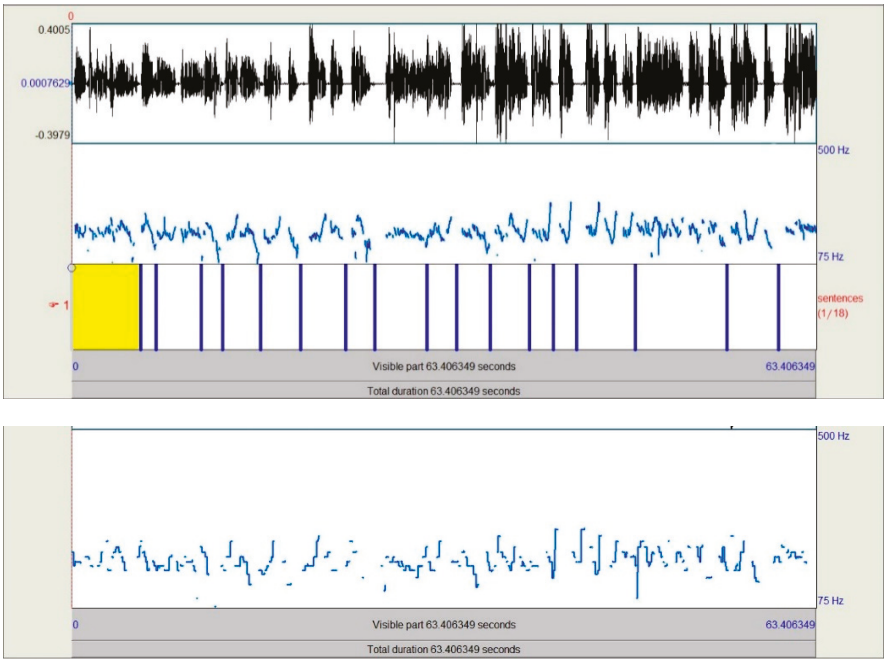


Figure A1. Stimulus 1—Pitch structure.

Table A1. Stimulus 1—Transcription.

	Portuguese	English
1.	O gravador . . . com os bichinhos todos, br-brinquedos, um cãozinho **anda** à volta dos brinquedos. //	The recorder . . . with all the little animals, to-toys, a puppy **goes** around the toys. //
2.	Os brinquedos **são** animaizinhos: //	The toys **are** little animals: //
3.	**Tem** uma boneca, um pintainho... uma foca, // e o cãozinho anda à volta deles. //	**There's** a doll, a chick . . . a seal, // and the puppy goes around them. //
4.	Olha, **começaram a andar**, os bichinhos. //	Look, they **started to walk**, the animals. //
5.	**Ficou** só o, a andar à volta à beira do gravador. //	Only the car **stayed**, going around next to the recorder. //
6.	A mão **veio**, // apanhou o carro... // e amassou-o. //	The hand **came**, // caught the car . . . // and mashed it. //
7.	**Veio** agora uma rãzinha... e uma bonequinha. //	Just now **came** a little frog . . . and a little doll. //
8.	A bonequinha **foi amassada**. //	The little doll **was mashed**. //
9.	E foi... com o pincel, **varreu** a bonequinha para o lado, para ela desaparecer. //	And went... with the paintbrush, **brushed** the little doll to the side to make her disappear. //
10.	**Veio** o pintainho, // [e] passou-lhe a bola por cima . . . //	The chick **came**, // [and] the ball passed over it . . . //
11.	Com o apanhador e com a vassoura, **apanharam** o pintainho. //	With the dustpan and the broom, they **got** the chick. //
12.	Depois **veio** outro patinho, // [e] também levou com o apanhador... //	Then **came** another little duck, // [and] the dustpan also hit it... //
13.	Outro bonequinho **foi** com a mão... //	Another little doll **was** with the hand . . . //
14.	**Apanhou** os dois.... //	It **caught** both . . . //
15.	**Veio** uma foquinha... // [e] foi apanhada . . . a foquinha . . . //	Then a little seal **came** . . . // [and] it was caught . . . the little seal . . . //
16.	**Veio** agora outro bichinho, // foi varrido, // e a rãzinha virou-se. //	Now another little animal **came**, // it was brushed away, // and the little frog turned, //
17.	**Cortaram-lhe** a cabeça . . . // tiraram-lhe a cabeça . . . // meteram os dedos, // [e] puxaram. //	They **cut off** its head . . . // took off its head . . . // put their fingers in, // [and] pulled. //
18.	**Mostrou** dois dedos - um dedo a avisar . . . Três dedos. //	It showed two fingers – one finger warning . . . Three fingers. //
19.	**[Diz:]** "Futebol Clube do Porto, tricampeão, noventa e quatro, noventa e sete". //	**[It says:]** Porto Football Club, tri-champion, ninety four, ninety seven. //

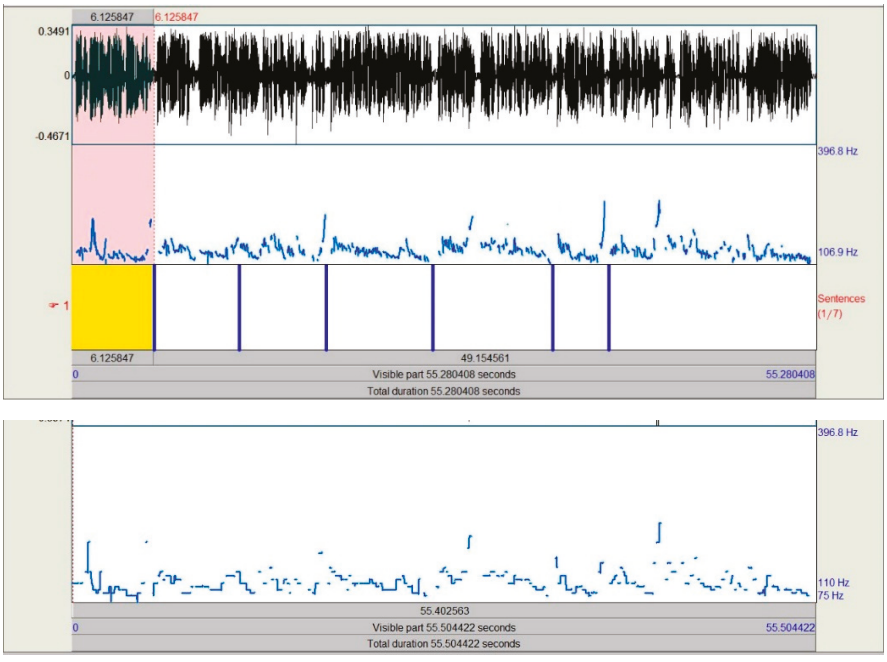


Figure A2. Stimulus 2—Pitch structure.

Table A2. Stimulus 2 - Transcription.

Portuguese	English
1. **Era** um grupo muito organizado, hã, desde a base até, hã... ao topo. //	1. It **was** a very organized group, uhh, from the base to, uhh... the top. //
2. hã, Um grupo que, numa primeira fase … // como eu referi, // **é** um grupo de estrangeiros. //	2. uhh A group that, in a first phase … // as I mentioned, // **is** a group of foreigners. //
3. Alguns deles, hã … no passado, já **foram** funcionários de empresas // que prestam serviços à Portugal Telecom. //	3. Some of them, uhh … in the past, **had been** employees of companies // that provide services to Portugal Telecom. //
4. hã- E, como tal, **tinham** acesso, muitas vezes privilegiado, a localização de- de- de linhas inativas da- da Portugal Telecom. //	4. uhh- And, as such, they **had** access, often privileged, to the location of- of- of inactive lines of- of Portugal Telecom. //
5. hã- Eles, identificando, digamos assim, estas localizações … // **procediam** então ao furto, hã, de determinados troços, hã, das linhas de cobre. //	5. uhh- They, by identifying, as it were, these locations … // they **would proceed** to the theft, uhh, of specific sections, uhh, of the copper lines. //
6. Para tal, **recorriam**, hã, à utilização de viaturas tipo furgon … //	6. For this, they **resorted,** uhh, to the use of vehicles like Furgon vans … //
7. [As] viaturas estas [**estavam**] preparadas com um alçapão, hã, no seu fundo … que colocavam em cima das caixas, procedendo então à entrada nas respetivas caixas e ao corte de pequenos troços destas linhas // que encaminhavam para- que encaminhavam para um interior da- da viatura. //	7. Vehicles [**were**] prepared with a trapdoor, uhh, in its bottom … that they would place on top of the boxes, // proceeding then to enter in the respective boxes and to cut small sections of these lines, // that they redirected to- they redirected to the inside of- of the vehicle. //

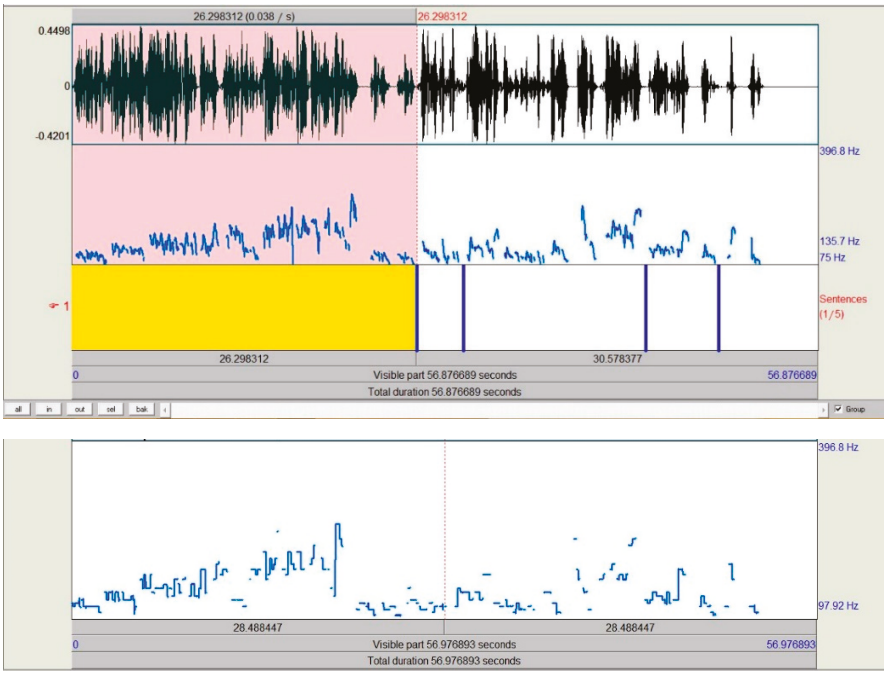


Figure A3. Stimulus 3—Pitch structure.

Table A3. Stimulus 3—Transcription.

Portuguese	English
1. Se vivendo entre o povo és virtuoso e nobre, //se vivendo entre os reis conservas a humildade, // se inimigo ou amigo, o poderoso e o pobre, são iguais para ti, à luz da eternidade, // se quem conta contigo // encontra mais que a conta, // se podes empregar os sessenta segundos do minuto //que passa// em obra de tal monta que o minuto se espraie em séculos fecundos,// então, ó ser sublime, o mundo inteiro **é** teu. //	1. If while living amongst the people // you are virtuous and noble, // if while living amongst kings // you keep your humility, // if foe or friend, powerful or poor, are equal to you in the light of eternity, //If the ones who count on you // find more than the bill, //if you can fill the sixty seconds of the minute // that passes in work of such amount, // in a way that the minute spreads in fertile centuries, // then, oh sublime being, the whole world **is** yours. //
2. Já **dominaste** os reis, os espaços! //	2. You'**ve conquered** kings, spaces! //
3. Mas, ainda para além, um novo sol **rompeu**, abrindo o infinito aos rumos dos teus passos, pairando numa esfera... acima deste plano... // sem recear jamais // que os erros te retomem . . . //	3. Yet, still beyond, a new sun **has risen,** opening infinity towards your steps, hovering in a sphere . . . above this plane... //never fearing // that mistakes return to you . . . //
4. Quando já nada houver em ti // que seja humano . . . // **alegra-te**, meu filho. //	4. When there is nothing left in you that is human . . . //**rejoice**, my son. //
5. Então... **serás** um homem! //	5. Then . . . you **will be** a man. //

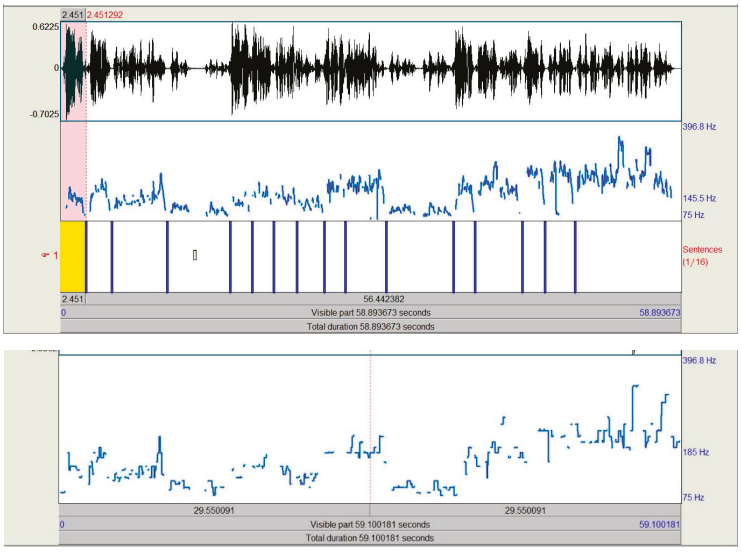


Figure A4. Stimulus 4—Pitch structure.

Table A4. Stimulus 4—Transcription.

	Portuguese		English
1.	À roda da nau **voou** três vezes. //	1.	Around the ship's wheel it **flew** three times, //
2.	**Voou** três vezes a chiar // e disse: //	2.	It **flew** three times creaking // and said: //
3.	"Quem é que **ousou entrar** nas minhas cavernas //que não desvendo, meus tetos negros do fim do mundo?"//	3.	"Who **has dared** to enter my caverns // that I don't unravel, my dark ceilings of the end of the world?" //
4.	E o homem do leme **disse**, tremendo: "El-Rei D. João Segundo". //	4.	And the helmsman **said**, trembling: "The King D. João the Second". //
5.	"De quem **são** as velas // onde me roço? //	5.	"Whose **are** the sails // I'm rubbing? //
6.	De quem [**são**] as quilhas // que vejo // e ouço?"//	6.	Whose **are** the keels //that I see // and hear?"//
7.	**Disse** o mostrengo // e rodou três vezes. //	7.	The moster **said** // and turned three times. //
8.	Três vezes **rodou**, imundo e grosso. //	8.	Three times he **turned**, filthy and thick. //
9.	"Quem **vem poder** // o que só eu posso? //	9.	"Who **comes** to do // what only I can do? //
10.	[Eu] que **moro** // onde nunca ninguém me visse // e escorro os medos do mar sem fundo?" //	10.	I **live** // where no one has ever seen me // and I drain the fears of the bottomless sea?" //
11.	E o homem do leme **tremeu** // e disse: // "El-Rei D. João Segundo". //	11.	And the helmsman **trembled** // and said: // "The King D. João the Second". //
12.	Três vezes do leme as mãos **ergueu.** //	12.	Three times from the helm his hands **were raised.** //
13.	Três vezes ao leme as **reprendeu**, // e disse no fim de tremer três vezes: //	13.	Three times to the helm they w**ere reattached,** // and said he after trembling three times: //
14.	"Aqui ao leme **sou** mais do que eu! //	14.	"Here at the helm **I am** more than myself! //
15.	**Sou** um povo //que quer o mar que é teu! //	15.	I **am** a people that wants the sea that is yours! //
16.	E mais que o mostrengo //que me a alma teme // e roda nas trevas do fim, // **manda** a vontade //que me ata ao leme// de El Rei D. João Segundo!" //	16.	And more than the monster // that my soul fears, // and turns in the darkness at the end, // the will that commands and ties me to the helm **is** the one of King D. João the Second!" //

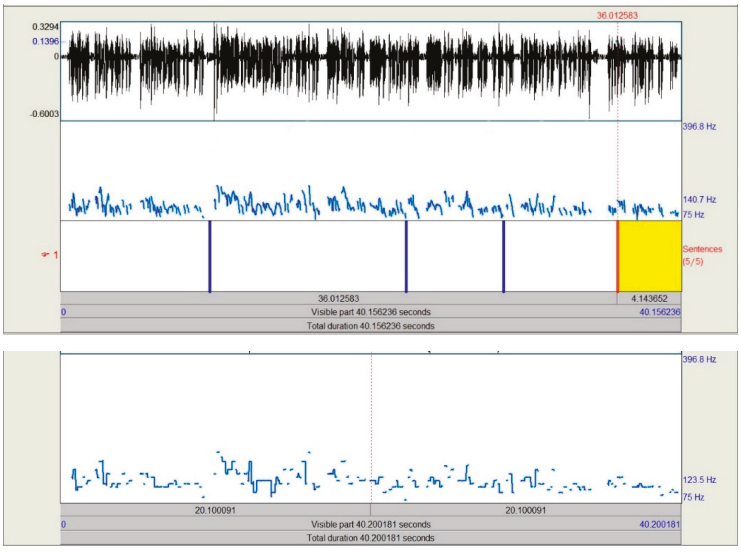


Figure A5. Stimulus 5—Pitch structure.

Table A5. Stimulus 5—Transcription.

Portuguese	English
1. A pedido da direção do canal, e depois do arranque do programa Curto Circuito, **é-nos solicitada** a divulgação de um comunicado, // que passo a citar: //	1. By request of the board of the channel, and after the beginning of the show Curto Circuito, we **are asked** to divulge a bulletin // that I hereby quote: //
2. "Após ter analisado em direto o conteúdo do programa denominado Curto Circuito, // a direção da SIC Radical **decidiu romper** unilateralmente o contrato // que a ligava à produtora Sigma 3. //	2. "After having analyzed the content of the show called Curto Circuito, // the board of SIC Radical has **decided to** unilaterally **break** the contract that connected it to the producer Sigma 3. //
3. Com esta medida, o programa denominado Curto Circuito **é retirado** imediatamente de grelha." //	3. With this action, the show called Curto Circuito **is** immediately **withdrawn** from the grid." //
4. **Assina** o diretor do canal, Francisco Penim, // que está comigo em estúdio para clarificar . . . esta posição. //	4. It is signed by the channel's director, Francisco Penim, // **who is** with me in the studio to clarify . . . this position. //
5. Francisco, boa tarde. A pergunta não **pode** ser outra: Porquê . . . esta decisão? //	5. Francisco, good afternoon. The question **cannot be** other: Why . . . this decision? //

References

1. Frazier, L.; Carlson, K.; Cliftonjr, C. Prosodic phrasing is central to language comprehension. *Trends Cogn. Sci.* **2006**, *10*, 244–249. [CrossRef] [PubMed]
2. Krumhansl, C.L.; Jusczyk, P.W. Infants' Perception of Phrase Structure in Music. *Psychol. Sci.* **1990**, *1*, 70–73. [CrossRef]
3. Palmer, C.; Krumhansl, C.L. Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Percept. Psychophys.* **1987**, *41*, 505–518. [CrossRef] [PubMed]
4. Stoffer, T.H. Representation of Phrase Structure in the Perception of Music. *Music. Perception: Interdiscip. J.* **1985**, *3*, 191–220. [CrossRef]

5. Hirotani, M.; Frazier, L.; Rayner, K. Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *J. Mem. Lang.* **2006**, *54*, 425–443. [CrossRef]
6. Steinhauer, K.; Friederici, A.D. Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *J. Psycholinguist. Res.* **2001**, *30*, 267–295. [CrossRef]
7. Rader, G. Creating printed music automatically. *Computer* **1996**, *29*, 61–68. [CrossRef]
8. Sloboda, J.A. Experimental Studies of Music Reading: A Review. *Music. Perception: Interdiscip. J.* **1984**, *2*, 222–236. [CrossRef]
9. Patel, A.D.; Peretz, I.; Tramo, M.; Labreque, R. Processing Prosodic and Musical Patterns: A Neuropsychological Investigation. *Brain Lang.* **1998**, *61*, 123–144. [CrossRef]
10. Steinhauer, K.; Alter, K.; Friederici, A.D. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* **1999**, *2*, 191–196. [CrossRef]
11. Brown, M.; Salverda, A.P.; Dilley, L.C.; Tanenhaus, M.K. Expectations from preceding prosody influence segmentation in online sentence processing. *Psychon. Bull. Rev.* **2011**, *18*, 1189–1196. [CrossRef] [PubMed]
12. de Carvalho, A.; Dautriche, I.; Christophe, A. Preschoolers use phrasal prosody online to constrain syntactic analysis. *Dev. Sci.* **2016**, *19*, 235–250. [CrossRef] [PubMed]
13. Friederici, A.D.; Mecklinger, A.; Spencer, K.M.; Steinhauer, K.; Donchin, E. Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cogn. Brain Res.* **2001**, *11*, 305–323. [CrossRef]
14. Knösche, T.R.; Neuhaus, C.; Haueisen, J.; Alter, K.; Maess, B.; Witte, O.W.; Friederici, A.D. Perception of phrase structure in music. *Hum. Brain Mapp.* **2005**, *24*, 259–273. [CrossRef] [PubMed]
15. Firestone, C.; Scholl, B.J. Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behav. Brain Sci.* **2015**, *39*, 1–77. [CrossRef] [PubMed]
16. Schuetze-Coburn, S.; Shapley, M.; Weber, E.G. Units of Intonation in Discourse: A Comparison of Acoustic and Auditory Analyses. *Lang. Speech* **1991**, *34*, 207–234. [CrossRef]
17. Wagner, M.; Watson, D.G. Experimental and theoretical advances in prosody: A review. *Lang. Cogn. Process.* **2010**, *25*, 905–945. [CrossRef] [PubMed]
18. Wightman, C.W.; Shattuck-Hufnagel, S.; Ostendorf, M.; Price, P.J. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* **1992**, *91*, 1707–1717. [CrossRef]
19. Li, W.; Yang, Y. Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience* **2009**, *158*, 1416–1425. [CrossRef]
20. Pannekamp, A.; Toepel, U.; Alter, K.; Hahne, A.; Friederici, A.D. Prosody-driven Sentence Processing: An Event-related Brain Potential Study. *J. Cogn. Neurosci.* **2005**, *17*, 407–421. [CrossRef]
21. Neuhaus, C.; Knösche, T.R.; Friederici, A.D. Effects of Musical Expertise and Boundary Markers on Phrase Perception in Music. *J. Cogn. Neurosci.* **2006**, *18*, 472–493. [CrossRef]
22. Silva, S.; Barbosa, F.; Marques-Teixeira, J.; Petersson, K.M.; Castro, S.L. You know when: Event-related potentials and theta/beta power indicate boundary prediction in music. *J. Integr. Neurosci.* **2014**, *13*, 19–34. [CrossRef] [PubMed]
23. Silva, S.; Branco, P.; Barbosa, F.; Marques-Teixeira, J.; Petersson, K.M.; Castro, S.L. Musical phrase boundaries, wrap-up and the closure positive shift. *Brain Res.* **2014**, *1585*, 99–107. [CrossRef] [PubMed]
24. Glushko, A.; Steinhauer, K.; DePriest, J.; Koelsch, S. Neurophysiological Correlates of Musical and Prosodic Phrasing: Shared Processing Mechanisms and Effects of Musical Expertise. *PLoS ONE* **2016**, *11*, 0155300. [CrossRef] [PubMed]
25. Nan, Y.; Knösche, T.R.; Zysset, S.; Friederici, A.D. Cross-cultural music phrase processing: An fMRI study. *Hum. Brain Mapp.* **2008**, *29*, 312–328. [CrossRef] [PubMed]
26. Buxó-Lugo, A.; Watson, D.G. Evidence for the Influence of Syntax on Prosodic Parsing. *J. Mem. Lang.* **2016**, *90*, 1–13. [CrossRef] [PubMed]
27. Cole, J.; Mo, Y.; Hasegawa-Johnson, M. Signal-based and expectation-based factors in the perception of prosodic prominence. *Lab. Phonol. J. Assoc. Lab. Phonol.* **2010**, *1*, 425–452. [CrossRef]
28. Cole, J.; Mo, Y.; Baek, S. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Lang. Cogn. Process.* **2010**, *25*, 1141–1177. [CrossRef]
29. Meyer, L.; Henry, M.J.; Gaston, P.; Schmuck, N.; Friederici, A.D. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cereb Cortex* **2017**, *27*, 4293–4302. [CrossRef]

30. Mattys, S.L.; White, L.; Melhorn, J.F. Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *J. Exp. Psychol. Gen.* **2005**, *134*, 477–500. [CrossRef]
31. Itzhak, I.; Pauker, E.; Drury, J.E.; Baum, S.R.; Steinhauer, K. Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport* **2010**, *21*, 8–13. [CrossRef] [PubMed]
32. Goswami, U. A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Curr. Dir. Psychol. Sci.* **2017**, *27*, 56–63. [CrossRef]
33. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [CrossRef] [PubMed]
34. Behrens, S. Characterizing sentence intonation in a right hemisphere-damaged population*1. *Brain Lang.* **1989**, *37*, 181–200. [CrossRef]
35. Vaissière, J.; Michaud, A. Prosodic constituents in French: a data-driven approach. In *Prosody and syntax*; Fónagy, Y.K.I., Moriguchi, T., Eds.; John Benjamins: Amsterdam, The Netherlands, 2006; pp. 47–64.
36. Jackson, D.A.; Somers, K.M.; Harvey, H.H. Similarity Coefficients: Measures of Co-Occurrence and Association or Simply Measures of Occurrence? *Am. Nat.* **1989**, *133*, 436–453. [CrossRef]
37. Mantell, J.T.; Pfordresher, P.Q. Vocal imitation of song and speech. *Cognition* **2013**, *127*, 177–202. [CrossRef]
38. Berkowska, M.; Bella, S.D. Reducing linguistic information enhances singing proficiency in occasional singers. *Annals N. Y. Acad. Sci.* **2009**, *1169*, 108–111. [CrossRef]
39. Hirst, D.; Di Cristo, A. French intonation: a parametric approach. *Die Neuer. Spr.* **1984**, *83*, 554–569.
40. Lerdahl, F.; Jackendoff, R.S. *A Generative Theory of Tonal Music*; MIT Press: London, England, 1996; ISBN 978-0-262-26091-6.
41. Jensen, K. Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP J. Adv. Signal Process* **2007**, *2007*, 159. [CrossRef]

brain sciences

MDPI

Article

Is It Speech or Song? Effect of Melody Priming on Pitch Perception of Modified Mandarin Speech

Chen-Gia Tsai [1,2] and Chia-Wei Li [3,*]

[1] Graduate Institute of Musicology, National Taiwan University, Taipei 106, Taiwan; tsaichengia@ntu.edu.tw
[2] Neurobiology and Cognitive Science Center, National Taiwan University, Taipei 106, Taiwan
[3] Department of Radiology, Wan Fang Hospital, Taipei Medical University, Taipei 116, Taiwan
* Correspondence: 799032@w.tmu.edu.tw

Received: 22 September 2019; Accepted: 21 October 2019; Published: 22 October 2019

Abstract: Tonal languages make use of pitch variation for distinguishing lexical semantics, and their melodic richness seems comparable to that of music. The present study investigated a novel priming effect of melody on the pitch processing of Mandarin speech. When a spoken Mandarin utterance is preceded by a musical melody, which mimics the melody of the utterance, the listener is likely to perceive this utterance as song. We used functional magnetic resonance imaging to examine the neural substrates of this speech-to-song transformation. Pitch contours of spoken utterances were modified so that these utterances can be perceived as either speech or song. When modified speech (target) was preceded by a musical melody (prime) that mimics the speech melody, a task of judging the melodic similarity between the target and prime was associated with increased activity in the inferior frontal gyrus (IFG) and superior/middle temporal gyrus (STG/MTG) during target perception. We suggest that the pars triangularis of the right IFG may allocate attentional resources to the multi-modal processing of speech melody, and the STG/MTG may integrate the phonological and musical (melodic) information of this stimulus. These results are discussed in relation to subvocal rehearsal, a speech-to-song illusion, and song perception.

Keywords: melody perception; tonal language; inferior frontal gyrus; priming effect

1. Introduction

Tonal languages are characterized by the use of lexical tones for distinguishing lexical semantics. Lexical tones include distinct level pitches and pitch-glide patterns. Owing to pitch variation, spoken utterances in tonal languages are rich in melody and sometimes comparable to music. Related to this, it has been acknowledged that tonal languages and music share similarity in the perceptual-cognitive processing of pitch. Recognition of lexical tones relies on relations between successive pitches [1–3], and thus the underlying neural substrates partially overlapped with those underlying recognition of musical pitch intervals [4]. Evidence of similarity between spoken utterances of tonal languages and music also comes from traditional music. The distinction between speech and song is blurred in many genres of Chinese musical theater. For example, it has been suggested that eight oral delivery types in Cantonese opera occupy different positions on a speech-music spectrum according to their tonal features, rhythmic features, and instrumental accompaniment [5]. In Chinese opera, a sung utterance may be perceived as somewhat speech-like because of high congruency between its musical melody and lexical tones. On the other hand, the pitches of a spoken utterance could be embedded into the musical scale provided by the accompaniment music. Using this musical scale as a tonal schema, listeners may perceive this utterance as song.

The fact that the tonal context provided by an instrumental accompaniment could perceptually transform speech of a tonal language to song raises a possibility that the listener could perceive the melody of a spoken utterance as a musical melody when he/she is primed by appropriate musical

cues. In the present study, we reported a novel priming effect of musical melody on pitch perception of spoken Mandarin utterances. The target was a speech-like stimulus. The melody of this target was mimicked by a musical melody, which served as the prime. When the listener was primed by this musical melody, he/she tended to perceive the target as song.

Previous studies have reported that acoustically identical English utterances can be perceived as either speech or song. Deutsch and colleagues found that when a spoken English phrase was repeated several times, listeners were likely to perceive this phrase as song [6]. The authors hypothesized that exposure to repetition of a speech fragment may be associated with greater activity in the neural substrates of pitch processing, relative to the condition in which a spoken phrase was presented once. Moreover, this repetition effect may result in a re-evaluation of prosodic features of this spoken phrase [7]. These two hypotheses for the speech-to-song illusion were supported by previous neuroimaging studies demonstrating that (1) the effect of perceiving a spoken phrase as song via repetition localized to the right mid-posterior superior temporal sulcus (STS) and middle temporal gyrus (MTG) implicated in pitch processing [8,9], and (2) the subjective vividness of the speech-to-song illusion was positively correlated with activity in a left frontotemporal loop implicated in evaluation of linguistic prosody. This left frontotemporal loop comprises the inferior frontal gyrus (IFG), frontal pole, and temporal pole [9].

Using functional magnetic resonance imaging (fMRI), the present study aimed at specifying the neural underpinnings of the perceptual transformation from Mandarin speech-like utterances to song across a musical prime that mimics the melody of speech. In light of the aforementioned studies of the speech-to-song illusion in English, we hypothesized that the effect of melody priming on the pitch processing of Mandarin speech-like utterances would be associated with increased activity in the IFG, which may contribute to the cognitive processes for attending to the melodic features of speech and for comparing speech with the musical prime. Specifically, the pars triangularis of the right IFG (IFGtri) seems to play a prominent role in evaluation of prosodic information of speech [10–12]. We expected to observe greater activity within the right IFGtri during listening to Mandarin speech-like utterances preceded by a melody prime, compared to listening to the same stimulus without melody priming. In addition, we hypothesized that the anterior insula and supplementary motor area (SMA) implicated in subvocal rehearsal [13] may co-activate with the right IFGtri because participants may engage subvocal rehearsal strategies for encoding the melodic features of Mandarin speech-like utterances.

In addition to attention control and sensorimotor mechanisms, the present study was also expected to shed new light on the perceptual processing of song. Sammler and colleagues employed a functional magnetic resonance adaptation paradigm to identify the neural correlates of binding lyrics and tunes in unfamiliar song [14]. Results revealed that the left mid-posterior STS showed an interaction of the adaptation effects for lyrics and tunes. The authors suggested that this region may contribute to an integrative processing of lyrics and tunes. Alonso and colleagues reported that binding lyrics and tunes for the encoding of new songs was associated with the involvement of the bilateral mid-posterior MTG [15]. In the present study, a Mandarin speech-like utterance could be perceived as song when it was preceded by a musical melody mimicking the melody of the utterance. We hypothesized that melody priming would lead to increased activity in the STS/MTG implicated in binding lyrics and tunes during song perception.

2. Materials and Methods

2.1. Participants

Twenty native Mandarin speakers (age range 20–43 years; six males) participated in the fMRI experiment, in which three fMRI scanning runs for the present linguistic study (focusing on a tonal language) alternated with two fMRI scanning runs for a musical study (focusing on symphonies and concertos). This design was used to minimize affective habituation that could occur with repeated exposure to the same emotional music. The selection and recruitment of participants are mentioned

in the next paragraph. Other methods and results of the musical study are not mentioned further in this paper.

Participants were recruited via a public announcement on the internet, which stated the requirement of a high familiarity with Western classical music. In a pre-scan test, volunteers were asked to write down their feelings in response to the musical stimuli in the musical study. Eight musical stimuli with duration of 30 s were presented in a fixed order. After listening to each 30-s stimulus, the volunteers were asked to write down their feelings in response to the passage just before the theme recurrence and their feelings in response to the theme recurrence. They were also asked to explain their feelings in terms of musical features. The first author of this article (a musicologist) selected participants for the fMRI experiment according to the following inclusion criteria: (1) more than five passages just prior to the theme recurrence evoked his/her anticipation; (2) the recurrence of more than five themes evoked a feeling of resolution; and (3) their feelings were appropriately explained in terms of musical features for more than five excerpts. Thirty-one adult volunteers completed this questionnaire. Twenty-seven volunteers met our screening criteria. Twenty of them completed the fMRI experiment. They were free from neurological, psychiatric, or auditory problems. Fifteen participants studied musical instruments for six years or more. The participants were compensated with approximately 16 USD after the completion of fMRI scan.

In the present linguistic study, participants were excluded from analyses of fMRI data if they were unable to discriminate between matched and mismatched trials at better than chance levels (see 2.5. Data Analysis). A female participant was excluded in this way. The data of another female participant were discarded because of incomplete behavioral data acquisition during the fMRI session. Thus, the final sample included in fMRI analyses consisted of 18 adults (mean age = 27.1 years; SD = 6.5 years; mean experience in most experienced instrument = 8.9 years; SD = 4.2 years; six males). Written informed consent was obtained from each participant prior to participation in the study. All research procedures were performed in accordance with a protocol approved by the Institutional Review Board of National Taiwan University (201611HM008). This study was conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

Auditory stimuli were noise, linguistic stimuli, and musical stimuli. The noise stimulus was white noise with a duration of 1.6 s. The duration of each linguistic and musical stimulus was 1.9–2.4 s. The linguistic stimuli were spoken sentences in Mandarin. Each sentence contained six characters. A female broadcaster was invited to recite 65 sentences with flat affect and natural prosody. These materials were recorded and saved as digital sound files.

To generate stimuli that can be perceived as either speech or song, three steps of pitch adjustment were applied on the spoken utterances. These steps are similar to the auto-tune processes used to "songify" news reports or any normal speech. The first step was "quantizing"; the pitches were adjusted to match the nearest note in the chromatic musical scale. The second step was "flattening"; pitch glides of these spoken utterances were flattened by approximately 95%. Figure 1 illustrates the effects of quantizing and flattening on a spoken Mandarin sentence. Third, the first author adjusted the melody of each utterance to match the C major or B-flat major scales by transposing some pitches by a semitone. These three steps were carried out using Cubase (Steinberg Media Technologies GmbH by VeriSign, Inc., HH, Germany) The modified spoken utterances can be perceived as speech because of high congruency between their pitch contours and lexical tones. They can also be perceived as song because their pitches can be embedded into the musical scale. Among the 65 utterances, 50 utterances for the scanning session and 6 utterances for the training session were selected by the first author.

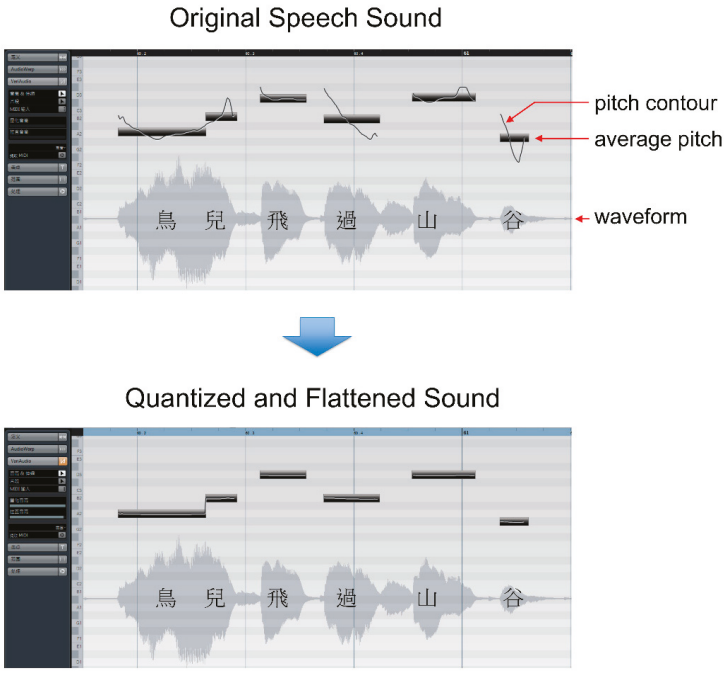


Figure 1. The first and second steps (quantizing and flattening) of pitch adjustment of spoken utterances using Cubase (Steinberg Media Technologies GmbH by VeriSign, Inc.) The pitches were adjusted to match the nearest note in the chromatic musical scale (quantizing). All pitch glides of the spoken utterances were flattened by approximately 95% (flattening).

All musical stimuli were melodies containing 6–10 notes, with the pitch ranging from E3 to F4 (fundamental frequency 164.8–349.2 Hz). There were three types of musical stimuli: Match, Mismatch, and Melody-Variation types. The musical stimuli of the Match type were melodies extracted from the linguistic stimuli using "MIDI extraction" in Cubase. As a result, each melody of the Match type closely resembles the melody of a linguistic stimulus. Each melody of the Mismatch type was generated from each melody of the Match type by elevating or lowering the pitches of 4–7 notes by 2–9 semitones while keeping the rhythm and tonality unchanged. The musical stimuli of the Melody-Variation type were paired melodies; the first melody (prime) was composed by the first author of this article, and the second melody (target) was generated from the first melody by elevating or lowering the pitches of 4–7 notes by 2–9 semitones while keeping the rhythm and tonality unchanged. All musical stimuli were in the C major or B-flat major tonalities and generated with a virtual musical instrument named "oboe" using Reason 7.0 (Propellerhead Inc., STH, Sweden).

There were five conditions in this study, as depicted in Figure 2. The experimental condition was melody-language-match (ML-match) condition, in which the prime was a musical stimulus mimicking the melody of the target linguistic stimulus. In the noise-language (NL) condition, the prime was noise, and the target was a linguistic stimulus. In the melody-melody-match (MM-match) condition, the prime and the target were the same musical melody. The NL and MM-match conditions were two control conditions of this study. Compared to NL, ML-match additionally demanded attention to the melodic features of speech, pitch processing, and tonal working memory. Both ML-match and MM-match demanded pitch processing and tonal working memory, as participants were asked to judge the melodic similarity between the prime and target. Compared to MM-match, ML-match

additionally demanded selective allocation of attention to the melodic features of speech. For audio examples of the stimuli for NL and ML-match, see Supplementary Materials.

The stimuli of the Mismatch and Melody-Variation types were used in the melody-language-mismatch (ML-mismatch) and melody-melody-mismatch (MM-mismatch) conditions, respectively. In the ML-mismatch condition, the prime was a musical melody that mismatched the melody of the target linguistic stimulus. In the MM-mismatch condition, the prime was a musical melody that mismatched the target musical melody. There were 20 trials in each of the experimental and control conditions, whereas there were 10 trials in each of the two mismatch conditions. The fMRI data for the two mismatch conditions were not analyzed. The target stimuli for ML-match and NL were counter-balanced across participants; the spoken utterances in ML-match were used as the target stimuli in NL for the other 10 participants, and vice versa.

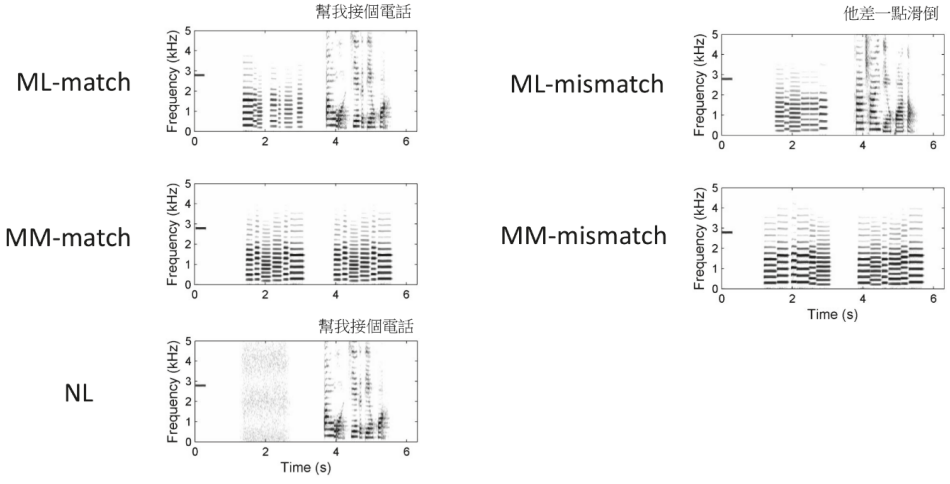


Figure 2. Examples of spectrograms of the stimuli in the five conditions. (ML-match: melody-language-match; ML-mismatch: melody-language-mismatch; MM-match: melody-melody-match; MM-mismatch: melody-melody-mismatch; NL: noise-language).

2.3. Procedure

The current study included a training session and a scanning session, which were separated by 5–10 min. In the training session, the participants were trained outside the MRI scanner room to familiarize themselves with the tasks. The experimenter explained to each participant with a PowerPoint presentation and sound examples that the melody of a Mandarin utterance can be compared to a musical melody. For demonstration, the experimenter rated the melodic similarity between a target and a prime on a 4-point Likert scale for five trials (one trial for each condition) by pressing a button (a right-most button for "very similar", a right button for "slightly similar", a left button for "slightly dissimilar", and a left-most button for "very dissimilar"). In a similar manner, the participant practiced six trials, including two trials for ML-match and one trial for each of the other four conditions. This training session lasted approximately 15 min. None of the stimuli used in the training session were presented in the scanning session.

Schematic description of the procedure of the fMRI experiment is illustrated in Figure 3. There were five runs for the whole fMRI design, with three musical runs alternating with two linguistic (speech–melody) runs. The duration of each run was approximately 450 s. In the musical runs, participants were instructed to listen to famous symphonies and concertos or atonal random sequences. Methods and results for these musical runs will be detailed in another article.

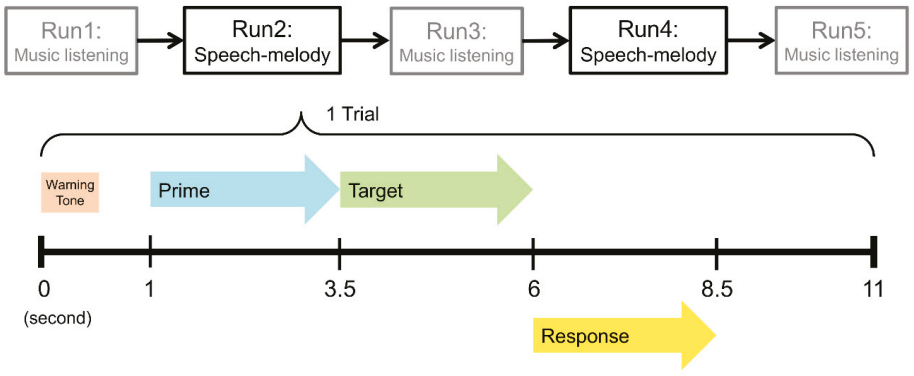


Figure 3. Schematic description of the procedure of the functional magnetic resonance imaging (fMRI) experiment.

Auditory stimuli in the linguistic runs were delivered through scanner-compatible headphones at a volume sufficiently loud enough that participants could readily perceive the stimuli over the scanner noise. Eighty trials of the five conditions were presented in two linguistic runs in a pseudorandom order. Each trial began with a warning tone (2.8 kHz, 0.3 s). Then, the prime and target stimuli were sequentially presented. The participants were instructed to listen to them and to rate the similarity of their melodies by pressing a button (a right-most button for "very similar" and similarity-score of 4, a right button for "slightly similar" and similarity-score of 3, a left button for "slightly dissimilar" and similarity-score of 2, and a left-most button for "very dissimilar" and similarity-score of 1). This task was used to assess whether participants were attending to the task. In total, the pre-scan training session and the scanning session took approximately 85 min for each participant.

Twenty-four to twenty-seven months after the fMRI experiment, the participants were asked to fill out a short online questionnaire. Fourteen participants completed this questionnaire. In this online questionnaire, auditory stimuli of five trials of the ML-match condition and five trials of the NL condition were randomly selected and presented in a random order. The participants rated each of the utterance (target stimulus) using a sliding scale to indicate how speech-like or song-like it was. Then, they rated on a sliding scale how much they agreed or disagreed these two statements: 'I paid more attention to the musical melodies of the utterances that were preceded by matched melodies, compared to those preceded by noise'; 'I more tended to covertly imitate the utterances that were preceded by matched melodies, compared to those preceded by noise.' The results of this questionnaire were expected to reveal how the melody prime affected the processing of the target utterance.

2.4. MRI Data Acquisition

For imaging data collection, participants were scanned using a 3T MR system (MAGNETOM Prisma, Siemens, Erlangen, Germany) and a 20-channel array head coil at the Imaging Center for Integrated Body, Mind, and Culture Research, National Taiwan University. In the functional scanning, about 2.5 mm slices of axial images were acquired using a gradient echo planar imaging (EPI) with the following parameters: time to repetition = 2500 ms, echo time = 30 ms, flip angle = 87°, in-plane field of view = 192 × 192 mm, and acquisition matrix = 78 × 78 × 45 to cover whole cerebral area. For spatial individual-to-template normalization in preprocessing, a Magnetization Prepared Rapid Gradient Echo T1-weighted imaging with spatial resolution of 0.9 mm isotropic was acquired for each participant.

2.5. Data Analyses

One-sample one-tailed *t*-tests were used to determine whether each participant's ratings of prime-target similarity for the 20 trials in ML-match were significantly higher than the chance-level

score of 2.5. Participants were excluded from analyses of fMRI data if their ratings for ML-match were not significantly higher than this chance-level score. For the final sample included in fMRI analyses, paired-sample *t*-tests were performed to assess differences in the similarity ratings between ML-match and ML-mismatch, as well as between MM-match and MM-mismatch. For the ratings of speech-like or song-like traits of the target stimuli in ML-match and NL, a paired-sample two-tailed *t*-test was used to assess the effect of the melody prime. For the ratings of participants' agreement with the two statements about auditory attention and subvocal imitation, one-sample two-tailed *t*-tests were used to determine whether these ratings were significantly greater than the neutral midpoint of the scale (neither agreed nor disagreed).

Preprocessing and analyses of the fMRI data were performed using SPM12 (Wellcome Trust Centre for Neuroimaging, LDN, United Kingdom). The first four volumes of each run were discarded to allow for magnetic saturation effects. The remaining functional images were corrected for head movement artifacts and timing differences in slice acquisitions. Preprocessed functional images were coregistered to the individual's anatomical image, normalized to the standard Montreal Neurological Institute (MNI) brain template, and resampled to a 2-mm isotropic voxel size. Normalized images were spatially smoothed using a Gaussian kernel of 6-mm full width at half maximum to accommodate any anatomical variability across participants.

We performed an event-related analysis to recover the response evoked by each target stimulus. Statistical inference was based on a random effect approach at two levels. The data of each participant were analyzed using the general linear model via fitting the time series data with the canonical hemodynamic response function (HRF) modeled at the event (target). Linear contrasts were computed to characterize responses of interest, averaging across fMRI runs. The group-level analysis consisted of two paired *t*-tests for (1) the contrast of ML-match minus NL, and (2) the contrast of ML-match minus MM-match. We then identified regions that were significantly active for both the ML-match minus NL contrast and the ML-match minus MM-match contrast. This was done because both ML-match and MM-match involved pitch processing and tonal working memory processing, as the melody of the prime needed to be stored and compared to that of the target. To reveal activation clusters related to the perceptual transformation from speech to song across a musical prime, we applied an inclusive mask of the ML-match minus MM-match contrast on the ML-match minus NL contrast. In the fMRI analyses, statistical significance was thresholded at FDR-corrected $p < 0.05$ with a minimum cluster size of 10 voxels.

3. Results

Analysis of the subjective ratings of prime-target similarity showed that one participant's ratings for ML-match were not significantly higher than chance level ($p = 0.44$). This participant was excluded from the analyses of fMRI data because she was unable to discriminate between matched and mismatched trials at better than chance levels. The data of another participant were discarded because of incomplete behavioral data acquisition during the fMRI session. The final sample for fMRI analyses was therefore 18 participants, whose similarity ratings for ML-match were significantly higher than the chance-level score ($p < 0.001$). Figure 4 displays their rating data for five conditions. The similarity ratings for ML-match were significantly higher than ML-mismatch ($p < 0.0001$), and those for MM-match were significantly higher than MM-mismatch ($p < 0.0001$).

Analysis of the ratings of speech-like or song-like traits of the target stimuli in ML-match and NL showed that the target stimuli in ML-match were perceived as significantly more song-like than those in NL ($p < 0.001$). Analysis of the ratings of participants' agreement with the two statements showed that the participants paid significantly more attention to the musical melodies of the utterances that were preceded by matched melodies, compared to those preceded by noise ($p < 0.005$). The participants significantly more tended to covertly imitate the utterances that were preceded by matched melodies, compared to those preceded by noise ($p < 0.01$), as shown in Figure 5.

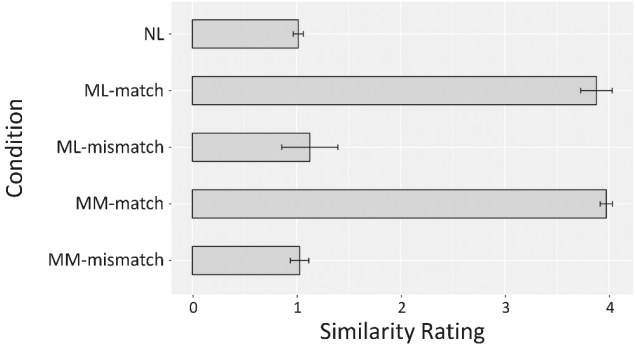


Figure 4. Participants' ratings of prime-target similarity for five conditions. Error bars indicate standard deviation.

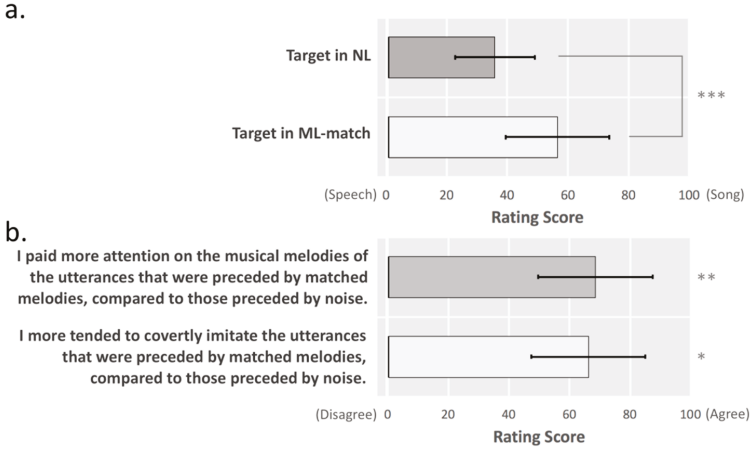


Figure 5. Results of the online questionnaire after the fMRI experiment. (**a**) Rating scores of speech-like or song-like traits of the target stimuli in NL and ML-match showed that the target stimuli in ML-match were perceived as more song-like than those in NL. (**b**) Rating scores of participants' agreement with the two statements showed that the participants paid more attention to the musical melodies of the utterances that were preceded by matched melodies and more tended to covertly imitate these utterances, compared to the target utterances that were preceded by noise. Error bars indicate standard deviation. Note: * $p < 0.01$, ** $p < 0.005$, *** $p < 0.001$.

Results of the whole-brain analyses of fMRI data were summarized by Table 1, Table 2, and Figure 6. Compared to the NL condition, the ML-match condition was associated with significantly increased activity in a number of regions, including the motor/premotor cortex, superior parietal areas, Rolandic operculum, temporal pole, anterior insula, IFG, superior/middle temporal gyrus (STG/MTG), SMA, caudate, thalamus, and cerebellum. Compared to the MM-match condition, the ML-match condition was associated with significantly increased activity in STG/MTG, temporal pole, IFG, anterior insula, superior parietal areas, hippocampus, SMA, putamen, caudate, and cerebellum. The intersection of ML-match minus NL and ML-match minus MM-match yielded activity in IFG, STG/MTG, dorsal premotor cortex, temporal pole, anterior insula, SMA, caudate, and thalamus.

Table 1. Activation clusters for the contrasts of ML-match minus NL and ML-match minus MM-match. (MNI: Montreal Neurological Institute; ML-match: melody-language-match; ML-mismatch: melody-language-mismatch; MM-match: melody-melody-match; MM-mismatch: melody-melody-mismatch; NL: noise-language).

Volume Information	MNI Coordinate			t-Value	Cluster (voxel)
	X	Y	Z		
ML-match minus NL					
Precentral gyrus	38	−24	52	11.81	1979
Postcentral gyrus	44	−32	62	8.53	
Dorsal premotor cortex	38	−2	64	7.26	
Inferior parietal lobule	36	−42	52	5.86	
Dorsal premotor cortex	54	−2	50	5.14	132
Thalamus	14	−18	10	8.08	
Superior temporal pole	62	6	−6	7.24	1310
Anterior insula	40	24	0	7.01	
Superior temporal gyrus	68	−18	6	6.65	
Inferior frontal gyrus (pars orbitalis)	46	20	−12	6.19	
Rolandic Operculum	46	−20	20	6.14	
Inferior frontal gyrus (pars triangularis)	48	32	−2	6.04	
Caudate	−10	8	2	6.87	260
Thalamus	4	−8	10	4.98	
Anterior insula	−34	20	0	6.31	365
Inferior frontal gyrus (pars orbitalis)	−42	18	−8	5.87	
Inferior frontal gyrus (pars triangularis)	−32	30	6	5.62	
Supplementary motor area	4	10	64	6.12	158
Cerebellar lobule IV–V	−12	−52	−20	5.76	375
Cerebellar lobule VI	−22	−52	−22	5.3	
Middle temporal gyrus	−68	−22	2	5.43	77
Superior temporal gyrus	56	−36	14	5.34	80
Cerebellar lobule Crus II	−4	−84	−30	4.79	39
Supplementary motor area	4	−16	56	4.68	91
Supplementary motor area	6	22	46	4.62	116
Rolandic operculum	48	6	10	4.5	43
ML-match minus MM-match					
Superior temporal gyrus	62	−4	−10	11.83	6545
Middle temporal gyrus	64	−24	−4	9.86	
Inferior frontal gyrus (pars triangularis)	48	30	−2	8.31	
Inferior frontal gyrus (pars orbitalis)	40	26	−8	8.20	
Anterior insula	46	20	−12	8.05	
Superior/middle temporal pole	50	14	−24	7.69	
Middle temporal gyrus	−58	−6	−10	11.58	3274
Superior temporal gyrus	−54	−12	−4	11.19	
Anterior insula	−32	20	0	6.73	
Inferior frontal gyrus (pars orbitalis)	−48	36	−10	6.23	
Inferior frontal gyrus (pars triangularis)	−46	30	10	5.49	
Hippocampus	−20	−18	−16	5.98	68
Supplementary motor area	4	22	48	5.52	359
Supragenual anterior cingulate cortex	−4	30	42	3.99	
Putamen	16	8	2	5.26	123
Thalamus	6	−6	10	3.95	
Caudate	−12	10	12	4.98	95
Inferior parietal lobule	34	−48	44	4.87	194
Angular gyrus	32	−58	48	3.79	
Parahippocampal gyrus	14	−2	−20	4.68	49
Cerebellar lobule Crus II	0	−84	−22	4.00	35
Superior parietal lobule	−32	−60	54	3.92	41
Inferior parietal lobule	−30	−58	42	3.59	

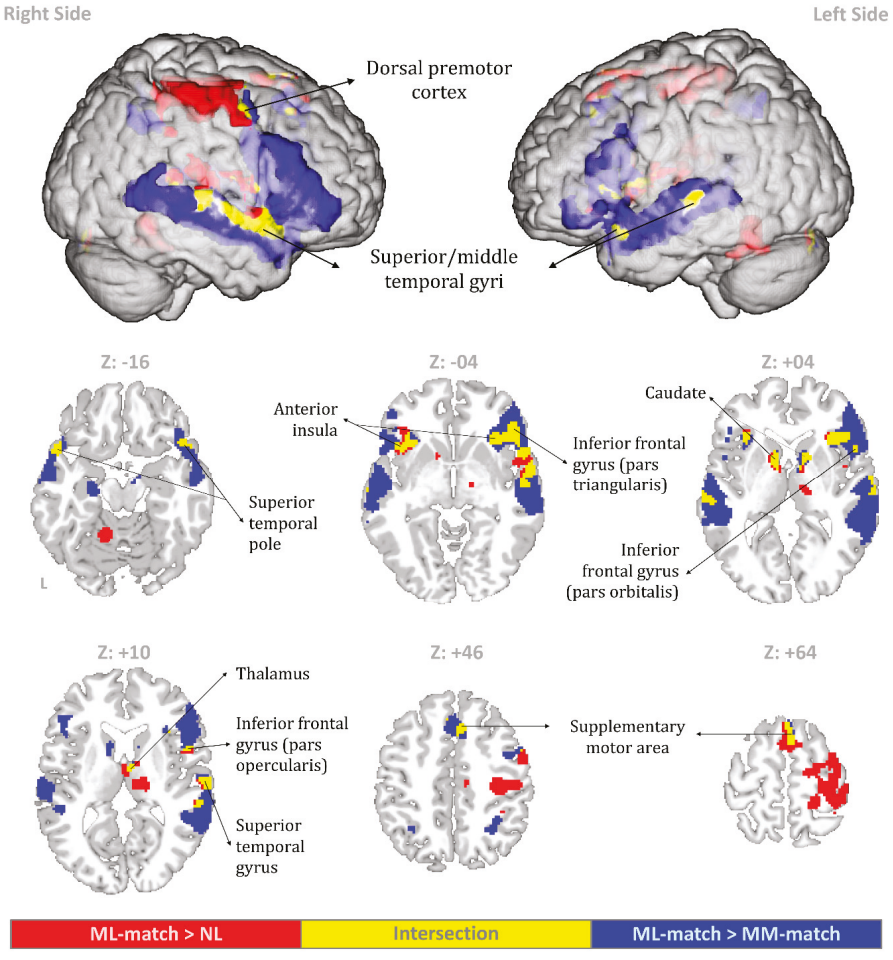


Figure 6. Group-level activation maps for ML-match minus NL (red), ML-match minus MM-match (blue), and their intersection (yellow).

Table 2. Activation clusters for the intersection of ML-match minus NL and ML-match minus MM-match.

Volume Information	MNI Coordinate			t-Value	Cluster (voxel)
	X	Y	Z		
Middle temporal gyrus	−62	−22	4	10.98	77
Superior temporal gyrus	−58	−26	2	7.95	
Superior/middle temporal gyrus	66	−8	−2	10.44	
Inferior frontal gyrus (pars orbitalis)	40	26	−8	8.20	
Anterior insula	34	24	−4	7.88	1017
Inferior frontal gyrus (pars triangularis extending into pars opercularis)	52	16	6	7.45	
Superior temporal pole	60	10	−10	6.58	
Supplementary motor area	4	22	48	5.52	61
Supragenual anterior cingulate cortex	4	8	62	3.23	38
	−4	30	42	3.99	13

Table 2. *Cont.*

Volume Information	MNI Coordinate			t-Value	Cluster (voxel)
	X	Y	Z		
Superior temporal pole, superior temporal gyrus	−52	14	−20	5.31	
Anterior insula	−32	20	0	4.64	238
Inferior frontal gyrus (pars orbitalis)	−44	20	−6	3.34	
Caudate	16	8	2	5.26	63
Thalamus	4	−8	10	3.95	
Caudate	−8	0	4	4.98	48
Superior temporal gyrus	56	−36	10	4.63	28
Inferior frontal gyrus (pars opercularis)	50	8	8	4.46	18
Dorsal premotor cortex	48	2	52	4.39	29
Cerebellar lobule Crus II	−6	−84	−32	4.00	23

4. Discussion

Spoken utterances in tonal languages are intrinsically rich in melodic content, and therefore the differentiation between speech and song in tonal languages is sometimes difficult to make. When a Mandarin speech-like utterance is preceded by a musical melody that mimics the speech melody, the listener may perceive this utterance as if it is being sung. In the present study, we used fMRI to explore this melody priming effect on pitch processing of Mandarin speech. Pitch contours of spoken utterances were modified so that the utterances can be perceived as either speech or song. Participants were asked to rate the melodic similarity between the prime and target. Analyses of fMRI data revealed increased activity in a number of regions for the intersection of speech preceded by matched music minus speech preceded by noise (ML-match > NL) and speech preceded by matched music minus music preceded by identical music (ML-match > MM-match), including the bilateral IFG, anterior insula, SMA, and STG/MTG. This finding echoes previous hypotheses and results of the speech-to-song illusion that exposure to repetition of a speech fragment is associated with greater activity in the neural substrates of pitch processing and re-evaluation of melodic features of this speech fragment [6,7,9].

The task of judging the melodic similarity between the prime and target in ML-match demanded the processing of melodic features of the target. Based on prior research on the neural correlates of prosody processing, we speculate that the right IFGtri, which showed activity for the intersection of ML-match minus NL and ML-match minus MM-match, may support the melodic processing of the speech-like target in ML-match. It has been reported that listening to "prosodic" speech (speech with no linguistic meaning, but retaining the slow prosodic modulations of speech) was associated with enhanced activity in the right IFGtri (extending into the pars opercularis of IFG) compared to normal speech [10]. The right IFGtri also responded to pitch patterns in song [16]. Moreover, a study of sarcasm comprehension in the auditory modality demonstrated that negative prosody incongruent with positive semantic content activated the right anterior insula extending into the IFGtri [12]. During perception of neutral, sad, and happy prosody, individuals with autism spectrum disorder displayed reduced activity in the right IFGtri compared to normal controls [10]. Taken together, we suggest that the right IFGtri may allocate attentional resources for the melodic processing of the target stimulus in ML-match. This view is supported by participants stating that they paid more attention to the melodic features of the target stimuli in ML-match, compared to those in NL.

One may speculate that the right IFGtri activity for ML-match minus MM-match reflects its role in working memory. However, both ML-match and MM-match involved a comparison of two melodies, a task demanding tonal working memory. One interpretation of increased activity in the right IFGtri for ML-match is that the task of melody comparison in ML-match preferentially relied on action-related sensorimotor coding of tonal information, whereas this coding played a lesser role in MM-match. There has been evidence indicating that the right IFGtri is engaged in the multi-modal processing of tonal or verbal information. For example, McCormick and colleagues investigated the neural

basis of the crossmodal correspondence between auditory pitch and visuospatial elevation, finding a modulatory effect of pitch-elevation congruency on activity in the IFGtri and anterior insula [17]. Golfinopoulos and colleagues demonstrated that the right IFGtri exhibited increased activity when speech production was perturbed by unpredictably blocking subjects' jaw movements [18]. Moreover, a study of sensory feedback to vocal motor control also reported that trained singers showed increased activation in the right IFGtri, anterior insula, and SMA in response to noise-masking [19]. This finding is especially relevant to our study, as we found co-activation of the right IFGtri, anterior insula, and SMA for the intersection of ML-match minus NL and ML-match minus MM-match. We suggest that the participants may use subvocal rehearsal to facilitate the task of melody comparison in ML-match. Indeed, our participants reported that they more tended to covertly imitate the target stimuli in ML-match compared to those in NL. During covert vocal imitation of the target stimulus, the anterior insula may be responsible for the laryngeal somatosensory functions and voice pitch control [20–22], the SMA may support motor planning and monitoring/evaluation of this plan [23–27], and the right IFGtri may allocate cognitive resources to integrate the auditory coding and action-related sensorimotor coding of the melodic pattern of the target.

Besides speech perception, the finding of the involvement of the bilateral STG/MTG in the melody priming effect on the pitch processing of Mandarin speech also provides an enriched perspective on song perception. Results of the online questionnaire showed that the target stimuli in ML-match were perceived as more song-like than those in NL. We found that the left mid-posterior STG/MTG was activated for the intersection of ML-match minus NL and ML-match minus MM-match. This cluster was effectively identical to that described by two previous neuroimaging studies on song perception. Sammler and colleagues reported mid-posterior STS activation for the interaction effect of lyrics and tunes during passive listening to unfamiliar songs, suggesting its role in the integrative processing of lyrics and tunes at prelexical, phonemic levels [14]. Alonso and colleagues reported that binding lyrics and tunes for the encoding of new songs was associated with the involvement of the bilateral mid-posterior MTG [15]. We suggest that the right STG/MTG may integrate the musical (melodic) and phonological information of the targets in ML-match. This view parallels an earlier report finding that STG/MTG was activated for the intersection of listening to sung words minus listening to "vocalize" (i.e., singing without words) and listening to sung words minus listening to speech [28].

A study of the speech-to-song illusion [9] showed a positive correlation between the subjective vividness of this illusion and activity in the pars orbitalis of the bilateral IFG, which also exhibited activation for the intersection of ML-match minus NL and ML-match minus MM-match in the present study. These regions have been implicated in a broad range of cognitive processes, such as response inhibition [29,30], response selection [31], working memory [32,33], semantic processing [34,35], and prosody processing [36]. The pars orbitalis of the bilateral IFG appeared to contribute to certain high-level cognitive processes necessary for the melody-similarity-judgment task. Its exact role remains to be specified by future research.

A few limitations of the present study should be noted. First, in the speech-to-song illusion [6] a spoken phrase was repeated without modification, whereas we modified the pitch contours of spoken Mandarin utterances so that they differed from normal speech. Caution should be exercised when comparison is made between the results of this study and those of the speech-to-song illusion. Future research could explore how the manipulations of pitch flattening and the clarity of tonality of spoken utterances impact the melody priming effect of on the pitch processing of Mandarin speech. It is also interesting to examine whether native speakers, non-native speakers (second language speakers), and non-speakers differ in the pitch processing of "songified" speech. A previous study compared speech-to-song illusions in tonal and non-tonal language speakers, finding that both non-tonal native language and inability to understand the speech stream as a verbal message predicted the speech-to-song illusion [37]. Second, the final sample included in fMRI analyses mainly consisted of amateur musicians. We cannot ascertain whether this melody priming effect can also be observed in

non-musicians. Specific musical or cognitive abilities may correlate with the tendency of perceptual transformation from Mandarin speech to song. However, this idea remains to be tested in future studies.

5. Conclusions

The present study has examined the neural underpinnings of the perceptual transformation from modified Mandarin speech to song across a musical prime that mimics the melody of speech. Based on our fMRI data and previous literature, we suggest that the right IFGtri may play a role in allocation of attentional resources to the multi-modal processing of the melodic pattern of this stimulus. Moreover, the STG/MTG may integrate its phonological and musical (melodic) information. While these findings corroborate and extend previous studies on the speech-to-song illusion, we believe that further exploration of the melodic characteristics of tonal and non-tonal languages would significantly advance our understanding of the relationship between speech and song.

Supplementary Materials: Audio examples of the stimuli for NL and ML-match are available online at http://www.mdpi.com/2076-3425/9/10/286/s1.

Author Contributions: Conceptualization, C.-G.T.; methodology, C.-G.T. and C.-W.L.; software, C.-W.L.; validation, C.-G.T. and C.-W.L.; formal analysis, C.-G.T. and C.-W.L.; investigation, C.-G.T. and C.-W.L.; resources, C.-G.T. and C.-W.L.; data curation, C.-G.T. and C.-W.L.; writing—original draft preparation, C.-G.T. and C.-W.L.; writing—review and editing, C.-G.T. and C.-W.L.; visualization, C.-W.L.; supervision, C.-G.T.; project administration, C.-G.T.; funding acquisition, C.-G.T.

Funding: This research was funded by grant projects (MOST 106-2420-H-002-009 and MOST 108-2410-H-002-216) from Ministry of Science and Technology, Taiwan.

Acknowledgments: The authors would like to express our gratitude to Prof. Tai-Li Chou for helpful discussion. We also thank Chao-Ju Chen for data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, C.; Peng, G.; Wang, W.S.-Y. Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *J. Acoust. Soc. Am.* **2012**, *132*, 1088–1099. [CrossRef] [PubMed]
2. Zhang, C.; Chen, S. Toward an integrative model of talker normalization. *J. Exp. Psychol. Hum. Percept. Perform.* **2016**, *42*, 1252–1268. [CrossRef] [PubMed]
3. Zhang, C.; Peng, G.; Shao, J.; Wang, W.S.-Y. Neural bases of congenital amusia in tonal language speakers. *Neuropsychologia* **2017**, *97*, 18–28. [CrossRef] [PubMed]
4. Tsai, C.-G.; Chou, T.-L.; Li, C.-W. Roles of posterior parietal and dorsal premotor cortices in relative pitch processing: Comparing musical intervals to lexical tones. *Neuropsychologia* **2018**, *119*, 118–127. [CrossRef]
5. Yung, B. *Cantonese Opera*; Cambridge University Press: Cambridge, UK, 1989.
6. Deutsch, D.; Henthorn, T.; Lapidis, R. Illusory transformation from speech to song. *J. Acoust. Soc. Am.* **2011**, *129*, 2245–2252. [CrossRef]
7. Falk, S.; Rathcke, T.; Dalla Bella, S. When speech sounds like music. *J. Exp. Psychol. Hum. Percept. Perform.* **2014**, *40*, 1491–1506. [CrossRef]
8. Tierney, A.; Dick, F.; Deutsch, D.; Sereno, M. Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* **2013**, *23*, 249–254. [CrossRef]
9. Hymers, M.; Prendergast, G.; Liu, C.; Schulze, A.; Young, M.L.; Wastling, S.J.; Barker, G.J.; Millman, R.E. Neural mechanisms underlying song and speech perception can be differentiated using an illusory percept. *NeuroImage* **2015**, *108*, 225–233. [CrossRef]
10. Kotz, S.A.; Meyer, M.; Alter, K.; Besson, M.; Von Cramon, D.Y.; Friederici, A.D. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang.* **2003**, *86*, 366–376. [CrossRef]
11. Gebauer, L.; Skewes, J.; Hørlyck, L.; Vuust, P. Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage Clin.* **2014**, *6*, 370–378. [CrossRef]
12. Matsui, T.; Nakamura, T.; Utsumi, A.; Sasaki, A.T.; Koike, T.; Yoshida, Y.; Harada, T.; Tanabe, H.C.; Sadato, N. The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia* **2016**, *87*, 74–84. [CrossRef] [PubMed]

13. Lima, C.F.; Krishnan, S.; Scott, S.K. Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends Neurosci.* **2016**, *39*, 527–542. [CrossRef] [PubMed]
14. Sammler, D.; Baird, A.; Valabrègue, R.; Clément, S.; Dupont, S.; Belin, P.; Samson, S. The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* **2010**, *30*, 3572–3578. [CrossRef] [PubMed]
15. Alonso, I.; Davachi, L.; Valabrègue, R.; Lambrecq, V.; Dupont, S.; Samson, S. Neural correlates of binding lyrics and melodies for the encoding of new songs. *NeuroImage* **2016**, *127*, 333–345. [CrossRef] [PubMed]
16. Merrill, J.; Sammler, D.; Bangert, M.; Goldhahn, D.; Lohmann, G.; Turner, R.; Friederici, A.D. Perception of words and pitch patterns in song and speech. *Front. Psychol.* **2012**, *3*, 76. [CrossRef] [PubMed]
17. McCormick, K.; Lacey, S.; Stilla, R.; Nygaard, L.C.; Sathian, K. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia* **2018**, *112*, 19–30. [CrossRef]
18. Golfinopoulos, E.; Tourville, J.A.; Bohland, J.W.; Ghosh, S.S.; Nieto-Castanon, A.; Guenther, F.H. fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage* **2011**, *55*, 1324–1338. [CrossRef]
19. Kleber, B.; Friberg, A.; Zeitouni, A.; Zatorre, R. Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage* **2017**, *147*, 97–110. [CrossRef]
20. Schulze, K.; Zysset, S.; Mueller, K.; Friederici, A.D.; Koelsch, S. Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* **2011**, *32*, 771–783. [CrossRef]
21. Zarate, J.M. The neural control of singing. *Front. Hum. Neurosci.* **2013**, *7*, 237. [CrossRef]
22. Behroozmand, R.; Ibrahim, N.; Korzyukov, O.; Robin, D.A.; Larson, C.R. Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain Cogn.* **2014**, *84*, 97–108. [CrossRef] [PubMed]
23. Alario, F.-X.; Chainay, H.; Lehéricy, S.; Cohen, L. The role of the supplementary motor area (SMA) in word production. *Brain Res.* **2006**, *1076*, 129–143. [CrossRef] [PubMed]
24. Ellis, R.J.; Norton, A.C.; Overy, K.; Winner, E.; Alsop, D.C.; Schlaug, G. Differentiating maturational and training influences on fMRI activation during music processing. *NeuroImage* **2012**, *60*, 1902–1912. [CrossRef] [PubMed]
25. Iannaccone, R.; Hauser, T.U.; Staempfli, P.; Walitza, S.; Brandeis, D.; Brem, S. Conflict monitoring and error processing: new insights from simultaneous EEG–fMRI. *NeuroImage* **2015**, *105*, 395–407. [CrossRef] [PubMed]
26. Sachs, M.; Kaplan, J.; Der Sarkissian, A.; Habibi, A. Increased engagement of the cognitive control network associated with music training in children during an fMRI Stroop task. *PLoS ONE* **2017**, *12*, e0187254. [CrossRef] [PubMed]
27. Rong, F.; Isenberg, A.L.; Sun, E.; Hickok, G. The neuroanatomy of speech sequencing at the syllable level. *PLoS ONE* **2018**, *13*, e0196381. [CrossRef] [PubMed]
28. Schön, D.; Gordon, R.; Campagne, A.; Magne, C.; Astésano, C.; Anton, J.-L.; Besson, M. Similar cerebral networks in language, music and song perception. *NeuroImage* **2010**, *51*, 450–461. [CrossRef]
29. Leung, H.-C.; Cai, W. Common and differential ventrolateral prefrontal activity during inhibition of hand and eye movements. *J. Neurosci.* **2007**, *27*, 9893–9900. [CrossRef]
30. Berkman, E.T.; Burklund, L.; Lieberman, M.D. Inhibitory spillover: intentional motor inhibition produces incidental limbic inhibition via right inferior frontal cortex. *NeuroImage* **2009**, *47*, 705–712. [CrossRef]
31. Goghari, V.M.; Macdonald, A.W. The neural basis of cognitive control: response selection and inhibition. *Brain Cogn.* **2009**, *71*, 72–83. [CrossRef]
32. Chen, S.A.; Desmond, J.E. Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *NeuroImage* **2005**, *24*, 332–338. [CrossRef] [PubMed]
33. Marklund, P.; Persson, J. Context-dependent switching between proactive and reactive working memory control mechanisms in the right inferior frontal gyrus. *NeuroImage* **2012**, *63*, 1552–1560. [CrossRef] [PubMed]
34. Sabb, F.W.; Bilder, R.M.; Chou, M.; Bookheimer, S.Y. Working memory effects on semantic processing: priming differences in pars orbitalis. *NeuroImage* **2007**, *37*, 311–322. [CrossRef] [PubMed]
35. Zhuang, J.; Randall, B.; Stamatakis, E.A.; Marslen-Wilson, W.D.; Tyler, L.K. The interaction of lexical semantics and cohort competition in spoken word recognition: an fMRI study. *J. Cogn. Neurosci.* **2011**, *23*, 3778–3790. [CrossRef]

36. Perrone-Bertolotti, M.; Dohen, M.; Lœvenbruck, H.; Sato, M.; Pichat, C.; Baciu, M. Neural correlates of the perception of contrastive prosodic focus in French: a functional magnetic resonance imaging study. *Hum. Brain Mapp.* **2013**, *34*, 2574–2591. [CrossRef]
37. Jaisin, K.; Suphanchaimat, R.; Figueroa Candia, M.A.; Warren, J.D. The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. *Front. Psychol.* **2016**, *7*, 662. [CrossRef]

brain sciences

MDPI

Article

Electrical Brain Responses Reveal Sequential Constraints on Planning during Music Performance

Brian Mathias [1,2,*], William J. Gehring [3] and Caroline Palmer [1,*]

1 Department of Psychology, McGill University, Montreal, QC H3A 1B1, Canada
2 Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany
3 Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA; wgehring@umich.edu
* Correspondence: bmathias@cbs.mpg.de (B.M.); caroline.palmer@mcgill.ca (C.P.); Tel.: +49-341-9940-2485 (B.M.); +1-514-398-6128 (C.P.)

Received: 11 January 2019; Accepted: 26 January 2019; Published: 28 January 2019

Abstract: Elements in speech and music unfold sequentially over time. To produce sentences and melodies quickly and accurately, individuals must plan upcoming sequence events, as well as monitor outcomes via auditory feedback. We investigated the neural correlates of sequential planning and monitoring processes by manipulating auditory feedback during music performance. Pianists performed isochronous melodies from memory at an initially cued rate while their electroencephalogram was recorded. Pitch feedback was occasionally altered to match either an immediately upcoming Near-Future pitch (next sequence event) or a more distant Far-Future pitch (two events ahead of the current event). Near-Future, but not Far-Future altered feedback perturbed the timing of pianists' performances, suggesting greater interference of Near-Future sequential events with current planning processes. Near-Future feedback triggered a greater reduction in auditory sensory suppression (enhanced response) than Far-Future feedback, reflected in the P2 component elicited by the pitch event following the unexpected pitch change. Greater timing perturbations were associated with enhanced cortical sensory processing of the pitch event following the Near-Future altered feedback. Both types of feedback alterations elicited feedback-related negativity (FRN) and P3a potentials and amplified spectral power in the theta frequency range. These findings suggest similar constraints on producers' sequential planning to those reported in speech production.

Keywords: sensorimotor learning; sequence production; sequence planning; feedback monitoring; EEG; N1; FRN; music performance; music cognition; altered auditory feedback

1. Introduction

Many everyday behaviors, such as having a conversation, writing a note, and driving a car, involve the production of action sequences. A core tenet of theories of sequential behavior is that, in order to produce action sequences quickly and accurately, individuals must plan appropriate movements prior to their execution [1]. Evidence for future-oriented planning during sequential tasks comes from anticipatory ordering errors, in which upcoming sequence events are produced earlier in the sequence than intended. Documented in both speech production [2,3] and music performance [4–6], anticipatory errors suggest that producers have access to a range of upcoming events in a sequence at any given time during production. The span of sequence positions between an event's correct (intended) position and its incorrect (error-produced) position is taken to indicate a producer's range of planning, due to the items' simultaneous accessibility [5]. Serial ordering errors during music performance tend to arise more often from closer sequence distances than from farther distances [7–9]. This tendency suggests that producers have increased access to events intended for nearer in the future compared to events that are intended for farther ahead in the future [8]. These proximity constraints on planning have

been attributed to interference of future events with current events during memory retrieval, decay of item information over time, and individual differences in producers' working memory spans [8,10,11]. Sequential models of sequence planning in both speech [12] and music [8] use the term "planning gradient" to refer to a decrease in memory activation of upcoming sequence events as the distance from the current event increases.

In addition to planning upcoming units of speech and music during production, speakers and musicians monitor the perceptual outcomes of their previous productions. In order to monitor perceptual outcomes during auditory-motor tasks, producers compare perceived auditory feedback with an intended auditory outcome [10]. Theoretical approaches to feedback monitoring have focused heavily on the concept of internal models, or representations that simulate a response to estimate an outcome (for a review, see [13]). Internal models are thought to arise from interactions between bottom-up, incoming sensory information and top-down expectations or predictions formed by the motor system during production [14]. A framework known as "predictive coding" assumes that the goal of production is to minimize prediction error (i.e., mismatches between predictions that are generated by an internal model and sensory information originating in the environment) [15,16]. Musicians possess strong associations between musical actions and their sensory outcomes [17], which may explain why the perception of inaccurate auditory feedback during the production of auditory-motor sequences can disrupt production [18]. Mismatches between auditory feedback from musician's planned movements [7] as well as nonmusicians' planned movements can generate prediction errors, evidenced by an increasing error-related negativity [19]. Experimentally altering the contents of pitch feedback during music performance can disrupt the regular timing of key presses [20] and increase pitch error rates [21,22]. The computer-controlled removal of auditory feedback in laboratory environments does not disrupt well-learned performance; musicians can continue performing well-learned music when auditory feedback is removed [23,24], and altering feedback so that it is highly different from expected feedback has little effect on a previously learned performance [25]. Performance is disrupted when the altered auditory feedback is similar to the planned events [25,26]. Thus, current evidence suggests that disruption caused by altered auditory feedback may depend on similarity-based interference with planned sequence representations, leading to a novel prediction: If the planning of future events occurs in a graded fashion (higher activation for immediately upcoming events compared to more distant future events), then altered feedback that matches immediately upcoming events should disrupt performance more than altered feedback matching sequentially distant events.

To our knowledge, no studies in the domain of speech production have tested neural effects of hearing upcoming linguistic content that is presented sooner than expected while speaking. This may be due to the difficulties of independently manipulating auditory feedback during concurrent speech production. Auditory feedback during speech production can be electronically delayed, so that instead of hearing current feedback, one hears feedback that matches previous utterances. Presenting linguistic content that matches upcoming utterances is more difficult, however, because it requires the presentation of speech content that has not yet been produced by the speaker. In electronic music performance, one can present auditory feedback that matches future keypresses, due to a simpler sound production apparatus. One study examined musicians' neural responses to future-oriented altered auditory feedback as they performed tone sequences on a piano [7]. Occasional alterations in auditory feedback were presented that matched upcoming (future) events as pianists performed melodies from memory. An example of future-oriented feedback is if a pianist was currently producing tone A and was planning to produce tone B later in the sequence, tone B would be presented auditorily when the pianist's hand struck the A-key on the keyboard. Future-oriented feedback pitches elicited larger (event-related potential (ERP)) responses than altered feedback that matched previous (past) events, and amplitudes of the ERPs elicited by the altered feedback pitches correlated with the amount of temporal disruption elicited in the pianists' key presses [7]. It is unknown, however, whether disruptive effects of altered auditory feedback that match future events depend on an individual's

planning gradient: If producers' plans are biased toward the activation of immediately upcoming events compared to events planned for the distant future, then we would expect pitch feedback that matches immediate future events to generate greater similarity-based interference, and in turn greater performance disruption, than future-oriented feedback that matches distant future events. Thus, future-oriented theories of planning predict greater performance disruption for altered feedback that matches near future events compared to far future events.

Several studies have suggested that sensory processing of altered auditory feedback during production is marked by early- to middle-latency ERP responses to tone onsets [27–29]. N1 and P2 ERP components in particular are sensitive to whether speech or tones are generated by oneself versus others [30–32]. The N1 is a negative-going ERP component that peaks at about 100 ms following sound onsets and is followed by the positive-going P2 component [33,34]. Amplitudes of these components are more negative when sounds are generated by others than when they are self-generated, which is thought to reflect motor-induced suppression of auditory cortical processing [35,36]. Perceptual studies have demonstrated that N1 and P2 amplitudes also become more negative (larger N1 and smaller P2) in response to tones that are selectively attended to, compared to unattended tones, suggesting a role of these components in early auditory sensory processing [37–40]. N1 and P2 waves occur in quick succession about 50–150 ms following sound onsets, and arise from several temporally overlapping, spatially-distributed sources, with primary generators in the auditory cortices and planum temporale [33,34,41–43]. Thus, N1 and P2 amplitudes may serve as a proxy for the degree to which sensory processing of auditory feedback is suppressed during sound production: A negative-going shift in amplitudes occurs when processing of auditory feedback is enhanced, and a positive-going shift occurs when processing of auditory feedback is suppressed. Combined with the notion that future-oriented feedback that matches near future events may generate greater similarity-based interference than feedback matching far future events, this principle leads to the prediction that altered auditory feedback that matches near future events should decrease the expected N1 and P2 suppression compared to altered feedback that matches far future events.

Additional ERP components linked to action-related expectations are elicited when sensory feedback indicates that an action has resulted in an unexpected outcome. Frontally maximal feedback-related negativities (FRNs) are elicited roughly 150–250 ms following the unexpected outcome in music performance tasks [44–47], as well as during other tasks, such as reward prediction and monetary gambling tasks [48]. FRN amplitudes may be associated with the degree to which unexpected feedback violates a producer's feedback-related expectations [49–53]. The FRN component often co-occurs with neural oscillations in the theta frequency range (4–8 Hz), thought to reflect the implementation of cognitive control [54–56]. The FRN is typically followed by a frontally-maximal P3a component, which peaks around 300–500 ms following the onset of unexpected feedback. The P3a may reflect the updating of stimulus memory representations [57,58], decision-making processes [59,60], and voluntary shifts of attention to unexpected stimuli [61,62]. If altered auditory feedback during music performance triggers the emergence of a more cognitively-controlled (e.g., deliberative, goal-directed, model-based, prefrontal) state, as opposed to a habitual (e.g., automatic, model-free, striatal) performance state [63], then we would expect theta frequency activity to be enhanced following any feedback alterations, and to be accompanied by FRN and P3a potentials. A benefit of extracting theta band activity related to the FRN is that it can account for potential overlap of neighboring FRN and P3a potentials in the ERP waveform [56,64].

The current study investigated the relationship between performers' planning and feedback monitoring processes by presenting altered auditory feedback corresponding to upcoming (future) sequence events during music performance. The timing of pianists' key presses in response to altered auditory feedback pitches was measured. Pianists memorized and performed isochronous melodic sequences on an electronic keyboard while hearing feedback triggered by their key presses over headphones. Altered pitch feedback was manipulated in four conditions: Future +1 ("near future"), future +2 ("far future"), noncontextual, and baseline. In the future +1 condition, participants heard an

altered pitch presented at the current location that matched the intended (memorized) pitch at the next location in the sequence. In the future +2 condition, participants heard an altered pitch presented at the current location that matched the intended (memorized) pitch at the location two events ahead of the current location. In the noncontextual condition, participants heard a pitch that was not present in the sequence; this control condition tested effects of hearing an altered feedback pitch that was unrelated to performers' planning processes. Finally, in the baseline condition, participants heard the expected auditory feedback with no pitch alterations.

We tested three predictions: First, near future (future +1) altered auditory feedback was expected to induce greater interference with the production of currently planned events than far future (future +2) altered auditory feedback. This prediction is based on producers' use of planning gradients, in which plans are weighted toward near compared to distant sequence events [8,12]. Pianists were therefore expected to show greater temporal disruption following future +1 altered auditory feedback compared to future +2 altered feedback. Second, we expected performance disruption to be associated with decreased N1 and P2 suppression following future +1 feedback compared to future +2 feedback. Third, future +1, future +2, and noncontextual altered feedback pitches were expected to elicit FRN and P3a ERP components (relative to the baseline condition), as well as corresponding theta oscillations within the timeframe of the FRN.

2. Materials and Methods

2.1. Participants

Twenty-eight right-handed adult pianists with at least 6 years of private piano instruction were recruited from the Montreal community. Four participants were excluded from analysis due to insufficient data after trials performed from memory that contained pitch errors (n = 3) or EEG artifacts (n = 1) were removed. The remaining 24 pianists (15 women, age M = 21.1 years, SD = 2.7 years) had between 6 and 20 years of piano lessons (M = 11.5 years, SD = 3.9 years). Participants reported having no hearing problems. Two of the pianists reported possessing absolute pitch. Participants provided written informed consent, and the study was reviewed by the McGill University Research Ethics Board.

2.2. Stimulus Materials

Four novel melodies that were notated in a binary meter (2/4 time signature), conforming to conventions of Western tonal music, were used in the study. An example of a melody is shown in Figure 1. All melodies were isochronous (containing only 8 quarter notes), were notated for the right hand, and were designed to be repeated without stopping 3 times in each trial (totaling 24 quarter-note events). Each stimulus melody was composed to have no repeating pitches within two sequence positions. The 4 melodies were composed in the keys of G major, D minor, C major, and B minor. Suggested fingering instructions were also notated.

During the experiment, auditory feedback pitches triggered by participants' key presses while performing the melodies were occasionally replaced by an altered pitch. The altered pitches were chosen from the same diatonic key as the original melody to maintain the melodic contour of the original melody, and to avoid tritone intervals. Altered feedback pitches occurred in one of 8 possible locations within each trial. As metrical accent strength has been found to influence both correct (error-free) music performance and the likelihood of performance errors [8,65,66] among performing musicians, half of the altered feedback locations occurred at odd-numbered serial positions in the tone sequence (aligning with strong metrical accents in the melody's binary time signature), and the other half occurred at even-numbered serial positions (aligning with weak metrical accents in the melody's time signature).

Examples of potential altered feedback pitches for one stimulus melody are shown in Figure 1. In the future +1 condition, participants heard the pitch that corresponded to the next intended (memorized) pitch in the melodic sequence when they pressed the piano key. In the future +2

condition, participants heard the pitch that corresponded to the intended pitch that was 2 events ahead in the melodic sequence. In the noncontextual condition, participants heard a pitch from the melody's diatonic key (determined by the key signature of each stimulus melody) that was not present in the melodic sequence. Noncontextual pitches were chosen to match the contour and interval size as closely as possible to that of the intended pitch. The noncontextual condition was intended to serve as a control condition, to test effects of hearing an altered feedback pitch that was unrelated to performers' planning processes. Finally, in a baseline condition, no auditory feedback pitches were altered (participants heard the intended auditory feedback).

Each stimulus trial contained three and a half continuous iterations (without pausing) of a repeated melody (described below in Procedure). Each trial began with a 12-beat metronome sounded every 500 ms; the first four beats indicated the intended pace and the remaining eight beats coincided with the pianists' first iteration of the melody, forming the synchronization phase of the trial (see Figure 2). The metronome then stopped and the pianists continued performing for two and a half more iterations of the melody, forming the continuation phase of the trial. Altered feedback pitches could occur during the continuation phase only. A minimum of zero and a maximum of two pitches were altered within a single trial, with a maximum of one altered pitch per melody iteration. When two altered pitches occurred in a single trial, they were always separated by at least three unaltered pitch events. No alterations occurred on the first pitch of any iteration or on the last four pitches of any trial.

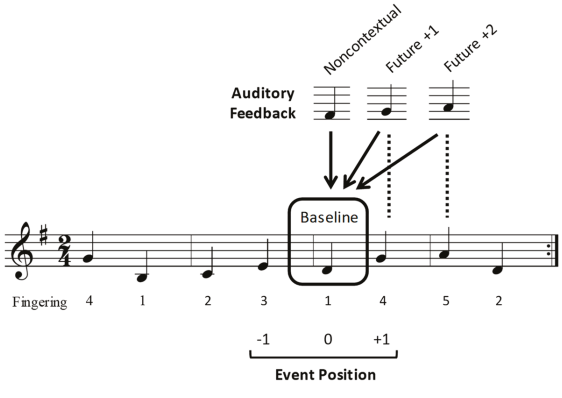


Figure 1. Example of a notated stimulus melody. Sample altered feedback pitches for the four auditory feedback conditions (baseline, noncontextual, future +1, and future +2), and the three target event positions (−1, 0, +1) over which interonset intervals (IOIs) and event-related potentials (ERPs) were analyzed are shown. Target event positions are numbered with respect to the distance of the altered feedback from its intended sequence position. Arrows show the location at which the altered feedback pitches occurred, and dashed lines indicate the origin of the altered feedback pitches.

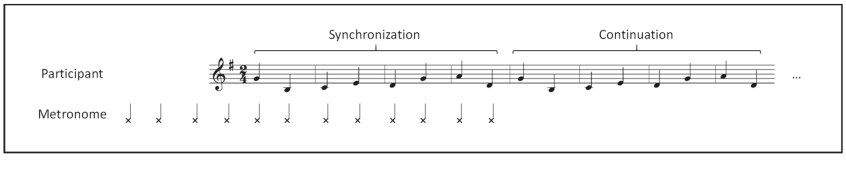


Figure 2. Synchronization-continuation trial. Participants synchronized the first iteration of each melody with a metronome ('Synchronization'), and then performed two and a half additional melody iterations without the metronome ('Continuation'). Four initial metronome beats set the performance tempo. The metronome sounded every 500 ms.

2.3. Equipment

Participants performed the stimulus melodies on a Roland RD-700SX musical instrument digital interface (MIDI) digital piano keyboard (Roland Corporation, Ontario, CA, USA) in a sound- and electrically-attenuated chamber while EEG was recorded. As pianists performed, sound was emitted from a Roland Edirol SD-50 system (Roland Corporation, Ontario, CA, USA) and delivered through EEG-compatible air-delivery earphones (ER1-14B, Etymotic Research). Two channels were used for auditory feedback: "GMT piano 002" for piano key press auditory feedback, and "Rhy 001" for the metronome that signaled the performance rate at the start of each trial. Auditory feedback pitches were controlled using FTAP version 2.1.06 [67]. FTAP presented pre-programmed pitches at the time that pianists pressed each key, and measured key press timing information with 1-ms resolution.

2.4. Design

The study used a repeated measures within-participant design in which altered auditory feedback pitches were manipulated in four conditions: Future +1, future +2, noncontextual, and baseline. Participants completed trials in three blocks, each corresponding to an altered auditory feedback type (future +1, future +2, and noncontextual). Each block contained 32 trials, 50% of which contained no altered auditory feedback (baseline condition), and 50% which contained an altered feedback pitch (future +1, future +2, or noncontextual). Each trial containing altered auditory feedback was unique across the entire experiment and therefore was heard only once by participants. Block and melody orders were counterbalanced across the 24 participants. Participants performed a total of 96 (3 blocks $\times$ 32) trials, equivalent to 192 continuation iterations (32 future +1, 32 future +2, 32 noncontextual, and 96 baseline), over the course of the entire experiment. The dependent variables of the tone interonset interval (IOI), ERP component amplitudes, and theta band power were analyzed at sequential positions, -1, 0, and +1, relative to the altered tone location (as shown in Figure 1).

2.5. Procedure

Participants first completed a musical background questionnaire, followed by a piano performance memory test. Participants were then presented with a short novel right-hand melody (not included in the experiment) to practice and memorize; those who were able to memorize and perform it to a note-perfect criterion within three attempts, after up to three minutes of practice with the music notation, were invited to participate in the experiment. All pianists met this criterion. Following completion of the memory test, participants were outfitted with EEG caps and electrodes.

Participants were then asked to complete three practice trials in order to become familiar with the task. At the start of the practice trials, the participants were again presented the music notation of the single-hand melody that they had previously performed in the memory test. They were asked to indicate when they had memorized the melody. The music notation was then removed and replaced with a fixation cross. Participants were then asked to perform the melody from memory at the rate indicated by four clicks of a metronome cue (500 ms per quarter note beat). They were told that they would sometimes hear a tone that did not match the key that they pressed, but that they should keep performing at the rate cued by the metronome and try not to stop or slow down. Participants were also instructed to view the fixation cross while they were performing. The purpose of the fixation cross was to inhibit large eye movements and control participants' gaze locations during the performance task, following other EEG studies [68,69]. During each of the three practice trials, a single feedback pitch was altered to correspond to the future +1, future +2, and noncontextual experimental conditions. The order of the three practice trials was counterbalanced across participants.

Following the three practice trials, participants were presented with the music notation of one of the four experimental stimulus melodies. They were asked to practice the melody for a maximum of three minutes, using the notated fingering, with the goal of performing it from memory. Following memorization, the notation was removed and replaced with a fixation cross. Participants

then performed the melody from memory in the synchronization-continuation trials. The first three synchronization-continuation trials contained no altered feedback, so that the experimenters could verify that participants had successfully memorized the melody; all participants were able to perform at least one of the three verification trials without producing any pitch errors.

In each synchronization-continuation trial, participants were instructed to perform the melody from memory at the rate indicated by the metronome (500 ms per quarter-note beat), to not stop or slow down if they heard a tone that did not match the key that they pressed, and to continuously repeat the melody until they stopped hearing auditory feedback from their key presses. The metronome stopped when the participant began the second iteration of the melody. Participants were asked to refrain from moving their head or body while performing in order to minimize movement-related EEG artifacts. Eyeblinks typically create artifacts in the EEG signal, which can be addressed using a variety of artifact rejection procedures (for a review, see [70]). In order to minimize eyeblink-related artifacts, participants in some studies may be asked to refrain from blinking during certain parts of EEG trials. Since the duration of each synchronization-continuation trial in the current study exceeded 15 s, participants were not asked to refrain from blinking during the trial. Following each trial, participants indicated when they were ready to proceed to the next trial. This procedure was repeated for each of the 4 stimulus melodies and for each of the 3 feedback blocks. The synchronization-continuation trials lasted approximately 45 min. At the end of the experiment, participants were asked if they noticed any specific aspects of the altered feedback or its manipulation across the experiment; none of the participants reported an awareness of any relationship between the altered feedback and performance.

2.6. Data Recording and Analysis

2.6.1. Behavioral Data

Behavioral disruption associated with the presentation of altered auditory feedback was evaluated by analyzing IOIs from the time of one key press to the next key press (in ms) for pitches that occurred before (position −1), during (position 0), and after (position +1) the altered auditory feedback pitch (position +1; see Figure 1). Errors in pitch accuracy were identified by computer comparison of pianists' performances with the information in the notated musical score (Large, 1993). Pitch errors were defined as pitch additions, deletions, and corrections (errors in which pianists stopped after an error and corrected their performance). A mean of 7.9% of trials (SD = 7.3%) across subjects and conditions contained pitch errors; these trials were excluded from analyses, since any error that added or subtracted a tone from the melodic sequence changed the relationship between the participants' key presses and the pre-programmed auditory feedback.

2.6.2. EEG Data

Electrical activity was recorded at the scalp using a 64-channel Ag/AgCl electrode BioSemi ActiveTwo System (BioSemi, Inc., Amsterdam, The Netherlands). A sampling rate of 1024 Hz, recording bandwidth of 0 to 205 Hz, and resolution of 24 bits were used. Electrode locations were prescribed by the 10–20 international electrode configuration system. Horizontal and vertical eye movements were monitored by electrodes placed adjacent to the outer canthi of the eyes and above and below the right eye, respectively.

EEG data were analyzed using BrainVision Analyzer 2.0.2 (Brain Products GmbH, Gilching, Germany). Activity was re-referenced off-line to the average of all scalp electrodes, and signals were bandpass-filtered between 0.1 and 30 Hz. The EEG data were then segmented into 500 ms epochs beginning 100 ms prior to and continuing 400 ms after pitch onsets at positions −1, 0, and +1. Activity during the 100 ms prior to pitch onsets served as a baseline. An epoch duration of 500 ms was selected since it included activity that was shorter than three standard deviations below the mean IOI (=487 ms) of key presses recorded during the continuation period, and therefore avoided contamination of the observed waveforms with ERPs related to the subsequent pitch onset. Artifact rejection was performed

automatically using a ±50 μV rejection threshold at the 64 scalp electrodes, as well as the horizontal and vertical right eye electrodes. Artifacts were considered excessive for a given subject when more than half of the epochs from a given condition of the experiment exceeded the ±50 μV rejection threshold at one of the 64 scalp electrodes or at the horizontal or vertical eye electrodes. Trials that contained pitch errors were also excluded from EEG analyses, resulting in the inclusion of 30.4/32 epochs (SD = 3.2) in the future +1 condition, 28.2/32 epochs (SD = 3.3) in the future +2 condition, 28.2/32 epochs (SD = 2.3) in the noncontextual condition, and 85.3/96 epochs (SD = 6.8) in the baseline condition (which contained three times as many stimuli as it was matched to the other conditions).

Average ERPs by participant and experimental condition were then computed for the 500-ms window time-locked to the 100 ms prior to pitch onsets. Mean ERP amplitudes were statistically evaluated at 3 topographical regions of interest (ROIs), based on related findings [7]: Anterior (electrodes Fz and FCz), central (electrodes Cz and CPz), and posterior (electrodes Pz and POz). ERP amplitudes were statistically evaluated over 40-ms time windows selected based on previous findings [7] as follows: 80–120 ms (labeled N1), 120–160 ms (labeled P2), 180–220 ms (labeled FRN), and 250–290 ms (labeled P3a). All of the ERP components were maximal at the anterior ROI; results are therefore reported for the anterior ROI only, following previous work [7,56,71]. Repeated-measures analyses of variance (ANOVAs) were conducted on ERP component amplitudes to analyze the effects of feedback type (future +1, future +2, noncontextual, and baseline). Scalp topographic maps showing ERP component distributions were generated by plotting amplitude values on the scalp. Activity was averaged across the time window used for the analysis of each component. Within-participant correlations between mean ERP amplitudes and behavioral measures for each participant were computed using simple linear regression.

Because increases in spectral power in the theta frequency range (4–8 Hz) typically accompany the FRN [63], we analyzed theta power at the anterior ROI within the 200–300 ms that followed pitch onsets at the three event positions [7,56,72]. To allow for the specification of a temporal baseline period as well as a temporal buffer, with the purpose of preventing edge artifacts within the 100-ms epoch of interest, time-frequency decompositions were calculated for each participant in a −1000 to +1000 ms time window centered on pitch onsets [73]. Our goal in using time-frequency analysis was to ensure that any potential ERP component overlap in the average ERP waveforms did not provide an alternative interpretation of our results. In order to eliminate influences of faster or slower components overlapping the FRN in the average ERP waveforms, decompositions were computed using a Morlet wavelet transform based on each participant's average ERP waveforms for each experimental condition [56,64,74]. To achieve sufficient temporal resolution for the theta frequency range, the number of Morlet wavelet cycles used for analysis of the theta band was set to $n = 7$ [75,76]. Mean power in a pre-stimulus baseline period of −100 to 0 ms was subtracted from the 2-s time-frequency analysis window to permit the assessment of event-related changes in theta activity [77]. Repeated-measures ANOVAs on mean theta power within the 200–300 ms following pitch onsets with factors' feedback type (future +1, future +2, noncontextual, baseline) and event position (0, +1) were conducted to analyze the effects of feedback conditions on theta power. Post-hoc pairwise comparisons were made using Tukey's honestly significant difference (HSD) test for both behavioral and neural measures. η_p^2 was used as a measure of effect size.

3. Results

3.1. Future +1 Altered Feedback Disrupts Key Press Timing

The mean performance rate, indicated by the mean IOI per trial, during the continuation phase of the synchronization-continuation trials was 486.5 ms (SE = 0.3 ms), slightly faster than the metronome-indicated rate of 500 ms from the earlier synchronization phase. An ANOVA on mean IOIs per trial within the continuation phase by feedback condition yielded no main effect of feedback, $F(3, 69) = 1.78$, $p = 0.16$, suggesting that performance rates did not differ across the four

conditions (future +1 M = 485.8 ms, SE = 0.5; future +2 M = 486.2 ms, SE = 0.6; noncontextual M = 486.4, SE = 0.5; baseline M = 487.7, SE = 0.5). Thus, performers successfully maintained the same tempo for all feedback conditions, with slightly faster rates than the prescribed rate overall, consistent with similar previous studies [7,78].

Figure 3 shows IOIs at melody positions preceding, at, and following at the altered feedback pitches and the same positions in the unchanged baseline pitches. An ANOVA on mean IOIs by feedback condition (future +1, future +2, noncontextual, baseline) and event position (−1, 0, +1) revealed a significant interaction of feedback condition with event position, F (6, 138) = 3.60, $p < 0.005$, $\eta_p^2 = 0.14$. IOIs at position 0 were significantly shorter than IOIs at positions −1 and +1 for the future +1 feedback condition only (Tukey HSD = 3.81, $p < 0.05$). IOIs did not significantly differ between positions −1, 0, and +1 for any other condition. There were no main effects of feedback type or position on IOIs. Thus, the only condition in which the altered auditory feedback temporally disrupted performance was the future +1 feedback condition, in which performers shortened the time interval during which they heard the altered feedback tone.

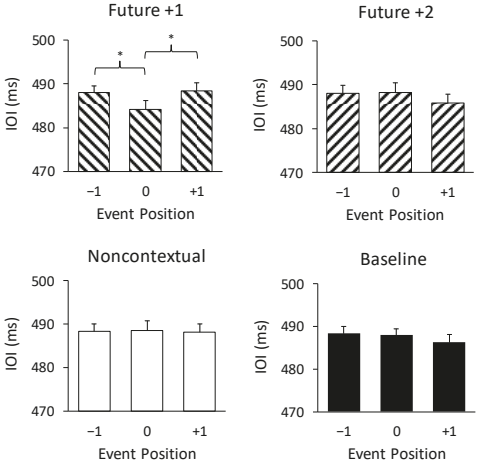


Figure 3. Pianists' mean interonset intervals (IOIs) by altered feedback condition (baseline, future +1, future +2, and noncontextual) by target event position (−1, 0, +1). Error bars represent one standard error. * $p < 0.05$.

We next analyzed participants' key press errors (7.9% of all trials) by feedback condition. There was no significant main effect of feedback condition on the mean proportion of trials that contained errors, F (3, 69) = 0.37, $p = 0.78$. Pitch errors occurred at roughly equivalent rates across trials in all four feedback conditions (future +1 M = 7.8%, SE = 1.7%; future +2 M = 9.0%, SE = 1.5%; noncontextual M = 7.5%, SE = 1.2%; baseline M = 7.1%, SE = 1.6%).

3.2. EEG Results

3.2.1. Event-Related Potentials

Figure 4 shows grand averaged ERP waveforms time-locked to key press onsets, averaged across error-free trials. ERP components are time-locked to key presses corresponding to the feedback pitch onset at position 0, as well as to the key presses at melody positions −1 (preceding location) and +1 (following location). N1 components and P2 ERP components, labeled in Figure 4, were observed at positions −1, 0, and +1 for all feedback conditions. Additionally, FRN and P3a components were observed at position 0 for the three altered feedback conditions. Scalp topographies corresponding to the N1 and P2 components at positions −1, 0, and +1 by feedback condition are shown in Figure 5.

Topographies corresponding to the FRN and P3a components at position 0 are shown in Figure 6. Analyses of each ERP component are reported in turn.

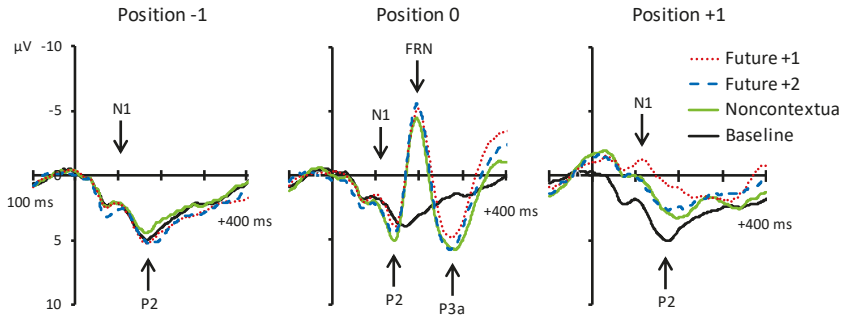


Figure 4. Grand average event-related potentials (ERPs) elicited by the four experimental conditions relative to target event positions −1, 0, and +1. Activity shown is averaged across all electrodes contained within the anterior region of interest (ROI). Negative is plotted upward.

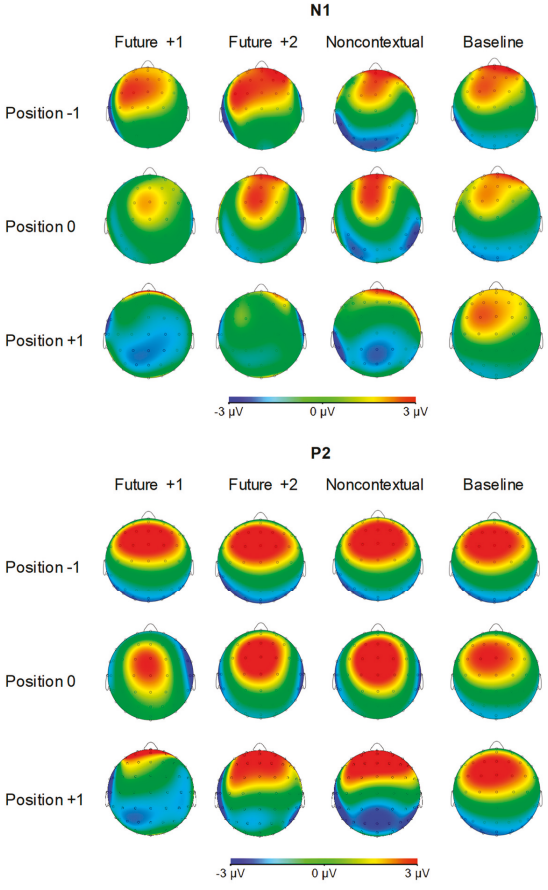


Figure 5. Voltage (in μV) scalp topographies of N1 and P2 components relative to target event positions −1, 0, and +1 by feedback condition. Activity averaged over 40 ms surrounding each component's grand average peak is shown.

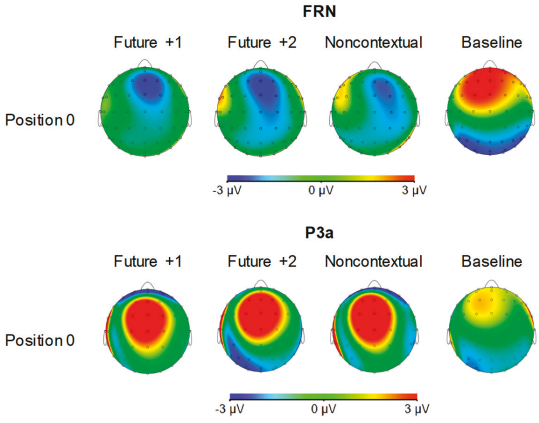


Figure 6. Voltage (in μV) scalp topographies of feedback-related negativity (FRN) and P3a components elicited by pitches at target event position 0 by feedback condition. Activity averaged over 40 ms surrounding each component's grand average peak is shown.

N1 component (80–120 ms). We first evaluated whether mean amplitudes within the N1 time window differed across auditory feedback conditions. We conducted one-way ANOVAs on N1 amplitudes at each event position with the factor feedback type. N1 amplitudes did not significantly differ across feedback conditions at position −1, $F(3, 69) = 0.18$, $p = 0.91$. N1 amplitudes also did not significantly differ across feedback conditions at position 0, $F(3, 69) = 1.47$, $p = 0.23$. Analysis of N1 amplitudes at position +1 yielded a significant main effect of feedback type, $F(3, 69) = 7.42$, $p < 0.001$, $\eta_p^2 = 0.24$. All three altered feedback types elicited a significantly more negative N1 than did baseline feedback pitches (Tukey HSD = 1.73, $p < 0.05$). Thus, N1 amplitudes at event position +1 were sensitive to whether altered auditory feedback was presented one tone earlier (altered feedback conditions) or not (baseline condition). Specifically, N1 amplitudes were more negative following altered compared to baseline feedback.

P2 component (120–160 ms). We next evaluated whether mean amplitudes within the P2 time window differed across auditory feedback conditions. We conducted one-way ANOVAs on P2 amplitudes at each event position with the factor feedback type. P2 amplitudes did not significantly differ across feedback conditions at position −1, $F(3, 69) = 0.25$, $p = 0.86$. P2 amplitudes also did not significantly differ across feedback conditions at position 0, $F(3, 69) = 1.04$, $p = 0.38$. Analysis of P2 amplitudes at position +1 yielded a significant main effect of feedback type, $F(3, 69) = 13.95$, $p < 0.001$, $\eta_p^2 = 0.38$. All three altered feedback types elicited a significantly less positive P2 than baseline feedback pitches (Tukey HSD = 2.12, $p < 0.01$). Furthermore, the P2 elicited by future +1 feedback was significantly less positive than the P2 elicited by future +2 and noncontextual altered feedback (Tukey HSD = 1.73, $p < 0.05$). Thus, like the N1 component, the P2 was sensitive to whether altered auditory feedback was presented one tone earlier or not. Critically, P2 amplitudes were more negative following future +1 feedback compared to future +2 feedback.

Correlation of N1 and P2 amplitudes. The temporal proximity of N1 and P2 components as well as their co-occurrence following both altered and unaltered feedback is consistent with their interpretation as joint indices of auditory sensory processing [34]. To test the relationship between N1 and P2 components, mean amplitudes within the N1 time window (80–120 ms) were compared with amplitudes within the adjacent P2 time window (120–160 ms) for each position and feedback condition. As shown in Table 1, amplitudes within the time windows of the N1 and P2 were significantly correlated for all feedback conditions at positions −1, 0, and +1 (all ps < 0.001).

FRN component (180–220 ms). Analysis of mean amplitudes within the FRN time window at position 0 yielded a significant main effect of feedback type, $F(3, 69) = 31.53$, $p < 0.001$, $\eta_p^2 = 0.58$.

All three altered feedback types elicited a significantly more negative FRN compared to the baseline condition (Tukey HSD = 2.58, $p < 0.05$). No other comparisons reached significance. Thus, all three altered auditory feedback types elicited an FRN response.

P3a component (250–290 ms). Analysis of mean amplitudes within the P3a time window at position 0 yielded a significant main effect of feedback type, F (3, 69) = 7.70, $p < 0.001$, $\eta_p^2 = 0.25$. All three altered feedback types elicited a significantly more positive P3a compared to the baseline condition (Tukey HSD = 2.44, $p < 0.05$). No other comparisons reached significance. Thus, as predicted, all three altered auditory feedback types elicited a P3a response.

Table 1. Correlations of mean N1 and P2 amplitudes at target event positions −1, 0, and +1 for each feedback condition. * $df = 22$, $p < 0.001$.

	Position −1	Position 0	Position +1
Future +1	0.84 *	0.90 *	0.79 *
Future +2	0.94 *	0.64 *	0.82 *
Noncontextual	0.81 *	0.96 *	0.96 *
Baseline	0.64 *	0.73 *	0.82 *

3.2.2. Evoked Oscillatory Responses

To assess whether altered auditory feedback influenced spectral power within the theta frequency range, we computed spectral power in the 4–8 Hz frequency range within the anterior ROI at each event position, as shown in Figure 7. Analysis of theta spectral power during the 200–300 ms following pitch onsets by feedback condition (future +1, future +2, noncontextual, and baseline) and position (−1, 0, and +1) yielded main effects of both feedback condition, F (3, 69) = 6.49, $p = 0.001$, $\eta_p^2 = 0.22$, and position, F (2, 46) = 8.24, $p = 0.001$, $\eta_p^2 = 0.26$. There was also a significant interaction between feedback condition and position, F (6, 138) = 7.68, $p < 0.001$, $\eta_p^2 = 0.25$. Theta power was greater for each of the three altered feedback conditions compared to the baseline feedback condition at position 0 (Tukey HSD = 157.2, $p < 0.01$). Theta power was also greater at position 0 compared to position +1 within each of the three altered feedback conditions (Tukey HSD = 157.2, $p < 0.01$). In sum, theta power increased only following altered feedback pitches that occurred at position 0, and not following (unaltered) feedback pitches that occurred at position +1. Thus, changes in theta power depended on whether the feedback was altered or not, and not on whether the feedback contents were repeated (future +1) or not (future +2).

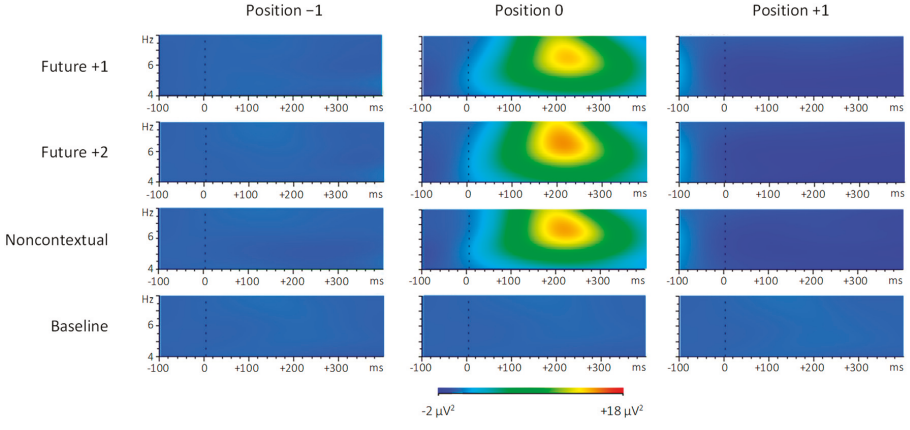


Figure 7. Evoked spectral power within the 4-8 Hz (theta) frequency range following pitch onsets at target event positions −1, 0, and +1. Brighter colors indicate greater spectral power.

3.3. Correlations of Neural and Behavioral Measures

ERP Amplitudes and IOIs

To examine the relationship between the temporal disruption to key press timing and the ERP components, we first tested whether the temporal disruption arising from future +1 auditory feedback—the shortening of the position 0 IOI—correlated with mean amplitudes of ERP components at position +1 that immediately followed the disrupted timing. As shown in Figure 8, the shortened mean IOIs at position 0 correlated significantly with mean amplitudes of the subsequent N1 in the future +1 condition, r (22) = 0.47, $p < 0.05$. Shorter IOIs at position 0 were associated with a larger N1 response to the pitch that followed the altered feedback. The correlation of mean IOIs at position 0 with amplitudes of the P2 at position +1 yielded a similar pattern of association, but the correlation did not reach significance, r (22) = 0.27, p = 0.22. Mean N1 and P2 amplitudes did not correlate with mean IOIs at position 0 for any other feedback condition (future +2, noncontextual, and baseline feedback). Thus, auditory sensory processing of the tone following the altered feedback, reflected in the N1, was associated with temporal disruption only when near future altered feedback was presented.

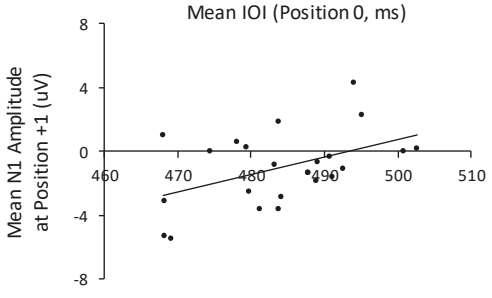


Figure 8. Correlation of mean IOIs at target event position 0 in the future +1 altered feedback condition with mean N1 amplitudes elicited by the tone that followed the altered auditory feedback pitch in the future +1 condition. Each dot represents one participant.

To examine the relation between temporal disruption and FRN responses to altered auditory feedback, we computed the interonset change (in ms) from the IOI at position 0 to the IOI at position +1. Participants' mean difference in IOIs between positions 0 and +1 across all three altered feedback conditions correlated significantly with mean FRN amplitudes time-locked to the altered feedback pitch (position 0), r (21) = 0.41, $p < 0.05$, shown in Figure 9. Mean amplitudes within the time window of the FRN were not correlated with the difference in IOIs across positions 0 and +1 for the baseline condition, r (21) = 0.27, p = 0.24. Therefore, amplitudes of the FRN elicited by altered auditory feedback were associated with changes in the performance rate that succeeded the altered feedback: More negative FRNs were associated with increases in the performance rate. No other ERP component amplitudes correlated significantly with IOIs or with IOI differences at event positions preceding or following altered auditory feedback.

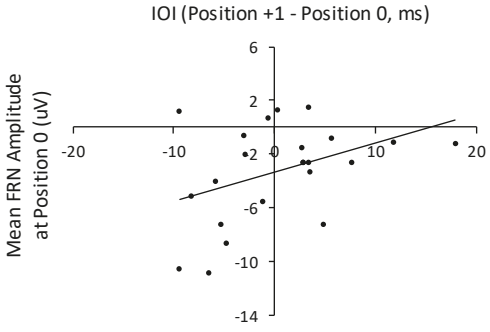


Figure 9. Correlation of mean IOI differences (target event position 1 minus position 0) from the three altered feedback conditions (future +1, future +2, and noncontextual) with mean FRN amplitudes elicited by altered feedback (target event position 0) across the three altered feedback conditions (future +1, future +2, and noncontextual). Each dot represents one participant.

4. Discussion

We examined the relationship between future-oriented planning processes and feedback monitoring during music performance. Skilled pianists performed short melodies from memory. Perceived auditory feedback was occasionally altered to match immediately upcoming sequence events (future +1), later future events (future +2), or unrelated pitches that were not contained within the performed sequences (noncontextual). There were several novel findings. First, only future +1 altered feedback—not future +2 or noncontextual altered feedback—perturbed the timing of pianists' key presses. Second, the length of time it took performers to initiate the pitch following the future +1 altered feedback pitch was associated with larger auditory sensory potentials to the post-altered feedback pitch. Third, all types of altered feedback elicited FRN and P3a potentials. Fourth, FRN amplitudes increased as performers sped up following the altered feedback pitch, in response to all types of altered auditory feedback. Together, these findings suggest that future-oriented planning during production influences how performers monitor their auditory feedback. The range of sequential planning may be constrained by distance: Events at nearby sequence positions had a greater influence on planning and monitoring processes than did events at farther positions, consistent with theories of sequence production in which planned events are activated along a gradient that is defined by sequential distance [8,9,12]. According to a predictive coding model [15], a cascade of forward models for upcoming movements may generate an error signal in response to altered auditory feedback that is stronger when the feedback matches nearby sequence positions than when it matches farther positions.

4.1. Behavioral Findings

The timing of pianists' performances was disrupted following the perception of altered auditory feedback that corresponded to near future, but not far future, events. According to future-oriented theories of planning during music and speech production, immediately upcoming events receive stronger activation than events that are farther ahead in a melody or utterance [8,12]. When pianists heard an altered feedback pitch that matched an event that was already strongly activated in memory, the altered pitch may have generated similarity-based interference with the event that was currently being produced. Thus, temporal perturbations observed in the future +1 condition may reflect the greater interference of near future altered feedback with currently planned pitch events compared to far future altered feedback. This interpretation is consistent with theories of sensorimotor production in which actions and their auditory effects share common cognitive representations [79,80], as well as theories in which actions are planned in terms of their sensory effects [81–83]. We previously demonstrated that future-oriented, but not past-oriented, altered auditory feedback induced compensatory adjustments in keystroke timing [7]. The current results

extend this finding by suggesting that future-oriented interference interacts with graded planning and monitoring processes during music performance.

Another important factor that constrains memory retrieval of sequence of events is the similarity between sequence elements. Evidence from production errors and priming paradigms has indicated that grammatical and phonological similarity influence lexical retrieval [84,85], and tonal and metrical accent relationships influence event retrieval during music performance [8,65]. For example, musicians are more likely to produce pitch errors in metrically weak than in metrically strong accent positions [65]; sequence events that align with greater metrical accent strength tend to be produced with greater intensity [86]. The melodies used in the current study were designed so that metrical accents of the future +2 altered feedback pitch were more similar to the currently planned pitch event than were the metrical accents of future +1 feedback [87]. This metrical similarity approach would predict that the future +2 altered feedback should generate greater interference and performance disruption than future +1 feedback. This prediction was not supported by the current results: Instead, altered feedback that contained serially proximal pitches was more disruptive to performance than altered feedback that contained metrically similar pitches. This suggests that serial proximity may play a greater role than metrical accent strength in generating interference with planned representations for the short sequences used in the current study. One explanation for the lesser contribution of metrical accent to the disruptive effects of altered auditory feedback could be that metrical relationships between sequence events tend to span longer timeframes than the timespans between serially proximal events [65].

Serially-shifted feedback, like the future-oriented altered auditory feedback presented in our study, is known to increase performers' overall key press error rates [21]. We observed heightened error rates in all altered auditory feedback conditions compared to baseline (unchanged) feedback. Error rates were relatively low compared to rates as high as 40% observed in other studies employing serially-shifted auditory feedback [22]. A likely explanation for this difference is that single pitches were altered at random sequence locations in the current study, which prevented performers from anticipating the alterations, unlike in previous studies, in which auditory feedback was continuously and consistently altered. When auditory feedback is predictably altered, performers can develop strategies to compensate for predictable deviations from expected feedback. Even under conditions in which every feedback tone is altered during music performance, pitch errors begin to occur only after several melody repetitions [88]. Future studies could further investigate interactions between hierarchical and distance constraints on sequence planning using musical materials that amplify differences between strongly and weakly accented events.

4.2. EEG Findings

Altered auditory feedback attenuated cortical sensory suppression compared to baseline feedback, reflected in amplitude-shifted N1 and P2 ERP components. Sensory suppression is widely believed to result from the congruence between sensory consequences of actions and sensory predictions generated by forward models of motor commands (for a review, see [29]). Theories of motor control have proposed that efference copies of motor commands are used to predict sensory outcomes of those commands, and that sensory suppression results from the subtraction of an efference copy from actual sensory input [14,89]. Sensory suppression is often used as an implicit measure of agency, as actions must be volitional in order to generate predictive models of motor commands [90,91]. Increased auditory sensory processing following altered feedback pitches could therefore indicate that altered feedback disrupted pianists' sense of agency or control over the sounds that they were producing (cf. [7]). This interpretation also fits with the proposal that sensory suppression during production may serve the purpose of allowing producers to differentiate self-generated from externally-generated sensations [36].

We observed a greater reduction of sensory suppression following future +1 altered feedback compared to future +2 altered feedback, reflected in the P2. This finding suggests that the post-altered feedback pitch received enhanced cortical sensory processing in the future +1 condition compared

to the future +2 condition. Further, reduced sensory suppression in the future +1 condition was associated with a quicker initiation of the tone following the future +1 feedback pitch. Together, these findings suggest that enhanced cortical sensory processing following the future +1 altered auditory feedback may have aided the recovery from perturbations caused by the unexpected feedback. Indeed, expectancy violations tend to receive enhanced neural processing compared to events that fulfill expectations, in line with a predictive coding view of cortical responses [92]. It is unlikely that differences in P2 amplitudes for future +1 and future +2 conditions were driven by differences in selective attention between altered feedback conditions, since sensory suppression during auditory production appears to be uninfluenced by whether attention is directed toward or away from one's own actions or their auditory effects [93]. It is also unlikely that this amplitude difference is due to differences between future +1 and future +2 conditions in terms of pitch repetition. From a repetition suppression perspective, we would expect decreased—not increased—cortical processing of the tone that followed the future +1 altered feedback tone, since this tone was repeated and stimulus repetition classically results in a decreased brain response due to sensory adaptation [94,95]. The fact that theta power did not distinguish future +1 from future +2 responses supports this interpretation. We propose that sensory suppression depended on the differences in interference generated by the future +1 and future +2 altered feedback pitches with concurrent planning processes. Amplitudes may indicate the degree of conflict or mismatch between perceived altered auditory feedback and concurrent planning processes, which are biased towards the immediate future.

Both N1 and P2 components are sensitive to a variety of acoustic features of incoming auditory signals, highlighting a role of these components in early auditory sensory processing. For example, pitch changes in vocal stimuli during active vocalization elicit larger N1 and P2 responses than pitch changes in non-voice complex stimuli, which in turn elicit larger amplitudes than pure tones [96]. Acoustic spectral complexity [42], pitch discrimination and speech-sound training [43,97], and the rate of speech formant transition [98] have all been shown to modulate N1 and P2 responses. The current results extend these findings by demonstrating that N1 and P2 amplitudes also take into account the relationship between pitch changes and planned events in an auditory sequence. Speech sounds are generally more spectrally complex than musical sounds [99]. An open question for future research is therefore whether alterations of auditory feedback during speech production are better detected by the auditory system than feedback alterations during music performance.

FRN and P3a ERP components were elicited by all altered auditory feedback (future +1, future +2, and noncontextual) pitches. ERP amplitudes were equivalent across all altered feedback conditions. FRN and P3a components have been elicited by altered auditory feedback during music performance in previous studies [45–47]. None of these studies compared neural responses to different types of altered auditory feedback, with the exception of Katahira and colleagues [45], who manipulated the diatonicity of altered feedback tones. The current finding suggests that performers identified and subsequently oriented toward all types of unexpected feedback. This finding fits with the principle that any alteration of feedback during auditory-motor tasks creates a mismatch between movements and expected auditory outcomes, which create larger violations for producers with higher skill levels [19] or with greater sequence familiarity [100]. Studies using flanker gambling tasks have demonstrated that the FRN is sensitive to the perceptual distinctiveness of unexpected stimuli [101–103]. The noncontextual control condition presented diatonically-related altered feedback pitches that were more distinct from the pitch set of the produced melodies than were the altered pitches in the future +1 and future +2 melodies. Yet, the FRN elicited by noncontextual altered feedback did not differ from that elicited by contextual (future +1 and future +2) feedback. The association between FRN amplitudes and speed of the altered feedback pitch for all three altered feedback conditions further supports this interpretation. Thus, FRN responses may be less affected by perceptual distinctiveness or by performers' planning processes and more dependent on action-related expectations. Future studies may address this possibility directly with manipulations of perceptual distinctiveness.

Theta power increases were also observed following all types of altered feedback tones, about 200–300 ms after the altered pitch onsets. The lack of differences in theta power across feedback conditions confirms the FRN results, and suggests that amplitudes of the FRN elicited by altered feedback were unaltered by overlapping ERP components. Increases in theta power within the approximate timeframe of the FRN component suggest that identification of expectancy violating pitches coincided with the emergence of a more cognitively controlled, deliberative mental state, as opposed to a mental state relying primarily on habit or performance routines [63]. Just as the FRN has been suggested to reflect surprising action-based outcomes [104], theta has been referred to as a "surprise signal" that leads to task-specific adjustments in cognitive control [63]. Theta frequency oscillations may coordinate the excitability of populations of mid-frontal neurons, thereby providing a temporal window in which cognitive control can be instantiated [105]. Orienting to altered auditory feedback during music performance may therefore involve a switch from a state of relatively automatic performance to performance that is more deliberative and goal-directed, characterized by transitory changes in the production rate. Finally, increases in theta power did not differentiate between altered feedback types. Similar to the notion that the FRN may depend more on action-related expectations than on performers' planning processes, equivalent increases in theta across feedback conditions suggest that any violation of any action-sound association is sufficient for invoking the need for cognitive control.

5. Conclusions

This study provides the first neural support for the finding in speech production and music performance that planning of upcoming events in a sequence is influenced by the serial proximity of the future events. Feedback monitoring processes interacted with planning processes: Performers' perception of altered feedback tones that matched immediately upcoming future events resulted in behavioral and neural adaptations, including temporal disruption (speeded IOIs), enhanced cortical sensory processing following the altered feedback (amplitude-shifted N1 and P2 responses), and increased theta frequency activity. These findings support models of sequence production in which the planning of future events is modulated by their serial distance from the current event [8,12], and contribute to our understanding of the link between sensory suppression and action planning during the performance of complex action sequences. The N1-P2 complex may serve as a neural marker for disruptive effects of altered auditory feedback in sensorimotor tasks.

Author Contributions: B.M., W.J.G. and C.P. conceived and designed the experiment; B.M. performed the experiment; B.M., W.J.G. and C.P. analyzed the data; B.M., W.J.G. and C.P. wrote the paper.

Funding: National Science Foundation Graduate Research Fellowship to B.M. Canada Research Chairs grant and Natural Sciences and Engineering Research Council of Canada grant 298173 to C.P.

Acknowledgments: We thank Pierre Gianferrara, Erik Koopmans, and Frances Spidle of the Sequence Production Lab for their assistance.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

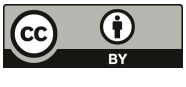
1. Lashley, K. The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior*; Jeffress, L.A., Ed.; Wiley: New York, NY, USA, 1951; pp. 112–136.
2. Fromkin, V.A. The non-anomalous nature of anomalous utterances. *Language* **1971**, *47*, 27–52. [CrossRef]
3. Garrett, M.F. Syntactic processes in sentence production. *New Approaches Lang. Mech.* **1976**, *30*, 231–256.
4. Palmer, C.; van de Sande, C. Units of knowledge in music performance. *J. Exp. Psychol. Learn. Mem. Cognit.* **1993**, *19*, 457–470. [CrossRef]
5. Palmer, C.; van de Sande, C. Range of planning in music performance. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 947–962. [CrossRef] [PubMed]

6. Palmer, C.; Drake, C. Monitoring and planning capacities in the acquisition of music performance skills. *Can. J. Exp. Psychol.* **1997**, *51*, 369. [CrossRef] [PubMed]
7. Mathias, B.; Gehring, W.J.; Palmer, C. Auditory N1 reveals planning and monitoring processes during music performance. *Psychophysiology* **2017**, *54*, 235–247. [CrossRef] [PubMed]
8. Palmer, C.; Pfordresher, P.Q. Incremental planning in sequence production. *Psychol. Rev.* **2003**, *110*, 683–712. [CrossRef]
9. Pfordresher, P.Q.; Palmer, C.; Jungers, M.K. Speed, accuracy, and serial order in sequence production. *Cognit. Sci.* **2007**, *31*, 63–98. [CrossRef]
10. Levelt, W.J. Monitoring and self-repair in speech. *Cognition* **1983**, *14*, 41–104. [CrossRef]
11. Palmer, C.; Schendel, Z.A. Working memory constraints in sequence production. *Psychon. Soc.* **2002**, *7*, 30.
12. Dell, G.S.; Burger, L.K.; Svec, W.R. Language production and serial order: A functional analysis and a model. *Psychol. Rev.* **1997**, *104*, 123. [CrossRef] [PubMed]
13. Pickering, M.J.; Clark, A. Getting ahead: Forward models and their place in cognitive architecture. *Trends Cognit. Sci.* **2014**, *18*, 451–456. [CrossRef] [PubMed]
14. Wolpert, D.M.; Ghahramani, Z.; Jordan, M.I. An internal model for sensorimotor integration. *Science* **1995**, *269*, 1880–1882. [CrossRef] [PubMed]
15. Friston, K. Prediction, perception and agency. *Int. J. Psychophysiol.* **2012**, *83*, 248–252. [CrossRef]
16. Vuust, P.; Dietz, M.J.; Witek, M.; Kringelbach, M.L. Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Ann. N. Y. Acad. Sci.* **2018**, *1423*, 19–29. [CrossRef]
17. Maes, P.J.; Leman, M.; Palmer, C.; Wanderley, M. Action-based effects on music perception. *Front. Psychol.* **2014**, *4*, 1008. [CrossRef]
18. Keller, P.E.; Koch, I. Action planning in sequential skills: Relations to music performance. *Q. J. Exp. Psychol.* **2008**, *61*, 275–291. [CrossRef]
19. Lutz, K.; Puorger, R.; Cheetham, M.; Jancke, L. Development of ERN together with an internal model of audio-motor associations. *Front. Hum. Neurosci.* **2013**, *7*, 471. [CrossRef]
20. Furuya, S.; Soechting, J.F. Role of auditory feedback in the control of successive keystrokes during piano playing. *Exp. Brain Res.* **2010**, *204*, 223–237. [CrossRef]
21. Pfordresher, P.Q. Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *J. Exp. Psychol. Hum. Percept. Perform.* **2003**, *29*, 949–964. [CrossRef]
22. Pfordresher, P.Q.; Palmer, C. Effects of hearing the past, present, or future during music performance. *Percept. Psychophys.* **2006**, *68*, 362–376. [CrossRef] [PubMed]
23. Finney, S.A. Auditory feedback and musical keyboard performance. *Music Percept.* **1997**, *15*, 153–174. [CrossRef]
24. Repp, B.H. Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Percept. Psychophys.* **1999**, *61*, 529–548. [CrossRef] [PubMed]
25. Pfordresher, P.Q. Auditory feedback in music performance: The role of melodic structure and musical skill. *J. Exp. Psychol. Hum. Percept. Perform.* **2005**, *31*, 1331–1345. [CrossRef] [PubMed]
26. Pfordresher, P.Q. Auditory feedback in music performance: The role of transition-based similarity. *J. Exp. Psychol. Hum. Percept. Perform.* **2008**, *34*, 708. [CrossRef]
27. Aliu, S.O.; Houde, J.F.; Nagarajan, S.S. Motor-induced suppression of the auditory cortex. *J. Cognit. Neurosci.* **2009**, *21*, 791–802. [CrossRef]
28. Baess, P.; Horváth, J.; Jacobsen, T.; Schröger, E. Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology* **2011**, *48*, 1276–1283. [CrossRef]
29. Bendixen, A.; Sanmiguel, I.; Schröger, E. Early electrophysiological indicators for predictive processing in audition: A review. *Int. J. Psychophysiol.* **2012**, *83*, 120–131. [CrossRef]
30. Christoffels, I.K.; van de Ven, V.; Waldorp, L.J.; Formisano, E.; Schiller, N.O. The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex. *PLoS ONE* **2011**, *6*, e18307. [CrossRef]
31. Horváth, J. Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Res.* **2015**, *1626*, 54–65. [CrossRef]
32. Sanmiguel, I.; Todd, J.; Schröger, E. Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology* **2013**, *50*, 334–343. [CrossRef] [PubMed]
33. Näätänen, R.; Picton, T. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* **1987**, *24*, 375–425. [CrossRef] [PubMed]

34. Näätänen, R.; Winkler, I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* **1999**, *125*, 826–859. [CrossRef] [PubMed]
35. Horváth, J.; Maess, B.; Baess, P.; Tóth, A. Action–sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *J. Cognit. Neurosci.* **2012**, *24*, 1919–1931. [CrossRef] [PubMed]
36. Sowman, P.F.; Kuusik, A.; Johnson, B.W. Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. *Exp. Brain Res.* **2012**, *222*, 149–157. [CrossRef] [PubMed]
37. Hillyard, S.A.; Hink, R.F.; Schwent, V.L.; Picton, T.W. Electrical signs of selective attention in the human brain. *Science* **1973**, *182*, 177–180. [CrossRef] [PubMed]
38. Okita, T. Event-related potentials and selective attention to auditory stimuli varying in pitch and localization. *Biol. Psychol.* **1979**, *9*, 271–284. [CrossRef]
39. Snyder, J.S.; Alain, C.; Picton, T.W. Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cognit. Neurosci.* **2006**, *18*, 1–13. [CrossRef] [PubMed]
40. Woldorff, M.G.; Gallen, C.C.; Hampson, S.A.; Hillyard, S.A.; Pantev, C.; Sobel, D.; Bloom, F.E. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8722–8726. [CrossRef] [PubMed]
41. Hari, R.; Aittoniemi, K.; Järvinen, M.L.; Katila, T.; Varpula, T. Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators. *Exp. Brain Res.* **1980**, *40*, 237–240. [CrossRef]
42. Shahin, A.; Bosnyak, D.J.; Trainor, L.J.; Roberts, L.E. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* **2003**, *23*, 5545–5552. [CrossRef] [PubMed]
43. Shahin, A.; Roberts, L.E.; Pantev, C.; Trainor, L.J.; Ross, B. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* **2005**, *16*, 1781–1785. [CrossRef] [PubMed]
44. Huberth, M.; Dauer, T.; Nanou, C.; Román, I.; Gang, N.; Reid, W.; Wright, M.; Fujioka, T. Performance monitoring of self and other in a turn-taking piano duet: A dual-EEG study. *Soc. Neurosci.* **2018**, *9*, 1–13. [CrossRef] [PubMed]
45. Katahira, K.; Abla, D.; Masuda, S.; Okanoya, K. Feedback-based error monitoring processes during musical performance: An ERP study. *Neurosci. Res.* **2008**, *61*, 120–128. [CrossRef] [PubMed]
46. Loehr, J.D.; Kourtis, D.; Vesper, C.; Sebanz, N.; Knoblich, G. Monitoring individual and joint action outcomes in duet music performance. *J. Cognit. Neurosci.* **2013**, *25*, 1049–1061. [CrossRef] [PubMed]
47. Maidhof, C.; Vavatzanidis, N.; Prinz, W.; Rieger, M.; Koelsch, S. Processing expectancy violations during music performance and perception: An ERP study. *J. Cognit. Neurosci.* **2010**, *22*, 2401–2413. [CrossRef] [PubMed]
48. San Martín, R.; Manes, F.; Hurtado, E.; Isla, P.; Ibañez, A. Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *Neuroimage* **2010**, *51*, 1194–1204. [CrossRef] [PubMed]
49. Carter, C.S.; Van Veen, V. Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognit. Affect. Behav. Neurosci.* **2007**, *7*, 367–379. [CrossRef]
50. Ferdinand, N.K.; Mecklinger, A.; Kray, J.; Gehring, W.J. The processing of unexpected positive response outcomes in the mediofrontal cortex. *J. Neurosci.* **2012**, *32*, 12087–12092. [CrossRef]
51. Ferdinand, N.K.; Opitz, B. Different aspects of performance feedback engage different brain areas: Disentangling valence and expectancy in feedback processing. *Sci. Rep.* **2014**, *4*, 5986. [CrossRef]
52. Oliveira, F.T.; McDonald, J.J.; Goodman, D. Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *J. Cognit. Neurosci.* **2007**, *19*, 1994–2004. [CrossRef] [PubMed]
53. Schaefer, A.; Buratto, L.G.; Goto, N.; Brotherhood, E.V. The feedback-related negativity and the P300 brain potential are sensitive to price expectation violations in a virtual shopping task. *PLoS ONE* **2016**, *11*, e0163150. [CrossRef] [PubMed]
54. Helfrich, R.F.; Knight, R.T. Oscillatory dynamics of prefrontal cognitive control. *Trends Cognit. Sci.* **2016**, *20*, 916–930. [CrossRef]
55. Navarro-Cebrian, A.; Knight, R.T.; Kayser, A.S. Frontal monitoring and parietal evidence: Mechanisms of error correction. *J. Cognit. Neurosci.* **2016**, *28*, 1166–1177. [CrossRef] [PubMed]
56. Gehring, W.J.; Willoughby, A.R. Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. In *Errors, Conflicts, and the Brain: Current Opinions on Performance Monitoring*; Falkenstein, M.U.M., Ed.; Max Planck Institute of Human Cognitive and Brain Science: Leipzig, Germany, 2004; pp. 14–20.

57. Donchin, E.; Coles, M.G. Is the P300 component a manifestation of context updating. *Behav. Brain Sci.* **1988**, *11*, 357–427. [CrossRef]
58. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* **2007**, *118*, 2128–2148. [CrossRef]
59. Nieuwenhuis, S.; Aston-Jones, G.; Cohen, J.D. Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychol. Bull.* **2005**, *131*, 510–532. [CrossRef] [PubMed]
60. Overbeek, T.J.; Nieuwenhuis, S.; Ridderinkhof, K.R. Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *J. Psychophysiol.* **2005**, *19*, 319–329. [CrossRef]
61. Berti, S. Switching attention within working memory is reflected in the P3a component of the human event-related brain potential. *Front. Hum. Neurosci.* **2015**, *9*, 701. [CrossRef] [PubMed]
62. Lange, F.; Seer, C.; Finke, M.; Dengler, R.; Kopp, B. Dual routes to cortical orienting responses: Novelty detection and uncertainty reduction. *Biol. Psychol.* **2015**, *105*, 66–71. [CrossRef] [PubMed]
63. Cavanagh, J.F.; Frank, M.J. Frontal theta as a mechanism for cognitive control. *Trends Cognit. Sci.* **2014**, *18*, 414–421. [CrossRef] [PubMed]
64. Bernat, E.M.; Nelson, L.D.; Steele, V.R.; Gehring, W.J.; Patrick, C.J. Externalizing psychopathology and gain–loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *J. Abnorm. Psychol.* **2011**, *120*, 352. [CrossRef] [PubMed]
65. Mathias, B.; Pfordresher, P.Q.; Palmer, C. Context and meter enhance long-range planning in music performance. *Front. Hum. Neurosci.* **2015**, *8*, 1040. [CrossRef] [PubMed]
66. Palmer, C.; Mathias, B.; Anderson, M. Sensorimotor mechanisms in music performance: Actions that go partially wrong. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 185–191. [CrossRef] [PubMed]
67. Finney, S.A. FTAP: A Linux-based program for tapping and music experiments. *Behav. Res. Methods Instrum. Comput.* **2001**, *33*, 65–72. [CrossRef]
68. Rihs, T.A.; Michel, C.M.; Thut, G. Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. *Eur. J. Neurosci.* **2007**, *25*, 603–610. [CrossRef] [PubMed]
69. Zhao, S.; Wang, Y.; Xu, H.; Feng, C.; Feng, W. Early cross-modal interactions underlie the audiovisual bounce-inducing effect. *NeuroImage* **2018**, *174*, 208–218. [CrossRef] [PubMed]
70. Croft, R.J.; Barry, R.J. Removal of ocular artifact from the EEG: A review. *Clin. Neurophysiol.* **2000**, *30*, 5–19. [CrossRef]
71. Schuermann, B.; Endrass, T.; Kathmann, N. Neural correlates of feedback processing in decision-making under risk. *Front. Hum. Neurosci.* **2012**, *6*, 204. [CrossRef]
72. Massi, B.; Luhmann, C.C. Fairness influences early signatures of reward-related neural processing. *Cognit. Affect. Behav. Neurosci.* **2015**, *15*, 768–775. [CrossRef]
73. Herrmann, C.S.; Grigutsch, M.; Busch, N.A. EEG oscillations and wavelet analysis. In *Event-Related Potentials: A Methods Handbook*; Handy, T.C., Ed.; MIT Press: Cambridge, MA, USA, 2005; pp. 229–259.
74. Bertrand, O.; Bohorquez, J.; Pernier, J. Time-frequency digital filtering based on an invertible wavelet transform: An application to evoked potentials. *IEEE Trans. Biomed. Eng.* **1994**, *41*, 77–88. [CrossRef] [PubMed]
75. Li, P.; Baker, T.E.; Warren, C.; Li, H. Oscillatory profiles of positive, negative and neutral feedback stimuli during adaptive decision making. *Int. J. Psychophysiol.* **2016**, *107*, 37–43. [CrossRef] [PubMed]
76. Mas-Herrero, E.; Marco-Pallarés, J. Frontal theta oscillatory activity is a common mechanism for the computation of unexpected outcomes and learning rate. *J. Cognit. Neurosci.* **2014**, *26*, 447–458. [CrossRef] [PubMed]
77. Makeig, S. Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalogr. Clin. Neurophysiol.* **1993**, *86*, 283–293. [CrossRef]
78. Pfordresher, P.; Palmer, C. Effects of delayed auditory feedback on timing of music performance. *Psychol. Res.* **2002**, *66*, 71–79. [CrossRef] [PubMed]
79. Hommel, B.; Müsseler, J.; Aschersleben, G.; Prinz, W. Codes and their vicissitudes. *Behav. Brain Sci.* **2001**, *24*, 910–926. [CrossRef]
80. Prinz, W. A common coding approach to perception and action. In *Relationships between Perception and Action*; Springer: Berlin, Germany, 1990; pp. 167–201.
81. Greenwald, A.G. Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychol. Rev.* **1970**, *77*, 73. [CrossRef]
82. Miall, R.C.; Wolpert, D.M. Forward models for physiological motor control. *Neural Netw.* **1996**, *9*, 1265–1279. [CrossRef]

83. Schubotz, R.I. Prediction of external events with our motor system: Towards a new framework. *Trends Cognit. Sci.* **2007**, *11*, 211–218. [CrossRef]
84. Dell, G.S. A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **1986**, *93*, 283. [CrossRef]
85. Friederici, A.D. Processing local transitions versus long-distance syntactic hierarchies. *Trends Cognit. Sci.* **2004**, *8*, 245–247. [CrossRef] [PubMed]
86. Palmer, C. Anatomy of a performance: Sources of musical expression. *Music Percept.* **1996**, *13*, 433–453. [CrossRef]
87. Lerdahl, F.; Jackendoff, R. An overview of hierarchical structure in music. *Music Percept.* **1983**, *1*, 229–252. [CrossRef]
88. Pfordresher, P.Q.; Mantell, J.T.; Brown, S.; Zivadinov, R.; Cox, J.L. Brain responses to altered auditory feedback during musical keyboard production: An fMRI study. *Brain Res.* **2014**, *1556*, 28–37. [CrossRef] [PubMed]
89. Bays, P.M.; Wolpert, D.M. Computational principles of sensorimotor control that minimize uncertainty and variability. *J. Physiol.* **2007**, *578*, 387–396. [CrossRef] [PubMed]
90. Dewey, J.A.; Knoblich, G. Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS ONE* **2014**, *9*, e110118. [CrossRef] [PubMed]
91. Gentsch, A.; Schütz-Bosbach, S. I did it: Unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. *J. Cognit. Neurosci.* **2011**, *23*, 3817–3828. [CrossRef] [PubMed]
92. Auksztulewicz, R.; Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **2016**, *80*, 125–140. [CrossRef] [PubMed]
93. Timm, J.; Sanmiguel, I.; Saupe, K.; Schröger, E. The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neurosci.* **2013**, *14*, 1. [CrossRef] [PubMed]
94. Brown, R.M.; Chen, J.L.; Hollinger, A.; Penhune, V.B.; Palmer, C.; Zatorre, R.J. Repetition suppression in auditory–motor regions to pitch and temporal structure in music. *J. Cognit. Neurosci.* **2013**, *25*, 313–328. [CrossRef] [PubMed]
95. Grill-Spector, K.; Henson, R.; Martin, A. Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cognit. Sci.* **2006**, *10*, 14–23. [CrossRef] [PubMed]
96. Behroozmand, R.; Korzyukov, O.; Larson, C.R. Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clin. Neurophysiol.* **2011**, *122*, 2408–2417. [CrossRef] [PubMed]
97. Tremblay, K.; Kraus, N.; McGee, T.; Ponton, C.; Otis, B. Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear Hear.* **2001**, *22*, 79–90. [CrossRef] [PubMed]
98. Carpenter, A.L.; Shahin, A.J. Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neurosci. Lett.* **2013**, *544*, 56–61. [CrossRef] [PubMed]
99. Zatorre, R.J.; Belin, P.; Penhune, V.B. Structure and function of auditory cortex: Music and speech. *Trends Cognit. Sci.* **2002**, *6*, 37–46. [CrossRef]
100. Mathias, B.; Palmer, C.; Perrin, F.; Tillmann, B. Sensorimotor learning enhances expectations during auditory perception. *Cereb. Cortex* **2015**, *25*, 2238–2254. [CrossRef]
101. Jia, S.; Li, H.; Luo, Y.; Chen, A.; Wang, B.; Zhou, X. Detecting perceptual conflict by the feedback-related negativity in brain potentials. *Neuroreport* **2007**, *18*, 1385–1388. [CrossRef]
102. Liu, Y.; Gehring, W.J. Loss feedback negativity elicited by single-versus conjoined-feature stimuli. *Neuroreport* **2009**, *20*, 632–636. [CrossRef]
103. Liu, Y.; Nelson, L.D.; Bernat, E.M.; Gehring, W.J. Perceptual properties of feedback stimuli influence the feedback-related negativity in the flanker gambling task. *Psychophysiology* **2014**, *51*, 782–788. [CrossRef]
104. Alexander, W.H.; Brown, J.W. Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* **2011**, *14*, 1338. [CrossRef]
105. Cavanagh, J.F.; Zambrano-Vazquez, L.; Allen, J.J. Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology* **2012**, *49*, 220–238. [CrossRef] [PubMed]



brain sciences

MDPI

Article

Attention Modulates Electrophysiological Responses to Simultaneous Music and Language Syntax Processing

Daniel J. Lee [1], Harim Jung [1] and Psyche Loui [1,2,*]

[1] Department of Psychology, Wesleyan University, Middletown, CT 06459, USA; jdlee@wesleyan.edu (D.J.L.); hjung01@wesleyan.edu (H.J.)
[2] Department of Music, Northeastern University, Boston, MA 02115, USA
* Correspondence: p.loui@northeastern.edu; Tel.: +01-617-373-6588

Received: 18 September 2019; Accepted: 29 October 2019; Published: 1 November 2019

Abstract: Music and language are hypothesized to engage the same neural resources, particularly at the level of syntax processing. Recent reports suggest that attention modulates the shared processing of music and language, but the time-course of the effects of attention on music and language syntax processing are yet unclear. In this EEG study we vary top-down attention to language and music, while manipulating the syntactic structure of simultaneously presented musical chord progressions and garden-path sentences in a modified rapid serial visual presentation paradigm. The Early Right Anterior Negativity (ERAN) was observed in response to both attended and unattended musical syntax violations. In contrast, an N400 was only observed in response to attended linguistic syntax violations, and a P3/P600 only in response to attended musical syntax violations. Results suggest that early processing of musical syntax, as indexed by the ERAN, is relatively automatic; however, top-down allocation of attention changes the processing of syntax in both music and language at later stages of cognitive processing.

Keywords: music; language; syntax; attention; comprehension; electroencephalography; event-related potentials

1. Introduction

Music and language are both fundamental to human experience. The two domains, while apparently different, rely on several notable similarities: both exhibit syntactic structure, and both rely on sensory, cognitive, and vocal-motor apparatus of the central nervous system. The nature of this relationship between syntactic structure in language and music, and their underlying neural substrates, is a topic of intense interest to the cognitive and brain sciences community.

The Shared Syntactic Integration Resource Hypothesis (SSIRH [1]) is an influential theoretical account of similarities and differences between cognitive processing for music and language. The SSIRH posits that neural resources for music and language overlap at the level of the syntax; in other words, processing of music and language should interact at the syntactic level, but not at other levels such as semantics or acoustic or phonemic structure.

Support for the SSIRH comes from a variety of behavioral and neural studies. Several studies have presented music and language simultaneously with and without syntactic violations, to test for effects of separate and simultaneous syntax violations on behavioral and neural measures [2–6]. Strong support for the SSIRH comes from a self-paced reading paradigm [2], where sentence segments were presented concurrently with musical chord progressions. One subset of the trials contained syntactic violations in language (garden path sentences), and another subset contained syntactic violations in music (out-of-key chords); a third subset contained simultaneous syntactic violations in both domains.

Reaction time results showed that during simultaneous violations of music and language, participants were slowest to respond to the double violation than they were to respond to each violation alone. This superadditive effect was not observed in a control experiment which manipulated the timbre of the music and the semantics of the language.

Although these results seem to offer convincing support for SSIRH, Perruchet and Poulin-Charronnat (2013) [7] showed that under semantic garden path manipulations (as opposed to syntactic garden path manipulations), violations of semantics can also yield the same pattern. Based on these results, Perruchet and Poulin-Charronnat (2013) suggested that increased attentional resources, rather than syntax processing per se, could lead to these statistical interactions.

The idea that attention can influence the pattern of interaction between music and language processing has since received more support. More recent work has argued that the processing resources of syntax for language and music might both rely on domain-general attentional resources, especially when simultaneously processing music and language in a dual-task situation [3,8]. In that regard, classic theories of attention distinguish between the endogenous, voluntary maintenance of a vigilant state, and the exogenous, involuntary orienting to stimulus events [9]. These attentional systems both affect reaction time, but involve different neural resources and unfold differentially over time [10]. Since reaction time during simultaneous linguistic and musical syntax processing may not readily differentiate between overlapping syntax-specific resources and the engagement of attentional resources, we turned to more time-sensitive measures of neural activity during the processing of music and language. This enables direct comparisons between neural responses to musical syntax violations and neural responses to language syntax violations at multiple time windows throughout the temporal cascade of attentional processes that are triggered during music and language processing. By comparing neural markers of syntax violations in the two domains of music and language, and testing for the interaction between attention and violations in each domain, we can clarify the roles that attentional resources might play in the processing of syntax in music and language.

The Early Right Anterior Negativity (ERAN) and the Early Left Anterior Negativity (ELAN) are reliably elicited event-related potential (ERP) markers of syntax processing in music and language respectively [5,11–14]. The ERAN is a frontally-generated negative waveform around 200 ms after the onset of musical syntax violations, whereas the ELAN is an analogous frontally-generated negativity after violations in linguistic syntax, such as violations in word category [15] or phrase structure [16]. Musical syntax processing has been localized to the inferior frontal gyrus (IFG) in magnetoencephalographic (MEG) and fMRI studies [17–20]. Additional results from ERAN of lesioned patients [21] and in children with Specific Language Impairment [22] have also provided evidence for the reliance of musical syntax processing on classic language-related areas such as the inferior frontal gyrus. The ERAN is also posited as an index of predictive processes in the brain, especially in the case of music, due to its reliance on the formation and subsequent violation of predictions that are learned from exposure to musical sound sequences [23]. Impaired ERAN is observed in adults with lesions in the left inferior frontal gyrus (Broca's area), which provides additional support for the SSIRH. Importantly, the ERAN is also sensitive to top-down task demands, such as attentional resources devoted to the musical stimuli in contrast to a concurrent, non-musical task [24]. When music and speech were simultaneously presented from different locations, irregular chords elicited an ERAN whereas irregular sentences elicited an ELAN; moreover, the ERAN was slightly reduced when irregular sentences were presented, but only when music was ignored, suggesting that the processing of musical syntax is partially automatic [25].

While ERAN and ELAN are markers of syntax processing, semantic processing is indicated by the N400, a negative-going centroparietal waveform beginning around 400–500 ms after a semantic anomaly [26,27]. In addition to being sensitive to semantic content of words, the N400 effect reflects the semantic associations between words and the expectancy for them more generally, showing a larger waveform as an incoming word is unexpected or semantically incongruous with the previous context. In response to ambiguities in linguistic syntax that violate the ongoing context, the P600 is another effect

that has also been observed [28]. The P600 is a positive waveform centered around the parietal channels and has been observed during garden path sentences, which are syntactically ambiguous sentences when a newly presented word or words require a reinterpretation of the preceding context [29]. Patel et al. (1998) tested the language-specificity of the P600 by presenting chord progressions in music and garden path sentences in language in separate experiments. Their results showed statistically indistinguishable positive waveforms in the P600 range; in addition, they observed an ERAN-like waveform specifically for music [30]

The P600 is similar in topography and latency to the P3, a complex of positive-going event-related potentials elicited from 300 ms and onwards following an unexpected and task-relevant event. The P3 is separable into two components: P3a, a fronto-central ERP, largest around FCz that reflects novelty processing, and P3b, a later parietally-centered ERP largest around Pz that is more sensitive to motivation and task demands [31]. Patients with frontal lobe lesions show altered habituation of the P3, suggesting that the amplitude of the P3 is subject to frontally-mediated processes [32]. The P3a and P3b have both been observed during top-down attention to syntactically incongruous events in music, and these waveforms are sensitive to different levels and genres of expertise [11,33]. Taken together, the literature suggests two main classes of ERPs during unexpectedness in music and language processing: one class within a relatively early time window of approximately 200 ms (ERAN) and another class during the later time window of 500–600 ms (P3, P600, N4). The earlier class of waveforms are thought to be partially automatic, that is, they are elicited even without top-down attention but their amplitude is modulated by attention. The later class of waveforms is highly sensitive to top-down demands including attention.

In this study we compare ERPs elicited by violations in musical and linguistic syntax, while attention was directed separately towards language or music. We used the stimulus materials from Slevc et al's (2009), but extended this study by adding a musical analog of the language comprehension task. Thus, across two experiments we were able to compare behavioral results during task-manipulated attention to language and music, while independently manipulating syntax in each domain at a finer timescale in order to test for effects in ERPs that are known markers of syntax processing and attention.

2. Materials and Methods

2.1. Subjects

Thirty-five undergraduate students from Wesleyan University participated in return for course credit. All participants reported normal hearing. Informed consent was obtained from all subjects as approved by the Ethics Board of Psychology at Wesleyan University. Sixteen students (11 males and 5 females, mean age = 19.63, SD = 2.03) were assigned to the Attend-language group: 15/16 participants in this group reported English as their first language, and 9/16 participants reported prior music training (total mean of training in years = 2.23, SD = 3.42). Nineteen students (8 males and 11 females, mean age = 19.40, SD = 2.03) were assigned to the Attend-music group. Background survey and baseline tests of one participant in this group was missing as the result of a technical error. Of the remaining 18 participants, 12/18 reported English as their first language, and 11/18 participants reported having prior music training (total mean of training in years = 3.11, SD = 4.01).

The two groups of subjects did not differ in terms of general intellectual ability, as measured by the Shipley Institute of Living scale for measuring intellectual impairment and deterioration [34]. Nor did they differ in low-level pitch discrimination abilities as assessed by a pitch-discrimination task [35]. (Two-up-one-down staircase procedure around the center frequency of 500 Hz showed similar thresholds between the two groups.) They also did not differ in musical ability as assessed using the Montreal Battery for Evaluation of Amusia [36], or in duration of musical training (years of musical training was not different between the two groups, $X^2 = 0.0215$, $p = 0.88$). Table 1 shows the demographics and baseline test performance of the participants in both conditions.

Table 1. Demographics and baseline test performance of the participants. Data are shown as mean (SD), range, or proportion. SD: Standard Deviation. *n*: Count in proportion.

Variable	Attend-Language (*N* = 16)	Attend-Music (*N* = 19)
Age in years, M (SD)	19.625 (2.029)	19.389 (2.033)
Male, *n*	11/16	8/19
Music Training, years, M (SD)	2.233 (3.422)	3.105 (4.012)
Musically trained, *n*	9/16	11/18
Full-Scale IQ (Estimated from Shipley-Hartford IQ scale, M (SD))	100 (10)	101 (7)
MBEA, M(SD)	23.375 (3.828)	25.11 (2.685)
Pitch Discrimination, ΔHz/500 Hz, M (SD)	12.469 (11.895)	11.087 (7.174)
Normal Hearing, %	100%	100%
English as First Language, *n*	15/16	12/18

2.2. Stimuli

The stimuli were adapted from Slevc, Rosenberg [2]. There were 144 trials in the study, including 48 congruent trials, 48 musically-incongruent trials, and 48 language-incongruent trials. In each trial, an English sentence was presented in segments simultaneously with a musical chord progression. Each segment of a sentence was paired with a chord that followed the rules of Western tonal harmony in the key of C major, played in a grand piano timbre. Linguistic syntax expectancy was manipulated through syntactic garden-path sentences, whereas musical syntax expectancy was manipulated through chords that were either in-key or out-of-key at the highlighted critical region (Figure 1). The chords and sentence segments were presented at the regular inter-onset interval of 1200 ms. At the end of each sentence and chord progression, a yes/no comprehension question was presented on the screen: In the Attend-Language group, this question was about the content of the sentence to direct participants' attention to language (e.g., "Did the attorney think that the defendant was guilty?"). For the Attend-Music group, the question at the end of the trial asked about the content of the music (e.g., "Did the music sound good?") to direct participants' attention to the music. Participants were randomly assigned to Attend-Language and Attend-Music groups. Participants' task, in both Attend-Language and Attend-Music groups, was always to respond to the question at the end of each trial by choosing "Yes" or "No" on the screen.

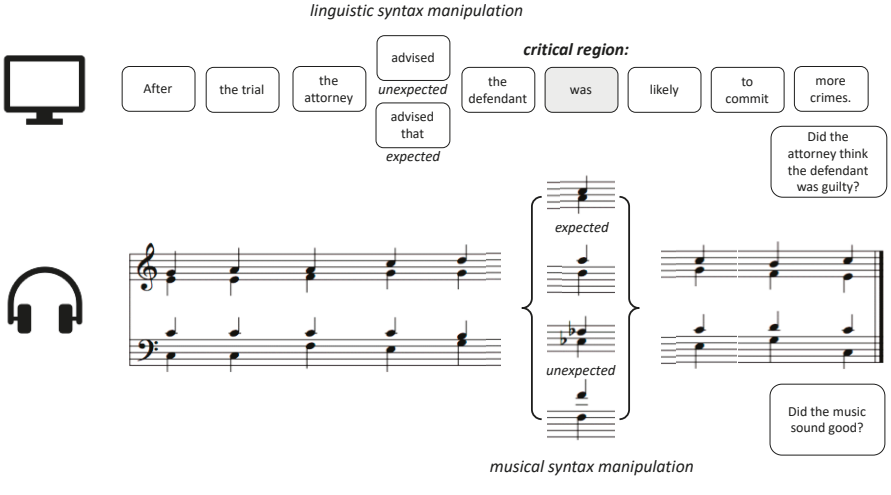


Figure 1. Example trials for Attend-language and Attend-music conditions.

2.3. Procedure

Participants first gave informed consent and filled out a background survey on their musical training, as well as a battery of behavioral tasks including the Shipley Institute of Living Scale to screen for impairments in intellectual functioning [34], the Montreal Battery for Evaluation of Amusia (MBEA) to screen for impairments in musical functioning [36], and a pitch discrimination test as a three-up-one-down psychophysical staircase procedure around the center frequency of 500 Hz to assess pitch discrimination accuracy [35]. The experiment was run on a Macbook Pro laptop computer using Max/MSP [37]. At the start of the experiment, participants were told to pay attention to every trial, and to answer a yes-or-no comprehension question about the language (Attend-Language condition) or about the music (Attend-Music condition) at the end of each trial. They were given a short practice run of 5 trials of the experiment in order to familiarize themselves with the task before EEG recording began. EEG was recorded using PyCorder software from a 64-channel BrainVision actiCHamp setup with electrodes corresponding to the international 10–20 EEG system. Impedance was kept below 10 kOhms. The recording was continuous with a raw sampling rate of 1000 Hz. EEG recording took place in a sound attenuated, electrically shielded chamber.

2.4. Behavioral Data Analysis

Behavioral data from Max/MSP were imported to Excel to compute the accuracy of each participant. Accuracy was evaluated against 50% chance-level in one-sample two-tailed t-tests in SPSS. For the Attend-Music condition, two subjects' behavioral data were lost due to technical error.

2.5. EEG Preprocessing

BrainVision Analyzer software (Brain Product Gmbh) 2.1 was used to preprocess raw data. EEG data were first re-referenced to TP9 and TP10 mastoid electrodes, and filtered with high-pass cutoff of 0.5 Hz, low-pass cutoff of 30 Hz, roll-off of 24 dB/oct, and a notch filter of 60 Hz. These filter settings were chosen based on previous work that looked at target ERPs similar to the current study [33,38], since filter settings introduce artifacts in ERP data [39]. Ocular correction ICA was applied to remove eye artifacts for each subject. Raw data inspection was done semi-automatically by first setting maximal allowed voltage step as 200 μV/ms, maximal difference of values over a 200 ms interval as 400 μV, and maximal absolute amplitude as 400 μV. Then, manual data inspection was performed to remove segments with noise due to physical movements.

2.6. Event-Related Potential Analysis

The preprocessed data were segmented into four conditions: music congruent, music incongruent, language congruent, and language incongruent. Each segment was 1200 ms long, spanning from a 200 ms baseline before the onset of the stimulus to 1000 ms after stimulus onset. The segments were averaged across trials, baseline-corrected, and grand-averaged across the subjects. To identify effects specific to syntax violations in each modality, a difference wave was created for each violation condition by subtracting ERPs for congruent conditions from ERPs for incongruent conditions, resulting in a Music-specific difference wave (Music violation minus no violation) and a Language-specific difference wave (Language violation minus no violation). From these difference waves we isolated ERP amplitudes at two recording sites, one at each time window of interest: E(L/R)AN from site FCz at 180–280 ms, and the N4 and P3 at site Pz at 500–600 ms. The mean amplitude of each ERP was exported for each participant from BrainVision Analyzer into SPSS for analysis.

Because both groups of participants experienced both types of syntactic violations (music and language), but each group of participants attended to only one modality (music or language), we used a mixed-effects analysis of variance (ANOVA) with the within-subjects factor of Violation (two levels: music and language) and the between-subjects factor of Attention (two levels: attend-music and

attend-language). This was separately tested for the two time windows: 1) the early ERAN/ELAN time window of 180–280 ms, and 2) the later N4/P3 time window of 500–600 ms.

3. Results

3.1. Behavioral Results

Participants performed well above the 50% chance level on language comprehension questions during the Attend-Language condition (M = 0.8457, SD = 0.0703, two-tailed t-test against chance level of 50% correct: $t(15) = 19.661$, $p < 0.001$), and on music comprehension questions during the Attend-Music condition (M = 0.6631, SD = 0.1253, two-tailed t-test against chance level of 50% correct: $t(16) = 5.371$, $p < 0.001$). This confirms that participants successfully attended to both language and music stimuli. Participants performed better on the Attend-Language than on the Attend-Music questions ($t(31) = 5.12$, $p < 0.001$).

3.2. Event-Related Potentials

Figure 2 shows each ERP and scalp topographies of difference waves. Figure 3 shows the specific effects for each ERP; statistics are shown in Table 2. A right anterior negative waveform was observed 180–210 ms following music violations, consistent with the ERAN. This ERAN was observed during both Attend-Language and Attend-Music conditions. During the Attend-Language condition, a centroparietal negative waveform was observed 500–600 ms following language violations, consistent with an N400 effect. This N400 effect was not observed during the Attend-Music condition. Instead, a posterior positive waveform was observed 500–600 ms after music violations during the Attend-Music condition, consistent with the Late Positive Complex or the P3 or P600 effect. These selective effects are tested in a mixed-model ANOVA for each time-point, with a between-subjects factor of attention (two levels: attend-music vs. attend-language) and a within-subjects factor of modality of syntax violation (two levels: music and language), as described below.

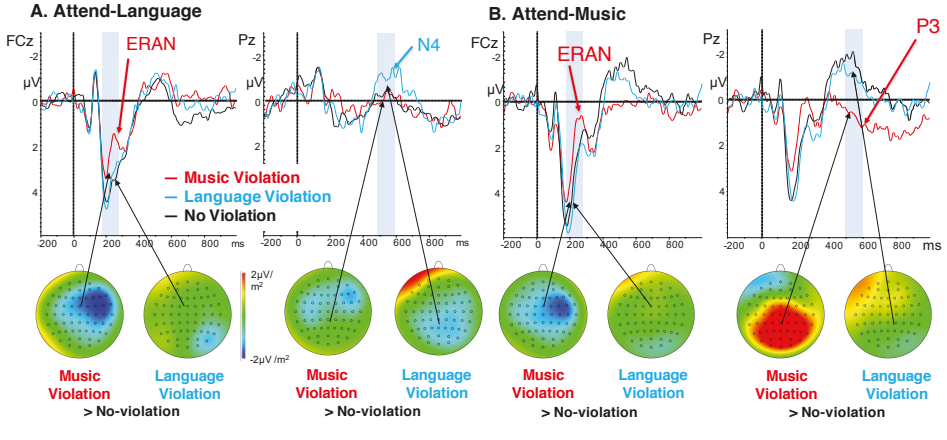


Figure 2. Overlays of ERPs from each condition with topographic maps of the difference wave between violation and no-violation conditions. Music syntax violation condition is shown in red and linguistic syntax violation condition is shown in blue. Black represents a condition when neither stimulus was violated. Topographic plots show difference waves between music violation and no-violation, or between language violation and no-violation. (**A**) When attending to language. (**B**) When attending to music.

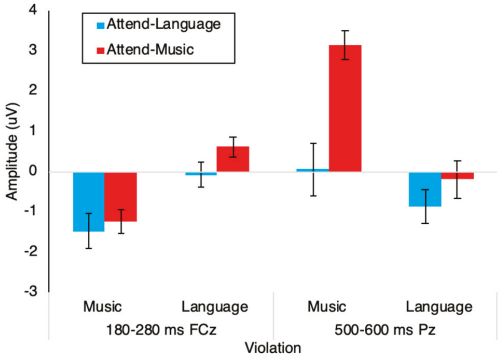


Figure 3. ERP effects of violation (amplitude of difference waves) across different conditions.

Table 2. ERP statistics.

Tests of Within-Subjects Contrasts					
Source	**Time-Window**	**df**	**F**	***p***	**Partial eta^2**
Violation	180–280 ms	1	33.198	< 0.001	0.501
	500–600 ms	1	31.317	< 0.001	0.487
Violation * Attend	180–280 ms	1	0.64	0.43	0.019
	500–600 ms	1	9.951	0.003	0.232
Tests of Between-Subjects Effects					
Source	**Time-Window**	**df**	**F**	***p***	**Partial eta^2**
Attend	180–280 ms	1	1.381	0.248	0.040
	500–600 ms	1	9.763	0.004	0.228

180–280 ms: A significant negative waveform was observed for the music violation but not for the language violation. The within-subjects effect of Violation showed a significant difference between music and language violations ($F(1,33) = 33.198$, $p < 0.001$). The between-subjects effect of Attention was not significant ($F(1,33) = 1.381$, $p = 0.248$). There was no significant interaction between the Violation and Attention factors (Figure 2). Tests of between-subjects effects showed no significant difference between the Attend-Music and the Attend-Language conditions (Figure 3).

500–600 ms. For the late time window, the within-subjects effect of Violation was significant ($F(1,33) = 31.317$, $p < 0.001$), and the between-subjects effect of Attention was significant ($F(1,33) = 9.763$, $p = 0.004$). Here, an Attention by Violation interaction was also significant ($F(1,33) = 9.951$, $p = 0.003$). This interaction is visible in the ERP traces and topographic plots in Figure 2 as well as in the amplitude results plotted in Figure 3: in the Attend-Language condition, only language violations elicited a negative waveform resembling an N400, whereas music violations were no different from the no-violation condition. The N400 shows a latency of 400–800 ms and a centro-parietal topography (Figure 2A), consistent with classic reports of the N400 effect (Kutas and Hillyard, 1984). In contrast, during the Attend-Music condition, only music violations elicited a large positive P3 waveform, whereas language violations showed no difference from the no-violation condition. The P3 shows a latency of 400–1000 ms and a centro-parietal topography (Figure 2B), consistent with the P3b subcomponent of the P3 complex (Polich, 2007). The P3 was only observed for music violations when attention was directed to music, and the N400 was only observed for language violations when attention was directed to language. This attention-dependent double dissociation between P3 and N400 is visible in Figures 2 and 3.

While musical violations elicited an ERAN in the early time window and a P3 in the late time window, language violations only showed an N400 in the late time window, and no effect in the early

time window. One potential explanation is that a minority of participants were not first-language English speakers; these participants may have been less sensitive to syntax violations as manipulated by the garden path sentences. Removing the subjects whose first language was not English resulted in a smaller sample size of all first-language English speakers: n = 15 in the Attend-Language condition, and n = 12 in the Attend-Music condition. Repeating the above behavioral and ERP analyses on these smaller samples showed the same pattern of results. Behavioral results showed significantly above-chance performance on both Attend-language condition ($t(14) = 28.66$, $p < 0.001$) and Attend-music conditions ($t(11) = 4.93$, $p = 0.002$). ERP statistics for first-language English speakers are shown in Table 3.

Table 3. ERP statistics for first-language English speakers only.

First Language English Speakers Only		Tests of Within-Subjects Contrasts			
Source	**Time-Window**	**df**	**F**	***p***	**Partial eta^2**
Violation	180–280 ms	1	14.216	< 0.001	0.428
	500–600 ms	1	30.722	< 0.001	0.618
Violation * Attend	180–280 ms	1	0.195	0.664	0.01
	500–600 ms	1	11.075	0.004	0.368
		Tests of Between-Subjects Effects			
Source	**Time-Window**	**df**	**F**	***p***	**Partial eta^2**
Attend	180–280 ms	1	0.971	0.337	0.049
	500–600 ms	1	13.99	< 0.001	0.424

4. Discussion

By separately manipulating linguistic syntax, musical syntax, and attention via task demands during simultaneous music and language processing, we were able to disentangle the effects of top-down attention on bottom-up processing of syntax and syntactic violations. Three main findings come from the current results: 1) For both music and language, syntactic violation processing activates a cascade of neural events, indexed by early and late ERP components as seen using time-sensitive methods. This replicates prior work (Koelsch et al., 2000 [11,12], Hahne and Friederici, 1999 [13,14], and many others). 2) Early components are less sensitive to attentional manipulation than late components, also replicating prior work [40,41]. 3) Attention affects musical and linguistic syntax processing differently at late time windows. This finding is novel as it extends previous work that identify early and late components in music and language syntax processing, by showing that the late components are most affected by attention, whereas the earlier stages of processing are less so. Taken together, results expand on the SSIRH by showing that top-down manipulations of attention differently affect the bottom-up processing of music and language, with effects of attention becoming more prominent throughout the temporal cascade of neural events that is engaged during music and language processing. We posit that the early stages of processing includes mismatch detection between the perceived and the expected events, with the expectation being core to syntactical knowledge in both language and music. In contrast, the late attention-dependent processes may include cognitive reanalysis, integration, and/or updating processes, which may require general attentional resources but are not specific to linguistic or musical syntax.

In some respects, the present results add to a modern revision of the classic debate on early- vs. late-selection theories of attention. While early-selection theories (Broadbent, 1958) posited that attention functions as a perceptual filter to select for task-relevant features in the stimulus stream, late-selection theories have provided evidence for relatively intact feature processing until semantic processing [42,43] or until feature integration [44]. Due to their fine temporal resolution, ERP studies provide an ideal window into this debate, allowing researchers to quantify the temporal cascade of neural events that subserve perceptual-cognitive events such as pitch and phoneme perception, and syntax and semantics processing. ERP results from dual-task paradigms such as dichotic listening have

shown that attention modulates a broad array of neural processes from early sensory events [45,46] to late cognitive events [47,48]. Here we observe the ERAN in response to musical syntax violations regardless of whether attention was directed to language or to music. The ERAN was elicited for music violations even when in the attend-language condition; furthermore its amplitude was not significantly larger during the attend-music condition. This result differs from previous work showing that the ERAN is larger during attended than during unattended conditions [24]. The difference likely stems from the fact that while in the previous study the visual task and the musical task were temporally uncorrelated, in the present study the language stimuli (sentence segments) and musical stimuli (chords) were simultaneously presented, with each language-music pair appearing in a time-locked fashion. Thus, when in the attend-language condition, the onset of musical chords became predictably coupled with the onset of task-relevant stimuli (sentence segments), even though the musical chords themselves were not task-relevant. This predictable coupling of task-irrelevant musical onsets with task-relevant linguistic stimulus onsets meant that it became more advantageous for subjects to allocate some bottom-up attentional resources to the music, or to allocate attentional resources to all incoming sensory stimuli at precisely those moments in time when stimuli were expected [49], as one modality could help predict the other. The fact that the ERAN was observed even when only slightly attended provides some support for a partially automatic processing of musical syntax, as posited in previous work [24]. When musical syntax violations were not task-relevant but were temporally correlated with task-relevant stimuli, they elicited intact early anterior negativity but no late differences from no-violation conditions. This early-intact and late-attenuated pattern of ERP results is also consistent with the relative attenuation model of attention, which posits that unselected stimulus features are processed with decreasing intensity [50].

One remaining question concerns whether the ERAN is driven by music-syntax violations, or whether the effects may be due to sensory violations alone. Indeed, musical syntax violations often co-occur with low-level sensory violations, such as changes in roughness or sensory dissonance. In that regard, the musical syntax violations used in the present study are carefully constructed to avoid sensory dissonance and roughness (see supplementary materials of Slevc et al., 2009 for a list of chord stimuli used). Thus the effects cannot be explained by sensory violations. Furthermore, Koelsch et al. (2007) had shown that ERAN is elicited even when irregular chords are not detectable based on sensory violations, which supports the role of ERAN in music-syntax violations. Given our stimuli as well as previous evidence, we believe that the currently observed ERAN reflects music-syntax violations rather than sensory violations. In contrast, no ELAN was observed in response to language violations. This may be because we used garden path stimuli for language violations, while previous studies that elicited early negative-going ERPs used word category violations [14] and phrase structure violations [16] rather than garden path sentences. The introduction of the linguistic garden path requires that participants re-parse the syntactic tree structure during the critical region of the trial; this effort to re-parse the tree likely elicited the N4 at the later time window, but lacks the more perceptual aspect of the violation that likely elicited the ELAN in prior studies (Hahne et al., 1999). Thus, the garden-path sentences and music-syntactic violations used in the present study may have tapped into distinct sub-processes of syntax processing.

It is remarkable that linguistic syntax violations only elicited a significant N400 effect, and no significant effects over any other time windows, even when language was attended. In contrast, musical syntax violations elicited the ERAN as well as the P3 in the attended condition, with the ERAN being observed even when musical syntax was unattended. Note that the P3 effect in this experiment is similar in topography and latency to the P600, which has been observed for semantic processing during garden path sentences. It could also be the Central-Parietal Positivity (CPP), which reflects accumulating evidence for perceptual decisions [51], which can resemble the P3 [52]. During the attend-music condition, linguistic syntax violations elicited no significant ERP components compared to no-violation conditions. This suggests a strong effect of attention on language processing. It is also worth noting that we saw a clear N400 and not a P600 or a P3 in response to garden path sentences in

language. The relationship between experimental conditions and N400 vs. P600 or P3 is an ongoing debate in neurolinguistics: Kuperberg (2007) posits that the N400 reflects semantic memory-based mechanisms whereas the P600 reflects prolonged processing of the combinatorial mechanism involved in resolving ambiguities [28]. Others argue that whether an N400 or a P600 is observed may in fact depend on the same latent component structure; in other words, the presence and absence of N400 and P600 may reflect two sides of the same cognitive continuum, rather than two different processes per se [53–55]. If the N400 and P600 are indeed two sides of the same coin, then this could mean that language and music processing are also more related than the different effects would otherwise suggest.

5. Limitations

One caveat is that, similar to the original paradigm from which we borrow in this study [2], music was always presented auditorily, whereas language was always presented visually. Thus, the differences we observe between musical and linguistic syntax violation processing could also be due to differences in the modality of presentation. In future studies it may be possible to reverse the modality of presentation, such as by visually presenting musical notation or images of hand positions on a piano [56] with spoken sentence segments. Although doing so would require a more musically trained subject pool who can read musical notation or understand the images of hand positions, prior ERP studies suggest that visually presented musical-syntactic anomalies would still elicit ERP effects of musical syntax violation, albeit with different topography and latency [56]. Furthermore, although participants performed above chance on both attend-language and attend-music comprehension questions, they did perform better on the attend-language task; this imposes a behavioral confound that may affect these results. Future testing on expert musicians may address this behavioral confound. Future studies may also work to increase the sample size, and to validate and match the samples with sensitive baseline measures in both behavioral and EEG testing in order to minimize confounding factors arising from potential differences between participant groups. Importantly, garden path sentences are only one type of syntactic violation; it remains to be seen how other types of violations in linguistic syntax, such as word category violations, may affect the results. Finally, it is yet unclear how syntax and semantics could be independently manipulated in music, or indeed the degree to which syntax and semantics are fully independent, in music as well as in language [57]. In fact, changing musical syntax most likely affects the meaning participants derive from the music; however, specifically composed pieces with target words in mind might be a way to get at a musical semantics task without overtly manipulating syntax [58]. Nevertheless, by separately manipulating music and language during their simultaneous processing, and crossing these manipulations experimentally with top-down manipulations of attention via task demands, we observe a progressive influence of attention on the temporal cascade of neural events for the processing of music and language.

Author Contributions: Conceptualization, H.J. and P.L.; Data curation, D.J.L., H.J. and P.L.; Formal analysis, D.J.L. and P.L.; Funding acquisition P.L.; Investigation, P.L.; Methodology, H.J. and P.L.; Project administration, D.J.L. and H.J.; Resources, P.L.; Software, D.J.L. and P.L.; Visualization, D.J.L. and P.L.; Writing – original draft, D.J.L.; Writing – review & editing, P.L.

Funding: This research was funded by grants from NSF-STTR #1720698, Imagination Institute RFP-15-15, and Grammy Foundation to P.L., D.J.L. and H.J. acknowledge support from Ronald E. McNair Post-Baccalaureate Achievement Program.

Acknowledgments: We thank C.J. Mathew and Emily Przysinda for assistance with data collection, and all the participants of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patel, A.D. Language, music, syntax and the brain. *Nat. Neurosci.* **2003**, *6*, 674–681. [CrossRef]

2. Slevc, L.R.; Rosenberg, J.C.; Patel, A.D. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* **2009**, *16*, 374–381. [CrossRef] [PubMed]
3. Roncaglia-Denissen, M.P.; Bouwer, F.L.; Honing, H. Decision Making Strategy and the Simultaneous Processing of Syntactic Dependencies in Language and Music. *Front. Psychol.* **2018**, *9*, 38. [CrossRef] [PubMed]
4. Jentschke, S.; Koelsch, S.; Friederici, A.D. Investigating the relationship of music and language in children: Influences of musical training and language impairment. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 231–242. [CrossRef] [PubMed]
5. Koelsch, S.; Gunter, T.C.; Wittfoth, M.; Sammler, D. Interaction between Syntax Processing in Language and in Music: An ERP Study. *J. Cogn. Neurosci.* **2005**, *17*, 1565–1577. [CrossRef]
6. Fedorenko, E.; Patel, A.; Casasanto, D.; Winawer, J.; Gibson, E. Structural integration in language and music: Evidence for a shared system. *Mem. Cogn.* **2009**, *37*, 1–9. [CrossRef]
7. Perruchet, P.; Poulin-Charronnat, B. Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* **2012**, *20*, 310–317. [CrossRef]
8. Slevc, L.R.; Okada, B.M. Processing structure in language and music: A case for shared reliance on cognitive control. *Psychon. Bull. Rev.* **2015**, *22*, 637–652. [CrossRef]
9. Posner, M.I.; Petersen, S.E. The Attention System of the Human Brain. *Annu. Rev. Neurosci.* **1990**, *13*, 25–42. [CrossRef]
10. Coull, J.T.; Nobre, A.C. Where and When to Pay Attention: The Neural Systems for Directing Attention to Spatial Locations and to Time Intervals as Revealed by Both PET and fMRI. *J. Neurosci.* **1998**, *18*, 7426–7435. [CrossRef]
11. Koelsch, S.; Schmidt, B.-H.; Kansok, J. Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology* **2002**, *39*, 657–663. [CrossRef] [PubMed]
12. Koelsch, S.; Gunter, T.; Friederici, A.D.; Schröger, E. Brain Indices of Music Processing: "Nonmusicians" are Musical. *J. Cogn. Neurosci.* **2000**, *12*, 520–541. [CrossRef] [PubMed]
13. Sammler, D.; Koelsch, S.; Ball, T.; Brandt, A.; Grigutsch, M.; Huppertz, H.-J.; Knösche, T.R.; Wellmer, J.; Widman, G.; Elger, C.E.; et al. Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage* **2013**, *64*, 134–146. [CrossRef] [PubMed]
14. Hahne, A.B.; Friederici, A.D. Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *J. Cogn. Neurosci.* **1999**, *11*, 194–205. [PubMed]
15. Friederici, A.D. Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* **2002**, *6*, 78–84. [CrossRef]
16. Neville, H.; Nicol, J.L.; Barss, A.; Forster, K.I.; Garrett, M.F. Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. *J. Cogn. Neurosci.* **1991**, *3*, 151–165. [CrossRef]
17. Maess, B.; Koelsch, S.; Gunter, T.C.; Friederici, A.D. Musical syntax is processed in Broca's area: An MEG study. *Nat. Neurosci.* **2001**, *4*, 540–545. [CrossRef]
18. Cheung, V.K.M.; Meyer, L.; Friederici, A.D.; Koelsch, S. The right inferior frontal gyrus processes nested non-local dependencies in music. *Sci. Rep.* **2018**, *8*, 3822. [CrossRef]
19. Tillmann, B.; Koelsch, S.; Escoffier, N.; Bigand, E.; Lalitte, P.; Friederici, A.; Von Cramon, D. Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage* **2006**, *31*, 1771–1782. [CrossRef]
20. Bianco, R.; Novembre, G.; Keller, P.; Kim, S.-G.; Scharf, F.; Friederici, A.; Villringer, A.; Sammler, D.; Keller, P. Neural networks for harmonic structure in music perception and action. *NeuroImage* **2016**, *142*, 454–464. [CrossRef]
21. Sammler, D.; Koelsch, S.; Friederici, A.D. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* **2011**, *47*, 659–673. [CrossRef] [PubMed]
22. Jentschke, S.; Koelsch, S.; Sallat, S.; Friederici, A.D. Children with Specific Language Impairment Also Show Impairment of Music-syntactic Processing. *J. Cogn. Neurosci.* **2008**, *20*, 1940–1951. [CrossRef] [PubMed]
23. Koelsch, S.; Vuust, P.; Friston, K. Predictive Processes and the Peculiar Case of Music. *Trends Cogn. Sci.* **2018**, *23*, 63–77. [CrossRef] [PubMed]
24. Loui, P.; Grent-'t-Jong, T.; Torpey, D.; Woldorff, M. Effects of attention on the neural processing of harmonic syntax in Western music. *Brain Res. Cogn. Brain Res.* **2005**, *25*, 678–687. [CrossRef]

25. Maidhof, C.; Koelsch, S. Effects of Selective Attention on Syntax Processing in Music and Language. *J. Cogn. Neurosci.* **2011**, *23*, 2252–2267. [CrossRef] [PubMed]
26. Kutas, M.; Hillyard, S.A. Brain potentials during reading reflect word expectancy and semantic association. *Nature* **1984**, *307*, 161–163. [CrossRef]
27. Kutas, M.; Hillyard, S. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* **1980**, *207*, 203–205. [CrossRef]
28. Kuperberg, G.R. Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res.* **2007**, *1146*, 23–49. [CrossRef]
29. Osterhout, L.; Holcomb, P.J.; Swinney, D.A. Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *J. Exp. Psychol. Learn. Mem. Cogn.* **1994**, *20*, 786–803. [CrossRef]
30. Patel, A.D.; Gibson, E.; Ratner, J.; Besson, M.; Holcomb, P.J. Processing syntactic relations in language and music: An event-related potential study. *J. Cogn. Neurosci.* **1998**, *10*, 717–733. [CrossRef]
31. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* **2007**, *118*, 2128–2148. [CrossRef] [PubMed]
32. Knight, R.T.; Grabowecky, M.F.; Scabini, D. Role of human prefrontal cortex in attention control. *Adv. Neurol.* **1995**, *66*, 21–36. [PubMed]
33. Przysinda, E.; Zeng, T.; Maves, K.; Arkin, C.; Loui, P. Jazz musicians reveal role of expectancy in human creativity. *Brain Cogn.* **2017**, *119*, 45–53. [CrossRef] [PubMed]
34. Shipley, W.C. A Self-Administering Scale for Measuring Intellectual Impairment and Deterioration. *J. Psychol.* **1940**, *9*, 371–377. [CrossRef]
35. Loui, P.; Guenther, F.H.; Mathys, C.; Schlaug, G. Action-perception mismatch in tone-deafness. *Curr. Boil.* **2008**, *18*, R331–R332. [CrossRef] [PubMed]
36. Peretz, I.; Champod, A.S.; Hyde, K. Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Ann. N. Y. Acad. Sci.* **2003**, *999*, 58–75. [CrossRef] [PubMed]
37. Zicarelli, D. An extensible real-time signal processing environment for Max. In Proceedings of the International Computer Music Conference, University of Michigan, Ann Arbor, MI, USA, 1–6 October 1998. Available online: https://quod.lib.umich.edu/i/icmc/bbp2372.1998.274/1 (accessed on 25 October 2019).
38. Loui, P.; Wu, E.H.; Wessel, D.L.; Knight, R.T. A generalized mechanism for perception of pitch patterns. *J. Neurosci.* **2009**, *29*, 454–459. [CrossRef]
39. Widmann, A.; Schröger, E.; Maess, B. Digital filter design for electrophysiological data—A practical approach. *J. Neurosci. Methods* **2015**, *250*, 34–46. [CrossRef]
40. Hillyard, S.A.; Hink, R.F.; Schwent, V.L.; Picton, T.W. Electrical Signs of Selective Attention in the Human Brain. *Science* **1973**, *182*, 177–180. [CrossRef]
41. Donchin, E.; Heffley, E.; Hillyard, S.A.; Loveless, N.; Maltzman, I.; Ohman, A.; Rosler, F.; Ruchkin, D.; Siddle, D. Cognition and event-related potentials. II. The orienting reflex and P300. *Ann. N. Y. Acad. Sci.* **1984**, *425*, 39–57. [CrossRef]
42. Gray, J.A.; Wedderburn, A.A.I. Shorter articles and notes grouping strategies with simultaneous stimuli. *Q. J. Exp. Psychol.* **1960**, *12*, 180–184. [CrossRef]
43. Deutsch, J.A.; Deutsch, D. Attention: Some theoretical considerations. *Psychol. Rev.* **1963**, *70*, 51–60. [CrossRef]
44. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [CrossRef]
45. Woldorff, M.G.; Hillyard, S.A. Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.* **1991**, *79*, 170–191. [CrossRef]
46. Woldorff, M.G.; Gallen, C.C.; Hampson, S.A.; Hillyard, S.A.; Pantev, C.; Sobel, D.; Bloom, F.E. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8722–8726. [CrossRef]
47. Näätänen, R.; Gaillard, A.; Mäntysalo, S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* **1978**, *42*, 313–329. [CrossRef]
48. Falkenstein, M.; Hohnsbein, J.; Hoormann, J.; Blanke, L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr. Clin. Neurophysiol.* **1991**, *78*, 447–455. [CrossRef]

49. Large, E.W.; Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* **1999**, *106*, 119–159. [CrossRef]
50. Treisman, A.M. Contextual cues in selective listening. *Q. J. Exp. Psychol.* **1960**, *12*, 242–248. [CrossRef]
51. O'Connell, R.G.; Dockree, P.M.; Kelly, S.P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **2012**, *15*, 1729–1735. [CrossRef]
52. Van Vugt, M.K.; Beulen, M.A.; Taatgen, N.A. Relation between centro-parietal positivity and diffusion model parameters in both perceptual and memory-based decision making. *Brain Res.* **2019**, *1715*, 1–12. [CrossRef] [PubMed]
53. Brouwer, H.; Hoeks, J.C.J. A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Front. Hum. Neurosci.* **2013**, *7*, 758. [CrossRef] [PubMed]
54. Brouwer, H.; Crocker, M.W. On the Proper Treatment of the N400 and P600 in Language Comprehension. *Front. Psychol.* **2017**, *8*, 1327. [CrossRef] [PubMed]
55. Brouwer, H.; Crocker, M.W.; Venhuizen, N.J.; Hoeks, J.C.J. A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cogn. Sci.* **2017**, *41* (Suppl. 6), 1318–1352. [CrossRef]
56. Bianco, R.; Novembre, G.; Keller, P.E.; Scharf, F.; Friederici, A.D.; Villringer, A.; Sammler, D. Syntax in Action Has Priority over Movement Selection in Piano Playing: An ERP Study. *J. Cogn. Neurosci.* **2016**, *28*, 41–54. [CrossRef] [PubMed]
57. Hagoort, P. Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *J. Cogn. Neurosci.* **2003**, *15*, 883–899. [CrossRef]
58. Koelsch, S.; Kasper, E.; Sammler, D.; Schulze, K.; Gunter, T.; Friederici, A.D. Music, language and meaning: Brain signatures of semantic processing. *Nat. Neurosci.* **2004**, *7*, 302–307. [CrossRef]

brain sciences

MDPI

Article

Early Influence of Musical Abilities and Working Memory on Speech Imitation Abilities: Study with Pre-School Children

Markus Christiner [1,*,†] and Susanne Maria Reiterer [2]

[1] Department of Linguistics, Unit for Language Learning and Teaching Research, University of Vienna, 1090 Vienna, Austria
[2] Centre for Teacher Education, Unit for Language Learning and Teaching Research, University of Vienna, 1090 Vienna, Austria; susanne.reiterer@univie.ac.at
* Correspondence: markus.christiner@univie.ac.at
† Recipient of a DOC-team-fellowship of the Austrian Academy of Sciences.

Received: 25 May 2018; Accepted: 29 August 2018; Published: 1 September 2018

Abstract: Musical aptitude and language talent are highly intertwined when it comes to phonetic language ability. Research on pre-school children's musical abilities and foreign language abilities are rare but give further insights into the relationship between language and musical aptitude. We tested pre-school children's abilities to imitate unknown languages, to remember strings of digits, to sing, to discriminate musical statements and their intrinsic (spontaneous) singing behavior ("singing-lovers versus singing nerds"). The findings revealed that having an ear for music is linked to phonetic language abilities. The results of this investigation show that a working memory capacity and phonetic aptitude are linked to high musical perception and production ability already at around the age of 5. This suggests that music and (foreign) language learning capacity may be linked from childhood on. Furthermore, the findings put emphasis on the possibility that early developed abilities may be responsible for individual differences in both linguistic and musical performances.

Keywords: phonetic language aptitude; intrinsic singing; singing ability; musical aptitude; working memory

1. Introduction

Musical abilities and the link to language functions have gained considerable scientific interest in the past decade. Music and language are highly intertwined, but despite their similarity remarkably different in many respects. Music, song, and language are all to a large degree acoustic and sensory-motor phenomena, perceived and executed similarly, which might be one of the reasons why investigations have started to compare the three faculties intensively [1–4]. In language research, understanding positive transfer effects from music to language, which might be induced by musical input/training or may stem from enhanced musical abilities/aptitude, has been of remarkable interest [5–10]. Interdisciplinary research comparing and trying to account for the differences and commonalities between music, song and language functions ranges from brain to behavioral and evolutionary to ethological research [2,11–14]. Comparing musical abilities with language functions often focuses on testing foreign language learning rather than on first language acquisition [15–17]. This allows us to observe individual differences more effectively especially when it comes to the link between music and foreign language learning by analyzing the acoustic levels of speech, such as phonetics and pronunciation. New language material, which is unfamiliar to and imitated by the participants informs about individual differences in pronunciation performances illustrating how fast and accurately individuals can adapt to new languages which they have not been exposed to [18–20].

Foreign language learning capacity is also influenced by musical training and musicians seem to detect speech incongruities much faster and more accurately than non-musicians [21]. Furthermore, musical training partly influences novel speech processing and learning [7–9,22].

Even though language experts have increasingly provided more evidence for individual differences among native speakers' language proficiency [23,24], inter-learner variation in phonetic abilities is more difficult to observe within the mother tongue compared to other domains like grammar or vocabulary knowledge. In this research we use the term "phonetic language abilities", "phonetic aptitude" or "speech imitation ability", interchangeably and what we mean is the capacity to imitate, mimic and pronounce spoken speech based on holistic judgments of human native speaker raters, judging imitated prosody as well as phonetic (segmental) aspects [18]. Referring to aptitude as a more stable "trait" demands developing tasks which are untrained or testing requires minimizing educational/training/experiential influence. This is achieved best by choosing participants who lack experience in what they are tested for. In language aptitude research, pre-school children tested in foreign language capacity and musical abilities are ideal participants because they fulfill both above mentioned criteria. Interdisciplinary research on language aptitude and musical abilities has mostly focused on adults and pupils [3,4,15,16]. Research on pre-school children has largely been neglected so far, despite the fact, that the latter is informative in terms of (phonetic and musical) aptitude, since younger children are still less influenced by education and training (environmental/social influences). Education might be one of the important driving forces supporting children's progress in cognitive abilities, linguistic development and musical abilities which in turn are related to the social environment in which children grow up. The input children receive correlates to some degree with the output produced [25,26] suggesting that the less formal educational input children receive, the more other factors than training might impact on their performance in foreign languages and music. Even though individual differences will also depend on the input given by caretakers and parents, pre-school children may be most naïve in terms of educational influence compared to older participants and rely on their aptitude while solving problems or learning new skills. In psychology aptitude has been described as a raw material allowing individuals acquire new abilities or adapting behavior faster and more accurately than their peers [27,28]. Aptitude is often considered to be a domain-specific skill and individuals with particular aptitudes show genuine, exceptional and outstanding abilities compared to the general population [29]. This suggests that people who demonstrate certain aptitudes might have at least the potential for outperforming their less talented peers. Research on aptitude is diverse ranging from giftedness with sports, playing chess, composition or writing, to name but a few. Language aptitude has been studied behaviorally [3,4,15,16,28] and less intensively neuro-physiologically, but recent research accumulates knowledge about individual differences in brain structure and function on different levels of linguistic expertise for phonetics and grammar learning [30–34]. More general studies on the bilingual brain investigating polyglots, multi-linguals and bilinguals have revealed evidence of individual differences in language learning and brain processing based on working memory capacity, intelligence, different musical abilities, language exposure and age of acquisition [3,4,18–20,35,36]. Similar to language aptitude studies well-known cases of exceptional musicians such as Mozart or Bach do not leave any doubt about the role of aptitude in individual differences in abilities. First evidence that aptitude may also be gene-related and contributes to individual differences in either language or musical abilities has been provided recently. Genetic differences in the auditory pathway have been found to be responsible for differences in music perception [37]. Another longitudinal study has detected that the basic forms and shapes of Heschl's gyri, which are seen as markers for high musical aptitude in brain research, do not change over time [5,38] and differences in brain structure of musicians have been reported in multiple investigations [39–41]. Musicians' auditory stimulation evokes enhanced activity in a number of originally non-auditory regions in musicians' brains, such as the sensorimotor, the parietal, the dorsolateral prefrontal cortex, as well as pre-motorand supplementary motor areas [42–48]. Furthermore, musical training induces plastic changes and influences the complexity in white matter

architecture of the cortico-spinal tract [49] and the arcuate fasciculus [50]. Generally speaking, individual differences in musical abilities are said to be based on both nature and nurture related influence. The earlier infants find themselves in a music-rich environment, the better their musical abilities may develop [51]. Regarding nurture related effects it has been reported that musical training during childhood has a significant effect on motor and auditory skills and may lead to structural brain differences in a relatively short period of time [52]. Music training can explain structural brain differences in adult musicians [52] but also directly improves speech segmentation [7], duration perception [6] and pitch perception ability in children [9]. This suggests that both linguistic and musical skills are based on shared neural mechanisms [2,17]. In this interdisciplinary context, there is a rarely mentioned analogy to language acquisition processes. Infants can undoubtedly learn virtually all languages and, for instance, growing up bilingually leads to similar language skills in two languages which they learn without difficulty as a matter of environmental contact [7]. It therefore is an accepted notion that the earlier someone is exposed to a language, the better the achievement [53,54]. Exactly the same seems to be the case for music acquisition processes, even though most investigations that compared music and language focused on different analyses and did not directly compare the acquisition processes per se.

Music and speech are recognized as separate capacities but are perceived by the same auditory system requiring similar cognitive skills [1,2,5]. Achievement in musical abilities improves working memory (WM) capacity, which is again important for multiple cognitive abilities (being neither language- nor music-specific), found to be trainable with transfer effects to general intelligence, executive control and problem solving [55]. WM capacity and its link to musical aptitude has been observed among adults and children [56] and suggestions that verbal and tonal processing and execution show large overlaps (plus subtle differences as a function of musical expertise) have already been shown in a series of neurocognitive investigations e.g., [57–60]. In addition, research on WM training programs has noted remarkable improvements after training sessions with children suffering from attention deficits, not only in what was trained, but also in new unrelated tasks [61]. WM ability is age-related showing that 3 digits of strings of numbers in a forward order are recalled at around five years of age [62]. Following language aptitude research, it has often been argued that WM has some potential to replace the idea of aptitude and indeed many investigations have been able to detect that WM capacity is related to processing, retaining, and repeating unfamiliar language material [18–20,63], placing WM amongst one of the strongest predictors of linguistic success. As already mentioned, WM is age-related, following developmental steps from simpler to more complex. Likewise, this could be similar when learning a new language. In previous research on adults it has been illustrated that 9 to 11 syllable long language material allows us to observe individual differences [18,19]. For pre-school children it can be suggested that 5 to 6 syllable long unfamiliar language material might be appropriate to test their phonetic aptitude.

Apart from WM overlaps of music and speech, song represents a transitional or hybrid faculty, which comprises both linguistic and musical features. Studies focusing on singing capacity and language functions are still underrepresented in recent literature. Some investigations focused on comparing language learning and singing ability [18] and singing as a learning tool [4,64]. Research has also demonstrated the effects of vocal long-term training [50] and found an improved connectivity between the kinesthetic and auditory feedback system and the anterior insular cortex [65], which also contributes to voice motor/somatosensory control and expertise in singers [66]. Furthermore, structural adaptations in singers lead to changes in the complexity and volume of white matter tracts [50].

Comparisons between speech and song [67] concluded that vocalization of speech and song largely shares the same neural network and bilateral activation in the superior temporal sulcus, the inferior pre- and post-central gyrus and the superior temporal gyrus. Speaking and singing draw on common grounds, as body posture, emission, resonance or articulation are based on the same principles [68]. Singing compared to speech is slower in production and trains the motor ability and

the vocal apparatus. This is one fundamental reason why singing (intoned word production) is often used for therapeutic purposes to regain motor ability to vocalize and indeed children's language progress develops alongside motor control [69,70].

This investigation focuses on aptitude for acquiring phonetic patterns of unfamiliar languages and its relationship to musical abilities. We sought to uncover the link between phonetic aptitude, singing and musical abilities in pre-school children to better understand music and language acquisition processes from a developmental perspective. We hypothesized that if pre-school children already performed differently in music, singing and phonetic language tasks, like adults do when tested behaviorally, it would be evidence that language or speech imitation aptitude is either developed very early or at least a very stable trait. This could open new discussions and accumulate evidence of the distribution of language aptitude within the general population and thus suggest how aptitude and individual differences can be detected, used and integrated in learning settings to support language acquisition processes. Understanding learners' needs and individual differences in aptitude may eventually change educational programs which could improve language learning as well as positively affect other related cognitive abilities. 35 pre-school children were tested for their musical abilities (music perception and singing), their ability to imitate Turkish, Tagalog, Russian, and Chinese (phonetic aptitude, speech imitation), their WM (digit span) and social- environmental variables, such as the influence of caretakers and caretakers' musical activities. This was done with a view to ascertain whether foreign language and musical abilities of pre-school children were comparable to what had been found in adults.

2. Methods

2.1. Participants

In this investigation we selected 35 (16 female and 19 male) pre-school children at the age of 5 to 6 (mean age = 5.66 and *SD* = 0.48). All of them visited a private kindergarten, were German monolinguals and naïve in formal language and musical training, apart from counting numbers in English and simple singing activities like Happy Birthday. None of them grew up bilingually or participated in a specific language program. A questionnaire revealed that neither the participants nor the parents had had contact to the language material (stimuli) which was tested. The parents belonged to a higher socio-economic background and gave informed consent and agreed to the participation of their children in this investigation. The children were also orally asked whether they liked to participate in the study and all happily agreed because they had already been familiarized with the experimenter on several music teaching occasions. They were instructed to stop at any time, if they felt uncomfortable or if they wished to withdraw their consent. The testing frame took place within two weeks and all tests were performed separately at different times within this time window. The experimenter was well integrated into the kindergarten and started work half a year before their testing to make sure that the children knew him well. For analyzing the background information of the children and the parents, a questionnaire had been designed which focused on musical profiling, singing behavior and language contact.

2.2. Speech Imitation

For testing the children's ability to remember and repeat unfamiliar language material we selected threephrases for each of the 4 different languages which had been taken (Turkish, Tagalog, Russian, and Chinese). The language material was fiveand sixsyllables long. The original phrases had been spoken by native speakers and recorded in a sound proof room. The children were tested in a separate room in the kindergarten where they had to listen to the phrases threetimes before they repeated the language stimuli, which were recorded and rated by sixnative speakers for Tagalog and Turkish, by fourraters for Russian and sevenraters for Chinese. Ratings were performed on a scale between 0 and 10, where 10 was the highest and 0 the lowest score. Native speakers are said to make judgments comparable to

those of phonetic experts [53,54] and multiple investigations used the same methodology to analyze individual differences in phonetic abilities [18–20,63]. We instructed the raters to immediately rate the overall performance (spontaneous global judgment) as well as to use headphones while rating the files.

The analysis of the participants' singing ability was based on two criteria. First of all, one was how well the children sang according to fourmusic teachers who regularly visited the kindergarten and analyzed the children's singing ability. They had to rate their singing ability on a scale between 0 and 10, where 0 was the lowest and 10 the highest number. This measurement focused on accuracy, intonation, timing and how well the children sang. For the second criterion the same scale was in use for the ratings. The kindergarten teachers were instructed to observe how intuitively the children started singing without having been instructed to do so (intrinsic motivation) over the period of 14 days. This aimed at isolating the children's inner needs and intrinsic motivation to sing which should reveal whether those who sing without being instructed may perform differently in language and musicality tasks compared to those who show less motivation to sing. Additionally, the parents were asked to estimate how many hours their children were singing during the week as well as to indicate how many hours they were singing with their children and playing a musical instrument.

2.3. Music Perception

The music perception abilities of the children were tested by employing the (Primary Measures of Music Audiation) PMMA [71], a test still widely used in research to measure musicality. This test measures the ability to discriminate tonal and rhythmical changes of paired musical statements and has been designed for children from kindergarten to third grade. The test is subdivided into two sections. While the first one analyzes children's ability to detect tonal changes, the second one focuses on their ability to discriminate rhythmical changes. Even though this test is widely used for measuring musical aptitude, there are some limitations regarding the validity of the test. For instance, studies reported inconsistent results for the two subtests which show deviations from the published norms [72,73]. Another investigation noted that especially the internal reliability of the rhythm subtest should be treated with caution for grade 1 students and kindergarten children [73].

2.4. Working Memory

For testing the working memory abilities of the pre-school children we used strings of numbers/digit span [74]. The numbers were recorded and the children had to listen to the numbers and repeat them in the same chronological order. As a familiarization task, two numbers were given in a string for testing whether the children understood their task. The strings of numbers increased in length and the test stopped after the children could not accurately repeat the strings of numbers a second time.

2.5. Ethical Approval

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the University Hospital and the Faculty of Medicine Tübingen, Project identification code 529/2009BO2.

3. Results

3.1. Descriptives and Correlations

For illustration of the relationships between the individual variables, tables are shown in the following sections. Table 1 contains the descriptive of the variables under consideration. The units are the actual scores of the variables measured. Table 2 shows the correlations of the variables.

Table 1. The descriptive of the variables under consideration.

Descriptive				
Variables	***M***	***SD***	***Min***	***Max***
Working memory forward	3.71	1.20	2.00	8.00
PMMA total	52.94	6.93	38.00	65.00
PMMA tonal	28.34	4.46	18.00	36.00
PMMA rhythm	24.60	3.89	17.00	34.00
Singing ability	6.53	2.30	0.25	10.00
Singing behavior	5.35	2.39	0.25	10.00
Speech imitation	2.99	1.07	0.87	5.80
Tagalog Mean	3.84	1.15	1.06	7.17
Chinese Mean	2.77	0.84	1.09	4.23
Turkish Mean	3.01	1.76	0.33	7.06
Russian Mean	2.32	1.33	0.17	4.92

PMMA: Primary Measures of Music Audiation.

Table 2. Correlations between the individual variables.

Correlations (Spearman)											
Variables	**PMMA Total**	**PMMA Tonal**	**PMMA Rhythm**	**Working Memory F**	**Singing Ability**	**Singing Behavior**	**Speech Imitation**	**Tagalog Mean**	**Chinese Mean**	**Turkish Mean**	**Russian Mean**
PMMA total	1	0.84 **†	0.77 **†	0.53 **†	0.44 **†	0.32	0.44 **†	0.45 **†	0.40 *†	0.39 *†	0.37 *
PMMA tonal	0.84 **†	1	0.35 *	0.58 **†	0.39 *†	0.23	0.38 *†	0.40 *†	0.42 *†	0.34 *	0.28
PMMA rhythm	0.77 **†	0.35 *	1	0.26	0.34 *	0.36 *	0.31	0.35 *	0.19	0.28	0.29
Working memory F	0.53 **†	0.58 **†	0.26	1	0.32	0.29	0.56 **†	0.42 *†	0.41 *†	0.58 **†	0.38 *†
Singing ability	0.44 **†	0.39 *†	0.34 *	0.32	1	0.80 **†	0.25	0.44 **†	−0.04	0.21	0.23
Singing behavior	0.32	0.23	0.36 *	0.29	0.80 **†	1	0.39 *†	0.53 **†	−0.09	0.36 *	0.38 *†
Speech imitation	0.44 **†	0.38 *†	0.31	0.56 **†	0.25	0.39 *†	1	0.69 **†	0.61 **†	0.90 **†	0.89 **†
Tagalog Mean	0.45 **†	0.40 *†	0.35 *	0.42 *†	0.44 **†	0.53 **†	0.69 **†	1	0.23	0.57 **†	0.53 **†
Chinese Mean	0.40 *†	0.42 *†	0.19	0.41 *†	−0.04	−0.09	0.61 **†	0.23	1	0.56 **†	0.47 **†
Turkish Mean	0.39 *†	0.34 *	0.28	0.58 **†	0.21	0.36 *	0.90 **†	0.57 **†	0.56 **†	1	0.72 **†
Russian Mean	0.37 *	0.28	0.29	0.38 *†	0.23	0.38 *†	0.89 **†	0.53 **†	0.47 **†	0.72 **†	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed). † Correlation is significant after Benjamini-Hochberg correction for overall false discovery rate $p \leq 0.05$. PMMA: Primary Measures of Music Audiation.

3.2. Musicality Test PMMA

The PMMA total score was significantly correlated with the working memory test (strings of numbers which were repeated in forward order), $r_s = 0.53$, p (two-tailed) < 0.01, and the singing parameters singing ability, $r_s = 0.44$, p (two-tailed) < 0.01. The PMMA total score was also significantly correlated with all language imitation tasks, Tagalog ($r_s = 0.45$, p (two-tailed) < 0.01), Chinese ($r_s = 0.40$, p (two-tailed) < 0.05), Turkish ($r_s = 0.39$, p (two-tailed) < 0.05), Russian ($r_s = 0.37$, p (two-tailed) < 0.05) and with the overall speech imitation ability which comprisesof all language imitation tasks ($r_s = 0.44$, p (two-tailed) < 0.01).

3.3. Speech Imitation (ComprisesTagalog, Chinese, Turkish and Russian)

Speech imitation showed a significant correlation with the PMMA total score, $r_s = 0.44$, p (two-tailed) < 0.01, with the tonal subtest, $r_s = 0.38$, p (two-tailed) < 0.05. Speech imitation also was significantly related to how accurately the children were repeating strings of numbers in forward order (working memory), $r_s = 0.56$, p (two-tailed) < 0.01 and the singing parameter "singing behavior" which revealed how intuitively the children sang without being instructed to sing, $r_s = 0.39$, p (two-tailed) < 0.05.

3.4. Working Memory

The working memory ability correlated with the musicality test PMMA, the total score, $r_s = 0.53$, p (two-tailed) < 0.01, the tonal subtest, $r_s = 0.58$, p (two-tailed) < 0.01. Further correlations of the working memory capacity to all language imitation tasks, Tagalog ($r_s = 0.42$, p (two-tailed) < 0.05), Chinese ($r_s = 0.41$, p (two-tailed) < 0.05), Turkish ($r_s = 0.58$, p (two-tailed) < 0.01), Russian ($r_s = 0.38$, p (two-tailed) < 0.05) and with the overall speech imitation ability ($r_s = 0.56$, p (two-tailed) < 0.01) were also observed.

3.5. Singing Ability

Singing ability, which measures how accurately the children sang, was correlated with the PMMA total score ($r_s = 0.44$, p (two-tailed) < 0.01), the tonal PMMA subtest ($r_s = 0.39$, p (two-tailed) < 0.05) and the rhythm PMMA subtest ($r_s = 0.34$, p (two-tailed) < 0.05). Singing ability also showed a significant relationship to singing behavior ($r_s = 0.80$, p (two-tailed) < 0.01) and singing ability also correlated with the language imitation task Tagalog ($r_s = 0.44$, p (two-tailed) < 0.01).

3.6. Singing Behavior

Singing behavior which should reveal the intrinsic motivation of the children to sing, correlated with the rhythm PMMA subtest ($r_s = 0.36$, p (two-tailed) < 0.05) and singing ability, $r_s = 0.80$, p (two-tailed) < 0.01 and showed a significant relationship to Tagalog $r_s = 0.49$, p (two-tailed) < 0.01, and to the global speech imitation ability, $r_s = 0.36$, p (two-tailed) < 0.05 as well.

3.7. Inter-Rater Reliability

The inter-rater reliability Cronbach's alpha was applied and was 0.88 for Tagalog, 0.85 for Chinese, 0.94 for Turkish and 0.86 for Russian which are all in the acceptable range above 0.70.

3.8. Whitney–Mann Test (Group Comparisons)

We divided our group into high and low musical aptitude based on the music perception task PMMA (tonal and rhythmic discrimination ability). The high musical aptitude group (*Median* = 4) performed significantly better than the low musical aptitude group (*Median* = 3) in the working memory task, U = 51.00, z = −3.58, $p < 0.001$, r = −0.61. The high musical aptitude group (*Median* = 3.25) also performed significantly better than the low musical aptitude group (*Median* = 2.75) in speech imitation, U = 83.00, z = −2.28, $p < 0.005$, r = −0.39. The high musical aptitude group

(*Median* = 7.88) performed significantly better than the low musical aptitude group (*Median* = 4) in singing (singing ability), U = 91.50, z = −2.01, $p < 0.005$, r = −0.34. The significance of all three group comparisons was also given after Bonferroni–Holm-Correction. Significance was inferred at $p < 0.05$ after Bonferroni–Holm-Correction for the three variables working memory, singing ability and speech imitation. For illustration see Figure 1 below.

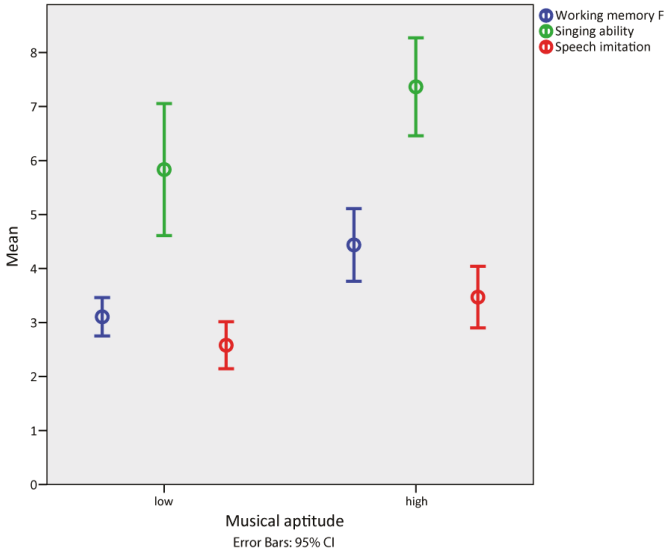


Figure 1. Working memory, singing ability and speech imitation was higher in children with high (compared to low) musical aptitude.

4. Discussion

The music perception task (PMMA) was used to create group membership of the children and consists of rhythm and tonal discrimination tasks. Even though this test is widely used for measuring musical aptitude, there are some limitations concerning the validity of the test. According to Stamouet et al. [73], the results of the rhythm subtest should be regarded with caution for pre-school children and grade 1 students as a result of cross-cultural issues which may also be relevant for this investigation as the children were German native speakers. However, in this study similarly to research on adults, effects of individual differences in music perception, working memory capacity, speech imitation and singing have been found. For statistical analysis we split the groups based on music perception abilities, measured by the PMMA and we created two groups of high and low musical aptitude. Working memory, speech imitation and singing ability have been found to be significantly different between the two musical groups created, while singing behavior has failed to reach statistical significance within the model, even though there was a tendency.

4.1. Musical Expertise, Plasticity, Musical Abilities and Working Memory

Several investigations have found that music perception abilities improve the ability to remember, imitate and retrieve unfamiliar language material [5,15]. As shown in this investigation, the same seems to hold true for pre-school children: The higher their music perception ability, the better their speech imitation ability to memorize and imitate new, unfamiliar language material. This puts emphasis on the importance of addressing the relationship between musical and linguistic abilities. The building of expertise in the areas of linguistic or musical abilities is, in essence, the recurrent problem of the

relationship between "nature and nurture", difficult to investigate experimentally, given a lack of testable definitions of the "talent-ability" terms. There is considerable debate as to whether differences in behavior are due either to "innate talent" or to the quantity and quality of practice in a given domain. Indeed, the viewpoint of interaction between genetically and epi-genetically driven abilities modified by experience is perhaps dependent on the skill domain, but in general widely accepted in the domains of language or music acquisition.

Musicians' neurophysiological, auditory enhancements and beneficial transfer effects on language functions are often related to the years of training [50,52,75–78], and ontogenetically speaking, discrimination of rhythm is an early developed mechanism that starts prenatally and continues during infancy [77]. The effect of musical training and the impact of culture on music acquisition is undeniable, even though individual differences in high achievement may be based on other aspects as well. In this study the children were naïve in terms of individual musical training which might suggest that other factors than proper music training, such as very early developmental or pre-, peri- or post-natal influencescontribute to individual differences in their musical abilities. Auditory models have already proposed that primary capacities influence musical aptitude, while secondary musical skills are environmentally shaped by the culture and individual training received [79]. First evidence that primary capacities for musical aptitude may also be gene-related, like processing of auditory signals, which alter the auditory pathway crucial for discriminating musical input, has already been reported [37]. Multigenerational family studies have also revealed that several predisposing genes or variants contribute to musical aptitude e.g., [80] and evidence for alterations of the brain structure of musicians, which improve music perception and performance, are diverse e.g., [5,20,40,49,51]. Inter-individual changes and structural differences in the auditory cortex cannot be ascribed to training effects only and particular brain areas, seen as markers for high musical abilities (e.g., Heschl's gyrus) seem to be rather stable in shape [5,38,40,41]. For instance, duplications of Heschl's gyrus occur more often in musicians rather than in non-musicians [38]. Anatomical alterations of the gray matter [33,34] or volume and complexity differences in white matter tracts of singers [50] have been reported. While increasingly more evidence manifests that musical aptitude could be gene-related or at least a stable trait over life-time, studies that identify biological markers, such as genetic or neuroanatomical markers for phonetic aptitude, have been largely neglected so far within the area of second language acquisition or language learning research.

The underlying reasons for the link between musical and phonetic aptitude are also based on shared cognitive functions as well as on shared mechanisms in the execution and processing of music and speech [5–10]. Music and language functions require the recruitment of similar cognitive processes, attention control, anatomical and neuroanatomical endowment [5]. WM capacity, an elaborate cognitive skill, crucial for multiple abilities, is related to phonetic aptitude and musical abilities. The ability to remember rapid and temporary information is important for the learning of new language material which is poor in linguistic content [81]. There is an analogy to remembering melodies or discriminating different musical pieces. Language learners in the beginning phase will benefitmost from higher WM capacity. The basic acoustic signals of musical sounds (pitch, timing and timbre) play a key role in both speech and music, especially in conveying information [82]. Pitch, the property of sounds that is organized by a scale, can be judged as lower or higher. Timing refers to temporal events of sounds and timbre to the perceived sound quality also referred to as tone quality [82]. Evidence has been provided that the improved processing of timbre and pitch in musicians is based on functional adaptations, but seems to play a general role in musical development from infancy onwards [83]. Furthermore, musicians' improvements in detecting pitch and timbre cues largely rely on plastic adaptations [49,82,84]. Language acquisition processes are also based on differentiating timbres of speech that are meaningful and/or meaningless [84] and for musicians it is also necessary to discriminate timbre differences between various instruments which in turn seem to have an effect on speech sound discrimination as well [82]. Evidence for an overlap between processing tonal and verbal material, especially comes from brain research e.g., [51,53]. Higher WM abilities are also associated

with musical aptitude [18], but are also age related [85]. Children (aged 4 to 6) can remember around twowords in a forward order, and threedigits of strings of numbers in a forward order [62]. The results of this as well as earlier research corroborate the findings of previous investigations and suggest that WM is highly important for learning new languages. Thus, it can be assumed that people with high WM capacity are faster learners [86], showing that WM predicts not only overall language aptitude scores, but is related to the language analysis components of the aptitude construct. Therefore, the link between WM, language acquisition and musical aptitude is a very promising research area and the overlaps of WM capacity for music and language functions may lead or has led researchers to argue that WM has the potential to replace the whole language aptitude construct (equaling language aptitude with working memory capacity). This claim, however, should be treated with caution, since one limitation of direct imitation tasks, like used in this investigation, is that it always requires high working memory loads. Indirect imitation tasks retrieved from long-term memory may reduce WM influence and inform about phonemic awareness. Future research on phonetic aptitude should include both direct and indirect imitation tasks to get a multidimensional impression of phonetic language aptitude. The measurements used in this investigation provided evidence on how fast and accurately someone imitates, retrieves and memorizes unfamiliar speech material on an "ad-hoc" basis, while indirect imitation tasks would also yield information about achievements in meta-cognitive awareness (phonemic awareness).

4.2. Singing Behavior, Singing Ability and Speech Imitation

Singing behavior was not significantly different in the music groups created which is in line with previous research where music perception ability of singers and instrumentalists did not differ, while their ability to reproduce new language material was significantly better in the singer group [19]. The result of this investigation, however, could also be due to the fact that pre-school children do not have comprehensive musical skills. Even though sensory consonance and pitch discrimination ability develops relatively early during infancy, harmonic knowledge, for instance, develops significantly later between 6 and 12 years [87]. Furthermore, limitations of the study design must be mentioned here as well. The ratings of the singing behavior were rated based on mere observations by the caretakers in the kindergarten. Our reasons for choosing this design lay in keeping testing time appropriately short for pre-school children and in rendering the testing situation as natural as possible.

Individual differences in the performances of children were also based on intrinsic motivation to produce vocalizations, which might be driven by their inner needs to express their feelings to the outer world. The variable singing behavior correlated more with to the global speech imitation score than mere singing ability or singing accuracy. The reason for this might be that children's fine motor abilities are still under development and this affects both, the singing and the articulatory skills [3,4]. First language acquisition develops along motor control development [88] and singing as vocal behavior largely shares the same mechanisms. Evidence that vocal motor commands can alter and influence speech perception has already been provided e.g., [89]. Intrinsic singing behavior and the desire to sing, reflecting intrinsic motivation to sing, are maybe more important than accuracy to playfully expand vocal flexibility in this period of life. However, since correlations differed only between one another in the case of the single contributions of singing ability and singing behavior, these results should be interpreted with caution.

4.3. Typologically Different Languages and Musical Abilities

Tone languages and non-tone languages are different in many respects and even though causality cannot be explained, based on the correlations, there seems to be a tendency towards differences in non-tone languages and tone-languages in relation to singing. Future investigations may need to consider typologically different languages and their relationship to musical aspects which cannot be explained within this limited research design. Recent research has isolated different transfer effects from music to language, where pitch discrimination contributed to tone-languages, while rhythmic

discrimination contributed to non-tone languages [63]. Goswami and colleagues also showed that "novel remediation strategies on the basis of rhythm and music may offer benefits for phonological and linguistic development" [90] and early musical training during childhood supports foreign language perception, memory and later foreign language acquisition processes [91]. As a general rule, it has been accepted that during infancy, basically between six and twelve months, language specific phonetic contrasts are perceived, followed by a decline in the ability to perceive non-native contrasts [92]. The first twoyears of development, therefore, seem to leave a deep cognitive imprint on children's language performances also later in life [93]. Twelve months old infants' speech segmentation and speech processing abilities seem to be predictive measurements for observing individual differences in language development between fourand 6 years of age [94]. The children in this investigation were five-to six-year-olds and had already acquired native phonetic contrasts for the German language, a non-tone language. Singing ability and singing behavior did not seem to relate to the Chinese performances. Chinese only correlated with the PMMA total and tonal score and working memory capacity, but did not show correlations to the rhythm PMMA subtest, or to any of the singing criteria. Although correlations do not inform about causality, further research on tone language imitation and its relation to musical measurements should be investigated in more detail. This could include various musical measurements focusing on pitch, timing and timbre perception tasks, musical instrument playing or singing to better understand the impact of musical abilities on second language learning of non-tone language speakers.

An analysis of Chinese imitation performances of school children at the age of 9 have revealed that around 40% of the variance of Chinese imitation of non-tone (German) language native speakers could be explained by singing ability, tonal perception ability together with WM capacity [63]. Learning Chinese as a second language may require precise musical knowledge an ability which is developed at the age of 9 [95]. Research has shown that tone language speakers are better at discriminating tone contrasts in any language compared to non-tone language speakers [96] and evidence that musicians and tone-language speakers share cognitive and perceptual skills for pitch discrimination has also been reported recently [97]. Musicians and tone-language speakers largely share cognitive and perceptual skills for pitch discrimination which illustrates a bidirectional path between music and language [97]. Evidence has been provided that high melodic ability of non-tone language speakers leads to better performances in detecting tonal variations in Mandarin [98]. This shows that musical ability may be highly relevant for learning tone-languages as a non-tone language speaker. This may be the reason as to why Chinese native speakers who learn a non-tone language during adulthood face pronunciation difficulties, and vice versa, non-tone language speakers have difficulties to discriminate Chinese sounds. Singing should be more closely integrated in future research designs to isolate the influence of singing on the acquisition of tone languages as a second language of non-tone language speakers.

5. Conclusions

Musical ear, singing ability/behavior and working memory capacity are linked to speech imitation abilities already at a very early stage in development. Comparable to research on adults, we have found similar effects and links in pre-school children, who are naïve in terms of education and music training, suggesting that individual differences might also be based on very early developed factors. Group comparisons of children with high versus low music perception abilities reveal, that the high musicality groups perform better in novel speech imitation, can sing better and show enhanced WM capacity. Children at this particular age are less influenced by educational input than adults, which hints at early developmental factors contributing to individual differences in musical abilities and (novel) speech learning abilities. This shows that music and language capacities are ultimately linked in children and adults. Singing behavior did not yield statistically significant differences between groups, which could show that this behavioral measure displayed lower reliability. On the other hand, it is in line with research on adults that singing ability, singing motivation and music perception are

different, non-overlapping entities, sometimes leading to similar and sometimes to different transfer effects [19].

Author Contributions: M.C.: data curation, formal analysis, investigation, methodology, project idea, project administration, software, writing—original draft, resources. S.M.R.: resources, supervision, conceptualisation, methodology, discussion, writing—reviewing and editing.

Funding: This research received no external funding.

Acknowledgments: We want to thank the study participants.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Patel, A.D. *Music, Language and the Brain*; Oxford University Press: Oxford, UK, 2008.
2. Patel, A.D. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* **2011**, *2*, 142. [CrossRef] [PubMed]
3. Fonseca-Mora, M.C.; Jara-Jiménez, P.; Gómez-Domínguez, M. Musical plus phonological input for young foreign language readers. *Front. Psychol.* **2015**, *6*, 286. [CrossRef] [PubMed]
4. Fonseca-Mora, M.C.; Toscano-Fuentes, C.; Wermke, K. Melodies that help: The relation between language aptitude and musical intelligence. *Angl. Int. J. Engl. Stud.* **2011**, *22*, 101–118.
5. Seither-Preisler, A.; Parncutt, R.; Schneider, P. Size and synchronization of auditory cortex promotes musical, literacy and attentional skills in children. *J. Neurosci.* **2014**, *34*, 10937–10949. [CrossRef] [PubMed]
6. Chobert, J.; François, C.; Velay, J.-L.; Besson, M. Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* **2012**, *24*, 956–967. [CrossRef] [PubMed]
7. François, C.; Chobert, J.; Besson, M.; Schön, D. Music training for the development of speech segmentation. *Cereb. Cortex* **2012**, *23*, 2038–2043. [CrossRef] [PubMed]
8. Besson, M.; Chobert, J.; Marie, C. Transfer of training between music and speech: Common processing, attention, and memory. *Front. Psychol.* **2011**, *2*, 94. [CrossRef] [PubMed]
9. Moreno, S.; Marques, C.; Santos, A.; Santos, M.; Castro, S.L.; Besson, M. Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb. Cortex* **2008**, *19*, 712–723. [CrossRef] [PubMed]
10. Moreno, S.; Bialystok, E.; Barac, R.; Schellenberg, E.G.; Cepeda, N.J.; Chau, T. Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* **2012**, *22*, 1425–1433. [CrossRef] [PubMed]
11. Sun, Y.; Lu, X.; Ho, H.T.; Johnson, B.W.; Sammler, D.; Thompson, W.F. Syntactic processing in music and language: Parallel abnormalities observed in congenital amusia. *Neuroimage Clin.* **2018**, *19*, 640–651. [CrossRef] [PubMed]
12. Fitch, W.T. The evolution of music in comparative perspective. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 29–49. [CrossRef] [PubMed]
13. Fitch, W.T.; Martins, M.D. Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N. Y. Acad. Sci.* **2014**, *1316*, 87–104. [CrossRef] [PubMed]
14. Theofanopoulou, C.; Boeckx, C.; Jarvis, E.D. A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning. *Proc. Biol. Sci. R. B* **2017**, *284*, 20170988. [CrossRef] [PubMed]
15. Milovanov, R. Musical aptitude and foreign language learning skills: Neural and behavioural evidence about their connections. In Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009), Jyväskylä, Finland, 12–16 August 2009; pp. 338–342.
16. Slevc, R.L.; Myake, A. Differences in second-language proficiency—Does musical ability matter? *Psychol. Sci.* **2006**, *8*, 675–681. [CrossRef] [PubMed]
17. Milovanov, R.; Huotilainen, M.; Välimäki, V.; Esquef, P.A.A.; Tervaniemi, M. Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Res.* **2008**, *1194*, 81–89. [CrossRef] [PubMed]
18. Christiner, M.; Reiterer, S.M. Song and speech: Examining the link between singing talent and speech imitation ability. *Front. Psychol.* **2013**, *4*, 874. [CrossRef] [PubMed]

19. Christiner, M.; Reiterer, S.M. A Mozart in not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Front. Hum. Neurosc.* **2015**, *9*, 482. [CrossRef] [PubMed]
20. Christiner, M.; Reiterer, S.M. Music, song and speech: A closer look at the interfaces between musicality, singing and individual differences in phonetic language aptitude. In *Cognitive Individual Differences in Second Language Processing and Acquisition*; Granena, G., Jackson, D.O., Yilmaz, Y., Eds.; John Benjamins: Amsterdam, The Netherlands, 2016; pp. 131–156.
21. Schön, D.; Magne, C.; Besson, M. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* **2004**, *41*, 341–349. [CrossRef] [PubMed]
22. Wong, P.C.M.; Perrachione, T.K. Learning pitch patterns in lexical identification by native english-speaking adults. *Appl. Psycholinguist.* **2007**, *28*, 565–585. [CrossRef]
23. Pakulak, E.; Neville, H.J. Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *J. Cogn. Neurosci.* **2010**, *22*, 2728–2744. [CrossRef] [PubMed]
24. Andringa, S.J. The use of native speaker norms in critical period hypothesis research. *Stud. Second Lang. Acqu.* **2014**, *36*, 565–596. [CrossRef]
25. Korecky-Kröll, K.; Uzunkaya-Sharma, K.; Czinglar, C.; Sommer-Lolei, S.; Yanagida, T.; Dressler, W.U. The lower the slower: Parental SES and input affect speed of development of vocabulary and morphology. In Proceedings of the Child Language Symposium at the Meeting of University of Warwick, Coventry, UK, 20–21 July 2015.
26. Korecky-Kröll, K.; Uzunkaya-Sharma, K.; Dressler, W.U. Requests in Turkish and German child-directed and child speech: Evidence from different socio-economic backgrounds. In *Social Environment and Cognition in Language Development: Studies in Honor of AyhanAksu-Koç*; Ketrez, N., Küntay, A.C., Özçalýþkan, S., Özyürek, A., Eds.; Benjamins: Amsterdam, The Netherlands, 2017; pp. 53–68.
27. Gagné, F. From gifts to talents: The DMGT as a developmental model. In *Conceptions of Giftedness*, 2nd ed.; Sternberg, R.J., Davidson, J.E., Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 98–119.
28. Wen, Z.; Biedroń, A.; Skehan, P. Foreign language aptitude theory: Yesterday, today and tomorrow. *Lang. Teach.* **2017**, *50*, 1–31. [CrossRef]
29. Vinkhuyzen, A.A.E.; Van Der Sluis, S.; Posthuma, D.; Boomsma, D.I. The heritability of aptitude and exceptional talent across different domains in adolescents and young adults. *Behav. Genet.* **2009**, *39*, 380–392. [CrossRef] [PubMed]
30. Elmer, S.; Jänggi, J.; Jäncke, L. Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: Insights from simultaneous interpreters. *Cortex* **2014**, *54*, 179–189. [CrossRef] [PubMed]
31. Vandermosten, M.; Price, C.J.; Golestani, N. Plasticity of white matter connectivity in phonetics experts. *Brain Struct. Funct.* **2016**, *221*, 3825–3833. [CrossRef] [PubMed]
32. Kepinska, O.; de Rover, M.; Caspers, J.; Schiller, N.O. On neural correlates of individual differences in novel grammar learning: An fMRI study. *Neuropsychologia* **2016**, *98*, 156–168. [CrossRef] [PubMed]
33. Vaquero, L.; Rodriguez-Fornells, A.; Reiterer, S. The left, the better: White-matter brain integrity predicts foreign language imitation ability. *Cereb. Cortex* **2017**, *27*, 3906–3917. [CrossRef] [PubMed]
34. Reiterer, S.M.; Hu, X.; Erb, M.; Rota, G.; Nardo, D.; Grodd, W.; Winkler, S.; Ackermann, H. Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Front. Psychol.* **2011**, *2*, 271. [CrossRef] [PubMed]
35. Abutalebi, J.; Cappa, S.F.; Perani, D. The bilingual brain as revealed by functional neuroimaging. *Biling. Lang. Cogn.* **2001**, *4*, 179–190. [CrossRef]
36. Perani, D.; Abutalebi, J.; Paulesu, E.; Brambati, S.; Scifo, P.; Cappa, S.F.; Fazio, F. The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Hum. Brain Mapp.* **2003**, *9*, 170–182. [CrossRef] [PubMed]
37. Oikkonen, J.; Huang, Y.; Onkamo, P.; Ukkola-Vuoti, L.; Raijas, P.; Karma, K.; Vieland, V.J. A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Mol. Psychiatry* **2015**, *20*, 275–282. [CrossRef] [PubMed]
38. Benner, J.; Wengenroth, M.; Reinhardt, J.; Stippich, C.; Schneider, P.; Blatow, M. Prevalence and function of Heschl's gyrus morphotypes in musicians. *Brain Struct. Funct.* **2017**, *222*, 3587–3603. [CrossRef] [PubMed]
39. Schneider, P.; Andermann, M.; Wengenroth, M.; Goebel, R.; Flor, H.; Rupp, A.; Diesch, E. Reduced volume of Heschl's gyrus in tinnitus. *Neuroimage* **2009**, *45*, 927–939. [CrossRef] [PubMed]

40. Schneider, P.; Scherg, M.; Dosch, G.H.; Specht, H.J.; Gutschalk, A.; Rupp, A. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* **2002**, *5*, 688–694. [CrossRef] [PubMed]
41. Schneider, P.; Sluming, V.; Roberts, N.; Bleeck, S.; Rupp, A. Structural, functional and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 387–394. [CrossRef] [PubMed]
42. Baumann, S.; Koeneke, S.; Meyer, M.; Lutz, K.; Jancke, L. A network for sensory-motor integration: What happens in the auditory cortex during piano playing without acoustic feedback? *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 186–188. [CrossRef] [PubMed]
43. Koelsch, S.; Fritz, T.; Schulze, K.; Alsop, D.; Schlaug, G. Adults and children processing music: An fMRI study. *Neuroimage* **2005**, *25*, 1068–1076. [CrossRef] [PubMed]
44. Bangert, M.; Peschel, T.; Schlaug, G.; Rotte, M.; Drescher, D.; Hinrichs, H.; Heinze, H.J.; Altenmüller, E. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage* **2006**, *30*, 917–926. [CrossRef] [PubMed]
45. Zatorre, R.J.; Chen, J.L.; Penhune, V.B. When the brain plays music: Auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* **2007**, *8*, 547–558. [CrossRef] [PubMed]
46. Altenmüller, E. Neurology of musical performance. *Clin. Med.* **2008**, *8*, 410–413. [CrossRef]
47. Chen, J.L.; Penhune, V.B.; Zatorre, R.J. Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* **2008**, *20*, 226–239. [CrossRef] [PubMed]
48. Brown, R.M.; Chen, J.L.; Hollinger, A.; Penhune, V.B.; Palmer, C.; Zatorre, R.J. Repetition suppression in auditory-motor regions to pitch and temporal structure in music. *J. Cogn. Neurosci.* **2013**, *25*, 313–328. [CrossRef] [PubMed]
49. Imfeld, A.; Oechslin, M.S.; Meyer, M.; Loenneker, T.; Jancke, L. White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *Neuroimage* **2009**, *46*, 600–607. [CrossRef] [PubMed]
50. Halwani, G.F.; Loui, P.; Rüber, T.; Schlaug, G. Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists and non-musicians. *Front. Psychol.* **2011**, *2*, 156. [CrossRef] [PubMed]
51. Gordon, E.E. *A Music Learning Theory for Newborn and Young Children*; GIA: Chicago, IL, USA, 2003.
52. Hyde, K.L.; Lerch, J.; Norton, A.; Forgeard, M.; Winner, E.; Evans, A.C.; Schlaug, G. The Effects of Musical Training on Structural Brain Development. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 182–186. [CrossRef] [PubMed]
53. Bongaerts, T. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *Second Language Acquisition and the Critical Period Hypothesis*; Birdsong, D., Ed.; Lawrence Erlbaum: Mahwah, NJ, USA, 1999; pp. 133–159.
54. Bongaerts, T.; Planken, B.; Schils, E. Can late starters attain a native accent in foreign language? A test of the critical period hypothesis. In *The Age Factor in Second Language Acquisition*; Singleton, D., Lengyel, Z., Eds.; Multilingual Matters: Clevedon, UK, 1995; pp. 30–50.
55. Jaeggi, S.M.; Buschkuehl, M.; Jonides, J.; Shah, P. Short- and longterm benefits of cognitive training. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10081–10086. [CrossRef] [PubMed]
56. Strait, D.L.; Hornickel, J.; Kraus, N. Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav. Brain Funct.* **2011**, *7*, 44. [CrossRef] [PubMed]
57. Koelsch, S.; Schulze, K.; Sammler, D.; Fritz, T.; Müller, K.; Gruber, O. Functional architecture of verbal and tonal working memory: An fMRI study. *Hum. Brain Mapp.* **2009**, *30*, 859–873. [CrossRef] [PubMed]
58. Schulze, K.; Koelsch, S. Working memory for speech and music. *Ann. N. Y. Acad. Sci.* **2012**, *1252*, 229–236. [CrossRef] [PubMed]
59. Schulze, K.; Zysset, S.; Mueller, K.; Friederici, A.D.; Koelsch, S. Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* **2011**, *32*, 771–783. [CrossRef] [PubMed]
60. Williamson, V.J.; Baddeley, A.D.; Hitch, G.J. Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Mem. Cogn.* **2010**, *38*, 163–175. [CrossRef] [PubMed]
61. Klingberg, T.; Forssberg, H.; Westerberg, H. Training of working memory in children with ADHD. *J. Clin. Exp. Neuropsychol.* **2002**, *24*, 781–791. [CrossRef] [PubMed]
62. Roman, A.S.; Pisoni, D.B.; Kronenberger, W.G. Assessment of working memory capacity in preschool children using the missing scan task. *Infant Child Dev.* **2014**, *23*, 575–587. [CrossRef] [PubMed]

63. Christiner, M.; Rüdegger, S.; Reiterer, S.M. Sing Chinese and tap Tagalog? Predicting individual differences in musical and phonetic aptitude using language families differing by sound-typology. *Int. J. Multiling.* **2018**, in press. [CrossRef]
64. Ludke, K.M.; Ferreira, F.; Overy, K. Singing can facilitate foreign language learning. *Mem. Cogn.* **2014**, *42*, 41–52. [CrossRef] [PubMed]
65. Kleber, B.; Veit, R.; Birbaumer, N.; Gruzelier, J.; Lotze, M. The brain of opera singers: Experience-dependent changes in functional activation. *Cereb. Cortex* **2010**, *20*, 1144–1152. [CrossRef] [PubMed]
66. Kleber, B.; Zeitouni, A.G.; Friberg, A.; Zatorre, R.J. Experience-dependent modulation of feedback integration during singing: Role of the right anterior insula. *J. Neurosci.* **2013**, *33*, 6070–6080. [CrossRef] [PubMed]
67. Özdemir, E.; Norton, A.; Schlaug, G. Shared and distinct neural correlates of singing and speaking. *Neuroimage* **2006**, *33*, 628–635. [CrossRef] [PubMed]
68. García-López, I.; GavilánBouzas, J. The singing voice. *Acta Otorrinolaringol.* **2010**, *61*, 441–451. [CrossRef]
69. Krishnan, S.; Alcock, K.; Carey, D.; Bergström, L.; Karmiloff-Smith, A.; Dick, F. Fractionating nonword repetition: The contributions of short-term memory and oromotor praxis are different. *PLoS ONE* **2017**, *12*, e0178356. [CrossRef] [PubMed]
70. D'Souza, D.; D'Souza, H.; Karmiloff-Smith, A. Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *J. Child Lang.* **2017**, *44*, 591–627. [CrossRef] [PubMed]
71. Gordon, E.E. *Primary Measures of Music Audiation*; GIA: Chicago, IL, USA, 2006.
72. Gouzouasis, P.; Guhn, M.; Kishor, N. The predictive relationship between achievement and participation in music and achievement in core Grade 12 academic subjects. *Music Educ. Res.* **2007**, *9*, 81–92. [CrossRef]
73. Stamou, L.; Schmidt, C.P.; Humphreys, J.T. Standardization of the Gordon Primary Measures of Music Audiation in Greece. *J. Res. Music Educ.* **2010**, *58*, 75–89. [CrossRef]
74. Wechsler, D. *The Measurement of Adult Intelligence*; Williams and Wilkins: Baltimore, MD, USA, 1939.
75. Hyde, K.L.; Lerch, J.; Norton, A.; Forgeard, M.; Winner, E.; Evans, A.C.; Schlaug, G. Musical training shapes structural brain development. *J. Neurosci.* **2009**, *29*, 3019–3025. [CrossRef] [PubMed]
76. Moreno, S.; Bidelman, G.M. Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* **2014**, *308*, 84–97. [CrossRef] [PubMed]
77. Hannon, E.E.; Trainor, L.J. Music acquisition: Effects of enculturation and formal training on development. *Trends Cogn. Sci.* **2007**, *11*, 466–472. [CrossRef] [PubMed]
78. Elmer, S.; Jäncke, L. Relationships between music training, speech processing, and word learning: A network perspective. *Ann. N. Y. Acad. Sci.* **2018**. [CrossRef] [PubMed]
79. Karma, K. Auditory and Visual Temporal Structuring: How Important is Sound to Musical Thinking? *Psychol. Music* **1994**, *22*, 20–30. [CrossRef]
80. Pulli, K.; Karma, K.; Norio, R.; Sistonen, P.; Göring, H.H.H.; Järvelä, I. Genome-wide linkage scan for loci of musical aptitude in Finnish families: Evidence for a major locus at 4q22. *J. Med. Genet.* **2008**, *45*, 451–456. [CrossRef] [PubMed]
81. Perkins, J.M.; Baran, J.A.; Gandour, J. Hemispheric specialization in processing intonation contours. *Aphasiology* **1996**, *10*, 343–362. [CrossRef]
82. Kraus, N.; Chandrasekaran, B. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **2010**, *11*, 599–605. [CrossRef] [PubMed]
83. Musacchia, G.; Sams, M.; Skoe, E.; Kraus, N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15894–15898. [CrossRef] [PubMed]
84. Trollinger, V.L. The Brain in Singing and Language. *Gen. Music Today* **2012**, *23*, 20–30. [CrossRef]
85. Loosli, S.V.; Buschkuehl, M.; Perrig, W.J.; Jaeggi, S.M. Working memory training improves reading processes in typically developing children. *Child Neuropsychol.* **2012**, *18*, 62–78. [CrossRef] [PubMed]
86. Yoshimura, Y. The role of working memory in language aptitude. In *The Past, Present, and Future of Second Language Research*; Bonch-Bruevich, X., Crawford, W.J., Hellermann, J., Higgins, C., Nguyen, H., Eds.; Cascadilla Press: Sommerville, MA, USA, 2001; pp. 144–163.
87. Trainor, L.J. Are there critical periods for musical development? *Dev. Psychobiol.* **2005**, *46*, 262–278. [CrossRef] [PubMed]
88. Iverson, J.M. Developing language in a developing body: The relationship between motor development and language development. *J. Child Lang.* **2010**, *37*, 229–261. [CrossRef] [PubMed]

89. Nasir, S.M.; Ostry, D.J. Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20470–20475. [CrossRef] [PubMed]
90. Goswami, U.; Wang, H.-L.; Cruz, A.; Fosker, T.; Mead, N.; Huss, M. Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *J. Cogn. Neurosci.* **2011**, *23*, 325–337. [CrossRef] [PubMed]
91. Dittinger, E.; Barbaroux, M.; D'Imperio, M.; Jäncke, L.; Elmer, S.; Besson, M. Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *J. Cogn. Neurosci.* **2016**, *28*, 1584–1602. [CrossRef] [PubMed]
92. Yeung, H.H.; Chen, K.H.; Werker, J.F. When does native language input affect phonetic perception? The precocious case of lexical tone. *J. Mem. Lang.* **2013**, *68*, 123–139. [CrossRef]
93. Dick, F.; Krishnan, S.; Leech, R.; Curtin, S. Language Development. In *Neurobiology of Language*; Hickok, G., Small, S., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 373–388.
94. Newman, R.; Ratner, N.B.; Jusczyk, A.M.; Jusczyk, P.W.; Dow, K.A. Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Dev. Psychol.* **2006**, *42*, 643–655. [CrossRef] [PubMed]
95. Sloboda, J.A. *Exploring the Musical Mind: Cognition, Emotion, Ability, Function*; Oxford University Press: Oxford, UK, 2005.
96. Wayland, R.P.; Guion, S.G. Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Lang. Learn.* **2004**, *54*, 681–712. [CrossRef]
97. Bidelman, G.M.; Hutka, S.; Moreno, S. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS ONE* **2013**, *8*, e60676. [CrossRef] [PubMed]
98. Delogu, F.; Lampis, G.; Olivetti Belardinelli, M. Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cogn. Process.* **2006**, *7*, 203–207. [CrossRef] [PubMed]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Brain Sciences Editorial Office
E-mail: brainsci@mdpi.com
www.mdpi.com/journal/brainsci

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-03943-127-4