

Article

A Simulation and Open-Data Workflow for Detecting Temporal Confounding in Repeated-Stimulus fMRI Pattern Similarity

Seweryn Lipiński

Faculty of Technical Sciences, University of Warmia and Mazury in Olsztyn, 10-036 Olsztyn, Poland; seweryn.lipinski@uwm.edu.pl

Abstract

Pattern similarity analysis is widely used to assess response reliability in repeated-stimulus fMRI designs, but pairwise similarity estimates may be confounded by temporal structure, including slow drift, autocorrelation, hemodynamic smoothing, and fixed repetition order. This study presents a simulation-supported workflow for detecting and reducing such confounding. A fixed design with 40 events, corresponding to ten stimuli repeated four times, was simulated with and without stable stimulus-specific structure. Naive within-versus-between similarity was compared with a time-neighborhood-matched contrast and a distance-adjusted regression estimator. In 500 null simulations, temporal structure alone produced a biased naive contrast (mean $\pm$ SD = -0.1309 ± 0.0021), whereas the time-neighborhood-matched contrast was near zero (-0.0013 ± 0.0018). The time-neighborhood-matched contrast increased monotonically with stimulus-effect strength and remained robust after introducing hemodynamic convolution, spatial dependence, structured nuisance components, and variation in autocorrelation, voxel count, and drift specification. In an OpenNeuro music-fMRI case study, the naive contrast was strongly negative, whereas temporal matching yielded a small positive residual contrast. Cross-validated selection of functionally responsive voxels increased this residual effect, indicating partial dilution in the whole-EPI analysis, while the naive contrast remained negative under all mask specifications. The workflow provides a practical diagnostic framework for repeated-stimulus fMRI pattern similarity analyses.

Keywords: fMRI; pattern similarity; temporal confounding; repeated-stimulus design; simulation; time-neighborhood matching; autocorrelation; BOLD signal; OpenNeuro

Academic Editor: James C. L. Chow

Received: 24 May 2026

Revised: 3 July 2026

Accepted: 7 July 2026

Published: 10 July 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Pattern similarity analysis is widely used in functional magnetic resonance imaging (fMRI) to examine whether repeated experimental conditions evoke reproducible multi-voxel response patterns. In repeated-stimulus designs, a common assumption is that presentations of the same stimulus should produce more similar neural response patterns than presentations of different stimuli. This logic is intuitive and underlies many applications of representational similarity analysis, pattern-information analysis, test–retest reliability assessment, and stimulus-specific response characterization in cognitive and sensory neuroscience [1–5].

However, fMRI pattern similarity estimates are not determined solely by stimulus identity. The blood oxygenation level-dependent (BOLD) signal is temporally structured, reflecting hemodynamic smoothing, physiological and scanner-related fluctuations, slow signal drift, and autocorrelation across adjacent time points [6–9]. In repeated-stimulus paradigms, these temporal properties can interact with the experimental design. If stimuli are repeated in a fixed or partially structured order, within-stimulus and between-stimulus comparisons may differ systematically in their temporal distance within the run. Under such conditions, a naive within-versus-between similarity contrast may reflect temporal proximity, repetition order, or run-level dynamics rather than genuine stimulus-specific reliability [9,10].

This problem is particularly relevant for naturalistic fMRI experiments. Naturalistic auditory, visual, and audiovisual paradigms are increasingly used because they approximate real-world perceptual conditions more closely than highly simplified block or event-related designs [11–13]. Music listening is a representative example: music contains rich spectral, temporal, rhythmic, timbral, and categorical structure, making it useful for studying auditory and higher-level representational processes [14–17]. At the same time, naturalistic music-fMRI designs may rely on repeated stimulus presentation, averaging across test repetitions, or structured stimulus sequences to improve signal-to-noise ratio. These features can make direct pattern similarity analyses vulnerable to temporal confounding.

The issue is not that temporal autocorrelation exists in BOLD data, which is already well recognized [9,18]. The methodological problem is more specific: temporal structure can induce apparent stimulus-identity effects in pairwise pattern similarity analyses even when no true stimulus-specific response pattern is present. Related concerns have been raised in representational similarity analysis, where task structure, repetition structure, and sampling properties can bias estimates of representational organization if not explicitly controlled [1,4]. Conversely, temporal correction procedures should reduce design-induced bias without eliminating genuine stimulus-related structure. A practical workflow for repeated-stimulus fMRI should therefore address both questions: whether a naive similarity contrast is confounded by temporal structure, and whether a corrected contrast remains sensitive to true stimulus-specific information.

The presented workflow specifically concerns fixed-order repeated-stimulus designs in which the same stimulus sequence is presented repeatedly at predetermined event positions. In the simulated design used here, for example, stimulus 1 occurred at positions 1, 11, 21, and 31, so that same-stimulus comparisons were restricted to temporal distances of 10, 20, and 30 events. This dependence between stimulus identity and temporal distance is the central design feature examined in the study.

The present study develops and demonstrates this workflow using simulations and an open-data case study. First, a fixed repeated-stimulus design was simulated, with 40 events corresponding to ten stimuli repeated four times. Event-level voxel patterns were generated with slow drift, autocorrelated temporal components, and noise. In the null simulation, no stimulus-specific pattern was included, allowing the effect of temporal structure alone to be evaluated. Second, a sensitivity simulation introduced increasing levels of true stimulus-specific structure to test whether time-neighborhood matching preserves genuine stimulus effects. Third, the workflow was applied to an open music-evoked fMRI dataset from OpenNeuro, originally designed to study cortical representations of music genre [19–21].

The aim of this study is therefore not to draw population-level conclusions about music perception from a small fMRI sample, but to provide a reproducible methodological demonstration. Specifically, the study evaluates whether temporal structure can bias naive repeated-stimulus pattern similarity estimates, whether time-neighborhood match-

ing reduces this bias, and whether the same workflow can be applied to real repeated-stimulus fMRI data. By combining simulation and open-data analysis, this work provides a practical diagnostic procedure for assessing temporal confounding in repeated-stimulus fMRI pattern similarity workflows.

2. Materials and Methods

2.1. General Workflow

The study was designed as a simulation-supported methodological demonstration followed by an open-data case study. The workflow consisted of four main steps. First, a repeated-stimulus fMRI-like design was simulated using a fixed sequence of ten stimuli repeated four times. Second, event-level multivoxel response patterns were generated with temporal drift, autocorrelated components, and noise, either with or without a true stimulus-specific pattern component. Third, naive within-versus-between pattern similarity was compared with a time-neighborhood-matched contrast designed to reduce temporal-distance confounding. Fourth, the same analytical workflow was applied to an open music-evoked fMRI dataset to illustrate its behavior in real repeated-stimulus data.

The simulations were intended to isolate the effect of temporal structure under controlled conditions. The open-data analysis was not used to support population-level inference, but to demonstrate how the proposed workflow can be applied to real fMRI data with repeated naturalistic stimuli.

2.2. Simulated Repeated-Stimulus Design

A fixed repeated-stimulus design was simulated to match the structure of a typical test-retest fMRI design. Each simulated run consisted of 40 events. Ten stimulus identities were presented four times in the same order. Thus, event positions 1–10 represented the first repetition, event positions 11–20 the second repetition, event positions 21–30 the third repetition, and event positions 31–40 the fourth repetition.

Stimulus identity was assigned periodically across the run, so that stimulus 1 appeared at event positions 1, 11, 21, and 31; stimulus 2 appeared at event positions 2, 12, 22, and 32; and so forth. This structure produced within-stimulus comparisons only at temporal distances of 10, 20, and 30 events. Between-stimulus comparisons occurred across a wider range of temporal distances.

This design was selected because it creates a controlled situation in which stimulus repetition and temporal distance are not independent.

2.3. Simulation of Event-Level Voxel Patterns

Each event was represented by a high-dimensional voxel pattern. The number of simulated voxel features was set to 3000. Event-level patterns were generated as the sum of temporal components, independent noise, and, in selected simulations, a true stimulus-specific component.

The temporal component consisted of slow drift and autocorrelated variation. Slow drift was modeled using a small set of smooth basis functions across event order, including a linear term, a quadratic term, and a sinusoidal component. Voxel-specific weights were assigned to each basis function, allowing the drift profile to vary across simulated voxels. Autocorrelated temporal variation was simulated using a first-order autoregressive process with autocorrelation coefficient $\rho = 0.90$ [9]. Independent Gaussian noise was then added to each event-level pattern.

After pattern generation, each voxel feature was standardized across events to zero mean and unit variance to reduce the influence of scale differences and disproportionately

variable features. Similarity between event patterns was then quantified using Pearson correlation across voxels.

2.4. Null Simulation Without Stimulus-Specific Structure

The null simulation was designed to test whether temporal structure alone can produce an apparent stimulus-identity effect in pattern similarity analysis. In this condition, no true stimulus-specific voxel pattern was included. Therefore, repeated presentations of the same stimulus did not share any stable stimulus-related component. Any difference between within-stimulus and between-stimulus similarity could only arise from temporal drift, autocorrelation, noise, and the fixed repetition structure of the design.

A total of 500 null simulations were performed. For each simulation, new temporal components and noise realizations were generated. Pairwise similarity, naive contrast, and time-neighborhood-matched contrast were then calculated.

2.5. Sensitivity Simulation with Increasing Stimulus-Specific Structure

A second simulation series was performed to verify whether time-neighborhood matching reduces temporal confounding without eliminating true stimulus-specific structure. In these simulations, each of the ten stimulus identities was assigned a stable high-dimensional voxel pattern. This stimulus-specific pattern was added to the temporal components and noise before voxel standardization.

The strength of the true stimulus-specific component varied across six levels: 0.00, 0.10, 0.25, 0.50, 0.75, and 1.00. The value of 0.00 corresponded to the null condition, whereas larger values represent progressively stronger stimulus-specific structure. For each strength level, 500 simulations were performed. The main outcome of this sensitivity analysis was the relationship between stimulus-effect strength and the time-neighborhood-matched similarity contrast.

This analysis was used as a sanity check for the workflow. If time-neighborhood matching merely removed all structured differences between repeated and non-repeated stimuli, the time-neighborhood-matched contrast would remain near zero even when a true stimulus-specific component was present. Conversely, a monotonic increase in time-neighborhood-matched contrast with stimulus-effect strength would indicate that the correction reduces temporal confounding while preserving genuine stimulus-related similarity.

2.6. Extended BOLD-like Sensitivity Simulation

To assess whether the main findings remained valid under a more realistic fMRI signal model, an additional sensitivity simulation was performed using continuous BOLD-like time series. The experimental design matched the OpenNeuro case study and consisted of 40 events, representing ten stimuli repeated four times in a fixed order. Each stimulus lasted 15 s, the repetition time was set to 1.5 s, and each simulated run contained 410 volumes.

For each stimulus, a stable spatial response pattern was generated on a regular $15 \times 20 \times 10$ voxel grid, corresponding to 3000 voxel features. The spatial maps were generated from Gaussian random fields and smoothed with a Gaussian kernel to introduce local spatial dependence. In addition to the stable stimulus-specific component, each event contained a random event-specific spatial component and a common event-related component.

Continuous neural response time series were constructed using 15 s boxcar functions and convolved with a canonical double-gamma hemodynamic response function. This procedure introduced temporal smoothing and hemodynamic carryover between consecutive events. The simulated BOLD-like signal additionally included slow temporal drift,

an AR(1) component with $\rho = 0.90$, spatially correlated Gaussian noise, and low-frequency physiological-like oscillatory nuisance components.

Event-level multivoxel patterns were extracted using the same procedure as in the OpenNeuro case study. For each event, the signal was averaged across an eight-volume window beginning two TRs after stimulus onset. Voxel features were subsequently standardized across events, and pairwise Pearson correlations were calculated between all event-level patterns.

The strength of the stable stimulus-specific component varied across the same six levels as in the original sensitivity analysis: 0.00, 0.10, 0.25, 0.50, 0.75, and 1.00. A total of 500 simulations were performed for each level. The naive within-minus-between contrast and the time-neighborhood-matched contrast were calculated for each simulation.

2.7. Pairwise Pattern Similarity

For each simulated run, Pearson correlation coefficients were computed between all pairs of event-level voxel patterns. Each pair was labeled as a within-stimulus comparison if both events corresponded to the same stimulus identity, and as a between-stimulus comparison if the two events corresponded to different stimuli.

For every event pair, temporal distance was calculated as the absolute difference between event positions within the run. Because the simulated design used ten stimuli repeated four times in the same order, within-stimulus comparisons occurred only at temporal distances of 10, 20, and 30 events. Between-stimulus comparisons occurred at most other temporal distances. This difference in temporal-distance distributions motivated the use of a time-neighborhood-matched contrast in addition to the naive contrast.

2.8. Naive Similarity Contrast and Time-Neighborhood-Matched Contrast

The naive similarity contrast was calculated as the difference between the mean within-stimulus similarity and the mean between-stimulus similarity:

$$C_{\text{naive}} = \bar{r}_{\text{within}} - \bar{r}_{\text{between}}. \quad (1)$$

This contrast represents the simplest repeated-stimulus pattern similarity estimate. However, it can be confounded when within-stimulus and between-stimulus pairs differ in their temporal-distance distributions.

To reduce this confounding, a time-neighborhood-matched contrast was calculated. For within-stimulus comparisons at temporal distance d , the comparison baseline consisted of between-stimulus comparisons at adjacent temporal distances $d - 1$ and $d + 1$. In the fixed-order design used here, this procedure was applied to within-stimulus distances of 10, 20, and 30 events. For each distance, the time-neighborhood-matched contrast was defined as:

$$C_{\text{matched},d,w} = \bar{r}_{\text{within},d} - \bar{r}_{\text{between},0 < |d_{ij}-d| \leq w}. \quad (2)$$

For the primary specification $w = 1$, the second term is the pooled mean across all between-stimulus pairs at distances $d - 1$ and $d + 1$.

The overall time-neighborhood-matched contrast was obtained by averaging across the three within-stimulus distances. In the OpenNeuro case study, the same calculation was performed separately for each participant and run before summarization.

Between-stimulus pairs at the two adjacent temporal distances were pooled to provide a local temporal baseline for each within-stimulus distance. This procedure does not assume a globally linear similarity-distance relationship; it assumes only that the temporal component varies sufficiently smoothly over the narrow interval from $d - 1$ to $d + 1$. The use of the nearest available distances was intended to minimize temporal mismatch while retaining between-stimulus comparisons on both sides of the target distance.

To assess whether the results depended on the width of the local temporal neighborhood, an ablation analysis was performed using neighborhood widths $w \in \{1, 2, 3\}$. For a target within-stimulus distance (d), the local between-stimulus baseline included all between-stimulus pairs satisfying $0 < |d_{ij} - d| \leq w$. Thus, the ± 1 specification included distances $d - 1$ and $d + 1$, the ± 2 specification included $d - 2$, $d - 1$, $d + 1$, and $d + 2$, and the ± 3 specification additionally included $d - 3$ and $d + 3$. The baseline was calculated as the pooled mean across all eligible between-stimulus pairs.

To visualize the temporal structure of the similarity estimates, mean pattern similarity was calculated as a function of temporal distance between events. This analysis was performed separately for within-stimulus and between-stimulus pairs. The aim was to determine whether similarity depended systematically on temporal distance, independently of stimulus identity.

2.9. Parameter-Sensitivity Analysis

To determine whether the null bias depended on a single simulation configuration, an additional parameter-sensitivity analysis was performed. The AR(1) autocorrelation coefficient was varied across $\rho = 0.3, 0.6$, and 0.9 , and the number of simulated voxel features was varied across 500, 3000, and 10,000. Each combination was evaluated using two drift specifications: a linear-plus-quadratic basis and the same basis supplemented with a sinusoidal component.

This produced 18 parameter combinations. For each combination, 500 null simulations were performed without a stable stimulus-specific component. The naive within-minus-between contrast and the time-neighborhood-matched contrast using the ± 1 neighborhood were calculated for each simulation. Mean values, standard deviations, 95% confidence intervals, and the percentage reduction in absolute null bias were then summarized across simulations.

2.10. Comparison with Distance-Adjusted Regression

To place the time-neighborhood-matched contrast within the broader set of approaches for temporal-bias control, it was compared with a distance-adjusted regression estimator. For each simulation, pairwise similarity was modeled as:

$$r_{ij} = \beta_0 + \beta_1 I(\text{same stimulus}_{ij}) + \gamma_1 z_{ij} + \gamma_2 z_{ij}^2 + \gamma_3 z_{ij}^3 + \varepsilon_{ij}, \quad (3)$$

where r_{ij} denotes pairwise pattern similarity, $I(\text{same stimulus}_{ij})$ indicates whether the two events represented the same stimulus, and z_{ij} denotes standardized temporal distance. The coefficient β_1 was interpreted as the stimulus-similarity contrast after adjustment for the global nonlinear relationship between temporal distance and similarity.

A saturated categorical representation of temporal distance was not identifiable in the fixed-order design because same-stimulus pairs occurred exclusively at distances 10, 20, and 30. The cubic polynomial therefore provided a flexible but identifiable adjustment for temporal distance. The naive contrast, the ± 1 time-neighborhood-matched contrast, and the distance-adjusted regression coefficient were compared across 500 null simulations.

A cubic specification was chosen as a parsimonious nonlinear model that could capture curvature while limiting overfitting across the relatively small number of distinct temporal distances. More flexible alternatives, such as spline models or higher-order polynomials, would require additional model-selection choices and were beyond the scope of this comparison.

2.11. OpenNeuro Music-fMRI Case Study

The workflow was applied to the open music-evoked fMRI dataset described by Nakai et al. [19–21]. The present case study used only the six test runs from each of five participants, because these runs contained repeated presentations of the same stimuli. Each test run contained ten 15 s music clips repeated four times in a fixed order. Functional images were acquired with TR = 1.5 s, TE = 30 ms, 68 axial slices, 2 mm isotropic voxels, and 410 volumes per run.

Raw BOLD images and event files were downloaded from OpenNeuro [21]. Each test-run event file contained 41 entries, with the final entry repeating the first stimulus. The accompanying dataset documentation did not provide a separate rationale for this terminal repetition. It was therefore treated as an unbalanced additional event, and only the first 40 entries were retained to obtain a balanced 10-stimulus $\times$ 4-repetition structure. Because the available derivatives did not include preprocessed 4D BOLD time series or confound time-series files, a lightweight native-space preprocessing procedure was applied. For each run, an EPI-derived brain mask was generated using Nilearn's `compute_epi_mask` function. Voxel time series within the resulting mask were linearly detrended and standardized to zero mean and unit variance across time. No spatial normalization, smoothing, or anatomical ROI restriction was applied.

For each event, a voxel-wise response pattern was extracted by averaging the preprocessed BOLD signal across an eight-volume window beginning two TRs after stimulus onset. This yielded 240 response patterns per participant and 1200 event-level response patterns across the five participants. Pairwise similarity, naive contrast, time-neighborhood-matched contrast, and temporal-distance summaries were then computed using the same workflow as in the simulations.

Cross-Validated Functionally Selected Voxel Analysis

To assess whether the whole-EPI analysis diluted music-related pattern similarity by including large numbers of weakly responsive voxel features, an additional cross-validated top-k analysis was performed. Because the test runs differed slightly in their image geometry, all runs for a given participant were resampled to the spatial grid of the first test run. Run-specific EPI masks were then calculated, and their intersection was used to define a common within-participant voxel space.

Voxel selection was conducted separately for each participant using a leave-one-run-out procedure. For each held-out test run, the remaining five runs were used to calculate the mean event-related response of every voxel. Voxel time series were linearly detrended and standardized across volumes, and the activation score was defined as the mean signal across the same eight-volume response windows used for pattern extraction. Voxels were ranked according to this training-run activation score, and the top 500, 1000, and 3000 voxels were retained.

The selected voxel indices were then applied only to the held-out run, in which the naive within-minus-between contrast and the ± 1 time-neighborhood-matched contrast were calculated. This procedure was repeated across all six test runs and five participants. A common whole-EPI-mask analysis performed on the same resampled data served as the internal reference for the top-k comparison.

2.12. Statistical Summaries and Software

For the simulation analyses, results were summarized across 500 independent simulations per condition. The main reported outcomes were the mean within-stimulus similarity, mean between-stimulus similarity, naive contrast, time-neighborhood-matched contrast, and time-neighborhood-matched contrasts at temporal distances of 10, 20, and

30 events. For the sensitivity analysis, these outcomes were summarized as a function of stimulus-effect strength.

For the OpenNeuro case study, results were summarized descriptively across participants, runs, and temporal-distance cells. Because the dataset included only five participants, no population-level inference was drawn from these data. The emphasis was placed on the behavior of the workflow and on the comparison between naive and temporally controlled similarity estimates.

All analyses were conducted in Python 3.11.14 using NumPy 2.4.2, pandas 3.0.0, SciPy 1.17.0, Nilearn 0.13.1, NiBabel 5.4.2, and Matplotlib 3.10.8.

3. Results

3.1. Null Simulation

The null simulation was used to test whether temporal structure alone can generate an apparent stimulus-identity effect in repeated-stimulus pattern similarity analysis. In this condition, no true stimulus-specific voxel pattern was included. Therefore, any difference between within-stimulus and between-stimulus similarity had to arise from temporal drift, autocorrelated variation, independent noise, and the fixed repetition structure of the simulated design.

Across 500 null simulations, the naive within-minus-between contrast was strongly displaced from zero. The mean within-stimulus similarity was -0.1460 , whereas the mean between-stimulus similarity was -0.0151 , yielding a mean naive contrast of -0.1309 . Thus, even though repeated presentations of the same simulated stimulus shared no stable stimulus-specific component, the naive analysis suggested substantially lower similarity for repeated stimuli than for different stimuli.

Time-neighborhood matching largely removed this apparent effect. After comparing within-stimulus pairs at distances of 10, 20, and 30 events with nearby between-stimulus pairs at adjacent temporal distances, the mean time-neighborhood-matched contrast was reduced to -0.0013 . These results are visualized in Figure 1.

This indicates that the negative naive contrast was primarily driven by the temporal-distance structure of the repeated-stimulus design rather than by stimulus identity.

3.2. Sensitivity Simulation

The sensitivity simulation was used to verify whether time-neighborhood matching removes only temporal-distance bias or whether it also suppresses genuine stimulus-specific structure. To test this, a stable stimulus-specific voxel pattern was added to the simulated event-level patterns with increasing effect strength. The stimulus-effect strength was varied across six levels: 0.00, 0.10, 0.25, 0.50, 0.75, and 1.00.

The time-neighborhood-matched contrast increased monotonically with stimulus-effect strength (Figure 2). At zero effect strength, the time-neighborhood-matched contrast was near zero (-0.0013), consistent with the null simulation. When a weak stimulus-specific component was introduced, the time-neighborhood-matched contrast increased slightly to 0.0032 at strength 0.10 and to 0.0263 at strength 0.25. For stronger stimulus-specific components, the time-neighborhood-matched contrast increased more clearly, reaching 0.0999 at strength 0.50, 0.1983 at strength 0.75, and 0.3024 at strength 1.00.

In contrast, the naive within-minus-between contrast remained strongly influenced by temporal structure. At zero effect strength, the naive contrast was -0.1309 . It remained negative at strengths 0.10 (-0.1258), 0.25 (-0.0999), and 0.50 (-0.0179), despite the presence of a true stimulus-specific component. Only at stronger effect levels did the naive contrast become positive, reaching 0.0920 at strength 0.75 and 0.2082 at strength 1.00.

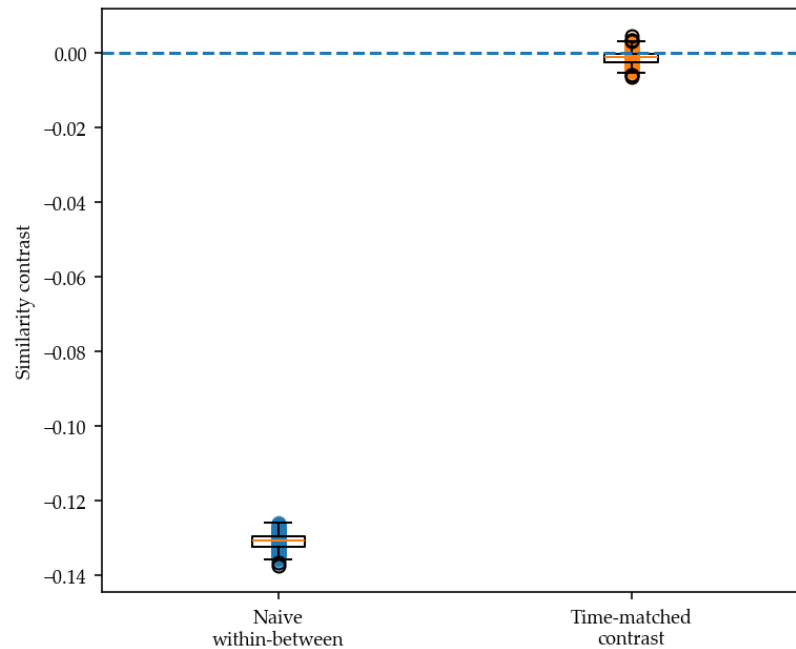


Figure 1. Null simulation: naive and time-neighborhood-matched similarity contrasts. Boxplots with individual simulation values are shown across 500 simulations with no true stimulus-specific voxel pattern. The naive within-minus-between contrast was consistently displaced from zero, demonstrating that temporal structure alone can generate a misleading apparent stimulus-identity effect. After time-neighborhood matching, the contrast was reduced to near zero.

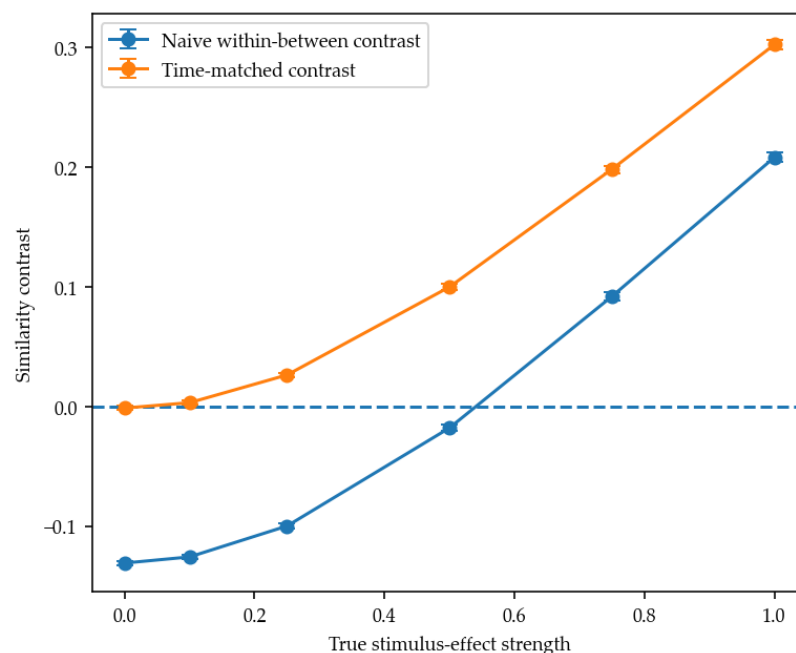


Figure 2. Sensitivity analysis of naive and time-neighborhood-matched similarity contrasts. Naive within-minus-between contrast and time-neighborhood-matched contrast are shown as a function of true stimulus-effect strength in simulated repeated-stimulus data. Error bars indicate standard deviations across 500 simulations. At zero stimulus-effect strength, the naive contrast was strongly negative, whereas the time-neighborhood-matched contrast was close to zero. As true stimulus-specific structure increased, the time-neighborhood-matched contrast increased monotonically.

These results show that time-neighborhood matching did not simply eliminate all structured differences between repeated and non-repeated stimuli. Instead, it reduced the temporal-distance bias observed in the null simulation while preserving sensitivity to true stimulus-specific structure. The monotonic relationship between stimulus-effect strength and time-neighborhood-matched contrast supports the use of the time-neighborhood-matched contrast as a diagnostic measure in repeated-stimulus fMRI pattern similarity workflows.

At the strongest effect levels, the matched contrast exceeded the naive contrast because the temporal component opposed part of the true stimulus-specific signal in the uncorrected estimate. Removing this opposing temporal contribution therefore increased the recovered contrast.

3.3. Extended BOLD-like Sensitivity Simulation

The extended simulation was used to determine whether the behavior of the time-neighborhood-matched contrast was preserved after introducing hemodynamic convolution, spatial dependence, and additional structured nuisance variation.

Under the extended null condition, the naive within-minus-between contrast remained strongly biased. Across 500 simulations, its mean value was -0.1231 ± 0.0071 , whereas the mean time-neighborhood-matched contrast was reduced to -0.0053 ± 0.0048 . Thus, temporal matching reduced the absolute null bias by approximately 95.7%, although a small residual negative bias remained.

The time-neighborhood-matched contrast increased strictly monotonically with stimulus-effect strength (Figure 3). Its mean value was -0.0053 at strength 0.00, -0.0025 at strength 0.10, 0.0132 at strength 0.25, 0.0633 at strength 0.50, 0.1330 at strength 0.75, and 0.2091 at strength 1.00. In contrast, the naive contrast remained negative up to strength 0.50 and became positive only at stronger stimulus-effect levels.

The same pattern was observed in the extended BOLD-like simulation, where temporal adjustment recovered a larger stimulus-specific contrast than the naive estimator at high effect strength.

These results show that the principal behavior of the workflow was preserved under a more realistic BOLD-like signal model. Hemodynamic convolution and spatially correlated noise introduced a small residual bias, but they did not eliminate the monotonic sensitivity of the time-neighborhood-matched contrast to true stimulus-specific structure.

3.4. Matching-Window Ablation

The matching-window ablation showed that the main conclusions were robust to the width of the local temporal neighborhood. Under the null condition, the time-neighborhood-matched contrast was -0.0013 ± 0.0018 for the ± 1 window, -0.0035 ± 0.0017 for ± 2 , and -0.0065 ± 0.0017 for ± 3 . These values corresponded to reductions of 99.0%, 97.4%, and 95.1%, respectively, relative to the naive null contrast of -0.1310 .

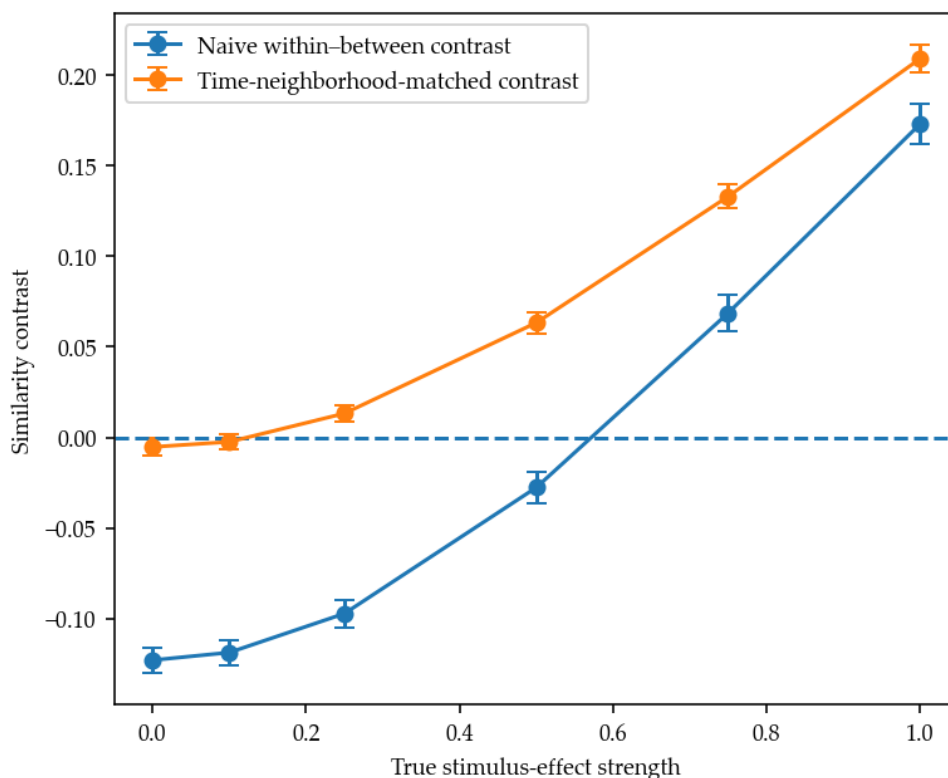


Figure 3. Extended BOLD-like sensitivity analysis including hemodynamic convolution and spatially correlated noise. Naive within-minus-between contrast and time-neighborhood-matched contrast are shown as a function of true stimulus-effect strength. Continuous event-related responses were convolved with a canonical double-gamma hemodynamic response function and combined with spatially correlated noise, AR(1) temporal variation, slow drift, and physiological-like nuisance components. Error bars indicate standard deviations across 500 simulations.

For all three neighborhood widths, the time-neighborhood-matched contrast increased strictly monotonically across the six stimulus-effect strengths (Figure 4). At strength 1.00, the mean time-neighborhood-matched contrasts were 0.3021 for ± 1 , 0.3001 for ± 2 , and 0.2973 for ± 3 . Thus, wider neighborhoods did not improve sensitivity to the stimulus-specific component but produced progressively larger residual null bias. The ± 1 neighborhood was therefore retained as the primary specification because it minimized temporal mismatch and provided the smallest residual bias.

3.5. Parameter-Sensitivity Analysis

The parameter-sensitivity analysis showed that the temporal-confounding effect was not specific to the original simulation setting. Across all 18 combinations of AR(1) coefficient, voxel count, and drift specification, the naive within-minus-between contrast remained displaced from zero. Its mean value ranged from -0.0487 to -0.1311 .

The magnitude of the naive bias increased with stronger temporal autocorrelation and with the inclusion of the sinusoidal drift component. Averaged across voxel counts and drift specifications, the mean naive contrast was -0.0655 for $\rho = 0.3$, -0.0852 for $\rho = 0.6$, and -0.1167 for $\rho = 0.9$.

In contrast, the time-neighborhood-matched contrast remained close to zero across the full parameter grid, ranging from -0.0017 to -0.0004 . This corresponded to a reduction of 98.0% to 99.6% in absolute null bias relative to the naive contrast.

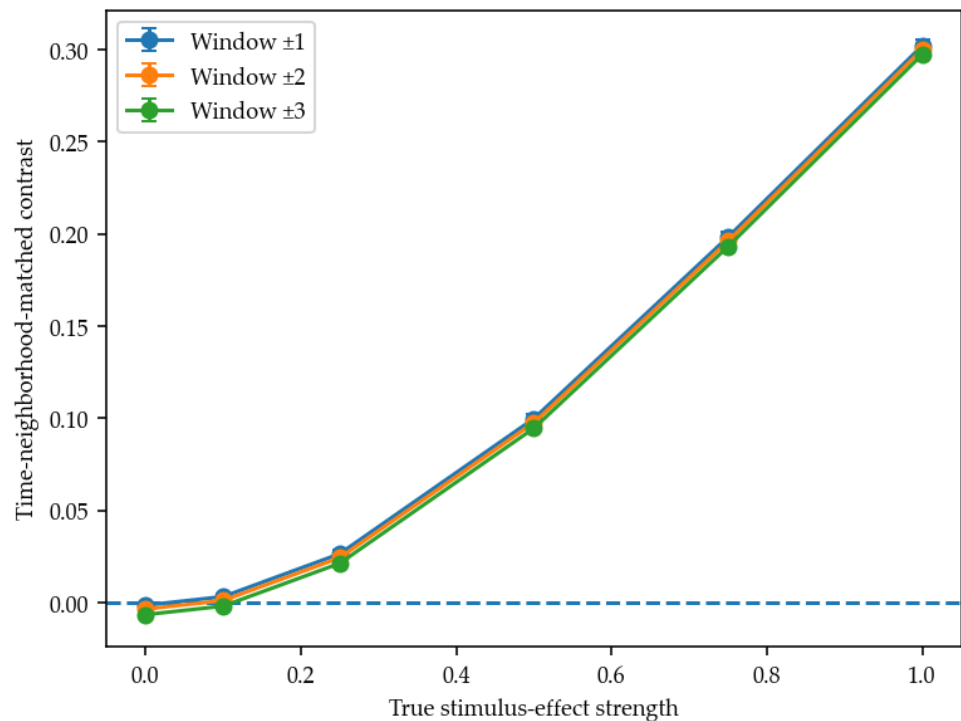


Figure 4. Matching-window ablation analysis. Mean time-neighborhood-matched similarity contrast is shown as a function of true stimulus-effect strength for neighborhood widths of ± 1 , ± 2 , and ± 3 events. For a target within-stimulus distance d , each baseline was calculated from pooled between-stimulus pairs within the corresponding local temporal neighborhood, excluding the target distance d itself. Error bars indicate standard deviations across 500 simulations.

The number of voxel features had little influence on the mean contrasts but affected their variability. Increasing the voxel count from 500 to 10,000 reduced the mean standard deviation of the naive contrast from approximately 0.0053 to 0.0012 and the mean standard deviation of the time-neighborhood-matched contrast from approximately 0.0047 to 0.0011. These results, visualized in Figure 5, indicate that the magnitude of the naive bias depends on temporal autocorrelation and drift structure, whereas the effectiveness of the time-neighborhood-matched contrast is robust across a broad range of parameter settings.

3.6. Comparison with Distance-Adjusted Regression

The naive contrast showed a substantial negative bias under the null condition, with a mean value of -0.1308 ± 0.0021 . Both temporal-bias-control approaches substantially reduced this displacement. The mean time-neighborhood-matched contrast was -0.0014 ± 0.0018 , whereas the mean coefficient from the distance-adjusted regression model was 0.0042 ± 0.0017 .

Relative to the naive contrast, the matched estimator reduced the absolute null bias by 98.95%, while the regression-based estimator reduced it by 96.79%. The corresponding 95% confidence intervals were $[-0.0015, -0.0012]$ for the time-neighborhood-matched contrast and $[0.0040, 0.0044]$ for the regression coefficient. The two estimators showed small residual biases in opposite directions. The local matching procedure retained a slight negative residual, whereas the global cubic regression retained a slight positive residual, suggesting that the two approaches approximate the nonlinear similarity–distance relationship differently.

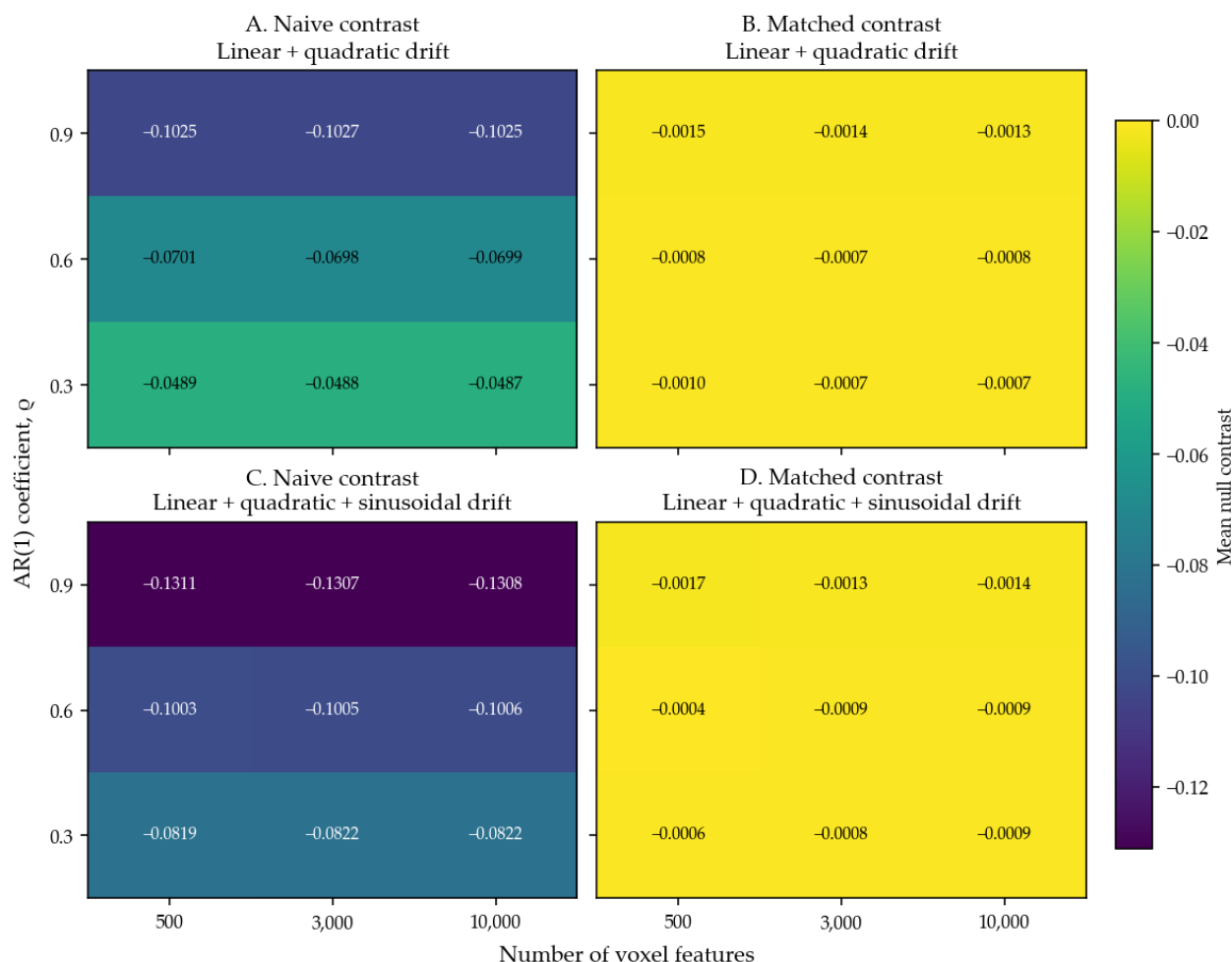


Figure 5. Parameter-sensitivity analysis of null contrast estimates across AR(1) coefficients, voxel counts, and drift specifications. Panels (A,B) show the naive and time-neighborhood-matched contrasts, respectively, for simulations including linear and quadratic drift terms. Panels (C,D) show the corresponding contrasts after addition of a sinusoidal drift component. Each cell represents the mean across 500 null simulations.

Table 1 summarizes the comparison between naive contrast, time-neighborhood-matched contrast, and distance-adjusted regression across 500 null simulations.

Table 1. Comparison of the naive contrast, time-neighborhood-matched contrast, and distance-adjusted regression.

Method	Mean Null Estimate	SD	Absolute Bias	Bias Reduction
Naive contrast	-0.1308	0.0021	0.1308	—
Time-neighborhood-matched contrast, $w = 1$	-0.0014	0.0018	0.0014	98.95%
Distance-adjusted regression	0.0042	0.0017	0.0042	96.79%

The regression model explained a large proportion of the pairwise-similarity variation ($R^2 = 0.876$ on average), but it retained a small positive residual bias. In this simulation, the time-neighborhood-matched contrast produced the smallest absolute residual bias. These results show that time-neighborhood matching is not the only available strategy for temporal-bias control, but rather a simple local alternative that performs comparably to global distance-adjusted regression.

Local matching may be preferable when only narrow temporal neighborhoods are considered credible, whereas regression adjustment may be useful when a global similarity–distance function can be specified adequately.

3.7. OpenNeuro Case Study

The aim of this case study was to examine whether the temporal-confounding pattern observed in simulation also appeared in real repeated-stimulus fMRI data. The analysis was therefore treated as a descriptive real-data demonstration of the workflow, not as a population-level test of music-evoked representational reliability in the five-participant dataset.

3.7.1. Naive Pattern Similarity

Across all five participants and six test runs, the mean within-stimulus similarity was -0.0963 , whereas the mean between-stimulus similarity was -0.0141 , yielding a naive within-minus-between contrast of -0.0822 . Thus, without temporal control, repeated presentations of the same stimulus appeared less similar to each other than different stimuli.

This result indicates that a simple test–retest pattern similarity approach is not sufficient for this repeated naturalistic fMRI design. Because within-stimulus comparisons occurred only at specific temporal distances, the naive contrast was potentially confounded by the temporal structure of the run.

3.7.2. Pattern Similarity

Mean pattern similarity varied systematically as a function of temporal distance between events (Figure 6). Between-stimulus similarity was high for temporally adjacent events, decreased at intermediate distances, and increased again at longer distances. For example, mean between-stimulus similarity was 0.3985 at distance 1, decreased to -0.2342 at distance 15, and increased to 0.2521 at distance 34.

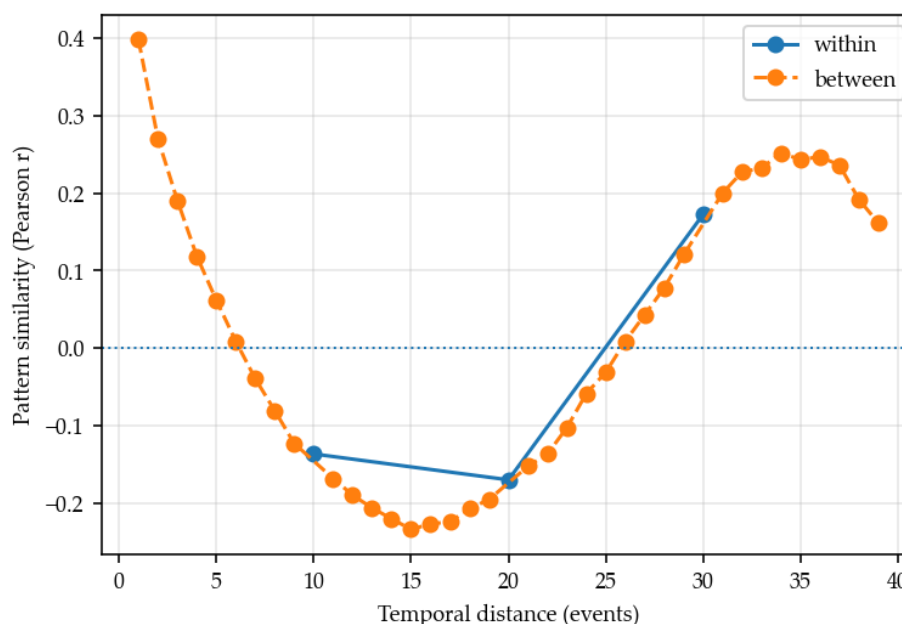


Figure 6. Pattern similarity as a function of temporal distance between events. Mean Pearson correlation coefficients are shown separately for within-stimulus comparisons and between-stimulus comparisons. Within-stimulus similarities were available only at temporal distances of 10, 20, and 30 events because the same ten test stimuli were repeated in a fixed order within each run.

The non-monotonic between-stimulus profile is consistent with an interaction between the fixed event order and slow run-level temporal structure, including drift and autocorrelation. Because individual temporal-distance cells sample different combinations of early, middle, and late run positions, low-frequency run-level variation can produce an oscillating similarity-versus-distance profile rather than a simple monotonic decay.

Within-stimulus similarities were available only at distances 10, 20, and 30 events because the same ten stimuli were repeated in the same order. Mean within-stimulus similarity was -0.1367 at distance 10, -0.1701 at distance 20, and 0.1727 at distance 30. These results show that pattern similarity was shaped by temporal position and run structure, not only by stimulus identity.

3.7.3. Time-Neighborhood Matching

To reduce temporal-distance confounding, within-stimulus comparisons were contrasted with nearby between-stimulus comparisons at adjacent temporal distances (Figure 7). After this correction, the mean time-neighborhood-matched contrast became small and positive: 0.0099 across all participant-run-distance cells.

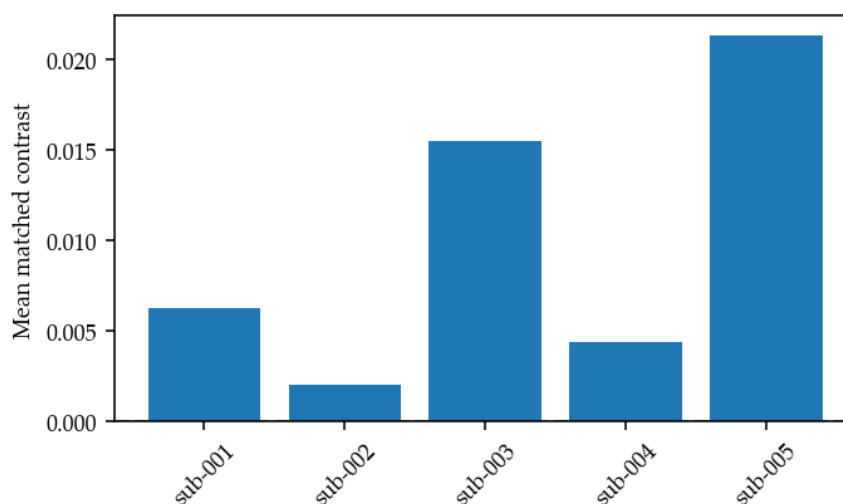


Figure 7. Subject-level time-neighborhood-matched similarity contrast. Bars show the mean time-neighborhood-matched contrast for each participant, calculated as within-stimulus similarity minus nearby between-stimulus similarity after matching temporal distance. Positive values indicate higher similarity for repeated presentations than for temporally adjacent non-repeated stimuli.

The mean time-neighborhood-matched contrast was positive for all five participants: sub-001 = 0.0063 , sub-002 = 0.0020 , sub-003 = 0.0155 , sub-004 = 0.0043 , and sub-005 = 0.0213 . By temporal distance, the mean time-neighborhood-matched contrast was 0.0083 for distance 10, 0.0050 for distance 20, and 0.0163 for distance 30.

The inter-participant spread was treated descriptively because it may reflect both genuine heterogeneity and residual noise associated with the small sample and light-weight preprocessing. With only five participants, these contributions could not be separated reliably.

At the subject level, the mean time-neighborhood-matched contrast was 0.0099 (range: 0.0020 – 0.0213). The subject-level values are displayed after rounding to four decimal places, whereas the reported group mean was calculated from the unrounded subject-level estimates. The mean time-neighborhood-matched contrast was positive for all five participants, but the case study was treated descriptively and not used for population-level inference.

3.7.4. Cross-Validated Functionally Selected Voxel Analysis

The cross-validated top-k analysis (Figure 8) indicated that inclusion of large numbers of weakly responsive voxel features attenuated the residual time-neighborhood-matched contrast. In the common whole-EPI mask, which contained a mean of 126,867 voxels across participants, the mean naive contrast was -0.0735 , whereas the mean time-neighborhood-matched contrast was 0.0111 . This value differs slightly from the original whole-EPI estimate of 0.0099 because the top-k sensitivity analysis used resampled runs and a common intersection mask as an internal reference, whereas the primary analysis used run-specific EPI masks.

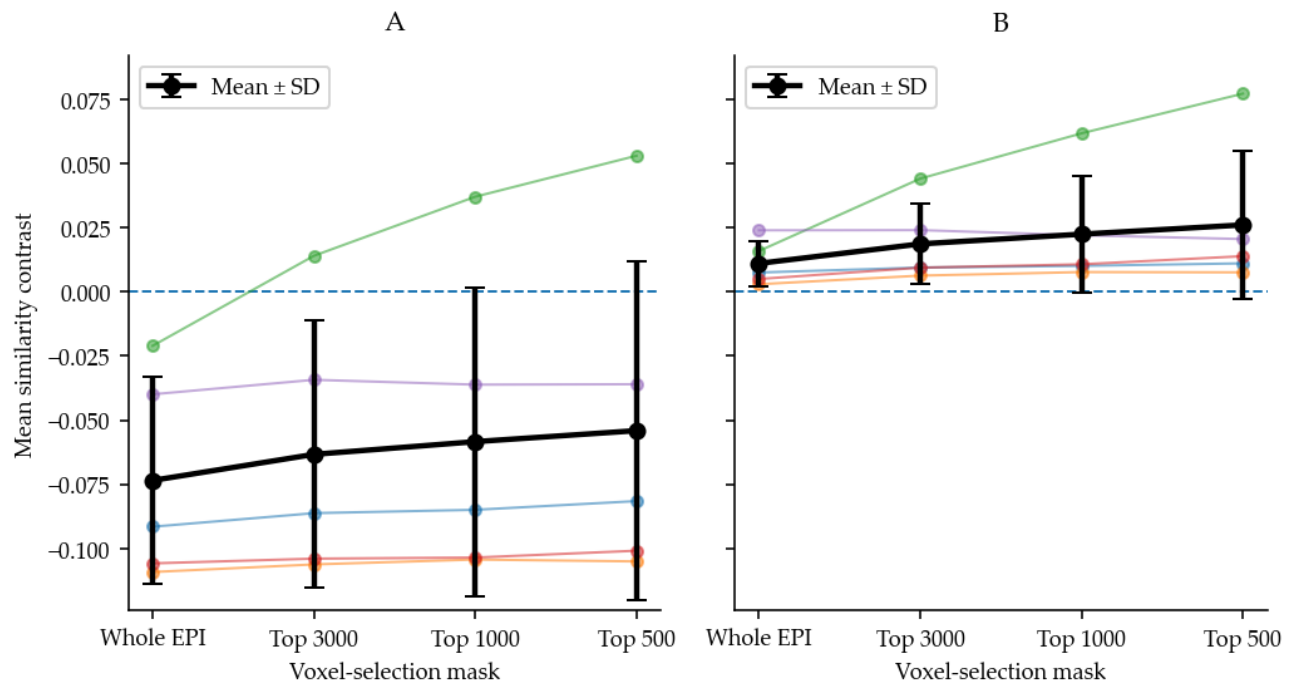


Figure 8. Cross-validated functional voxel-selection analysis in the OpenNeuro case study. Participant-level and group-mean naive within-minus-between contrasts (A) and time-neighborhood-matched contrasts (B) are shown for the common whole-EPI mask and for cross-validated masks containing the top 3000, 1000, and 500 most responsive voxels. Voxel ranking was performed using five training runs and evaluated in the held-out run. Thin colored lines represent participant-level estimates averaged across six held-out runs, whereas the black line and error bars represent the mean $\pm$ standard deviation across five participants. Restricting the analysis to functionally responsive voxels increased the time-neighborhood-matched contrast, while the naive contrast remained negative under every mask specification.

The time-neighborhood-matched contrast increased as the analysis was restricted to the most responsive voxels selected from independent training runs. Mean time-neighborhood-matched contrasts were 0.0187 for the top 3000 voxels, 0.0225 for the top 1000 voxels, and 0.0261 for the top 500 voxels. The corresponding naive contrasts remained negative at -0.0634 , -0.0584 , and -0.0541 , respectively. The time-neighborhood-matched contrast was positive for all five participants under every mask specification.

Thus, cross-validated functional voxel selection increased the magnitude of the residual stimulus-specific contrast, with the top-500 estimate approximately 2.4 times larger than the common whole-EPI estimate. Nevertheless, the absolute magnitude remained small. This value should be interpreted as a residual diagnostic signal under a simplified direct similarity analysis, not as an estimate of representational reliability or as directly

comparable with model-based RSA effect sizes reported by Nakai et al. [19]. However, the persistent negative naive contrast across all masks showed that restricting the analysis to responsive voxels did not eliminate temporal-distance confounding.

3.7.5. Summary

Overall, the case-study analyses showed that temporal structure strongly influenced pattern-similarity estimates in this repeated music-fMRI design. Naive within-versus-between similarity produced a strongly negative contrast, whereas time-neighborhood matching yielded a small positive residual contrast. Because the case study included only five participants and the fixed-order design does not support a straightforward exchangeable permutation scheme, the residual effect was interpreted descriptively rather than inferentially. Descriptive case-study statistics are provided in Table 2.

Table 2. Descriptive summary of the OpenNeuro case-study analysis.

Analysis	Metric	Value
Naive similarity	Within-stimulus mean	−0.0963
	Between-stimulus mean	−0.0141
	Within-minus-between contrast	−0.0822
Time-neighborhood-matched similarity	Mean contrast	0.0099
	Distance 10 contrast	0.0083
	Distance 20 contrast	0.0050
	Distance 30 contrast	0.0163
Subject-level contrast	Mean	0.0099
	Range	0.0020–0.0213
	Positive participants	5/5

3.8. Summary of Workflow Behavior

Across simulations and the open-data case study, the workflow consistently showed that naive repeated-stimulus pattern similarity can be strongly affected by temporal structure. In the null simulation, where no true stimulus-specific pattern was present, the naive contrast was markedly negative (−0.1309), whereas the time-neighborhood-matched contrast was close to zero (−0.0013). This confirmed that fixed repetition order, drift, and autocorrelated temporal components can generate misleading within-versus-between similarity effects in the absence of stimulus-specific structure.

The sensitivity simulation showed that time-neighborhood matching did not remove genuine stimulus-specific information. As stimulus-effect strength increased from 0.00 to 1.00, the time-neighborhood-matched contrast increased monotonically from −0.0013 to 0.3024. This indicates that the time-neighborhood-matched contrast reduced temporal confounding while remaining sensitive to true repeated-stimulus structure.

The OpenNeuro music-fMRI case study showed a pattern closer to the null or weak-effect simulations than to the strong-effect simulations. The naive contrast was strongly negative (−0.0822), whereas the time-neighborhood-matched contrast was small and positive (0.0099). Cross-validated functional voxel selection increased the time-neighborhood-matched contrast to 0.0187–0.0261, indicating partial dilution in the whole-EPI analysis, while the naive contrast remained negative across all mask specifications. Because the analysis involved only five participants and the fixed-order stimulus sequence prevented construction of a straightforward exchangeable permutation test, the residual time-neighborhood-matched contrast was interpreted descriptively rather than as population-level evidence of stimulus-specific reliability.

A summary of the main results is provided in Table 3.

Table 3. Summary of simulation and open-data case-study results.

Analysis Block	Condition/Metric	Naive Contrast	Time-Neighborhood-Matched Contrast	Additional Result	Interpretation
Null simulation	No stimulus-specific pattern	−0.1309	−0.0013	Bias reduction: 99.0%	Temporal structure alone produced a misleading naive contrast; matching reduced it to near zero.
Sensitivity simulation	Stimulus strength 0.10	−0.1258	0.0032	—	Very weak true structure was mostly obscured in the naive contrast.
Sensitivity simulation	Stimulus strength 0.25	−0.0999	0.0263	—	Time-neighborhood-matched contrast began to track true stimulus structure.
Sensitivity simulation	Stimulus strength 0.50	−0.0179	0.0999	—	Matching preserved a clear stimulus-specific component.
Sensitivity simulation	Stimulus strength 0.75	0.0920	0.1983	—	Both contrasts became positive, but time-neighborhood-matched contrast remained larger.
Sensitivity simulation	Stimulus strength 1.00	0.2082	0.3024	—	Strong stimulus structure was clearly preserved after matching.
Extended BOLD-like simulation	Null condition	−0.1231	−0.0053	Bias reduction: 95.7%	HRF convolution and spatial dependence left only a small residual bias after matching.
Extended BOLD-like simulation	Stimulus strength 1.00	0.1729	0.2091	Strictly monotonic across all six levels	Matching remained sensitive to true stimulus-specific structure.
OpenNeuro case study	Real music-fMRI data	−0.0822	0.0099	Positive matched contrast in all 5 participants	Strong temporal confounding remained, with only weak residual stimulus-specific similarity after matching.
OpenNeuro top- <i>k</i> sensitivity	Top 500 voxels	−0.0541	0.0261	Positive in 5/5 participants	Functional selection increased the residual time-neighborhood-matched contrast but did not remove temporal confounding.

4. Discussion

This study developed and tested a simulation-supported workflow for detecting temporal confounding in repeated-stimulus fMRI pattern similarity analysis. The main finding is that temporal structure alone can generate a misleading naive within-versus-between similarity contrast. In the null simulation, where no true stimulus-specific voxel pattern was present, the naive contrast was strongly negative. After time-neighborhood matching, this apparent effect was reduced to near zero. The sensitivity simulation further showed that the time-neighborhood-matched contrast increased monotonically when true stimulus-specific structure was introduced. Finally, the OpenNeuro music-fMRI case study demonstrated that the same issue can arise in real repeated naturalistic fMRI data: the naive contrast was strongly negative, whereas the time-neighborhood-matched contrast was small and positive. Because the case study was limited to five participants and the fixed-order design did not support a straightforward exchangeable permutation test, the real-data result was interpreted descriptively.

4.1. Temporal Structure Can Mimic or Distort Stimulus-Specific Similarity

The null simulation provides the clearest methodological result. Because the simulated data contained temporal drift, autocorrelated components, and noise but no stable stimulus-specific pattern, any non-zero within-versus-between contrast had to arise from

the interaction between temporal structure and the fixed repetition order. The strongly negative naive contrast therefore demonstrates that direct repeated-stimulus similarity estimates can be biased even when stimulus identity has no effect on the generated patterns.

This result is consistent with the broader understanding that the BOLD signal is temporally structured and that autocorrelation, drift, and hemodynamic smoothing affect fMRI time series [6–9,18]. However, the present analysis focuses on a more specific problem: temporal structure can distort pairwise pattern similarity contrasts when the distribution of temporal distances differs between within-stimulus and between-stimulus comparisons. In such designs, a naive similarity contrast may reflect event spacing and run-level dynamics rather than stimulus-specific reliability. This is conceptually related to prior concerns that representational similarity estimates can be biased by task structure, repetition structure, and sampling properties [1,4].

4.2. Time-Neighborhood Matching Reduced Temporal Bias While Preserving True Structure

The sensitivity simulation showed that time-neighborhood matching did not simply remove all structured differences between repeated and non-repeated stimuli. When the true stimulus-effect strength was zero, the time-neighborhood-matched contrast was close to zero. As the true stimulus-specific component increased, the time-neighborhood-matched contrast increased monotonically. This pattern indicates that the matching procedure reduced temporal-distance bias while preserving sensitivity to genuine stimulus-specific structure.

The extended BOLD-like simulation further supported this interpretation. After introducing hemodynamic convolution, spatially correlated voxel structure, spatially correlated noise, AR(1) variation, slow drift, and physiological-like nuisance components, the naive contrast remained strongly biased under the null condition. Time-neighborhood matching reduced this bias by approximately 95.7%, from -0.1231 to -0.0053 , and the time-neighborhood-matched contrast continued to increase monotonically with stimulus-effect strength. The small residual null bias indicates that local temporal matching does not fully model the physiological dynamics of the BOLD response. It should therefore be interpreted as a bias-reduction and diagnostic procedure rather than as a complete correction for hemodynamic carryover.

This is important because a temporal correction method should not merely suppress all repeated-stimulus effects. A useful diagnostic workflow should distinguish between two situations: a contrast produced by temporal structure alone and a contrast that remains when a true stable stimulus component is present. The sensitivity analysis supports this distinction. In contrast, the naive within-minus-between contrast remained distorted at weak and moderate stimulus-effect strengths, demonstrating that uncorrected similarity measures may be difficult to interpret even when some true stimulus-specific structure exists.

Temporal matching may therefore increase, rather than merely reduce, a contrast when run-level temporal structure opposes the true stimulus-specific component.

4.3. OpenNeuro Case Study Illustrates the Practical Relevance of the Workflow

The OpenNeuro music-fMRI case study showed a pattern consistent with the methodological concern identified in simulation. The naive within-minus-between contrast was strongly negative, implying lower similarity for repeated presentations of the same music clip than for different clips. This result is counterintuitive if interpreted as a simple reliability measure. However, the temporal-distance analysis showed that similarity varied strongly with event spacing and run position. After time-neighborhood matching, the

contrast became small and positive and was therefore interpreted as a weak descriptive residual effect.

The cross-validated functional voxel-selection analysis further showed that the small residual time-neighborhood-matched contrast was partly attenuated by inclusion of weakly responsive voxels. Restricting the analysis to independently selected responsive voxels increased the time-neighborhood-matched contrast from 0.0111 in the common whole-EPI reference to 0.0187–0.0261 in the top-k masks. However, the naive contrast remained negative on average under every mask specification. Thus, functional targeting increased the apparent stimulus-specific component but did not remove the temporal-distance confounding identified by the workflow.

This result should not be interpreted as evidence against the original findings of Nakai et al. [19,20]. The original study used model-based analyses to characterize categorical and feature-based representations of music in the brain, whereas the present case study applied a deliberately simplified direct pattern-similarity workflow. The current analysis therefore addresses a different methodological question: how naive repeated-stimulus similarity behaves when temporal distance is not explicitly controlled. The OpenNeuro dataset served as a real-data demonstration of the workflow, not as a basis for population-level inference about music perception.

4.4. Implications for Repeated-Stimulus fMRI Workflows

The results suggest that repeated-stimulus fMRI pattern similarity analyses should routinely inspect temporal-distance effects before interpreting within-versus-between contrasts. In fixed-order or partially structured designs, within-stimulus comparisons may occur at restricted temporal distances, whereas between-stimulus comparisons may span a broader temporal range. This imbalance can lead to misleading estimates of stimulus-specific reliability.

A practical workflow should therefore include at least three diagnostic steps:

1. Similarity should be plotted as a function of temporal distance separately for within-stimulus and between-stimulus comparisons;
2. Naive within-versus-between similarity should be compared with a time-neighborhood-matched contrast;
3. Inferential procedures should be selected only after verifying that their exchangeability or model assumptions are compatible with the experimental design. In strictly fixed-order designs, conventional label permutation may be invalid because it alters the temporal-distance structure of same-stimulus pairs.

The proposed workflow is intentionally simple. It is not intended to replace full fMRI modeling, hemodynamic response estimation, prewhitening, or more advanced representational modeling approaches. Rather, it provides a diagnostic layer that can be applied before drawing conclusions from direct pattern similarity contrasts. This is particularly relevant for reanalyses of open datasets and naturalistic fMRI datasets, where researchers may apply secondary analyses that differ from the original modeling strategy.

4.5. Limitations

Several limitations should be noted.

First, although the extended simulation incorporated hemodynamic convolution, spatially correlated noise, AR(1) variation, slow drift, and physiological-like oscillatory components, it remained a simplified generative model. It did not reproduce the full complexity of fMRI acquisition, subject-specific hemodynamic variability, motion-related artifacts, tissue-specific autocorrelation, or standard preprocessing procedures. The extended model was intended to test whether the main workflow behavior remained stable

after introducing several important temporal and spatial properties of BOLD data, rather than to provide a complete simulator of fMRI acquisition.

Second, the OpenNeuro case study used lightweight native-space preprocessing because the available derivatives did not include preprocessed 4D BOLD time series or confound time-series files. The analysis therefore should not be interpreted as an optimized neural decoding or representational analysis pipeline. Its purpose was to demonstrate how temporal confounding appears under a transparent and reproducible direct similarity workflow. The cross-validated top-k analysis additionally required resampling all test runs to a common within-participant grid. This harmonized voxel indexing across runs but did not constitute full inter-run registration or motion correction. Consequently, spatial correspondence of individual voxel features across runs should be regarded as approximate. More comprehensive nuisance regression, including motion parameters and physiological-noise components such as aCompCor, might reduce part of the observed run-level temporal structure and attenuate the magnitude of the naive contrast. However, such preprocessing would not by itself remove the imbalance in temporal-distance distributions imposed by the fixed stimulus order.

Third, the real-data case study included only five participants. For this reason, the OpenNeuro analysis was treated descriptively and was not used to support population-level inference. The inferential strength of the paper comes primarily from the simulations, where the presence or absence of stimulus-specific structure was controlled by design. The real-data analysis illustrates practical relevance rather than providing a definitive empirical estimate of music-evoked reliability.

Fourth, time-neighborhood matching is only one possible correction strategy. It reduces temporal-distance imbalance by comparing within-stimulus pairs with nearby between-stimulus pairs, but it does not model the full nonlinear temporal structure of the BOLD signal. Alternative approaches could include regression with nonlinear temporal basis functions, cross-validated distance measures, or full model-based estimation of event responses. The appropriate strategy should be selected according to experimental design and the scientific question.

Fifth, permutation-based inference is not straightforward in the strictly fixed-order design considered here. Same-stimulus pairs occur only at temporal distances 10, 20, and 30. Non-trivial permutations of stimulus labels alter this distance structure, whereas permutations that preserve the original within-stimulus distances leave the same-stimulus pairing structure unchanged. Consequently, the exchangeability assumptions required for a conventional permutation test are not satisfied. The time-neighborhood-matched contrast should therefore be interpreted as a diagnostic and bias-reduction estimator. Inferential testing would require a randomized or counterbalanced design, independent runs with varying stimulus orders, or a separately justified generative null model.

5. Conclusions

This study shows that naive repeated-stimulus fMRI pattern similarity can be strongly biased by temporal structure. In the null simulation, a misleading within-minus-between contrast appeared even though no true stimulus-specific pattern was present. Time-neighborhood matching reduced this bias to near zero, while the sensitivity simulation showed that the time-neighborhood-matched contrast remained responsive to increasing true stimulus-specific structure.

The OpenNeuro music-fMRI case study illustrated the same practical issue in real repeated-stimulus data: the naive contrast was strongly affected by temporal structure, whereas the time-neighborhood-matched contrast was small and positive. Because the fixed-order design did not permit a straightforward exchangeable permutation test, this residual effect was interpreted descriptively. Cross-validated selection of functionally re-

sponsive voxels increased the residual time-neighborhood-matched contrast, indicating that broad whole-EPI masking partly diluted stimulus-related information. These findings support the use of simulation-supported temporal diagnostics, including temporal-distance plots, time-neighborhood-matched contrasts, parameter-sensitivity analysis, and comparison with alternative adjustment methods, before interpreting repeated-stimulus fMRI pattern similarity as evidence of neural reliability or stimulus-specific representation.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The fMRI data analyzed in the OpenNeuro case study are publicly available in the OpenNeuro repository under dataset identifier ds003720, version 1.0.0 (DOI: 10.18112/openneuro.ds003720.v1.0.0), available at: <https://openneuro.org/datasets/ds003720/versions/1.0.0> (accessed on 26 April 2026). The simulation outputs and intermediate analysis tables were generated as part of the present study and are available from the author upon reasonable request.

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

References

1. Cai, M.B.; Schuck, N.W.; Pillow, J.W.; Niv, Y. Representational Structure or Task Structure? Bias in Neural Representational Similarity Analysis and a Bayesian Method for Reducing Bias. *PLoS Comput. Biol.* **2019**, *15*, e1006299. <https://doi.org/10.1371/journal.pcbi.1006299>.
2. Diedrichsen, J.; Kriegeskorte, N. Representational Models: A Common Framework for Understanding Encoding, Pattern-Component, and Representational-Similarity Analysis. *PLoS Comput. Biol.* **2017**, *13*, e1005508. <https://doi.org/10.1371/journal.pcbi.1005508>.
3. Kriegeskorte, N. Pattern-Information Analysis: From Stimulus Decoding to Computational-Model Testing. *Neuroimage* **2011**, *56*, 411–421. <https://doi.org/10.1016/j.neuroimage.2011.01.061>.
4. Mazurchuk, S.; Conant, L.L.; Tong, J.-Q.; Binder, J.R.; Fernandino, L. Stimulus Repetition and Sample Size Considerations in Item-Level Representational Similarity Analysis. *Lang. Cogn. Neurosci.* **2024**, *39*, 1161–1172. <https://doi.org/10.1080/23273798.2023.2232903>.
5. Yang, L.; Wei, J.; Li, Y.; Wang, B.; Guo, H.; Yang, Y.; Xiang, J. Test-Retest Reliability of Synchrony and Metastability in Resting State fMRI. *Brain Sci.* **2022**, *12*, 66. <https://doi.org/10.3390/brainsci12010066>.
6. Groen, I.I.; Piantoni, G.; Montenegro, S.; Flinker, A.; Devore, S.; Devinsky, O.; Doyle, W.; Dugan, P.; Friedman, D.; Ramsey, N.F.; et al. Temporal Dynamics of Neural Responses in Human Visual Cortex. *J. Neurosci.* **2022**, *42*, 7562–7580. <https://doi.org/10.1523/jneurosci.1812-21.2022>.
7. Laumann, T.O.; Snyder, A.Z.; Mitra, A.; Gordon, E.M.; Gratton, C.; Adeyemo, B.; Gilmore, A.W.; Nelson, S.M.; Berg, J.J.; Greene, D.J.; et al. On the Stability of BOLD fMRI Correlations. *Cereb. Cortex* **2017**, *27*, 4719–4732. <https://doi.org/10.1093/cercor/bhw265>.
8. Müller, K.; Mildner, T.; Lohmann, G.; Von Cramon, D.Y. Investigating the Stimulus-dependent Temporal Dynamics of the BOLD Signal Using Spectral Methods. *J. Magn. Reson. Imaging* **2003**, *17*, 375–382. <https://doi.org/10.1002/jmri.10268>.
9. Woolrich, M.W.; Ripley, B.D.; Brady, M.; Smith, S.M. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage* **2001**, *14*, 1370–1386. <https://doi.org/10.1006/nimg.2001.0931>.
10. Mumford, J.A.; Turner, B.O.; Ashby, F.G.; Poldrack, R.A. Deconvolving BOLD Activation in Event-Related Designs for Multi-voxel Pattern Classification Analyses. *NeuroImage* **2012**, *59*, 2636–2643. <https://doi.org/10.1016/j.neuroimage.2011.08.076>.
11. Saarimäki, H. Naturalistic Stimuli in Affective Neuroimaging: A Review. *Front. Hum. Neurosci.* **2021**, *15*, 675068. <https://doi.org/10.3389/fnhum.2021.675068>.
12. Sonkusare, S.; Breakspear, M.; Guo, C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn. Sci.* **2019**, *23*, 699–714. <https://doi.org/10.1016/j.tics.2019.05.004>.

13. Zhang, Y.; Kim, J.-H.; Brang, D.; Liu, Z. Naturalistic Stimuli: A Paradigm for Multiscale Functional Characterization of the Human Brain. *Curr. Opin. Biomed. Eng.* **2021**, *19*, 100298. <https://doi.org/10.1016/j.cobme.2021.100298>.
14. Haumann, N.T.; Petersen, B.; Vuust, P.; Brattico, E. Age Differences in Central Auditory System Responses to Naturalistic Music. *Biol. Psychol.* **2023**, *179*, 108566. <https://doi.org/10.1016/j.biopsycho.2023.108566>.
15. Tervaniemi, M. The Neuroscience of Music—towards Ecological Validity. *Trends Neurosci.* **2023**, *46*, 355–364. <https://doi.org/10.1016/j.tins.2023.03.001>.
16. Trost, W.; Trevor, C.; Fernandez, N.; Steiner, F.; Frühholz, S. Live Music Stimulates the Affective Brain and Emotionally Entrain Listeners in Real Time. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2316306121. <https://doi.org/10.1073/pnas.2316306121>.
17. Istók, E.; Brattico, E.; Jacobsen, T.; Ritter, A.; Tervaniemi, M. ‘I Love Rock ‘n’ Roll’—Music Genre Preference Modulates Brain Responses to Music. *Biol. Psychol.* **2013**, *92*, 142–151. <https://doi.org/10.1016/j.biopsycho.2012.11.005>.
18. Olszowy, W.; Aston, J.; Rua, C.; Williams, G.B. Accurate Autocorrelation Modeling Substantially Improves fMRI Reliability. *Nat. Commun.* **2019**, *10*, 1220. <https://doi.org/10.1038/s41467-019-09230-w>.
19. Nakai, T.; Koide-Majima, N.; Nishimoto, S. Correspondence of Categorical and Feature-Based Representations of Music in the Human Brain. *Brain Behav.* **2021**, *11*, e01936. <https://doi.org/10.1002/brb3.1936>.
20. Nakai, T.; Koide-Majima, N.; Nishimoto, S. Music Genre Neuroimaging Dataset. *Data Brief* **2022**, *40*, 107675. <https://doi.org/10.1016/j.dib.2021.107675>.
21. Nakai, T.; Koide-Majima, N.; Nishimoto, S. Music Genre fMRI Dataset. Available online: <https://openneuro.org/datasets/ds003720/versions/1.0.0> (accessed on 26 April 2026).

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.