Psychometrika (2026), **0**, 1–23
doi:10.1017/psy.2026.10131

PSYCHOMETRIC SOCIETY

APPLICATION AND CASE STUDIES - ORIGINAL

Bayesian Mixed Multidimensional Scaling for Auditory Processing

Giovanni Rebaudo[1,2], Fernando Llanos[3], Bharath Chandrasekaran[4] and Abhra Sarkar[5]

[1]University of Turin, Italy; [2]Collegio Carlo Alberto, Italy; [3]The University of Texas at Austin, USA; [4]Northwestern University, USA; [5]The University of Texas at Austin, USA

Corresponding author: Giovanni Rebaudo; Email: giovanni.rebaudo@unito.it

(Received 7 August 2025 UTC; revised 31 May 2026 UTC; accepted 12 June 2026 UTC)

Abstract

The human brain distinguishes speech sounds by mapping acoustic signals into a latent perceptual space. This space can be estimated via multidimensional scaling (MDS), preserving the similarity structure in lower dimensions. However, individual and group-level heterogeneity, especially between native and non-native listeners, remains poorly understood. Prior approaches often ignore such variability or cannot capture shared structure, limiting principled comparisons. Moreover, the literature often focuses on latent distances rather than the underlying features themselves. To address these issues, we develop a Bayesian mixed MDS method that accounts for both subject- and group-level heterogeneity, allows for the recovery of unique, identifiable latent features, facilitating their biological interpretability, while also determining the effective dimensionality of the latent space in an automated, data-adaptive manner. Simulations and an auditory neuroscience application demonstrate how these features reconstruct observed distances and vary with individual and language background, revealing novel insights.

Keywords: auditory neuroscience; dimensionality reduction; individual heterogeneity; multidimensional scaling; similarity analysis

1. Introduction

Understanding how the brain represents behaviorally relevant information such as speech is a fundamental question in auditory neuroscience. Speech signals are inherently multidimensional, with subtle variations in frequency, intensity, and timing information distinguishing different categories (Caclin et al., 2005, 2006). These acoustic subtleties need to be robustly encoded by the auditory system and mapped onto existing speech representations (e.g., in native speakers) or emergent representations (e.g., in non-native learners). A major goal for auditory neuroscientists is to decipher how these attributes are differentially represented in the brain and how these neural representations are shaped by different auditory experiences across the lifespan. One critical step to elucidate these questions is to quantify the degree of similarity (or dissimilarity) between neural representations of sounds within the same sensory-perceptual representational space. This step is critical because the degree of dissimilarity between neural representations of speech sounds must reflect linguistically relevant differences across languages. For instance, native speakers of Mandarin Chinese rely on subtle syllable-level pitch changes to decode word meanings, exhibiting a more nuanced neural encoding of pitch contours (or lexical

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

tones) than native speakers of languages that are not tonal (e.g., Bidelman et al., 2011; Feng, Gan, et al., 2021; Krishnan et al., 2010; Raizada et al., 2010).

Multidimensional scaling (MDS; Borg & Groenen, 2005; Carroll & Arabie, 1980; Torgerson, 1952, 1958) and its extension to individual difference scaling (INDSCAL; Carroll & Chang, 1970) are two dimensionality reduction techniques that have been extensively used to quantify the degree of similarity between neural representations of sensory stimuli captured with modern neuroimaging technologies. Given data on dissimilarity matrices between pairs of stimuli (e.g., in our work, speech sounds), MDS and INDSCAL can extract lower-dimensional latent features from which a denoised version of the observed dissimilarities can be reconstructed. In the classical literature, these quantities are usually described in geometric terms: the point coordinates of the stimuli in the low-dimensional space form the configuration, and the coordinate axes are called dimensions. In our Bayesian formulation, we treat these coordinates as unobserved random variables endowed with a prior and a posterior distribution, and we refer to them as latent features to emphasize this latent-variable interpretation. Throughout this article, latent features therefore denote the coordinate vectors representing the stimuli in the low-dimensional space, endowed with a probability model. More specifically, the latent features represent the original stimuli as points in a low-dimensional space such that the pairwise (usually Euclidean) distances computed in the reduced space (i.e., from the latent features) match the observed pairwise dissimilarities as well as possible.

Standard MDS methods, when applied to a single dissimilarity matrix, extract common latent features that allow the reconstruction of that matrix. Since they are restricted to handling two-way pairwise dissimilarity data, when multiple subject-specific dissimilarity matrices are available, they are often reduced to a single group-level matrix via complete pooling (e.g., by averaging the distances element-wise) and analyzed with standard MDS, or analyzed by running separate MDS for each subject, followed by post-hoc alignment of the resulting configurations across subjects. In contrast, INDSCAL can accommodate three-way data comprising pairwise dissimilarities for different individuals, allowing for the extraction of individual latent features and denoised distances by re-weighting the common latent features for each subject (Carroll & Chang, 1970). INDSCAL can therefore be used to simultaneously achieve dimensionality reduction, data visualization, and, when appropriate, meaningful interpretation of the inferred latent features. MDS-type models have also been extended to alternating least squares scaling (ALSCAL; Takane et al., 1977; Young et al., 1978) and proximity scaling (PROXSCAL; De Leeuw & Heiser, 1980), which minimize the standardized residual sum of squares (STRESS; Kruskal, 1964a, 1964b; Shepard, 1962), a measure of scaled difference between observed distances and their model reconstructed values. Originally introduced for continuous data, such extensions made MDS more broadly applicable to include non-metric data as well as missing values. MDS has gradually transformed from a descriptive geometric optimization method into a rigorous statistical inference framework, shifting the objective function from minimizing an empirical fit measure, like Kruskal's STRESS, to maximizing a formal likelihood function, which combines the underlying true distance in the latent MDS space and a random error term (Ramsay, 1977; Takane & Carroll, 1981). Asymmetric Mahalanobis distance-based formulations have also been considered (Bradlow & Schmittlein, 2000). Furthermore, MDS-type models based on latent utility functions have been extensively developed for binary choice and ordered preference data for psychometric analysis in marketing applications. Here, the utilities are constructed using choice and individual-specific location vectors in the MDS latent space. Two main strategies exist for this construction: the "vector" approach uses the inner product, and the "unfolding" approach uses the squared Euclidean distances (DeSarbo & Cho, 1989; DeSarbo et al., 1998, 2008; Fong et al., 2010; Park et al., 2008; Zinnes & Wolff, 1977, etc.).

For the rest of the article, we focus mainly on MDS methods for metric-valued continuous data, aligning with the requirements of our motivating application. More recently, such methods have seen widespread use in the neuroscience literature, becoming a staple technique over the last few decades. The notable appeal of MDS in this field may lie in its ability to process distance or dissimilarity matrices directly, thereby obviating the often-involved step of complex modeling of the raw, high-dimensional data sets. Di Liberto et al. (2015), Feng et al. (2019), Feng, Gan, et al. (2021), Khalighinejad et al. (2017),

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

Llanos et al. (2021), Mesgarani et al. (2014), and Zinszer et al. (2016) have used MDS to align acoustic and neural representations of sounds in a common space for comparison. In this body of literature, a higher degree of alignment between acoustic and neural representations is interpreted as a signature of a more robust neural encoding of stimulus signals. Here, by "alignment" we mean the correspondence between the space in which acoustic stimuli can be represented and the latent space in which the brain represents them. These works suggest that the degree of alignment between neural and/or acoustic representations of sounds can be investigated using a reduced set of latent features derived from the primary acoustic cues used to perceive sounds. For example, the perception of Mandarin tones seems to be based on two major acoustic dimensions: pitch height (high vs. low pitch) and pitch direction (rising vs. falling pitch) (Chandrasekaran, Gandour, et al., 2007; Gandour, 1978). Some work (Gandour & Harshman, 1978; Gandour, 1983) has also suggested a possibly relevant third dimension, namely, the magnitude of the slope in the pitch contour. In the last decades, MDS and INDSCAL have also become popular tools to assess the degree of structural alignment between acoustic and neural representations of speech sounds (Chandrasekaran, Gandour, et al., 2007; Chandrasekaran, Krishnan, et al., 2007; Raizada et al., 2010). Much of the attractiveness of these tools lies in their ability to represent global dissimilarity patterns between complex neural signals in a way that is easy to visualize and interpret. While MDS is not exclusively conceived as a technique for dimensionality reduction, using a reduced number of dimensions, prior neuroscience work (Feng, Li, et al., 2021; Llanos et al., 2021; Mesgarani et al., 2014) has been able to capture fine-grained linguistically-relevant differences in sound processing, which suggests that theoretical principles of dimensionality reduction inform the representation of speech sounds in the brain.

Current metric MDS approaches in neuroscience are, however, often used as mere visualization tools. Indeed, joint modeling and uncertainty quantification across groups (e.g., native vs. non-native listeners) are not addressed by the existing methods. Furthermore, most methods do not natively support inference on the number of latent features.

The Bayesian paradigm possesses the machinery to resolve these issues, though they have yet to all be comprehensively addressed in the current literature. Following the foundational work by Oh and Raftery (2001), who established a framework for performing inference and uncertainty quantification on latent coordinates via Markov chain Monte Carlo (MCMC), the Bayesian MDS literature has expanded to include scalable extensions for hyperbolic geometries (Liu et al., 2024; Praturu & Sharpee, 2022) and massive-scale phylogenetic applications (Holbrook et al., 2021). However, these models generally assume population homogeneity, rendering them insufficient for characterizing the complex neuroplasticity and varied speech-sound encoding found across diverse language groups. While domain-specific adaptations have emerged, such as the bioinformatics model by Nguyen and Holmes (2017), they are often restricted to a single latent dimension and fail to account for individual or group-specific variation in the underlying features. Even hierarchical approaches designed to capture heterogeneity, such as the model for orchestral interpretations of Beethoven symphonies by Yanchenko and Hoff (2020), situate variability at the level of denoised distances rather than within the unobserved latent features that induce those distances. Consequently, existing methods lack the multi-dimensional flexibility and feature-level hierarchical structure necessary to quantify the nuanced, group-specific latent representations required for our motivating neuroscience applications.

Indeed, inference about the latent features is challenging due to identifiability issues; while the reconstructed distances are identifiable, the latent features are unique up to transformations, that is, rotations, translations, and reflections. While these indeterminacies may be overlooked for simple visualization, anchoring the model becomes essential for meaningful interpretation and comparison. Specifically, in Bayesian frameworks, the posterior remains invariant under these transformations, necessitating explicit strategies for valid posterior summarization and uncertainty quantification. Common approaches to resolve this symmetry include imposition of strong priors to favor a specific orientation (e.g., DeSarbo et al., 1998), or utilizing post-hoc alignment to map posterior draws to a common reference configuration (e.g., Oh & Raftery, 2001); however, such strategies may not fully resolve more complex symmetries, such as signed permutations of the latent dimensions. Alternatively,

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

employing hard coordinate constraints to lock the configuration during estimation can provide a more concrete mathematical resolution (Bradlow & Schmittlein, 2000; Park et al., 2008); however, this rigidity may lead to sampler inefficiency, leading to poor MCMC mixing and slow convergence. Lin and Fong (2019) addressed identifiability by minimizing a loss function post-MCMC, allowing the sampler to move freely, avoiding the bottlenecks of hard constraints, while ensuring that the final latent features are unique and interpretable.

Building on these prior works, while addressing some of their aforementioned limitations, we propose a flexible Bayesian mixed MDS model for studying dissimilarity across subjects and groups. Our proposal allows performing biologically interpretable inference and straightforward uncertainty quantification for both individual and group-specific distances across stimuli, taking into account the individual idiosyncrasies as well as the heterogeneity between native and non-native Mandarin listeners at the level of the unobserved features. Our proposal thus takes into account two different sources of heterogeneity and does this at the latent feature level and not at the latent distance level. Moreover, by weighting the latent features, our proposal allows the latent axes to be identified uniquely and not be affected by rotation invariance, in contrast to standard MDS and in line with the classical INDSCAL. Importantly, this allows inferring biologically interpretable latent dimensions when such interpretation is meaningful.

Our proposed approach can also infer the effective number of relevant features in a semi-automated data-adaptive way. This is achieved by using cumulative shrinkage priors (Bhattacharya & Dunson, 2011) that impose increasing shrinkage on higher-dimensional latent features. By starting with a generous number of features, and then eliminating redundant higher-dimensional features which are shrunk toward near-zero values on-the-fly using an adaptive MCMC algorithm (following Roberts & Rosenthal, 2009), this approach obviates the need for computationally intensive methods, such as refitting models for different latent dimensions or executing trans-dimensional MCMC steps (e.g., Richardson & Green, 1997). Finally, we employ a post-processing procedure, adapting to recent developments to address similar issues in Bayesian latent factor models (LFMs; Papastamoulis & Ntzoufras, 2022), which maintains excellent sampling efficiency while fully resolving the identifiability issues, including translations and signed permutations of the values of the latent dimensions across MCMC samples, enabling meaningful inference on the latent features coherently across groups and subjects.

We apply the proposed model to assess group-level differences in the neural representation of Mandarin Chinese tones between native Mandarin speakers and monolingual native English speakers. Neural representations of Mandarin tones were extracted from a dataset of frequency-following responses (FFRs) to Mandarin tones (Llanos et al., 2017). The FFR is a scalp-recorded electrophysiological brain component that reflects phase-locked activity from ensembles of neurons along the central auditory nervous system. When the brain is stimulated with a periodic sound, neurons in the auditory system synchronize their oscillatory activity by firing at the same phase of each cycle in the stimulus waveform. This synchronized phase-locked activity is aggregated by the scalp-recorded FFR, which thus mimics the temporal structure of the sound with a high degree of fidelity. Prior cross-linguistic work (Krishnan et al., 2005; Llanos et al., 2017; Reetzke et al., 2018) has shown that the FFR reflects language experience-dependent plasticity. Specifically, Mandarin lexical tones are more faithfully represented in the FFR of native speakers of Mandarin Chinese, relative to native speakers of English (e.g., Krishnan et al., 2005). Therefore, Mandarin listeners are expected to convey bigger differences between FFRs to Mandarin tones than non-native speakers of Mandarin Chinese (Llanos et al., 2017; Reetzke et al., 2018). Motivated by these neuroscientific experiments and related prior literature, our proposed mixed MDS methodology allows scientists to map the observed pairwise FFR distances to a lower-dimensional common feature space, simultaneously enabling the evaluation of heterogeneity among language groups and subjects, presenting a principled statistical approach for comparing groups and individuals, while also allowing the visualization of the geometry. Notably, in our formulation, the latent axes of the reduced space are all uniquely identifiable, which makes inference and interpretation of the features biologically very meaningful.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

An alternative approach to obtaining a low-dimensional representation of the FFR time series might be through traditional LFMs. See, e.g., Aguilar and West (2000), albeit in a financial time-series context. LFMs assume that observed time series can be represented as linear combinations of latent sources plus noise. However, these assumptions may not hold when relationships are nonlinear, nonstationary, or fundamentally distance-based, as is often considered to be the case with the geometrically complex patterns of brain activity. Additionally, while LFMs can recover latent components, they usually lack meaningful spatial arrangement, making intuitive visualization challenging. In contrast, MDS does not rely on any such restrictive assumptions on a specific generative model for the observed time series. By operating directly on pairwise dissimilarities, MDS can be applied to a wide range of distance and dissimilarity matrices and provides low-dimensional embeddings that summarize complex similarity structures. It provides interpretable, low-dimensional visual embeddings that preserve large-scale geometric patterns, making it especially well-suited for exploratory and visualization-driven neurobiological analyses, which likely contributed to its popularity in such contexts. These features are particularly important for our application as well, where the focus is on characterizing and visualizing differences in underlying auditory processing mechanisms from pairwise dissimilarities between the corresponding FFRs. In particular, our proposal is tailored to dissimilarity data with multiple sources of heterogeneity (subjects and groups) and, motivated by the auditory neuroscientific application, targets low-dimensional Euclidean representations for dissimilarities.

The rest of this article is organized as follows. Section 2 provides additional background on our motivating scientific experiments. Section 3 details our novel multi-group mixed MDS models. Section 4 outlines statistical and computational challenges and solution strategies. In particular, Section 4.1 reports the proposed MCMC algorithm, Section 4.2 describes a strategy to select the number of features adaptively, and Section 4.3 outlines the post-processing algorithm to solve the identifiability issues and allow inference on the latent features. Section 5 discusses the results of the proposed method applied to some synthetic numerical experiments. Section 6 presents the results of the proposed method applied to the aforementioned two language groups' neural dissimilarities data. Section 7 contains concluding remarks.

2. Neural dissimilarities data

The neural dissimilarity matrices used in our analysis come from the previously published auditory neuroscience study by Llanos et al. (2017). The brain activities of $n = 28$ subjects, $n_1 = 14$ Mandarin speakers and $n_2 = 14$ English (non-Mandarin) speakers, were recorded under exposure to different Mandarin tones. Individual distance matrices between stimuli were computed as Euclidean distances between FFRs. As a tonal language, Mandarin Chinese has four syllabic pitch contours or tones that are used to convey different lexical meanings. For instance, the syllable "ma" can be interpreted as "mother," "hemp," "horse," or "scold" depending on whether it is pronounced with high level (T1), low-rising (T2), low-dipping (T3), or high-falling (T4) tones, respectively. These four tones, pronounced by native Mandarin speakers, represented the stimuli. Figure 1 provides context for the associated neural signals underlying our distance matrices. As FFRs reflect fast, phase-locked oscillations (often at hundreds of Hz), they are traditionally presented in the time domain to demonstrate neural tracking of fine temporal structure. In such representations, meaningful information is encoded in the rate of oscillation (neural pitch tracking) rather than peak amplitude. While qualitative cues exist, such as T3's slower overall rate and T4's terminal deceleration, these differences are naturally subtle.

This limited visual separability of FFRs justifies our reliance on quantitative, distance-based comparisons of the entire response rather than waveform inspection. Specifically, we focus on analyzing the Euclidean distances between neural representations of Mandarin tones computed from the subjects' brain activity measured via scalp-recorded FFRs, yielding for each subject an $S \times S$ (here, $S = 4$) symmetric distance matrix with zeros on the diagonal. FFRs were extracted from the dataset described in Llanos et al. (2017). They were acquired using three Ag-Cl electrodes placed on the vertex of the

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

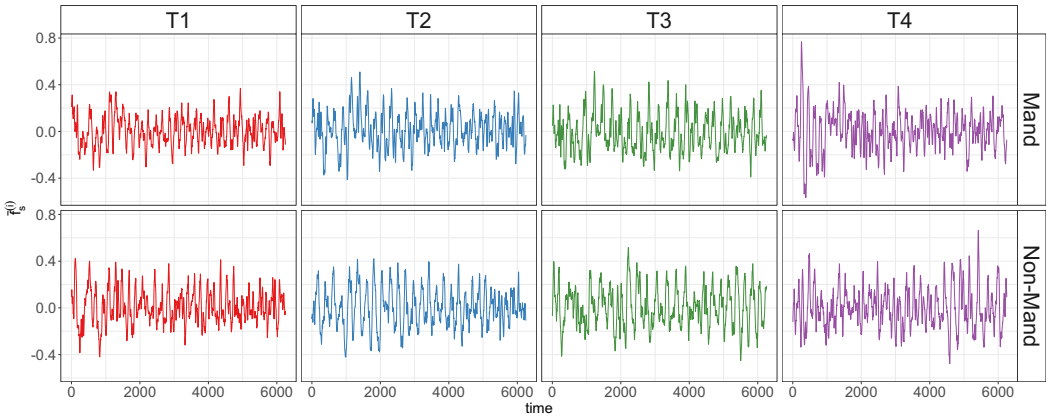


Figure 1. Tone neural data: Mean (intra-tone cross-measurements) FFR in the time domain from one Mandarin-speaking and one non-Mandarin-speaking listener. T1, T2, T3, and T4 denote the four Mandarin lexical tones: high level (T1), low-rising (T2), low-dipping (T3), and high-falling (T4).

scalp (active channel), the left mastoid (ground), and the right mastoid (reference channel). The neural activity captured by these electrodes was amplified and digitized with an actiCHamp system and a dedicated preamplifier (EP-preamp, gain 50 ×). The FFR dataset included 1, 000 artifact-free FFR trials per tone (i.e., 1, 000 FFRs in response to 1, 000 repetitions of each tone). To account for different trial-by-trial noise levels in the neural representation of the brain between the two language groups across subjects and time, some rescaling of the observations is necessary (see, e.g., Feng, Li, et al., 2021; Nili et al., 2014), which helps eliminate internal variability. Specifically, we compute the average FFR across 1, 000 repeated measurements (i.e., trials) rescaled by the average (across stimuli) of the within (repeated measurements) standard deviations.

Formally, let $f_{s,\,m}^{(i)}(t)$ denote the FFR in the time domain, representing the brain activity of subject i at time t under the exposure to stimulus s in the measurement m. We denote by $\bar{f}_s^{(i)}(t) = \sum_{m=1}^{1,\,000} f_{s,\,m}^{(i)}(t)/1,000$ the average over repeated measurements (as shown in Figure 1) and by $\mathrm{SSE}_s^{(i)}(t) = \sum_{m=1}^{1,\,000}\{f_{s,\,m}^{(i)}(t) - \bar{f}_s^{(i)}(t)\}^2$ the sum of squared errors at time t for subject i and stimulus s. Thus, as a measure of noise for the measurements at time t in subject i under the exposure to stimulus s, we use the estimated standard deviation $\widehat{\sigma}_s^{(i)}(t) = \{\mathrm{SSE}_s^{(i)}(t)/(1,000-1)\}^{1/2}$; and as a measure of individual dispersion at time t, the average of these standard deviations across stimuli $\widehat{\sigma}^{(i)}(t) = \sum_{s=1}^{S}\widehat{\sigma}_s^{(i)}(t)/S$. Finally, we rescale the average empirical distances at time t by these estimated noise variances that cannot be explained by the average FFR, that is, $\widetilde{f}_s^{(i)}(t) = \bar{f}_s^{(i)}(t)/\widehat{\sigma}^{(i)}(t)$. By rescaling the FFRs between average brain activities by the corresponding language group's inherent noise levels in this manner, we allow for biologically meaningful comparisons of the brain's ability to distinguish different auditory stimuli across subjects.

Figure 2 shows the average distances between stimuli in the two groups. We can see that, although there are group-specific characteristics due to language, for example, the recorded Mandarin speaker distances are better differentiated across pairs of tones than the ones from the non-Mandarin group, there are also strongly shared patterns across these groups, for example, the pairs $\{2, 3\}$ and $\{2, 4\}$ are well-separated pairs of tones while the pairs $\{1, 4\}$ and $\{1, 2\}$ are the closest in both groups.

The individual tone distances show that there is also substantial individual variability within groups. Note also a higher within-group variability in the non-Mandarin group, as expected. The corresponding data are presented alongside the analysis results later in the manuscript.

We are interested in learning the shared geometry of the lower-dimensional latent space representing these tones in the human brain across language groups and individuals, while also assessing how this space varies across these groups and individuals.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

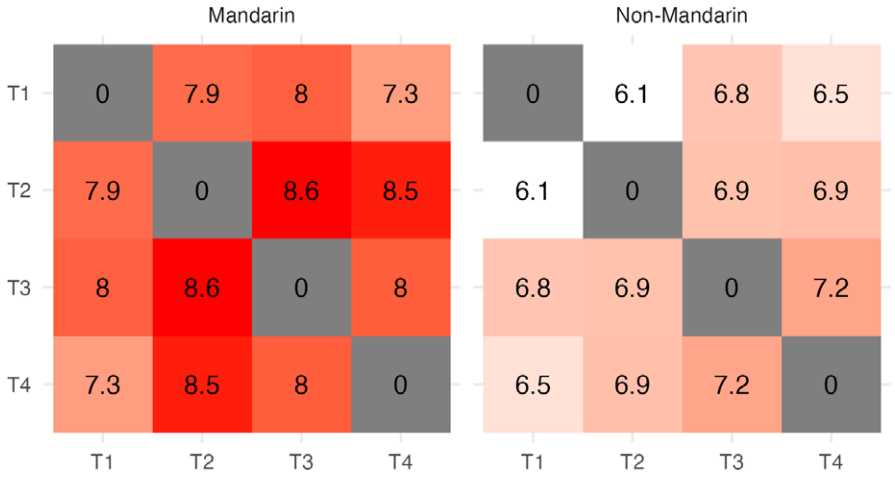


Figure 2. Tone neural distance data: Mean (intra-group cross-subject) FFR distances between different Mandarin tones, T1, T2, T3, and T4.

3. Mixed multidimensional scaling

Modeling the distances: In MDS, the observed distances between stimuli s and r, namely, $d_{s,r}$, $r < s, r = 1, \ldots, S-1, s = 2, \ldots, S$, are noisy measurements of latent distances $\delta_{s,r}$ computed in a lower H-dimensional feature space ($H < S$). We therefore treat the observed dissimilarities $d^{(i)}_{j,s,r}$ as noisy, strictly positive measurements of latent denoised distances $\delta^{(i)}_{j,s,r}$.

As in Nguyen and Holmes (2017) and Yanchenko and Hoff (2020), we assume a Gamma likelihood for the observed distances to perform our model-based MDS. However, motivated by the application, we allow for individual differences at the level of the latent features. We also consider different variances in the likelihood terms to take into account the heterogeneity of the noise levels between the two groups in our data.

More precisely, we denote by $d^{(i)}_{j,s,r}$ the distance between the s^{th} and the r^{th} stimuli ($r < s$) for the i^{th} individual from the j^{th} group and we assign a Gamma likelihood for the observed distances

$$d^{(i)}_{j,s,r} \overset{ind}{\sim} \text{Ga}\left(\texttt{mean} = \delta^{(i)}_{j,s,r}, \texttt{var} = \sigma^2_{\varepsilon,j}\right). \tag{1}$$

In the context of the experiment described in Section 2, $j = 1$ and $j = 2$ refer to the Mandarin-speaking group and the non-Mandarin-speaking group, respectively. The Gamma likelihood is parameterized in terms of mean and variance to ease the interpretation of the latent parameters, and it is centered around the underlying individual denoised distances $\delta^{(i)}_{j,s,r}$, which can vary across subjects within and across the two groups. The variance term $\sigma^2_{\varepsilon,j}$ varies across groups, allowing for different noise levels in different groups. The more standard shape and rate representation can be recovered as $\texttt{shape} = \texttt{mean}^2/\texttt{variance} = (\delta^{(i)}_{j,s,r})^2/\sigma^2_{\varepsilon,j}$ and $\texttt{rate} = \texttt{mean}/\texttt{variance} = \delta^{(i)}_{j,s,r}/\sigma^2_{\varepsilon,j}$.

The underlying individual distance $\delta^{(i)}_{j,s,r}$ is a function of H unobserved individual features $\eta^{(i)}_{j,s,h}$, such as

$$\delta^{(i)}_{j,s,r} = \sqrt{\sum_{h=1}^{H}\left(\eta^{(i)}_{j,s,h} - \eta^{(i)}_{j,r,h}\right)^2}, \tag{2}$$

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

where the number of latent features H is smaller than the number of original stimuli S. Here, $\boldsymbol{\eta}_{j,s}^{(i)} = (\eta_{j,s,1}^{(i)}, \ldots, \eta_{j,s,H}^{(i)})^{\mathrm{T}}$ denotes the latent vector with the feature values for the s^{th} stimulus for the i^{th} individual belonging to the j^{th} group.

On the choice of the Gamma likelihood: Note that in our hierarchical framework, the Gamma likelihood for the observed distances $d_{j,s,r}^{(i)}$ is conditional on latent distances $\delta_{j,s,r}^{(i)}$ derived from a lower-dimensional feature space. Under this specification, the Euclidean structure enters through the conditional mean of $d_{j,s,r}^{(i)}$, while the sampling variances $\sigma_{\varepsilon,j}^2$ accommodate varying noise levels across groups. The likelihood thus implicitly accounts for an error term, ε, characterizing the variability of the observed distances $d_{j,s,r}^{(i)}$ around the latent distances $\delta_{j,s,r}^{(i)}$, which is itself a function of the underlying feature parameters $\eta_{j,s,h}^{(i)}$, as defined in (2). The model for these features $\eta_{j,s,h}^{(i)}$ and their associated priors are detailed in Sections 3 and 3.2. By operating strictly within this conditional hierarchical framework, we fulfill our inference goals while avoiding the derivation of the induced marginal distribution, which remains analytically intractable.

The choice of the Gamma distribution for the conditional likelihood provides a good framework for modeling dissimilarities on $(0, \infty)$, consistent with established practice in Bayesian MDS (e.g., Nguyen & Holmes, 2017; Yanchenko & Hoff, 2020). Although parametric, the Gamma family offers some advantages over more rigid alternatives: unlike the log-Normal distribution (e.g., Ramsay, 1977, 1982), which is constrained by heavy tails and sparse mass near zero, or the truncated-Normal specification (e.g., Oh & Raftery, 2001), which is often dominated by its behavior at the origin, the Gamma distribution can smoothly transition from Exponential-like to Gaussian-like shapes (as $\texttt{shape} = \texttt{mean}^2/\texttt{variance}$ increases) while maintaining a robust intermediate tail behavior.

While more flexible nonparametric approaches, such as mixture models, could be considered when larger sample sizes are available, in the present application, the sample size is relatively modest, and the empirical data exhibit means that substantially exceed their variances (Figure 2). In this high-shape regime ($\texttt{shape} \gg 1$), the Gamma, log-Normal, and truncated-Normal distributions all converge toward a symmetric, Gaussian-like form, rendering the specific choice of parametric family largely inconsequential. Overall, the parametric Gamma likelihood thus represents a parsimonious and robust choice for our current modeling requirements.

Modeling the latent features: Our modeling efforts concentrate henceforth on flexibly characterizing the latent features $\eta_{j,s,h}^{(i)}$. Specifically, we model $\eta_{j,s,h}^{(i)}$ as a product of latent features shared across groups and subjects, $\eta_{s,h}$, and multiplicative coefficients specific to dimension h, group j, and subject i within group j. That is, we let

$$\eta_{j,s,h}^{(i)} = \widetilde{w}_{j,h}^{(i)} \eta_{s,h}. \tag{3}$$

The random coefficients $\widetilde{w}_{j,h}^{(i)}$ vary across features h, allowing variation, and thus importance, of the dimensions across individuals between and within groups.

Finally, in order to learn the subjects' and the two groups' variations, we specify an inverse gamma distribution with hyperparameters a_w and b_w such that the multiplicative terms have a mean equal to 1 and a large variance equal to 10. That is,

$$\begin{aligned} w_j^{-1} &\overset{iid}{\sim} \mathrm{Ga}\left(\texttt{shape} = a_w, \texttt{rate} = b_w\right), \\ w_{j,h}^{(i)^{-1}} &\overset{iid}{\sim} \mathrm{Ga}\left(\texttt{shape} = a_w, \texttt{rate} = b_w\right). \end{aligned} \tag{4}$$

Here, with some abuse of notation, $w_{j,h}^{(i)}$'s represent the individual-level weights that do not vary by group; however, the index j is still retained to distinguish individuals belonging to different groups (e.g., $w_{1,h}^{(1)}$ vs. $w_{2,h}^{(1)}$ denote the first individual in groups 1 and 2, respectively). Also, we parameterize the gamma distribution in terms of shape and rate hyperparameters to facilitate updating these

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

hyperparameters in the full conditionals. If the multiplicative terms w_j and $w_{j,h}^{(i)}$ are equal to 1, the MDS model ignores group and individual differences, entailing the same latent feature values for the stimulus s for all subjects within and across the two groups.

On the choice of shared features: Studies on Mandarin lexical tones in both native and non-native listeners have shown their perception to rely on core pitch-related dimensions across language backgrounds (Gandour, 1978, 1983; Gandour & Harshman, 1978). Neurophysiological work further demonstrates that these dimensions are robustly represented in early auditory responses, such as FFRs, in both Mandarin and non-Mandarin listeners (Chandrasekaran, Krishnan, et al., 2007).

This motivates our shared-feature MDS framework, where group and individual differences are characterized as shifts in representational geometry rather than the use of distinct feature sets. This approach aligns with the general consensus that these dimensions are universally relevant to auditory processing, while language experience modulates their relative weighting. While relaxing the shared-feature assumption may suit other domains, our model's use of shared latent features with group and individual-specific weighting provides an interpretable framework, grounded in neurobiological theory, for comparing neural encoding across different language backgrounds.

On the model for the weights: In model (3), the weights $\widetilde{w}_{j,h}^{(i)}$ are not identifiable from the $\eta_{s,h}$ parameters without the imposition of scale constraints. We discuss these constraints as part of our post-processing scheme in Section 4.3. Notwithstanding this specific form of non-identifiability, one may consider various strategies to impose further structure and interpretability on the weights. Specifically, we can consider:

(a) unrestricted: $\widetilde{w}_{j,h}^{(i)}$, or
(b) group-specific decomposition: $\widetilde{w}_{j,h}^{(i)} = w_j w_{j,h}^{(i)}$, or
(c) group-and-dimension-specific decomposition: $\widetilde{w}_{j,h}^{(i)} = w_{j,h} w_{j,h}^{(i)}$, or
(d) group-and-(dimension-within-group)-specific decomposition: $\widetilde{w}_{j,h}^{(i)} = w_j w_{j,h} w_{j,h}^{(i)}$.

As before, in models (b)–(d), $w_{j,h}^{(i)}$'s represent individual-level weights that do not vary directly by group j, but the index j is still retained to distinguish individuals from different groups.

Because the likelihood depends solely on latent distances, these specifications are stochastically equivalent, leaving internal structures unidentifiable without further constraints. However, by imposing consistently defined identifying constraints across these parameterizations, the individual weight components can be uniquely recovered from one another. For instance, the marginal effects remain well-defined in terms of the $\widetilde{w}_{j,h}^{(i)}$'s across all models: Group-and-dimension-specific effects can be defined as $\overline{w}_{j,h} = \sum_{i=1}^{n_j} \widetilde{w}_{j,h}^{(i)}/n_j$. Under model (c), this simplifies to $w_{j,h}$ given the constraint $\sum_{i=1}^{n_j} w_{j,h}^{(i)}/n_j = 1$. Likewise, overall group-specific effects can be defined as $\overline{w}_j = \sum_{h=1}^{H} \sum_{i=1}^{n_j} \widetilde{w}_{j,h}^{(i)}/(Hn_j)$. Under model (b), this equals w_j given the constraint $\sum_{h=1}^{H} \sum_{i=1}^{n_j} w_{j,h}^{(i)}/(Hn_j) = 1$.

In a Bayesian hierarchical framework, these distinct parameterizations do necessitate tailored prior specifications, which in turn dictate the mechanisms for information borrowing and shrinkage, and the efficiency of the MCMC sampling; generally, more granular decompositions will yield higher posterior uncertainty and require more robust implementation to manage the resulting computational complexity. However, as we will see in Section 4.3, using the constraints described above, the posterior distributions for weight components under alternative specifications of $\widetilde{w}_{j,h}^{(i)}$ can be readily derived (under implicit priors) from the MCMC samples of a primary model of choice. Consequently, irrespective of the initial specification, the underlying representational geometry remains accessible via straightforward reparameterization of the posterior samples.

Nevertheless, the practical choice between these models can be guided by considerations of meaningful borrowing of information and computational ease and efficiency. In this article, we adopted

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

model (b) because it keeps the MCMC and the post-processing schemes simple, while inducing a higher correlation between individual distances δ within the same group. As discussed above, the $w_{j,h}$'s can still be obtained using post-processing (see Section 4.3 and Figures 1 and 2 in the Supplementary Material).

3.1. Model identifiability

Our multiplicative individual effects make the latent feature space identifiable under rotations, enabling stronger interpretations of the latent features. To see this, note that reweighting the latent features is equivalent to multiplying the feature matrix by a diagonal matrix of weights. Therefore, if we rotate the axes (aside from the special case of permutations of the axes and the subclass of signed permutations that can be represented as rotation), the weight matrix would not be diagonal anymore, which is not permissible in our model. Therefore, it allows the axes of the latent features to be uniquely identified and unaffected by rotation invariance.

More precisely, let $\mathbf{H}_j^{(i)} = ((\eta_{j,s,h}^{(i)})) \in \mathbb{R}^{S\times H}$, $i = 1, \ldots, n_j$, $j = 1, \ldots, J$ be the individual latent feature matrices, $\mathbf{W}_j^{(i)} = \operatorname{diag}(\widetilde{w}_{j,1}^{(i)}, \ldots, \widetilde{w}_{j,H}^{(i)}) \in \mathbb{R}^{H\times H}$, $i = 1, \ldots, n_j$, $j = 1, \ldots, J$ be the corresponding diagonal weight matrices, and $\mathbf{H} = ((\eta_{s,h})) \in \mathbb{R}^{S\times H}$ be the shared latent feature matrix. Then, we can rewrite $\mathbf{H}_j^{(i)} = \mathbf{H}\mathbf{W}_j^{(i)}$. Now, let $\mathbf{R} \in \mathbb{R}^{H\times H}$ be a rotation matrix. If we rotate the individual features as $\mathbf{H}_j^{(i)}\mathbf{R}^{\mathrm{T}} = \mathbf{H}\mathbf{W}_j^{(i)}\mathbf{R}^{\mathrm{T}}$, it will not change the implied latent distances $\delta_{j,s,r}^{(i)}$, but the new individual weight matrices $\mathbf{W}_j^{(i),\,\mathrm{new}} = \mathbf{W}_j^{(i)}\mathbf{R}^{\mathrm{T}}$ are in general not diagonal anymore. This is why the individual weights, which are shared across stimuli, make the latent feature space identifiable in INDSCAL, contrary to standard MDS analysis. See Carroll and Chang (1970) for further discussion. Notably, the use of individual-specific dimension weights is not exclusive to INDSCAL; other prominent MDS frameworks, such as ALSCAL and PROXSCAL, can similarly accommodate individual differences by weighting different latent dimensions according to each subject's unique perceptual profile (see, e.g., Commandeur & Heiser, 1993; Takane et al., 1977).

That said, some non-identifiability persists since the features $\eta_{j,s,h}^{(i)}$ are still invariant to signed permutations and translations that preserve the latent distances δ (even if the axes are identifiable and cannot be arbitrarily rotated as discussed above). We address these issues separately via a post-processing scheme discussed later in Section 4.3.

Finally, we note that the latent distances $\delta_{j,s,r}^{(i)}$ and group-specific variance terms $\sigma_{\varepsilon,j}^2$ are identifiable in the traditional sense of Basu (2004) as stated in Swartz et al. (2004). Specifically, the identifiability of $\boldsymbol{\theta} = \{\eta_{j,s,h}^{(i)}, \sigma_{\varepsilon,j}^2, i = 1, \ldots, n_j, j = 1, \ldots, J, s = 1, \ldots, S, h = 1, \ldots, H\}$ up to the aforementioned transformations is equivalent to the identifiability of $\widetilde{\boldsymbol{\theta}} = \{\delta_{j,s,r}^{(i)}, \sigma_{\varepsilon,j}^2, i = 1, \ldots, n_j, j = 1, \ldots, J, s = 1, \ldots, S, r = 1, \ldots, S\}$. And the latter follows from the moment equations

$$\mathbb{E}(d_{j,s,r}^{(i)} \mid \widetilde{\boldsymbol{\theta}}) = \delta_{j,s,r}^{(i)}, \quad \operatorname{var}(d_{j,s,r}^{(i)} \mid \widetilde{\boldsymbol{\theta}}) = \sigma_{\varepsilon,j}^2.$$

3.2. Prior specification

To learn the shared latent features $\eta_{s,h}$, we adapt the multiplicative gamma process (MGP) prior (Bhattacharya & Dunson, 2011) for ordinary factor analysis that allows a stochastic ordering of the latent features in diminishing order of their importance.

Specifically, we assume a Gaussian distribution for the shared feature component $\eta_{s,h}$ whose variance varies across stimuli s and dimension h

$$\eta_{s,h} \overset{ind}{\sim} \text{Normal}\left(\texttt{mean} = 0,\ \texttt{sd} = \phi_{s,h}^{-1/2}\tau_h^{-1/2}\right). \tag{5}$$

The precision parameter of the shared component $\eta_{s,h}$ follows the MGP

$$\phi_{s,h} \overset{iid}{\sim} \text{Ga}\left(\texttt{shape} = \nu/2, \texttt{rate} = \nu/2\right), \tag{6}$$

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

$$\tau_h = \prod_{l=1}^{h} \delta_l, \quad \delta_l \overset{ind}{\sim} \text{Ga}(\texttt{shape} = a_l, \texttt{rate} = 1), \tag{7}$$

where $a_l = a_2$ for $l \geq 2$. The MGP prior shrinks the latent features increasingly toward zero as the dimension h increases. This implies that the prior favors a few relevant features, shrinking the rest to zero, while also inducing a probabilistic ranking of the relevant features such that, on average, the first feature explains more variability than the second one, and so on, coherent with the current understanding of the brain's behavior.

We follow Durante (2017) recommendations for the choice of the hyperparameters of the MGP. In particular, we set $a_1 = 2$ and $a_2 = 3$ since they induce the desired prior shrinkage. More precisely, they induce an order in the probability of the dimensions in a neighborhood around zero (Lemma 2 in Durante, 2017). Section 4 of the Supplementary Material provides further discussion on the choice of the hyperparameters.

Finally, we put an inverse-gamma prior to learn the group-specific variance terms

$$\sigma^2_{\varepsilon,j} \overset{ind}{\sim} \text{Inv-Ga}\left(\texttt{mean} = \mu_{\sigma^2_{\varepsilon,j}}, \texttt{var} = \sigma^2_{\sigma^2_{\varepsilon,j}}\right). \tag{8}$$

3.3. Model reparameterization

We can reparameterize the model in terms of the individual latent features $\eta^{(i)}_{j,s,h}$. Integrating out the shared dimensions $\eta_{s,h}$, we obtain

$$\eta^{(i)}_{j,s,h} \sim \text{Normal}\left(\texttt{mean} = 0, \texttt{sd} = w^{(i)}_{j,h} w_j \phi^{-1/2}_{s,h} \tau^{-1/2}_h\right). \tag{9}$$

This parameterization allows restating the probabilistic statements directly in terms of one of the main quantities of interest, namely, the individual-specific latent features. Note that, by the definition in (3), the $\eta^{(i)}_{j,s,h}$'s in (9) are dependent on each other since they share the same latent $\eta_{s,h}$'s. That is, $\left(\eta^{(i)}_{j,s,h}\right)_h$, $i = 1, \ldots, n_j, j = 1, \ldots, J, s = 1, \ldots, S$ belong to the span of the vectors $\eta_{s,h}, s = 1, \ldots, S$. This is consistent with our goal to infer a common latent feature's space across groups as well as subjects.

4. Posterior inference

Posterior inference is performed by sampling from an MCMC algorithm that exploits conditional conjugacy of the parameters when available. When sampling from the full conditionals is not straightforward, we exploit adaptive Metropolis schemes (Roberts & Rosenthal, 2009). The details are discussed in the next section.

4.1. MCMC algorithm

(1) Sample δ_1 from $p(\delta_1 \mid \cdots)$

$$= \text{Ga}\left(\texttt{shape} = a_1 + \frac{SH}{2}, \texttt{rate} = 1 + \frac{1}{2}\sum_{l=1}^{H} \tau^{(1)}_l \sum_{s=1}^{S} \phi_{s,l} \eta^2_{s,l}\right).$$

(2) Sample δ_h from $p(\delta_h \mid \cdots)$

$$= \text{Ga}\left(\texttt{shape} = a_h + \frac{S(H-h+1)}{2}, \texttt{rate} = 1 + \frac{1}{2}\sum_{l=h}^{H} \tau^{(h)}_l \sum_{s=1}^{S} \phi_{s,l} \eta^2_{s,l}\right),$$

where $\tau^{(h)}_l = \prod_{t=1,\ t\neq h}^{l} \delta_t$, for $l = 1, \ldots H$.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

(3) Sample $\phi_{s,h}$ from

$$p(\phi_{s,h} \mid \cdots) = \text{Ga}\left(\texttt{shape} = \frac{\nu+1}{2},\ \texttt{rate} = \frac{\nu + \tau_h \eta_{s,h}^2}{2}\right).$$

(4) Sample $\eta_{s,h}$ from

$$p(\eta_{s,h} \mid \cdots) \propto \mathcal{N}\left(\eta_{s,h} \mid \texttt{mean} = 0,\ \texttt{sd} = \phi_{s,h}^{-1/2}\tau_h^{-1/2}\right) \times \prod_{j=1}^{J}\prod_{i=1}^{n_j}\prod_{r:r\neq s} \text{Ga}\left(d_{j,s,r}^{(i)} \mid \texttt{mean} = \delta_{j,s,r}^{(i)},\ \texttt{var} = \sigma_{\varepsilon,j}^2\right).$$

We perform an adaptive MH step (Roberts & Rosenthal, 2009) with a random-walk univariate Gaussian proposal.

(5) Sample $\sigma_{\varepsilon,j}^2$ from

$$p\left(\sigma_{\varepsilon,j}^2 \mid \cdots\right) \propto \text{Inv-Ga}\left(\sigma_{\varepsilon,j}^2 \mid \texttt{mean} = \mu_{\sigma_{\varepsilon,j}^2},\ \texttt{var} = \sigma_{\sigma_{\varepsilon,j}^2}^2\right) \times \prod_{i=1}^{n_j}\prod_{s=2}^{S}\prod_{r=1}^{s-1} \text{Ga}\left(d_{j,s,r}^{(i)} \mid \texttt{mean} = \delta_{j,s,r}^{(i)},\ \texttt{var} = \sigma_{\varepsilon,j}^2\right).$$

We perform an adaptive MH step with a random-walk Gaussian proposal for $\log(\sigma_{\varepsilon,j}^2)$.

(6) Sample $w_{j,h}^{(i)}$ from

$$p\left(w_{j,h}^{(i)} \mid \cdots\right) \propto \text{Inv-Ga}\left(w_{j,h}^{(i)} \mid \texttt{shape} = a_w,\ \texttt{rate} = b_w\right) \times \prod_{s=2}^{S}\prod_{r=1}^{s-1} \text{Ga}\left(d_{j,s,r}^{(i)} \mid \texttt{mean} = \delta_{j,s,r}^{(i)},\ \texttt{var} = \sigma_{\varepsilon,j}^2\right).$$

We perform an adaptive MH step with a random-walk Gaussian proposal for $\log(w_{j,h}^{(i)})$.

(7) Sample w_j from

$$p(w_j \mid \cdots) \propto \text{Inv-Ga}\left(w_j \mid \texttt{shape} = a_w,\ \texttt{rate} = b_w\right) \times \prod_{i=1}^{n_j}\prod_{s=2}^{S}\prod_{r=1}^{s-1} \text{Ga}\left(d_{j,s,r}^{(i)} \mid \texttt{mean} = \delta_{j,s,r}^{(i)},\ \texttt{var} = \sigma_{\varepsilon,j}^2\right).$$

We perform an adaptive MH step with a random-walk Gaussian proposal for $\log(w_j)$. More precisely, in all adaptive MH steps, we use an adaptive random walk Metropolis kernel with Gaussian proposals, after transforming parameters to have support on $\mathbb{R}$ when needed. Every 50 iterations, we compute, for each parameter, the empirical acceptance rate over the last 50 iterations. If this acceptance rate is below the target value 0.44, we decrease the corresponding proposal standard deviation by a factor $e^{-\delta_m}$; if it is above 0.44, we increase it by a factor e^{δ_m}. Here, $\delta_m = \min\{0.01, 1/\sqrt{m}\}$, with m the current MCMC iteration, so the adaptations diminish over time. This scheme targets an acceptance rate of 0.44 for all scalar random walk proposals, which is a standard choice for efficient adaptive random walk Metropolis algorithms (Roberts & Rosenthal, 2009).

4.2. Selecting the number of features

Our Bayesian mixed MDS model allows performing dimensionality reduction by setting $H < S$. If we choose a conservative (i.e., larger than needed) upper bound H^+, the model allows us to give relevant importance to a few latent features $H < H^+$ and small importance to the remaining $H^+ - H$ latent features. Here, we think of H as the effective dimensionality of the latent feature space (i.e., the number of latent dimensions) so that the contribution from adding additional dimensions in reconstructing a denoised version of the observed distance matrices is negligible. However, running the MCMC for an

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

upper bound H^+ larger than needed can be computationally inefficient. Finding the number of relevant latent features H can therefore be helpful in reducing costs. Finding H can also be of inferential interest itself.

To address this, we use a relative Frobenius error $D(t)$, similar in spirit to Kruskal's STRESS, defined as the average, across subjects i and groups j, of the Frobenius distance between the observed $S \times S$ distance matrices $((d^{(i)}_{j,s,r}))$ and the corresponding $S \times S$ denoised latent distance matrices $((\delta^{(i)}_{j,s,r}))$ sampled at iteration t, divided by the Frobenius norm of the observed distance matrix. We set, according to the noise level in the measurements, a threshold D_T. In particular, we found that $D_T = 0.1$ favors a parsimonious latent space that can reconstruct well all the individual observable distances in our motivating neural tone distance experiment, as shown in Figure 8. More generally, D_T should be chosen to reflect the desired level of reconstruction accuracy for the observed distances and, when possible, calibrated using prior knowledge about the signal-to-noise ratio in the specific application. We perform the following steps with probability $p(t) = \exp\{-(\alpha_0 + \alpha_1 t)\}$ at iteration t, where $\alpha_0 \geq 0$ and $\alpha_1 > 0$ such that the adaptations occur often at the beginning of the chains, but decrease in frequency exponentially fast as the chain settles in. At the t^{th} iteration, if the current latent features are not sufficient to recover the distances well, that is, $D(t) > D_T$, we set $H(t+1) = H(t) + 1$ and add a latent feature $\eta_{s,H(t)+1}$. Otherwise, if $D(t) < D_T$, we set $H(t+1) = H(t) - 1$ and delete the feature $\eta_{s,H(t)}$. When we add a feature, we sample the parameters from the prior distribution.

The adaptive method allows, in a single run, to perform posterior inference on the individual and group-level latent distances, $\delta^{(i)}_{j,s,r}$ and $\delta_{j,s,r}$, together with selecting the number of features, with the convergence of the chain guaranteed by the diminishing probability condition (Roberts & Rosenthal, 2007). If, as in our motivating auditory neuroscience application, one is interested in performing inference on the actual latent features, we can fix the number of active dimensions H selected in a preliminary run of the MCMC chain or in an initial set of iterations of the chain after burn-in. In this way, we have the same number of features, H, sampled in the final stages of the chain that we can use to perform posterior inference on the latent features after solving the identifiability issues as described in the next section. Finally, as in Bhattacharya and Dunson (2011), we can perform uncertainty quantification on the number of features around the point estimate of H, in our case, the median of the sampled values $H(t)$ after burn-in, via credible intervals.

Additional evidence on the empirical performance of this adaptive procedure is reported in Section 5 of the Supplementary Material, where simulation studies show how it recovers the true latent dimension H in application-motivated scenarios.

4.3. Post-processing for feature identifiability

In MDS, inference on the latent features $\eta^{(i)}_{j,s,h}$ is challenging due to identifiability issues—while the reconstructed distances are identifiable, the features themselves are not. As discussed in Section 3.1, our construction overcomes the rotation invariance issue of the latent dimensions that affects standard MDS methods. However, while the set of the latent feature axes are now identifiable, the values of the latent features are still identifiable only up to translations and signed permutations of the axes (e.g., label switching of the dimensions and reflection). This makes the posterior summaries of the sampled $\eta^{(i)}_{j,s,h}$'s, for example, the posterior median, computed from the MCMC samples, not very meaningful.

To address this issue, we adapt recent post-processing ideas developed for Bayesian LFMs by Papastamoulis and Ntzoufras (2022). First, we remove translation invariance by centering the shared coordinates so that $\sum_{s=1}^{S} \eta_{s,h} = 0$ for every h, that is,

$$\eta_{s,h} \leftarrow \eta_{s,h} - \bar{\eta}_{\cdot,h}, \quad \text{where } \bar{\eta}_{\cdot,h} = \frac{1}{S}\sum_{s=1}^{S} \eta_{s,h}.$$

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

Under the parameterization $\widetilde{w}_{j,h}^{(i)} = w_j w_{j,h}^{(i)}$, we then apply the corresponding translation to the individual product coordinates as

$$\eta_{j,s,h}^{(i)} \leftarrow \eta_{j,s,h}^{(i)} - w_j w_{j,h}^{(i)} \bar{\eta}_{\cdot,h},$$

so that the fitted distances remain unchanged. If the goal is inference on the latent distances and on the individual latent dimensions inducing them, together with comparison of the individual configurations across subjects and groups, then this centering step, combined with the signed permutation alignment described below, is sufficient.

Indeed, after centering, we resolve the remaining label switching and sign ambiguity by applying signed permutation transformations that align the latent dimensions across posterior draws. These transformations are applied consistently to the shared coordinates $\eta_{s,h}$ and to the individual latent coordinates $\eta_{j,s,h}^{(i)}$, while the positive weights are affected only through the corresponding permutation of the dimension index h. Based on these adjusted samples, we can compute meaningful posterior summaries for the latent geometry, such as posterior medians and credible intervals for the individual latent coordinates.

As discussed in Section 3, for an additional summary of the contribution of each latent dimension within and across groups, it may also be useful to examine the weight structure directly. To this end, we further normalize the fitted weights $\widetilde{w}_{j,h}^{(i)}$. First, we remove the per-dimension scale indeterminacy induced by the transformation

$$\eta_{\cdot,h} \mapsto c_h \eta_{\cdot,h} \text{ and } \widetilde{w}_{j,h}^{(i)} \mapsto \frac{\widetilde{w}_{j,h}^{(i)}}{c_h}, \quad \text{where } c_h = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j} \widetilde{w}_{j,h}^{(i)} \text{ and } n = \sum_{j=1}^{J} n_j,$$

which enforces the normalization $\frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j} \widetilde{w}_{j,h}^{(i)} = 1$ via the rescaling $\eta_{\cdot,h} \leftarrow c_h \eta_{\cdot,h}$ and $\widetilde{w}_{j,h}^{(i)} \leftarrow \widetilde{w}_{j,h}^{(i)}/c_h$. Then, we reconstruct group-and-dimension-specific summaries as

$$w_{j,h} = \frac{1}{n_j}\sum_{i=1}^{n_j} \widetilde{w}_{j,h}^{(i)}, \qquad w_{j,h}^{(i)} = \frac{\widetilde{w}_{j,h}^{(i)}}{w_{j,h}},$$

which implies the identifying constraint $\frac{1}{n_j}\sum_{i=1}^{n_j} w_{j,h}^{(i)} = 1$. Finally, we define the overall group-specific summary as the average of these dimension-specific effects as

$$w_j = \frac{1}{H}\sum_{h=1}^{H} w_{j,h}.$$

In this way, posterior draws obtained under the original model can be re-expressed post hoc in terms of group-and-dimension-specific weights $w_{j,h}$ and their group-level average w_j. These weights can approach zero if so warranted by the data, providing a practical way to assess the relative and overall utilization of different latent dimensions within and across groups.

Based on these adjusted samples, we can compute meaningful posterior summaries for inference for each subject that are comparable across subjects and groups, for example, median posterior values and credible intervals. Trace plots of the latent features before and after post-processing that solves the identifiability issues for synthetic data are shown in Figure 6. Similar results for the real data application, along with the estimated posterior distributions of the weights $w_{j,h}$ and w_j, are included in Section 1 of the Supplementary Material.

4.4. MCMC diagnostics

The results reported in this article are all based on 10^5 MCMC iterations with the initial 4,000 iterations discarded as burn-in. The remaining samples were further thinned by an interval of 10. We programmed everything in `R` (R Core Team, 2024). The analyses were performed on a MacBook Pro with Apple M2 Pro chip, 16 GB RAM, using `R` version 4.5.1. For the real data set, the MCMC

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

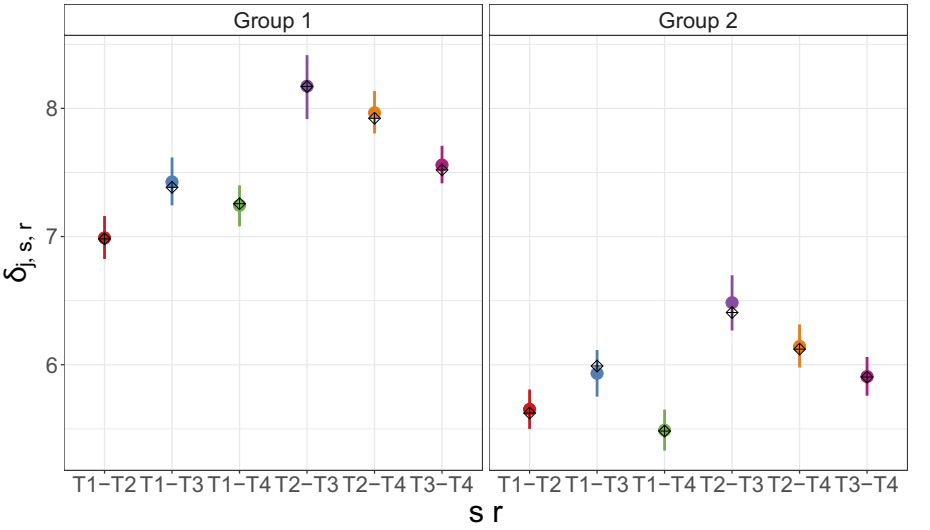


Figure 3. Results for synthetic data. Posterior medians and 90% credible intervals of the group-specific latent distances $\delta_{j,s,r}$. Black squares represent the median observed distances in each group.

algorithm took approximately 17 minutes, and the post-processing algorithm took approximately 11 minutes. The Supplementary Material provides further details and diagnostics. The code is available in the Supplementary Material.

5. Simulation studies

In this section, we discuss the results of some synthetic numerical experiments. In designing the simulation scenarios, we have tried to closely mimic our motivating "brain activities distances between tones" data set. We thus consider distances between $S = 4$ stimuli from $n = 28$ subjects recorded in two groups with cardinalities $n_1 = 14$ and $n_2 = 14$. We set the observed distances $d^{(i)}_{j,s,r}$ close to values that correspond to the estimated quantities (i.e., the posterior median of $\delta^{(i)}_{j,s,r}$) for the real data set, thus simulated distances can be reconstructed from a three-dimensional latent space.

Figure 3 shows the recovery of the distances at the group level. One can, in principle, define corresponding group-level distances based on the shared features $\eta_{j,s,h}$. However, due to the nonlinearity of the distance function, these would not correspond to the average of the individual-level distances within the group. For this reason, we believe it is more meaningful to define group-level distances directly as the average of the corresponding individual-level distances, integrating out the random effects from equation (2). Unfortunately, such integrals are not mathematically tractable in closed form, nor empirically calculable since the full populations are not accessible. To circumvent this, in our illustrations, we approximate the group-level distance parameters $\delta_{j,s,r}$ using the corresponding sample averages of the individual parameters, that is, $\delta_{j,s,r} = \sum_{i=1}^{n_j} \delta^{(i)}_{j,s,r}/n_j$, which serve as good proxies for the true $\delta_{j,s,r}$. Figure 3 suggests that the three shared latent features recover these $\delta_{j,s,r}$ quite well. Moreover, Figure 4 suggests that we recover the denoised version of the empirical distances quite well at the individual level. Figure 5 shows that our method allows us to recover the underlying latent features very efficiently. Here, we align the signs of the true and the estimated latent features for identifiability. When it comes to permuting the dimension labels, however, we leave them undisturbed. We can do this since our method assigns decreasing importance to the dimensions, consistent with the data-generating truth, and we are able to identify the order of the axes after post-processing. Importantly, contrary to standard MDS methods, we do not need to arbitrarily rotate the latent space, since the directions are identifiable.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

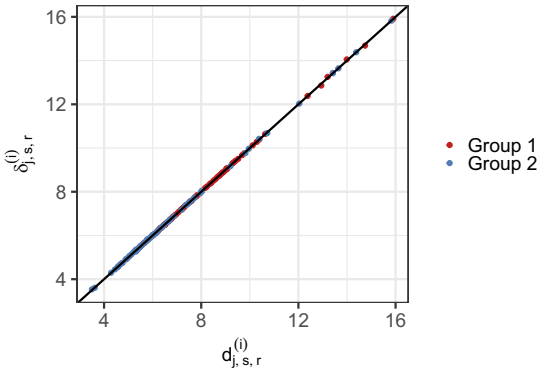


Figure 4. Results for synthetic data. Observed distances $d^{(i)}_{j,s,r}$ versus posterior medians of the denoised distances, $\delta^{(i)}_{j,s,r}$, reconstructed from three shared dimensions.

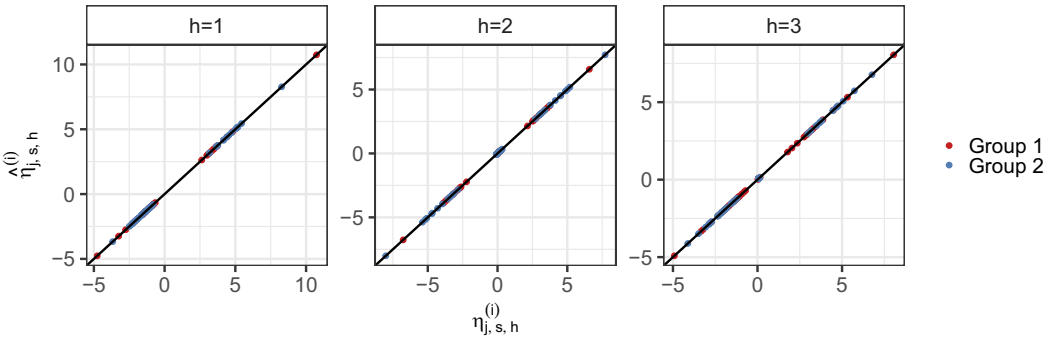


Figure 5. Results for synthetic data. True latent features $\eta^{(i)}_{j,s,h}$ versus their estimated posterior medians $\widehat{\eta}^{(i)}_{j,s,h}$.

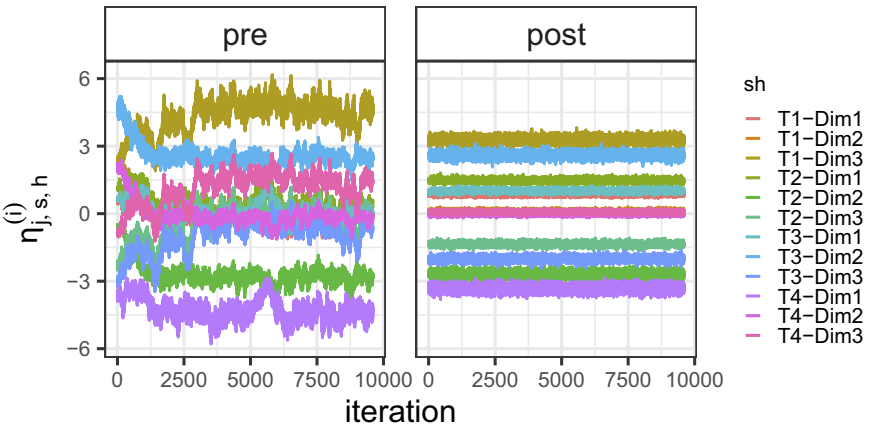


Figure 6. Diagnostics for synthetic data. Trace plots of the individual features $\eta^{(i)}_{j,s,h}$ sampled in the first subject of group 2 pre- and post-processing.

MCMC diagnostics for the simulation experiments are similar to those shown for the real data analysis presented in the next section and hence are omitted. Figure 6 shows the trace plots of the individual latent features before and after applying the algorithm to solve the identifiability issue in the first subject. We can see how the procedure described in Section 4.3 identifies the posterior samples of the different latent features, allowing meaningful inference and comparison at the level of the latent features $\eta^{(i)}_{j,s,h}$.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

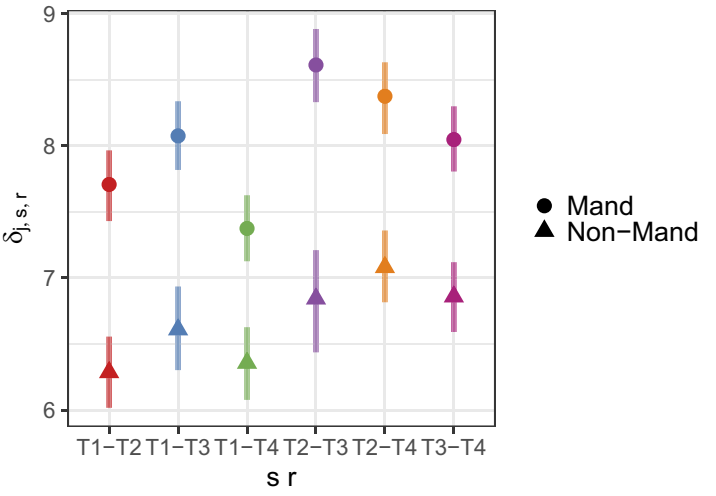


Figure 7. Results for real data. Posterior medians and 90% credible intervals of the group latent distances $\delta_{j,s,r}$ between stimuli.

Section 5 of the Supplementary Material provides additional simulations to systematically evaluate latent structure recovery in higher dimensions. Specifically, we consider scenarios with $S \in \{4, 8\}$ stimuli and $H_{\text{true}} \in \{2, 3\}$. Our model demonstrates robust performance, accurately recovering both the true latent distances and the correct latent dimensions across replications.

6. Application to neural tone distances

In this section, we discuss the results produced by our method applied to the two groups' tone distance data described in Section 2. Our inference goals, we recall, include understanding similarities in terms of stimulus distances between subjects within and across Mandarin and non-Mandarin listeners as well as comparing individual and group latent feature values. The selected number of dimensions H is 3 since, after the burn-in, the adaptive criterion discussed in Section 4.2 suggested sampling from $H = 2$ in 5% of the MCMC iterations and from $H = 3$ in 95% of the iterations.

Figure 7 shows the posterior point estimates and 90% credible intervals of the group distances $\delta_{j,s,r}$ between each of the $\binom{4}{2} = 6$ pairs of tones. We can see a similar ranking (across groups j) of the pairs of tones according to $\delta_{j,s,r}$. Indeed, after discarding individual differences, in both groups, the pairs of Mandarin tones $\{2, 3\}$ and $\{2, 4\}$ exhibited the strongest degree of neural dissimilarity, whereas pairs $\{1, 2\}$ and $\{1, 4\}$ exhibited the strongest degree of neural similarity. These findings are corroborated by empirical evidence from the preliminary analysis of the data in Section 2. Note that although the rankings between the point estimates are similar, they are not exactly the same across the two groups, and that there is more uncertainty, quantified via the posterior credible intervals, in the non-Mandarin group due to greater variability across these individuals. Moreover, consistent with recent neuroscience work (Llanos et al., 2017; Reetzke et al., 2018), we observe a better separation between neural representations of tones in the group of native listeners of Mandarin Chinese, relative to the group of non-native listeners. Figure 8 shows the posterior point estimates and 90% credible intervals of the individual distances $\delta^{(i)}_{j,s,r}$ as well as the corresponding observed distances. We can see that three latent common features reconstruct the observed distances very well. This indicates that the neural encoding of differences between multidimensional neural representations can be captured using three latent dimensions.

We also note greater uncertainty in the estimates of the denoised individual distances for the non-Mandarin subjects. Note that our Bayesian model allows us to quantify such uncertainty via posterior credible intervals while also allowing us to perform proper individual and group

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

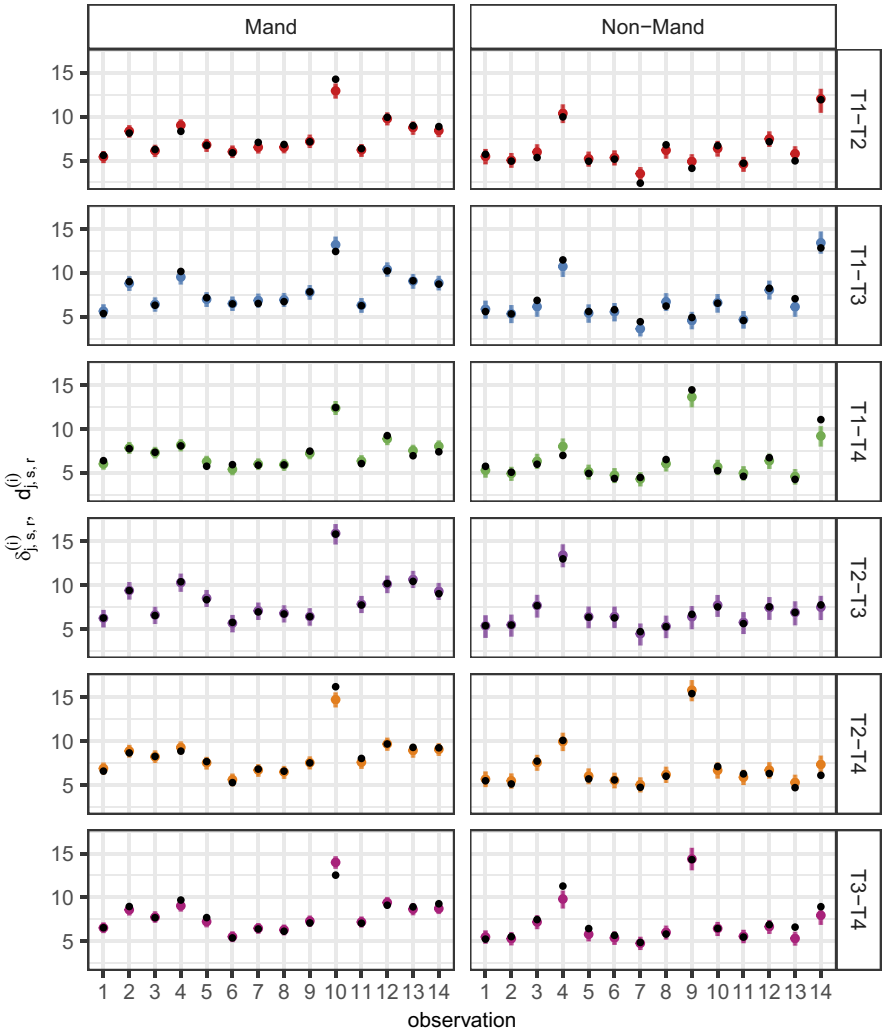


Figure 8. Results for real data. Posterior medians and 90% credible intervals of the individual latent distances $\delta^{(i)}_{j,s,r}$ between stimuli. Black points represent the observed distances $d^{(i)}_{j,s,r}$.

comparisons. Figure 8 also highlights the presence of some non-Mandarin speakers (e.g., subjects 7 and 11) who struggle to distinguish (i.e., smaller distances) between neural responses to specific pairs of tones (e.g., {T1, T2} and {T2, T3}) more than other subjects in the same group. Importantly, by addressing the identifiability issues, we are also able to perform inference on the group and individual latent feature values that reconstruct the distances. Notably, the model can also be used to infer structural differences at the level of the evoking stimuli with a high degree of specificity. These differences are depicted in Figure 9. In this figure, the first dimension encodes the difference between the neural representations of T4 (high-falling pitch) and the other tones: T1 (high-level pitch), T2 (low-rising pitch), and T3 (low-dipping pitch). Because T4 is the high-falling tone, this dimension reflects the neural encoding of differences in pitch direction (falling vs. level, rising, and dipping pitch). In contrast, the second dimension encodes the difference between the neural representations of T3 and T2. Because T3 and T2 are the low-dipping and low-rising tones, respectively, this dimension captures neural differences in pitch direction for low-onset tones (dipping vs. rising pitch). Lastly, the third dimension encodes the difference between the neural representations of T1 and T3. Because these tones are the ones with higher

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

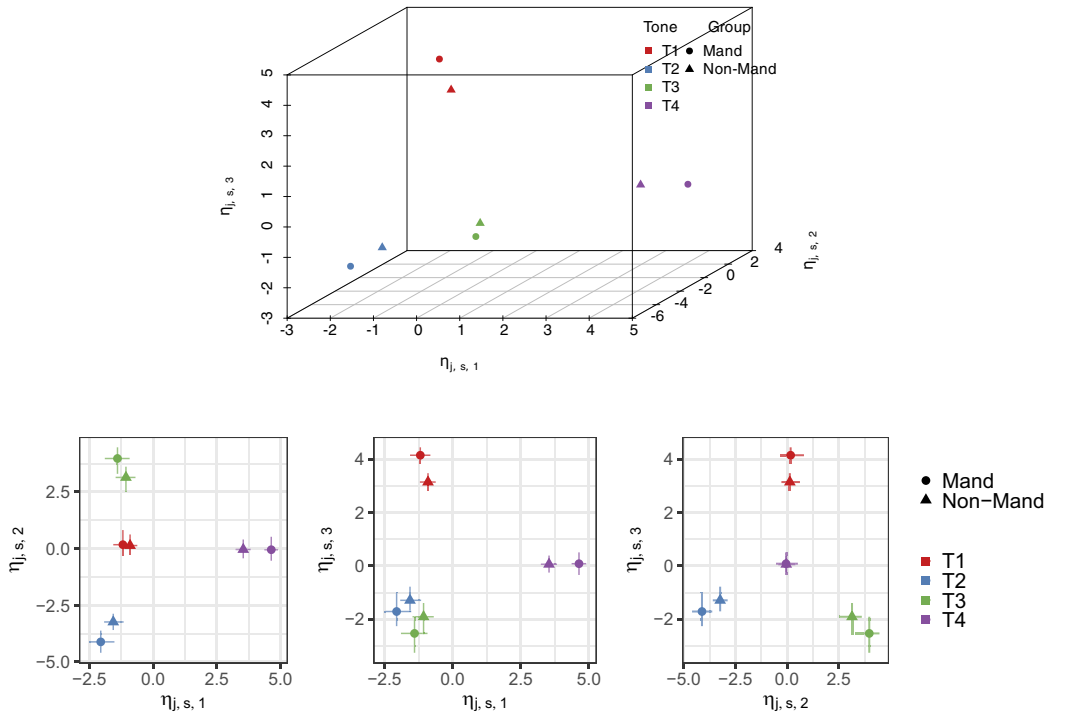


Figure 9. Results for real data. Upper panel: Three-dimensional scatter plot of posterior medians of the group latent feature values $\eta_{j,\,s,\,h}$. Lower panel: Two-dimensional representation of the posterior medians and 90% credible intervals of the group latent feature values $\eta_{j,\,s,\,h}$ in the different groups.

(T1) and lower (T3) pitch on average, this dimension may reflect the neural encoding of differences in pitch range (high vs. low range). Figure 9 also highlights the fact that the three latent features play the same role in the two groups and that the tones are more distinguished (i.e., $\eta_{j,\,s,\,h}$ more distant from 0) in all the features in Mandarin native speakers. See also the Supplementary Material for additional plots summarizing these results. Figure 10 shows the positions of different subjects on the feature space within and across the two groups and tones and the associated 90% credible intervals, presented separately for clarity. We see this at the individual level as well: the first feature is mainly useful for distinguishing tone 4. Aside from a few outlying large distances, the overall reconstruction is robust, confirming that the three-dimensional latent space effectively denoises and summarizes the individual dissimilarity matrices. The second and third features are mainly useful for representing other tones at both the group and individual levels. Moreover, although we note some similarities between Mandarin and non-Mandarin speakers, as expected, there is more individual variability in the latent features for non-native speakers. The individual latent features in Figure 10 also identify a few non-Mandarin speakers who struggle to distinguish the Mandarin tones, especially in the third dimension. Detailed convergence and identifiability diagnostics for the MCMC algorithm for the real data set are provided in Section 3 of the Supplementary Material.

7. Discussion

In this article, we proposed a novel MDS model for multi-group and multi-subject distances, motivated by auditory neuroscience research into the latent representation of speech sounds. Applying this model to Mandarin tone perception in native speakers and tone-naive English speakers, we characterized

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

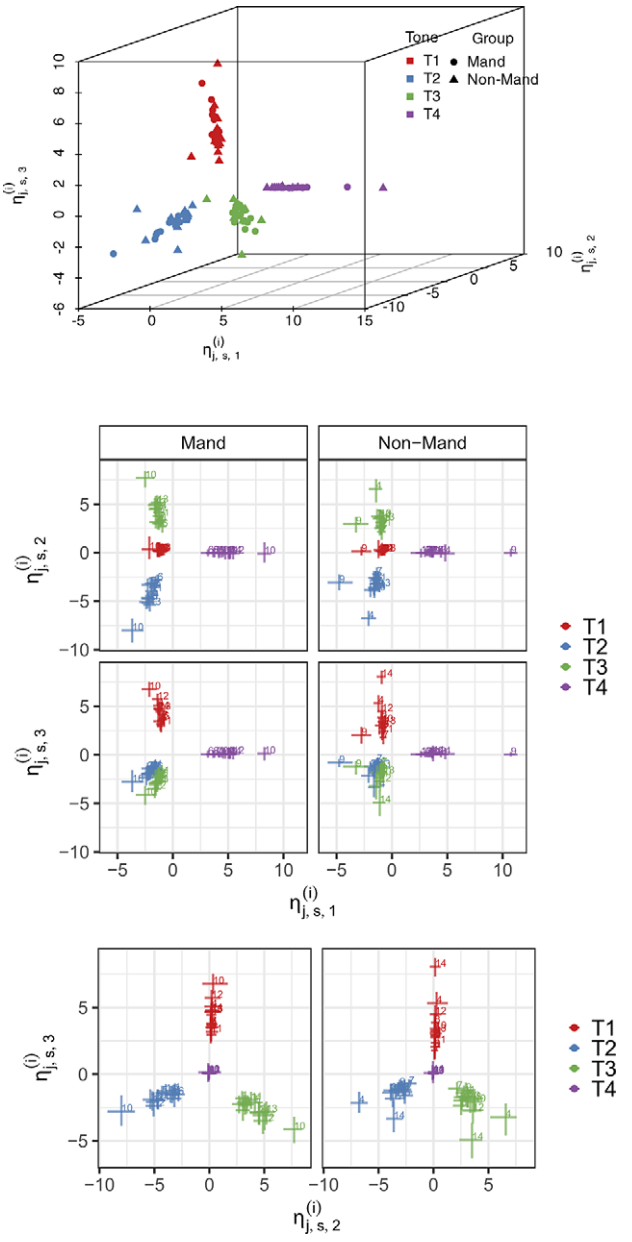


Figure 10. Results for real data. Upper panel: Three-dimensional scatter plot of posterior medians of the individual latent features $\eta^{(i)}_{j,\,s,\,h}$ between stimuli in the two groups. Lower panels: Two-dimensional representation of the posterior medians and 90% credible intervals of the individual latent feature values $\eta^{(i)}_{j,\,s,\,h}$ in the different groups.

latent features that reconstruct denoised empirical distances across both groups and individuals. Our framework identified structural differences in neural coding linked to long-term language exposure, while also revealing shared behaviors across groups.

Our proposed approach advances the MDS literature to enable neuroscientists to perform statistical comparisons across multi-subject and multi-group brain distances by mapping them to a biologically interpretable lower-dimensional common feature space. Our proposal allows borrowing of information

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

across groups and subjects, providing the means for a more comprehensive understanding of how the human brain processes speech signals. Moving beyond its role as a visualization tool, our proposed MDS framework incorporates formal uncertainty quantification and a data-driven approach for determining the optimal number of latent features.

Although our development is motivated by this auditory neuroscience application, the proposed model can be used to model structural differences between neural representations of other sensory stimuli (e.g., visual signals) captured by neuroimaging technologies, such as EEG, fMRI, and MEG. Prior neuroscience work (Kluender et al., 2019; Lewicki, 2002; Lotto & Holt, 2016; Stilp & Kluender, 2012) suggests that the neural coding of complex multidimensional sensory stimuli is informed by principles of dimensionality reduction. Our model could also be used to infer the latent neural dimensions needed by the human brain to encode differences between sensory signals and the impact of different kinds of experiences on the number of dimensions. Additionally, our approach can be used to address scientific problems in other fields, such as in bioinformatics to find lower-dimensional, biologically relevant features that summarize the recorded single-cell RNA gene expressions in several biomarkers, for example, to extend classical MDS used in a standard dimensionality reduction pipeline in bioinformatics research (Perraudeau et al., 2017). Such extensions might, however, require additional work to tailor our model to the specific scientific problem. Additionally, computations may need to be scaled up when datasets with a much larger number of stimuli are available. Efficient approximations to the posterior, for example, the variational Bayes algorithm developed for the simpler MDS model in Nguyen and Holmes (2017) or the Hamiltonian Monte Carlo-based sampler using representative subsamples for likelihood and gradient evaluations instead of the full distance matrix proposed recently by Sheth et al. (2026), may offer viable computational templates for such high-dimensional extensions of our work.

Some topics of our ongoing exploration include extensions to nonparametric frameworks utilizing more flexible likelihood functions, clustering listeners into latent subgroups, incorporating covariates, accommodating asymmetric distance matrices, and adaptations to longitudinal experiments capturing the dynamic evolution of the latent features.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2026.10131.

Data availability statement. The code and dissimilarity matrices necessary to reproduce the analyses are provided in the Supplementary Material.

Acknowledgements We thank an anonymous associate editor and three anonymous reviewers for comments on an earlier version leading to significant improvements.

Funding statement. This work was supported in part by the National Science Foundation grant NSF DMS-1953712 and the National Institute on Deafness and Other Communication Disorders grants R01DC013315 and R01DC015504. G.R. has been partially supported by the Italian grant MUR, PRIN project 2022CLTYP4.

Competing interests. The authors declare none.

References

Aguilar, O., & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, *18*, 338–357.

Basu, A. P. (2004). Identifiability. In *Encyclopedia of statistical sciences*. (eds S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson). John Wiley & Sons.

Bhattacharya, A., & Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, *98*, 291–306.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, *33*, 530–538.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications*. Springer.

Bradlow, E. T., & Schmittlein, D. C. (2000). The little engines that could: Modeling the performance of world wide web search engines. *Marketing Science*, *19*, 43–62.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*, 1959–1972.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*, 471–482.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607–649.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart–Young" decomposition. *Psychometrika*, *35*, 283–319.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, *25*, 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*, 148–156.

Commandeur, J. J. F., & Heiser, W. J. (1993). Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices. Department of Data Theory, University of Leiden Press, Leiden, The Netherlands.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. *Multivariate Analysis*, *5*, 501–522.

DeSarbo, W. S., & Cho, J. (1989). A stochastic multidimensional scaling vector threshold model for the spatial representation of "pick any/n" data. *Psychometrika*, *54*, 105–129.

DeSarbo, W. S., Kim, Y., Wedel, M., & Fong, D. K. H. (1998). A Bayesian approach to the spatial representation of market structure from consumer choice data. *European Journal of Operational Research*, *111*, 285–305.

DeSarbo, W. S., Park, J., & Scott, C. J. (2008). A model-based approach for visualizing the dimensional structure of ordered successive categories preference data. *Psychometrika*, *73*, 1–20.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*, 2457–2465.

Durante, D. (2017). A note on the multiplicative gamma process. *Statistics & Probability Letters*, *122*, 198–204.

Feng, G., Gan, Z., Llanos, F., Meng, D., Wang, S., Wong, P. C. M., & Chandrasekaran, B. (2021). A distributed dynamic brain network mediates linguistic tone representation and categorization. *NeuroImage*, *224*, 1–13.

Feng, G., Li, Y., Hsu, S.-M., Wong, P. C. M., Chou, T.-L., & Chandrasekaran, B. (2021). Emerging native-similar neural representations underlie non-native speech category learning success. *Neurobiology of Language*, *2*, 280–307.

Feng, G., Yi, H. G., & Chandrasekaran, B. (2019). The role of the human auditory corticostriatal network in speech learning. *Cerebral Cortex*, *29*, 4077–4089.

Fong, D. K. H., DeSarbo, W. S., Park, J., & Scott, C. J. (2010). A Bayesian vector multidimensional scaling procedure for the analysis of ordered preference data. *Journal of the American Statistical Association*, *105*, 482–492.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Gandour, J. T. (1978). The perception of tone. In *Tone* (ed V. A. Fromkin) (pp. 41–76). Elsevier.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*, 1–33.

Holbrook, A. J., Lemey, P., Baele, G., Dellicour, S., Brockmann, D., Rambaut, A., & Suchard, M. A. (2021). Massive parallelization boosts big Bayesian multidimensional scaling. *Journal of Computational and Graphical Statistics*, *30*, 11–24.

Khalighinejad, B., da Silva, G. C., & Mesgarani, N. (2017). Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience*, *37*, 2176–2185.

Kluender, K. R., Stilp, C. E., & Llanos, F. (2019). Long-standing problems in speech perception dissolve within an information-theoretic perspective. *Attention, Perception, & Psychophysics*, *81*, 861–883.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*, 81–95.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*, 161–168.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, *5*, 356–363.

Lin, L., & Fong, D. K. H. (2019). Bayesian multidimensional scaling procedure with variable selection. *Computational Statistics & Data Analysis*, *129*, 1–13.

Liu, B., Lubold, S., Raftery, A. E., & McCormick, T. H. (2024). Bayesian hyperbolic multidimensional scaling. *Journal of Computational and Graphical Statistics*, *33*, 869–882.

Llanos, F., German, J. S., Gnanateja, G. N., & Chandrasekaran, B. (2021). The neural processing of pitch accents in continuous speech. *Neuropsychologia*, *158*, 1–12.

Llanos, F., Xie, Z., & Chandrasekaran, B. (2017). Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. *Journal of Neuroscience Methods*, *291*, 101–112.

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.

Lotto, A. J., & Holt, L. L. (2016). Speech perception: The view from the auditory system. In *Neurobiology of language* (eds G. Hickok & S. L. Small) (pp. 185–194). Elsevier.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010.

Nguyen, L. H., & Holmes, S. (2017). Bayesian unidimensional scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations. *BMC Bioinformatics*, *18*, 65–79.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, *10*, 1–11.

Oh, M. S., & Raftery, A. E. (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, *96*, 1031–1044.

Papastamoulis, P., & Ntzoufras, I. (2022). On the identifiability of Bayesian factor analytic models. *Statistics and Computing*, *32*, 1–29.

Park, J., DeSarbo, W. S., & Liechty, J. (2008). A hierarchical Bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika*, *73*, 451–472.

Perraudeau, F., Risso, D., Street, K., Purdom, E., & Dudoit, S. (2017). Bioconductor workflow for single-cell RNA sequencing: Normalization, dimensionality reduction, clustering, and lineage inference. *F1000Research*, *6*, 1–28.

Praturu, A., & Sharpee, T. (2022). A Bayesian approach to hyperbolic multi-dimensional scaling. bioRxiv, pp. 1–13.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raizada, R. D. S., Tsao, F.-M., Liu, H.-M., Holloway, I. D., Ansari, D., & Kuhl, P. K. (2010). Linking brain-wide multivoxel activation patterns to behaviour: Examples from language and math. *NeuroImage*, *51*, 462–471.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, *42*, 241–266.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society: Series A*, *145*, 285–303.

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*, 1419–1427.

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *59*, 731–792.

Roberts, G. O., & Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, *44*, 458–475.

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*, 349–367.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*, 125–140.

Sheth, A., Smith, A., & Holbrook, A. J. (2026). Sparse Bayesian multidimensional scaling(s). *Computational Statistics*, *41*, 12.

Stilp, C. E., & Kluender, K. R. (2012). Efficient coding and statistically optimal weighting of covariance among acoustic attributes in novel sounds. *PLoS One*, *7*, 1–13.

Swartz, T. B., Haitovsky, Y., Vexler, A., & Yang, T. Y. (2004). Bayesian identifiability and misclassification in multinomial data. *Canadian Journal of Statistics*, *32*, 285–302.

Takane, Y., & Carroll, J. D. (1981). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, *46*, 389–405.

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, *17*, 401–419.

Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.

Yanchenko, A. K., & Hoff, P. D. (2020). Hierarchical multidimensional scaling for the comparison of musical performance styles. *Annals of Applied Statistics*, *14*, 1581–1603.

Young, F. W., Takane, Y., & Lewyckyj, R. (1978). ALSCAL: A nonmetric multidimensional scaling program with several individual-differences options. *Behavior Research Methods & Instrumentation*, *10*, 451–453.

Zinnes, J. L., & Wolff, R. P. (1977). Single and multidimensional same-different judgments. *Journal of Mathematical Psychology*, *16*, 30–50.

Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R. D. S. (2016). Semantic structural alignment of neural representational spaces enables translation between English and Chinese words. *Journal of Cognitive Neuroscience*, *28*, 1749–1759.

Cite this article: Rebaudo, G., Llanos, F., Chandrasekaran, B. and Sarkar, A. (2026). Bayesian Mixed Multidimensional Scaling for Auditory Processing. *Psychometrika*, 1–23. https://doi.org/10.1017/psy.2026.10131

Downloaded from https://www.cambridge.org/core. 10 Nov 2026 at 09:18:14, subject to the Cambridge Core terms of use.