

Human Recognition: The Utilization of Face, Voice, Name and Interactions—An Extended Editorial

Guido Gainotti

Institute of Neurology, Università Cattolica del Sacro Cuore, Fondazione Policlinico A. Gemelli, Istituto di Ricovero e Cura a Carattere Scientifico, 00168 Rome, Italy; guido.gainotti@unicatt.it; Tel.: +39-06-30156435

The many stimulating contributions to this Special Issue of *Brain Science* focused on some basic issues of particular interest in current research, with emphasis on human recognition using faces, voices, and names. I therefore thought that it could be useful to group these papers, after a short introduction, on the basis of the principal problem that they consider.

1. Introduction

Many of the earliest investigations into the neural substrates underlying human recognition using the face, voice, and name considered the basic problems of this research domain. In particular, emphasis was placed on the brain structures underlying these cognitive skills, the selectivity of work performed by these structures, and the main similarities and differences between these recognition modalities. These studies originally focused attention on the brain structures involved in facial recognition, showing that face-sensitive areas are located in the fusiform face area (FFA) [1,2] and the occipital face area (OFA) [3,4]. They also suggested that the occipital face area could be principally involved in the processing of visual sensory information. Conversely, they thought that the fusiform face area might be primarily implicated in facial identity recognition [4–6]. Similar observations were made when investigating the brain areas involved in voice recognition using the voice-sensitive regions found in the superior temporal gyrus/sulcus (STG/STS) [7]. In analogy with the distinction between the functions of the OFA and of the FFA, it was suggested that posterior STG/STS might be more involved in the processing of acoustic properties of voices, with the anterior STG/STS potentially thought to be more related to voice-based identity recognition [8–10]. A further similarity observed between the facial and vocal components of human recognition systems was that, even if face- and voice-selective areas were present in both hemispheres, the right-sided areas were more important for both facial [6] and vocal processing [11]. The right lateralization of these (non-verbal) human recognition modalities contrasted with the left lateralization of the (verbal) name recognition modality. This observation was consistent with the results of clinical studies carried out in patients with unilateral brain lesions [12,13] and of experimental investigations conducted in healthy subjects [14]. A final similarity between facial and vocal recognition modalities was that that, in addition to these core networks of components, both these human recognition modalities also comprise extended networks, permitting techniques to advance from unimodal processing into multimodal person representations. In contrast with these similarities, which inspired Belin et al. [8] and Yovel and Belin [15] to advance the notion that voice may be considered an ‘auditory face’, other authors have stressed the existence of important differences between these two modalities of human identification. Neurobiological investigations have, for instance, documented the lower degree of functional specificity shown by areas involved in voice perception in comparison to those involved in face perception [16,17]. The greater difficulty documented by

Citation: Gainotti, G. Human Recognition: People Recognition: The Utilization of Face, Voice, Name and Interactions—An Extended Editorial. *Brain Sci.* **2024**, *14*, 345. <https://doi.org/10.3390/brainsci14040345>

Received: 19 February 2024

Accepted: 21 March 2024

Published: 30 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

several authors (e.g., [18–20]) in vocal rather than facial human recognition is probably related to this difference. Another source of dissimilarity stems from the different temporal dimension of facial and vocal processing: faces can be discerned from a single snapshot [21], whereas voices constitute dynamic stimuli for which receivers must integrate information over time. Other investigations concerning facial, but not vocal, recognition modalities have highlighted two problems: (a) the face inversion effect, namely the observation that turning a picture of a face upside-down makes its identification much more difficult than the recognition of other objects [3,6]; and (b) the question of which relations between the pathways allow for, respectively, facial identification and the recognition of facial emotional expressions. The classical views on this subject assumed that two distinct pathways could allow for the recognition of facial identity and of emotional facial expressions. This was based on the assumption that the encoding of the structural aspects of facial identity should be mainly performed via the ventral visual pathway, whereas the view was held that the processing of facial expressions should be performed using the dorsal visual pathway [22–25]. However, Duchaine and Yovel [6] have proposed a revised framework in which OFA and FFA are engaged in processing facial shape, contributing to both facial identity and facial expression recognition. Furthermore, more recent animal studies have revealed the presence of information about facial identity and facial expression in the middle dorsal face area of macaque monkeys [26].

In more recent times, the attention of researchers has shifted from the basic components of human recognition systems to more complex functions. These investigations have mainly concerned the interactions that could exist between perceptual channels, which process faces and voices before the level of the ‘person identity nodes’ (PINs). The latter, according to cognitive models of recognizing familiar people (e.g., [27,28]), should allow for the identification of a person characterized by a given face and voice, giving access to the corresponding semantic (biographical) information. The earliest data to raise this problem were obtained by Schweinberger et al. [29,30], who observed that famous faces, but not famous names, caused a long-term repetition priming effect in the recognition of famous voices. Schweinberger et al. [29] concluded that these findings might be related to “perceptual links” between faces and voices, and this suggestion was strongly supported by data obtained by following authors. For instance, von Kriegstein et al. [31,32] measured, the brain activity of the FFA during identification tasks by means of functional magnetic resonance imaging (fMRI). These tasks saw subjects focus on either the speaker’s voice or the verbal content of sentences. These authors showed that familiar persons’ voices activated the FFA when the identification task was to focus on the speaker’s identity. Further, they demonstrated that, in these conditions, a functional connectivity between FFA and STS was obtained during familiar speaker recognition. In more recent years, other studies have tried to identify the main factors underlying the formation of these connections. Some of these studies have proposed the “metamodal” principle of brain organization [33], which maintains that the similarities in the cognitive and neural mechanisms underlying facial and vocal perception may lead to the recognition of identities across auditory and visual modalities. Other authors (e.g., [29–32]) have suggested that an important factor leading to the formation of interactions between facial and vocal processing channels could be the familiarity of stimuli. There are two important factors supporting this suggestion. The first is that a ‘familiarity feeling’ is automatically generated when a perceptual pattern matches a stored representation of the same stimulus and that the generation of the corresponding ‘familiarity feeling’ is the first step of the recognition process. The second factor is that the production of a familiarity feeling for a given face or voice is a right-lateralized component of the process of human recognition. The brain structures involved in facial and vocal recognition are just as right-lateralized. Clinical studies that have investigated the lateralization of early and late components of facial (prosopagnosia) and vocal (phonagnosia) recognition disorders have also shown that the disruption of feelings of facial and vocal familiarity usually occurs due to right brain lesions, whereas the unsettlement of name-related familiarity feelings is due to left brain

damage [34–36]. Even if all these investigations show that interactions exist between the perceptual channels processing faces and voices before the level of PINs, it remains possible that face–voice integrations may also be mediated by a later (post-perceptual) stage, as implied by classical models [27,28]. This proposal is, however, not supported by the results obtained by investigations that tried to evaluate the interactions between semantic information and facial or vocal learning.

Several aspects of these questions are taken into account in this Special Issue. They are considered here shortly. This is first performed by taking into account the problems relating to unimodal facial and vocal perceptual modalities and then the more general problems regarding the interactions between different perceptual channels. The importance of these connections in the functional reorganization of neural networks for human recognition in blind subjects and the role played by familiarity and by brain asymmetries in modality-specific and inter-modal interactions are also considered shortly.

2. Papers Specifically Dealing with the Brain Structures and Mechanisms Involved in Modality-Specific Aspects of Facial and Vocal Recognition

Since problems concerning facial processing have been more extensively studied than those relating to voice recognition, the number of papers dealing with facial recognition in our Special Issue surpasses that of articles concerning voice identification. Of the papers specifically considering face recognition, the contribution of Rossion et al. [37,38] is particularly interesting. This submission summarizes a 10-year research program that combined, in epileptic patients, the recording of intracerebral activity in the ventral occipito-temporal cortex, resulting in rapid periodic visual stimulation. Indeed, this study has increased our knowledge of the neural basis of face recognition, reconciling the wide distribution of neural face recognition activity with its (right) hemispheric and regional specialization and confirming the spatial dissociations in category selectivity between faces and other meaningful visual stimuli.

In another contribution, Hagen et al. [39] have tried to clarify the mechanisms underlying the ‘face inversion effect’ [3], evaluating the plasticity of the neural processes involved in this effect after extensive visual training in adulthood. At the behavioral level, the authors reported a significant reduction in the face inversion effect. This correlated with data obtained on a neural index of facial identity recognition, confirming the involvement of a substantial degree of plasticity in processes that are critical for facial identity recognition.

Two other important papers in this Special Issue factor in the problem of the relationship between the recognition of facial identity and the identification of emotional facial expressions. In particular, Schwartz et al. [40] have reconsidered the question of the neural pathways that enable the detection of facial identity and of facial expression, whereas Chauhan et al. [41] have tried to clarify if personal familiarity, which facilitates the rapid and optimized detection of faces, also promotes the detection of facial expressions. Schwartz et al. [40] have challenged the classical view, which assumes that facial identity and facial expression may be processed by segregated neural pathways [22–25], by testing the hypothesis that integrated representations of identity and expression arise spontaneously within deep neural networks. They used two different datasets to train a deep convolutional neural network in order to label face identity and facial expression, respectively, showing that information about identity and expression are encoded within common brain regions. The authors also showed, however, that features used to distinguish between identities and expressions become increasingly orthogonal from layer to layer, suggesting that deep neural networks can disentangle representational subspaces corresponding to facial identity and facial expression. On the other hand, Chauhan et al. [41] measured, in a visual search paradigm, accuracy and response time to two different emotional expressions (anger and happiness) displayed by familiar and unfamiliar faces in

order to determine local facial features contribute to both face identity and facial expression recognition. These authors showed that personal familiarity facilitates the detection of some (angry) facial expressions, but not of others (happy), and assumed that the advantage of familiar face recognition may rely on a feature-based type of processing, supporting the claim of Kaufmann and Schweinberger [42] that there is a relationship between the recognition of familiar faces and the recognition of facial expressions.

Only one article in this Special Issue concerns aspects of voice-based identity recognition specifically. In this paper, Stevenage et al. [43] examines voice processing in challenging listening conditions, starting from the premise that voice-based identity processing depends on both the ability to tell two instances of different speakers apart and the capacity to identify two instances of the same speaker. Since previous research only examines these vocal processing capabilities under relatively common listening conditions, the authors conduct two experiments that employ challenging listening tasks, determining just how good their subjects are in terms of these voice processing tasks. The authors show that vocal identity processing is a highly adaptable task that is assisted by familiarity with the speaker, suggesting that stored mental representations may be viewed as regions that capture and reflect vocal variability within a speaker, and also that voice processing is far better than was previously presumed.

3. Articles Dealing With the Interactions between Different Channels of Human Recognition and their Possible Underlying Mechanisms

Various problems relating to the interactions between different channels of human recognition are taken into account, sometimes with apparent conflicting results, in this Special issue of *Brain Science*. The hypothesis that interactions between different channels of human recognition may not only occur before the level of the 'PINs', but also at a later (post-perceptual) stage, is tackled by Fransson et al. [44]. These authors try to evaluate whether perceiving a person's face or voice enhances the encoding of his/her biographic data. They perform three experiments, in which subjects learn the biographic data of an individual, whether or not they are associated or not with vocal/facial perception. After learning, biographic data alone are presented to the same subjects who, in the test phase, are tested for both familiarity and their capacity to match biographic data to a name. The results show that the simultaneous processing of autobiographic and perceptual representations of humans does not modulate the encoding of biographic data, confirming that the interactions between autobiographic information and channels of human recognition do not occur after the level of the 'PINs'. The same research group [45], however, also present results that are at least partly inconsistent with models indicating that interactions between different channels of human recognition can occur before the level of the PINs. These authors show, in fact, that subjects who observe faces associated with vocal data, biographic data, or both, do not perform better in the retrieval phase, on either familiarity tasks or face-to-name identification tasks, than those who undergo the encoding of faces without additional information. Since these results contrast with most data reported in the literature and in other contributions to this Special Issue, I think that these inconsistencies could be due to two methodological reasons. The first and most important reason behind them could be that, in this experiment, all information about a character is learned and, therefore, those familiarity effects that certainly play an important role in the formation of these inter-modal interactions should also not work. The second reason could be that this experimental design, using a between-subject format with a rather low number of subjects in each group, probably does not have enough power to detect the small effects that can be observed in the absence of familiarity [29,31]. The stress I place in this commentary on the crucial role played by familiarity factors in the production of inter-sensory communications is, at least in part, because an important contribution to this Special Issue deals just with this question. Stevenage et al. [46] report the results of two experiments that examined face-voice correlations for familiar and unfamiliar stimuli. The

authors show that the correlation between facial and vocal processing was significant but small when recognizing unfamiliar individuals, whereas it was much larger when matching familiar subjects. According to the authors, these results support the view that facial and vocal processing are aligned as constituents of an overarching human perception system, but also that familiar and unfamiliar stimuli are processed in different ways. More specifically, the leading role played by familiarity could reflect the importance of a pre-existing mental representation and their cross-talk within the neural architectures when processing familiar faces and voices. In the last paper of this Special Issue to deal with interactions between different channels of human recognition, Zaske et al. [47] consider two problems. The first is that, even if models of inter-sensory integrations specify how visual and auditory information taken from faces and voices is combined, the exact timecourse of this audio-visual face-voice integration remains a matter of debate. The second is that the recognition of people from their voices may be facilitated by ‘voice distinctiveness’, in a manner similar to that which has been reported for faces. Since event-related potentials (ERPs) of an excellent temporal resolution are provided, Zaske et al. [47] use this technique to investigate the timecourse of face-voice-integration, also testing if the recognition of voices may be facilitated by audio-visual learning with distinctive vs. non-distinctive faces. They showed that voices that were previously learned with faces elicited an early (the N250-like) ERP component, which was similar in topography to that typically observed for facial stimuli and that, at the test, voices that were previously learned with distinctive faces were classified faster than those learned with non-distinctive faces. Since the preliminary source localization of the voice-induced N250 was compatible with a source in the fusiform gyrus, these findings provided support to the theory of an early interaction between vocal and facial processing areas during both learning and voice recognition.

4. Papers Evaluating the Role That the Voice-Face Connections Can have on the Functional Reorganization in Blind Subjects of the Neural Network for People Recognition

If the integration of vocal and facial signals plays a significant, but not critical role, in facilitating human recognition in subjects who can see, the role that these connections could have for blind subjects, who primarily rely on vocal cues to recognize a person’s identity, is much more relevant. If this role is not investigated, it will remain unclear how the neural systems for voice recognition reorganize themselves in the blind. Two papers have tackled different aspects of this problem in this Special Issue. In the first paper, Pang et al. [48] investigated the alterations in the resting-state functional connectivity among voice- and face-sensitive areas in blind subjects in comparison with controls. These authors found that, in blind subjects, the intranetwork connections among voice-sensitive areas are enhanced, whereas the connections between face- and voice-sensitive areas are diminished. They also suggested that, even if the visual deprivation decreases the inter-network connections between the voice- and face-sensitive areas, these connections may be still involved in the voice recognition processes through subcortical pathways. In particular, they stressed the importance of the connections between the core face-sensitive areas (e.g., FFA and OFA) and the amygdala and of those connecting the amygdala with the inferior frontal gyrus (IFG). Assuming that the face and voice show similar computational processing, and that this helps the functional reorganization [33] of blind individuals and their adaptation to the social environment in daily life, these findings may be consistent with the “metamodal” theory. In the second paper on this topic, Terruzzi et al. [49] tried to integrate neuropsychological and fMRI studies to clarify whether this functional reorganization could be performed by the right temporal lobe. Neuropsychological studies have, indeed, shown that the components of the task of famous voice recognition [20], on which blind subjects obtain their best results, are the same elements selectively impaired in patients with right temporal lesions [50], suggesting that the right temporal lobe

may play an important role in this reorganization. The surety of this hypothesis is strengthened using data obtained by Dormal et al. [51] when studying the functional preference for object sounds and voices in the brains of early blind and sighted individuals using fMRI: these authors reported evidence of a selective increase in the functional coupling between the left temporal voice area and the right fusiform gyrus. All these data could, therefore, point to a greater role of the right temporal lobe in the development of the vocal processing abilities of blind subjects than previously understood.

5. The Role Played by Familiarity and by Brain Asymmetries in Modality-Specific Recognition Modalities and in the Interactions between These Modalities

We have seen in the previous sections of this Editorial that several studies investigating specific recognition modalities have stressed the role played by familiarity factors in very different conditions, such as the study of Stevenage et al. [43] on voice processing in challenging listening conditions and the research of Chauhan et al. [41] on the relation between the recognition of familiar faces and the identification of emotional facial expressions. We have also reported the results obtained by Stevenage et al. [46] when investigating the role played by familiarity factors in the generation of face–voice interactions. We did so because, in this condition, the size of the face–voice interaction was also much greater when stimuli were familiar than when they concerned unfamiliar subjects. Castro-Laguardia et al. submitted a study specifically devoted to the identification of areas where local fMRI patterns could contribute to familiarity detection of faces, names, or both [52]. These authors identified cortical areas where it was possible to accurately cross-classify familiar stimuli from one category by using a classifier trained with the stimuli from the other ones. They also discovered several areas that supported the classification of familiar faces, but not of familiar names, in the fusiform gyrus, frontal, and temporal regions. They also showed, however, that areas which only contributed to familiarity detection for faces were primarily lateralized to the right hemisphere, and this could be interesting for at least two reasons. The first is that the right lateralization of areas which contributed only to familiarity detection for faces supports the suggestion, advanced in this Special Issue by Terruzzi et al. [49], that the right temporal lobe may play an important role in the development of voice processing abilities of blind subjects. The second is that the right lateralization of areas contributing to the familiarity detection of familiar faces, but not of familiar names, is consistent with the results of clinical studies conducted in patients with unilateral brain lesions and of experimental investigations carried out in healthy subjects. These studies, indeed, showed that (a) facial recognition disorders (prosopagnosia) and vocal recognition disorders (phonagnosia) are more frequent in the right than in left brain-damaged patients [(see [13], and [53] for reviews concerning the prosopagnosia and [54,55] for reviews concerning the phonagnosia) and that (b) these laterality effects can also be found in the recognition of familiar faces, voices, and names by normal subjects (see [14] for review). All these results could have more general implications for the problem of the interactions between different channels of human recognition and the possible underlying mechanisms. Several investigations (e.g., [29,56]) have, indeed, shown, that an interaction similar to that found between faces and voices is not observed between faces and names, suggesting that channels which process perceptual data are more closely integrated than those which process, respectively, perceptual and verbal data. In previous sections of this Editorial, it has been repeatedly stressed that both the facial and vocal perceptual modalities of person identification are more represented in the right than in the left hemisphere of the brain, whereas the verbal (name) modality is more represented in the left hemisphere. In any case, the existence of an integration between the perceptual modalities subserved by the right hemisphere, but not between those subserved, respectively, by the right and left hemispheres, provides further support to models which assume that an interaction between voice and face sensory systems occurs at the early processing stages, prior to accessing identity nodes.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Kanwisher, N.; McDermott, J.; Chun, M.M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* **1997**, *17*, 4302–4311. <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>.
2. McCarthy, G.; Puce, A.; Gore, J.C.; Allison, T. Face-Specific Processing in the Human Fusiform Gyrus. *J. Cogn. Neurosci.* **1997**, *9*, 605–610. <https://doi.org/10.1162/jocn.1997.9.5.605>.
3. Haxby, J.V.; Ungerleider, L.G.; Clark, V.P.; Schouten, J.L.; Hoffman, A.E.; Martin, A. The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* **1999**, *22*, 189–199. [https://doi.org/10.1016/S0896-6273\(00\)80690-X](https://doi.org/10.1016/S0896-6273(00)80690-X).
4. Gauthier, I.; Tarr, M.J.; Moylan, J.; Skudlarski, P.; Gore, J.C.; Anderson, A.W. The fusiform “face area” is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* **2000**, *12*, 495–504. <https://doi.org/10.1162/089892900562165>.
5. Blank, H.; Wieland, N.; von Kriegstein, K. Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neurosci. Biobehav. Rev.* **2014**, *47*, 717–734. <https://doi.org/10.1016/j.neubiorev.2014.10.022>.
6. Duchaine, B.; Yovel, G. A Revised Neural Framework for Face Processing. *Annu. Rev. Vis. Sci.* **2015**, *1*, 393–416. <https://doi.org/10.1146/annurev-vision-082114-035518>.
7. Belin, P.; Zatorre, R.J.; Lafaille, P.; Ahad, P.; Pike, B. Voice-selective areas in human auditory cortex. *Nature* **2000**, *403*, 309–312. <https://doi.org/10.1038/35002078>.
8. Belin, P.; Fecteau, S.; Bedard, C. Thinking the voice: Neural correlates of voice perception. *Trends Cogn. Sci.* **2004**, *8*, 129–135. <https://doi.org/10.1016/j.tics.2004.01.008>.
9. Andics, A.; McQueen, J.M.; Petersson, K.M.; Gál, V.; Rudas, G.; Vidnyánszky, Z. Neural mechanisms for voice recognition. *Neuroimage* **2010**, *52*, 1528–1540. <https://doi.org/10.1016/j.neuroimage.2010.05.048>.
10. Schall, S.; Kiebel, S.J.; Maess, B.; von Kriegstein, K. Voice identity recognition: Functional division of the right STS and its behavioral relevance. *J. Cogn. Neurosci.* **2015**, *27*, 280–291. https://doi.org/10.1162/jocn_a_00707.
11. Kriegstein, K.V.; Giraud, A.L. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* **2004**, *22*, 948–955. <https://doi.org/10.1016/j.neuroimage.2004.02.020>.
12. Sergent, J.; Signoret, J.L. Varieties of functional deficits in prosopagnosia. *Cereb. Cortex* **1992**, *2*, 375–388. <https://doi.org/10.1093/cercor/2.5.375>.
13. Gainotti, G.; Marra, C. Differential contribution of right and left temporo-occipital and anterior temporal lesions to face recognition disorders. *Front. Hum. Neurosci.* **2011**, *5*, 55. <https://doi.org/10.3389/fnhum.2011.00055>.
14. Gainotti, G. Laterality effects in normal subjects’ recognition of familiar faces, voices and names. Perceptual and representational components. *Neuropsychologia* **2013**, *5*, 1151–1160. <https://doi.org/10.1016/j.neuropsychologia.2013.03.009>.
15. Yove, G.; Belin, P. A unified coding strategy for processing faces and voices. *Trends Cogn. Sci.* **2013**, *17*, 263–271. <https://doi.org/10.1016/j.tics.2013.04.004>.
16. Young, A.W.; Frühholz, S.; Schweinberger, S.R. Face and voice perception: Understanding commonalities and differences. *Trends Cogn. Sci.* **2020**, *24*, 398–410. <https://doi.org/10.1016/j.tics.2020.02.001>.
17. Bestelmeyer, P.E.G.; Muehl, C. Neural dissociation of the acoustic and cognitive representation of voice identity. *NeuroImage* **2022**, *263*, 119647. <https://doi.org/10.1016/j.neuroimage.2022.119647>.
18. Damjanovic, L.; Hanley, J.R. Recalling episodic and semantic information about famous faces and voices. *Mem. Cogn.* **2007**, *35*, 1205–1210. <https://doi.org/10.3758/bf03193594>.
19. Barsics, C.; Brédart, S. Recalling episodic information about personally known faces and voices. *Conscious. Cogn.* **2011**, *20*, 303–308. <https://doi.org/10.1016/j.concog.2010.03.008>.
20. Quaranta, D.; Piccininni, C.; Carlesimo, G.A.; Luzzi, S.; Marra, C.; Papagno, C.; Trojano, L.; Gainotti, G. Recognition disorders for famous faces and voices: A review of the literature and normative data of a new test battery. *Neurol. Sci.* **2016**, *37*, 345–352. <https://doi.org/10.1007/s10072-015-2437-1>.
21. Sternberg, S. High speed scanning in human memory. *Science* **1966**, *153*, 652–654. <https://doi.org/10.1126/science.153.3736.652>.
22. Calder, A.J.; Young, A.W. Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* **2005**, *6*, 641. <https://doi.org/10.1038/nrn1724>.
23. Gobbini, M.I.; Haxby, J.V. Neural systems for recognition of familiar faces. *Neuropsychologia* **2007**, *45*, 32–41. <https://doi.org/10.1016/j.neuropsychologia.2006.04.015>.
24. Allison, T.; Puce, A.; McCarthy, G. Social perception from visual cues: Role of the STS region. *Trends Cogn. Sci.* **2000**, *4*, 267–278. [https://doi.org/10.1016/S1364-6613\(00\)01501-1](https://doi.org/10.1016/S1364-6613(00)01501-1).
25. Yovel, G.; O’Toole, A.J. Recognizing People in Motion. *Trends Cogn. Sci.* **2016**, *20*, 383–395. <https://doi.org/10.1016/j.tics.2016.02.005>.
26. Yang, Z.; Freiwald, W.A. Joint encoding of facial identity, orientation, gaze, and expression in the middle dorsal face area. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2108283118.
27. Burton, A.M.; Bruce, V.; Hancock, P.J. From pixels to people: A model of familiar face recognition. *Cogn. Sci.* **1999**, *23*, 1–31. [https://doi.org/10.1016/S0364-0213\(99\)80050-0](https://doi.org/10.1016/S0364-0213(99)80050-0).

28. Brédart, S.; Valentine, T.; Calder, A.; Gassi, L. An Interactive Activation Model of Face Naming. *Q. J. Exp. Psychol. Sect. A* **1995**, *48*, 466–486. <https://doi.org/10.1080/14640749508401400>.
29. Schweinberger, S.R.; Herholz, A.; Stief, V. Auditory Long term Memory: Repetition Priming of Voice Recognition. *Q. J. Exp. Psychol. Sect. A* **1997**, *50*, 498–517. <https://doi.org/10.1080/713755724>.
30. Schweinberger, S.R.; Robertson, D.M.; Kaufmann, J.M. Hearing facial identities. *Q. J. Exp. Psychol. Sect. A* **2007**, *60*, 1446–1456. <https://doi.org/10.1016/j.cortex.2010.11.011>.
31. von Kriegstein, K.; Kleinschmidt, A.; Sterzer, P.; Giraud, A.L. Interaction of face and voice areas during speaker recognition. *J. Cogn. Neurosci.* **2005**, *17*, 367–376.
32. von Kriegstein, K.; Giraud, A.L. Implicit multisensory associations influence voice recognition. *PLoS Biol.* **2006**, *4*, e326. <https://doi.org/10.1371/journal.pbio.0040326>.
33. Pascual-Leone, A.; Hamilton, R. The metamodal organization of the brain. *Prog. Brain Res.* **2001**, *134*, 427–445. [https://doi.org/10.1016/s0079-6123\(01\)34028-1](https://doi.org/10.1016/s0079-6123(01)34028-1).
34. Gainotti, G. Face familiarity feelings, the right temporal lobe and the possible underlying neural mechanisms. *Brain Res. Rev.* **2007**, *56*, 214–235. <https://doi.org/10.1016/j.brainresrev.2007.07.009>.
35. Borghesani, V.; Narvid, J.; Battistella, G.; Shwe, W.; Watson, C.; Binney, R.J.; Sturm, V.; Miller, Z.; Mandelli, M.L.; Miller, B.; et al. “Looks familiar, but I do not know who she is”: The role of the anterior right temporal lobe in famous face recognition. *Cortex* **2019**, *115*, 72–85. <https://doi.org/10.1016/j.cortex.2019.01.006>.
36. Piccininni, C.; Marra, C.; Quaranta, D.; Papagno, C.; Longo, C.; Sarubbo, S.; Zigiotta, L.; Luzzi, S.; Gainotti, G. Which components of famous people recognition are lateralized? A study of face, voice and name recognition disorders in patients with neoplastic or degenerative damage of the right or left anterior temporal lobes. *Neuropsychologia* **2023**, *181*, 108490. <https://doi.org/10.1016/j.neuropsychologia.2023.108490>.
37. Rossion, B.; Jacques, C.; Jonas, J. Intracerebral Electrophysiological Recordings to Understand the Neural Basis of Human Face Recognition. *Brain Sci.* **2023**, *13*, 354. <https://doi.org/10.3390/brainsci13020354>.
38. Ellis, H.D.; Jones, D.M.; Mosdell, N. Intra- and inter-modal repetition priming of familiar faces and voices. *Br. J. Psychol.* **1997**, *88*, 143–156. <https://doi.org/10.1111/j.2044-8295.1997.tb02625x>.
39. Hagen, S.; Laguesse, R.; Rossion, B. Extensive Visual Training in Adulthood Reduces an Implicit Neural Marker of the Face Inversion Effect. *Brain Sci.* **2024**, *14*, 146. <https://doi.org/10.3390/brainsci14020146>.
40. Schwartz, E.; O’nell, K.; Saxe, R.; Anzellotti, S. Challenging the Classical View: Recognition of Identity and Expression as Integrated Processes. *Brain Sci.* **2023**, *13*, 296. <https://doi.org/10.3390/brainsci13020296>.
41. Chauhan, V.; Castello, M.V.d.O.; Taylor, M.; Gobbini, M.I. Familiarity Facilitates Detection of Angry Expressions. *Brain Sci.* **2023**, *13*, 509. <https://doi.org/10.3390/brainsci13030509>.
42. Kaufmann, J.M.; Schweinberger, S.R. Expression influences the recognition of familiar faces. *Perception* **2004**, *33*, 399–408. <https://doi.org/10.1068/p5083>.
43. Stevenage, S.V.; Singh, L.; Dixey, P. The Curious Case of Impersonators and Singers: Telling Voices Apart and Telling Voices Together under Naturally Challenging Listening Conditions. *Brain Sci.* **2023**, *13*, 358. <https://doi.org/10.3390/brainsci13020358>.
44. Fransson, S.; Corrow, S.; Yeung, S.; Schaefer, H.; Barton, J.J.S. Effects of Faces and Voices on the Encoding of Biographic Information. *Brain Sci.* **2022**, *12*, 1716. <https://doi.org/10.3390/brainsci12121716>.
45. Karlsson, T.; Schaefer, H.; Barton, J.J.S.; Corrow, S.L. Effects of Voice and Biographic Data on Face Encoding. *Brain Sci.* **2023**, *13*, 148. <https://doi.org/10.3390/brainsci13010148>.
46. Stevenage, S.V.; Edey, R.; Keay, R.; Morrison, R.; Robertson, D.J. Familiarity Is Key: Exploring the Effect of Familiarity on the Face-Voice Correlation. *Brain Sci.* **2024**, *14*, 112. <https://doi.org/10.3390/brainsci14020112>.
47. Zäske, R.; Kaufmann, J.M.; Schweinberger, S.R. Neural Correlates of Voice Learning with Distinctive and Non-Distinctive Faces. *Brain Sci.* **2023**, *13*, 637. <https://doi.org/10.3390/brainsci13040637>.
48. Pang, W.; Zhou, W.; Ruan, Y.; Zhang, L.; Shu, H.; Zhang, Y.; Zhang, Y. Visual Deprivation Alters Functional Connectivity of Neural Networks for Voice Recognition: A Resting-State fMRI Study. *Brain Sci.* **2023**, *13*, 636. <https://doi.org/10.3390/brainsci13040636>.
49. Terruzzi, S.; Papagno, C.; Gainotti, G. The Right Temporal Lobe and the Enhancement of Voice Recognition in Congenitally Blind Subjects. *Brain Sci.* **2023**, *13*, 431. <https://doi.org/10.3390/brainsci13030431>.
50. Papagno, C.; Mattavelli, G.; Casarotti, A.; Bello, L.; Gainotti, G. Defective recognition and naming of famous people from voice in patients with unilateral temporal lobe tumors. *Neuropsychologia* **2018**, *116 Pt B*, 194–204. <https://doi.org/10.1016/j.neuropsychologia.2017.07.021>.
51. Dormal, G.; Pelland, M.; Rezk, M.; Yakobov, E.; Lepore, F.; Collignon, O. Functional Preference for Object Sounds Voices in the Brain of Early Blind Sighted Individuals. *J. Cogn. Neurosci.* **2018**, *30*, 86–106. https://doi.org/10.1162/jocn_a_01186.
52. Castro-Laguardia, A.M.; Ontivero-Ortega, M.; Morato, C.; Lucas, I.; Vila, J.; León, M.A.B.; Muñoz, P.G. Familiarity Processing through Faces and Names: Insights from Multivoxel Pattern Analysis. *Brain Sci.* **2023**, *14*, 39. <https://doi.org/10.3390/brainsci14010039>.
53. Mayer, E.; Rossion, B. Prosopagnosia. In *The Behavioral and Cognitive Neurology of Stroke*; Godefroy, O., Bogousslavsky, J., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 315–334. <https://doi.org/10.1017/CBO9780511544880.017>.
54. Gainotti, G. What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia* **2011**, *49*, 2273–2282. <https://doi.org/10.1016/j.neuropsychologia.2011.04.027>.

-
55. Gainotti, G.; Quaranta, D.; Luzzi, S. Apperceptive and Associative Forms of Phonagnosia. *Curr. Neurol. Neurosci. Rep.* **2023**, *23*, 327–333. <https://doi.org/10.1007/s11910-023-01271-5>.
 56. O'mahony, C.; Newell, F.N. Integration of faces and voices, but not faces and names, in person recognition. *Br. J. Psychol.* **2012**, *103*, 73–82 <https://doi.org/10.1111/j.2044-8295.2011.02044.x>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.