JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Temporally dissociable neural representations of pitch height and chroma

https://doi.org/10.1523/JNEUROSCI.1567-24.2024

Received: 18 August 2024
Revised: 18 November 2024
Accepted: 26 December 2024

Temporally dissociable neural representations of pitch height and chroma

Abbreviated title: Dissociable neural dynamics of pitch representations

Authors: Andrew Chang[1*], David Poeppel[1,2,3,4], Xiangbin Teng[5*]

[1] Department of Psychology, New York University, New York, NY, USA 10003

[2] Center for Language, Music, and Emotion (CLaME), New York University, New York, NY, USA 10003

[3] Music and Audio Research Lab (MARL), New York University, New York, NY, USA 10003

[4] Max Planck Society, Munich, Germany 80084

[5] Department of Psychology, Chinese University of Hong Kong, Hong Kong SAR, China

*Corresponding Authors

Andrew Chang

ac8888@nyu.edu

Department of Psychology, New York University

6 Washington Place, New York, NY 10003

Xiangbin Teng

xiangbinteng@cuhk.edu.hk

Department of Psychology, The Chinese University of Hong Kong

Sino Building, Shatin, N.T., Hong Kong SAR, China

Conflict of interest: The authors declare no competing financial interests.

Acknowledgments: A.C. was supported by a Ruth L. Kirschstein Postdoctoral Individual National Research Service Award, National Institute on Deafness and Other Communication Disorders/National National Institutes of Health (F32DC018205) and Leon Levy Scholarships in Neuroscience, Leon Levy Foundation/New York Academy of Sciences. X.T. was supported by Improvement on Competitiveness in Hiring New Faculties Funding Scheme, the Chinese University of Hong Kong (4937113). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the Poeppel Lab members at New York University, Max Planck Institute for Empirical Aesthetics, and Ernst Struengmann Institute for Neuroscience for their comments and support. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. During the preparation of this work the authors used ChatGPT for language editing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Abstract

The extraction and analysis of pitch underpin speech and music recognition, sound segregation, and other auditory tasks. Perceptually, pitch can be represented as a helix composed of two factors: height monotonically aligns with frequency, while chroma cyclically repeats at doubled frequencies. Although the early perceptual and neurophysiological mechanisms for extracting pitch from acoustic signals have been extensively investigated, the equally essential subsequent stages that bridge to high-level auditory cognition remain less well understood. How does the brain represent perceptual attributes of pitch at higher-order processing stages and how are the neural representations formed over time? We used a machine learning approach to decode time-resolved neural responses of human listeners (10 females and 7 males) measured by magnetoencephalography across different pitches, hypothesizing that different pitches sharing similar neural representations would result in reduced decoding performance. We show that pitch can be decoded from lower-frequency neural responses within auditory-frontal cortical regions. Specifically, linear mixed-effect modeling reveals that height and chroma explain decoding performance of delta band (0.5-4 Hz) neural activity at distinct latencies: a long-lasting height effect precedes a transient chroma effect, followed by a recurrence of height after chroma, indicating sequential processing stages associated with unique perceptual and neural characteristics. Furthermore, the localization analyses of the decoder demonstrate that height and chroma are associated with overlapping cortical regions, with differences observed in the right orbital and polar frontal cortex. The data provide a perspective motivating new hypotheses on the mechanisms of pitch representation.

Keywords: magnetoencephalography (MEG), machine learning, decoding, representational similarity analysis, auditory perception, octave equivalence

Significance Statement

Pitch is fundamental to various facets of human hearing, including music appreciation, speech comprehension, vocal learning, and sound source differentiation. How does the brain encode the perceptual features of pitch? By applying machine learning techniques to time-resolved neuroimaging data of individuals listening to different pitches, our findings reveal that pitch height and chroma-two distinct features of pitch-are associated with different neural dynamics within the auditory-frontal cortical network, with height playing a more prominent role. This offers a unified theoretical framework for understanding the perceptual and neural characteristics of pitch perception and opens new avenues for noninvasively decoding human auditory perception to develop brain-computer interfaces.

Introduction

Pitch is central to many aspects of human auditory cognition, including the appreciation of music, understanding speech, vocal learning, segregating sound sources, and forming auditory objects. In both psychoacoustics and auditory neuroscience, the functionally early mechanisms of extracting pitch from sounds have been extensively studied, yielding a range of foundational insights (e.g., de Cheveigné, 2010; Licklider 1951; Meddis & O'Mard, 1997; Oxenham, 2023). However, one question that is less typically addressed is this: once pitch has been extracted, how does the brain represent its perceptual attributes, given that pitch is not a monolithic concept? Understanding the neural representation of pitch will, this study conjectures, bridge a gap between low-level auditory processing and high-level auditory cognition.

Many behavioral experiments demonstrate that certain (often musical) tonal stimuli are perceived as more similar to each other than others. For example, the pitches of adjacent piano keys and keys an octave apart are perceptually more similar than other pairs. This perceptual pattern can be effectively represented by a helix model (Figure 1A), incorporating height and chroma factors, where a shorter distance between two pitches represents a higher perceptual similarity (Shepard, 1982; Ueda & Ohgushi, 1987). Height as a dimension log-linearly corresponds to the fundamental frequency (F0) of the tone. Two pitches with a smaller height difference are perceived as more similar. Height information is crucial for separating sound sources, such as distinguishing individuals' voices according to their different frequency ranges.

Chroma, on the other hand, signifies the cyclic and recurring patterns of perceptual similarities among pitches separated by an octave (doubling of F0 frequency), a phenomenon commonly referred to as octave equivalence (Révész, 1954; Shepard, 1964). Chroma plays an important role in recognizing acoustic patterns independent of the sound sources. In speech, for example, octave equivalence facilitates human vocal learning by equalizing the vocal range difference across ages and genders (Wagner & Hoeschele, 2022). In music, a melody shifted by an octave is recognized as being similar or even the same (Demany & Armand, 1984; Wright et al., 2000). Although the way one divides an octave interval into pitch classes (e.g., C, D, E or Do, Re, Mi) can vary across cultures, the recurrence of pitch classes per octave appears a commonly shared feature (Burns 1999; but see Jacoby et al., 2019). Octave equivalence cannot be simply explained by low-level acoustic matching, even though half of the harmonic frequencies inevitably overlap between pitches an octave apart (Jacoby et al., 2019; Wagner et al., 2022).

Despite important characterizations of the perceptual representation of pitch over many years, the neural representation of pitch remains incompletely characterized. Pitch is primarily represented in the auditory cortex (e.g., Allen et al., 2022; Bendor & Wang, 2005; Pantev et al., 1989; Penagos et al., 2011; Zatorre, 1988; Zatorre et al., 2002). Specifically, the posterior subregion (posterior planum temporale) is associated with height while the anterior subregion (extending from Heschl's gyrus into the planum polare) is associated with chroma (Briley et al., 2013; Moerel et al., 2015; Warren et al., 2003). Although these basic attributes of pitch seem spatially separable in cortex, several neurobiological questions remain unaddressed. The present experiment is designed with two questions in mind, motivated by issues of timing and neural dynamics. When do the representations of height and chroma emerge? Are they formed at the same time or in different processing stages? And how is the neural activity associated with perceptual similarity?

We apply a decoding analysis approach to the neural responses of pitch, investigated at a high temporal resolution, recorded by magnetoencephalography (MEG). If two pitches are perceptually more *dissimilar*, their neural representation should also be more dissimilar, and therefore the decoding performance is predicted to be higher. Therefore, by analyzing how the time course of the dissimilarity pattern among pitch-pairs resembles the perceptual dissimilarity predicted by pitch height and chroma factors, we can reveal the temporal neural dynamics of pitch representation.

Methods

Resource Availability

All stimuli, raw data, processed, and analysis codes have been deposited at an OSF repository (https://doi.org/10.17605/OSF.IO/NHCGJ) and are publicly available as of the date of publication.

Participants

This study incorporated data from 17 right-handed normal hearing, non-musician participants (self-reported; age range: 18-35 years; 10 female). Individual MRI scans were obtained for 14 participants. For the remaining three, who were unable to undergo MRI scanning, a standardized template was used. All participants provided written informed consent prior to their involvement and were compensated monetarily. The study was approved by the local ethics committee of the University Hospital Frankfurt.

Stimuli

Eight signals spanning three octaves in the pitch classes C, E, and G# were used: G#6 (415.3 Hz), C7 (523.3 Hz), E7 (659.3 Hz), G#7 (830.6 Hz), C8 (1046.5 Hz), E8 (1318.5 Hz), G#8 (1661.2 Hz), and C9 (2093.0 Hz). The fundamental frequencies of pitches are logarithmically equidistant, with a step corresponding to 1/3 octave (or 4 semitones). Each pitch was presented in three instrumental timbres (flute, violin, and piano), resulting in 24 individually recorded tones, retrieved from the NSynth Dataset (Engel et al., 2017). The duration of each recording was 0.4 seconds, and the sampling rate was 44.1 kHz. Rather than sinusoidally synthesized tones or a single timbre (as has been typical in previous studies), we used tones from various instruments to permit the generalizability of our findings. The amplitude of tones was pseudorandomly set between 67 and 73 dB-SPL for each presentation.

According to the Western music scale (i.e., 12-tone equal temperament, which divides an octave into 12 logarithmically equidistant pitch classes), we quantified the pairwise pitch relationships as height difference (measured in octaves, Figure 1B) and chroma equivalence (where 'True' indicates the same pitch class, Figure 1C).

A cochleagram representation was calculated for each stimulus (Figure 1D), with 128 filters between 50 Hz and 20 kHz and sample factor 2, constructed by the *pycochleagram* package (https://github.com/mcdermottLab/pycochleagram), and pairwise Spearman's correlations were calculated to assess cochleagram similarity (Figure 4B, where higher values indicate greater similarity).

Procedure

Each participant was exposed to 2400 stimuli (8 frequencies x 3 instrumental timbres x 100 repetitions per stimulus). The materials were presented in 100 runs, such that each run contained each stimulus, presented in a pseudorandom order, with half of the runs having consecutive tones in the same pitch (regardless of instrument). The inter-onset intervals within a run roved between 0.9 and 1.4 seconds. To ensure that participants were attending to the stimuli, after each run, they were asked to report whether there were any consecutive tones (i.e. stimuli of the same pitch) by making a yes/no button-press response using their right hand. The next run started 1.5-2 seconds after the response. (Although the behavioral data was unavailable due to technical failure, the post-block, offline behavioral responses are not of critical relevance to the goals and interpretation of this study.) In sum, each stimulus was presented 100 times, and each pitch was presented 300 times.

MEG Recording and Preprocessing

Each participant sits in a semi-dark room while their brain responses were recorded by an MEG with CTF 275 channels at 1200 Hz sampling rate (Figure 1E). They were required to look at a white cross in the center of a black screen throughout each trial.

We used Python package *MNE* (Gramfort et al., 2013) to implement the following signal processing. An independent component analysis (ICA) was used to remove artifact signals (Delorme et al., 2007). To speed up the performance of ICA, the raw sensor-space MEG data were first high-pass filtered at 1 Hz (zero-double phase), downsampled to 200 Hz, and segmented into -0.5 to 1 second epochs (time-locked to stimulus onset). The ICs reflecting artifact, including eye blinking, eye movement, electrocardiogram, and 50 Hz powerline noise, were identified based on visual inspection. Then we went back to the raw unfiltered sensor-space continuous data and removed the artifact ICs.

The artifact-clean unfiltered sensor-space continuous signal is then band-passed filtered (zero-double phase) into different delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-25 Hz) and low gamma (25-40 Hz) bands, downsampled to 200 Hz, segmented into -0.2 to 0.5 second epochs (tone onset time at 0 second), baseline corrected to prestimulus period, and rejected the epochs with maximum peak-to-peak signal amplitude exceeding 4e-12.

We converted signals from sensor-space to source-space in the follow steps. (1) The T1 images available from 14 participants were used to do participant-specific surface reconstruction and create BEM model and solution with FreeSurfer watershed algorithm, using FreeSurfer 7.1.1 (Fischl, 2012). For the 3 participants unavailable with T1 image, the fsaverage template brain was used. (2) The noise covariance matrices were empirically computed from the prestimulus epoched data. (3) The forward solutions were then estimated. (4) The inverse

operator was applied to epochs using dSPM method with regularization parameter lambda2 = 1/9. (5) The signals in the source-space were morphed and downsampled to the fsaverage ico3 template space. This step is to enhance the statistical power and mitigate overfitting of the decoding analysis by reducing the dimensions of the data and lower multicollinearity among the source signals. (6) To also mitigate overfitting, we reduced the dimensionality by only including signals from the bilateral early auditory cortex, auditory association cortex, and frontal lobe (inferior frontal cortex, insular and frontal opercular cortex, orbital and polar frontal cortex, dorsolateral prefrontal cortex), total in 316 vertices (Figure 1F), were used for subsequent decoding analysis, following the HCP-MMP1 atlas (Glasser et al. 2016), as these regions roughly covers the auditory ventral "what" pathway or have been reported to be associated with processing pitch (Hall & Plack, 2009; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000). In addition to auditory cortex, we included these broad frontal regions for several reasons. First, certain frontal areas are commonly involved in pitch processing (e.g., the inferior frontal area; Griffiths et al., 1999; Zatorre et al., 1994), whereas others are less clearly associated (e.g., the dorsolateral prefrontal cortex, which relates to pitch labeling, and the insular cortex, involved in processing pitch patterns; Bermudez & Zatorre, 2005; Wong et al., 2004). Consequently, there is no definitive guidance on which regions should be specifically included or excluded. Additionally, using decoding analyses instead of traditional fMRI contrasts may reveal more distributed and potentially unexpected regions involved in the task. Anyway, our machine learning model was designed to automatically shrink the weights of locations with minimal contribution, effectively excluding those regions.

Multivariate Pattern Decoding Analysis

To quantify the representational similarity of each pitch-pair, we used a Ridge classifier as the decoder, under Python packages *MNE* and *scikit-learn* (Pedregosa et al., 2011). The function *RidgeClassifierCV()* implements cross-validation procedure to automatically tune the optimal regularization (complexity) parameter (α) between 10^{-2} and 10^{3}, consisting of 41 logarithmically-spaced grid points. Compared to least square method, ridge method can avoid overfitting and multicollinearity by shrinking the coefficients (w) of the less important sources close to 0 (McDonald, 2009). It is very similar to the common logistic regression with L2 regularization, although it utilizes mean square error with L2 penalty instead of cross-entropy as the loss function (X: design matrix; y: dependent variable):

$$\min_{w} \left|\left|Xw - y\right|\right|_2^2 + \alpha \left|\left|w\right|\right|_2^2$$

The decoder was applied to each pitch-pair combination for each participant at each time point, irrespective of the instruments involved. The decoder used source brain responses at each frequency band from a single time point across trials as input, with each time point comprising 316 features across approximately 600 trials (300 trials per pitch). A stratified 5-fold cross-validation was performed, training on 80% of trials and evaluating on the remaining 20% in each iteration, with trials randomly partitioned in each fold. The magnitude of responses from each source was z-normalized based on the training data. Performance was assessed using the area under the curve of the receiver operating characteristic (ROC-AUC). If two pitches have a dissimilar neural representation at a certain time point, the decoding ROC-AUC will be high. In

the end, an 8x8 diagonally symmetric representational dissimilarity matrix was calculated for every subject at each time point (Figure 1G).

Statistical Analyses

To explain the representational dissimilarity matrix (Figure 1G) at each time point by pitch height difference (Figure 1B) and chroma equivalence (Figure 1C), we fitted a linear mixed-effects model (LMEM) at each time point using *statsmodels* in Python (Seabold & Perktold, 2010). The LMEM, an extension of linear regression, explores the predictors of interest (fixed effects), including height difference, chroma equivalence, and their interaction, while accounting for variability across participants (random effect) by including both random intercept and slope components (e.g., Harrison et al., 2018). We reported marginal and conditional R^2 as goodness-of-fit metrics, which are commonly used for LMEMs since standard linear regression R^2 does not apply (Nakagawa & Schielzeth, 2013). Though not directly comparable to traditional R^2, they offer a similar indication of model fit.

We assessed multicollinearity among the independent variables in the LMEMs using variance inflation factor (VIF) analysis. For the model involving height difference, chroma equivalence, and their interaction, the VIFs were 1.3, 7.7, and 8.5, respectively—all below the common cutoff of 10 (e.g., Thompson et al., 2017), indicating low multicollinearity and a robust model fit. Conversely, the model involving height difference, cochleagram similarity, and their interaction yielded VIFs of 58.6, 6.7, and 65.4, indicating multicollinearity and should be interpreted with caution. This multicollinearity is not unexpected, as both factors essentially refer to conceptually similar features—one defined in musicological terms (height) and the other in acoustical terms (cochleagram). It is particularly inevitable given the realistically recorded instrument tones used in this study. Nonetheless, we report the models including cochleagram for transparency and hope this information will aid future studies in synthesizing tones with dissociated cochleagram similarity and height difference.

To analyze ROC-AUC time-series data (Figure 2), we used a one-sample cluster-based permutation test (Maris & Oostenveld, 2007), flipping the signs based on whether values were above or below 0.5 within each participant. To conduct the cluster-based permutation test on LMEM (Figure 3A, Figure 4A and Figure 4C), the dependent and independent variables were randomly shuffled within each participant before fitting an LMEM for each iteration. (Additionally, a small uniform-random number between -0.0001 and 0.0001 was added to the ROC-AUC in cases where the LMEM failed to converge, which occurred in a very small number of iterations.) For all these tests, the number of permutations was set to 10k, the clustering threshold was set at an absolute t-value of 2.5 with a minimum width of 20 ms (5 consecutive time samples), and 2-tailed cluster-level p-values were reported.

Results

In a first step, we examined which frequency band(s) represent neural information in this paradigm to distinguish pitches. In the by-subject analysis, we collapsed all pitch-pairs per participant (*df* = 16), and one-sample cluster-based permutation tests showed that the ROC-AUC was significantly above chance in the delta, theta, and alpha bands (Figure 2A).

Orthogonally, in the by-item analysis, we collapsed all participants per pitch-pair (*df* = 27), and the ROC-AUC was also significantly above chance in the delta, theta, and alpha bands (Figure 2B). These findings suggest that neural encoding of pitch profiles is predominantly reflected in the lower frequency bands. Therefore, only delta, theta, and alpha bands are included in the subsequent analyses.

The ROC-AUC values were in the range of 0.50 to 0.55 (Figure 2). These relatively modest decoding values are likely caused by the diversity in sound types (different instruments) in our experimental design, which introduces additional variability in the neural responses, consequently impacting the classifier's performance. Despite this experiment-inherent challenge, the magnitude of our ROC-AUC aligns with findings from comparable pitch decoding MEG and fMRI studies that used single timbres (e.g., Czoschke et al., 2021; Sankaran et al., 2018, 2020), and our findings have a better generalizability. Importantly, the identification of statistically significant above-chance clusters implies that the signals are sufficiently robust for further analyses.

We next used linear mixed-effects modeling (LMEM) to examine the neural representational dissimilarity (ROC-AUC in Figure 2) at each time point can be explained by pitch height difference (Figure 1B), pitch chroma equivalence (Figure 1C), and their interaction. In the delta band, the time series of t-values for each coefficient of each independent variable are shown in Figure 3A. The cluster-based permutation test showed that the height difference was positively associated with ROC-AUC at approximately 95-285 ms and 335-415 ms. This suggests that neural responses were more different when listening to two pitches with a larger height difference, consistent with the height dimension of the helix model (Figure 1A). Chroma equivalence was negatively associated with ROC-AUC, at approximately 300-320 ms. This suggests that neural responses were harder to distinguish when two pitches belonged to the same pitch class, consistent with the chroma dimension of the helix model (Figure 1A). Finally, there was a significant interaction effect at approximately 300-320 ms. The fitted LMEM coefficients at each time point in the delta band are visualized in Figures 3B and C.

We also conducted, using the same approach, several control analyses. First, we applied the same LMEM to the ROC-AUC from the theta and alpha bands, but none of the factors were significant (Figure 4A). Second, we tested an alternative hypothesis: was the effect of chroma equivalence driven by the higher acoustic similarity between pitches that are an octave apart? We calculated the cochleagram of each tone, obtained the pairwise similarities (Figure 4B), replaced the chroma equivalence factor in the LMEM with this similarity matrix, and redid the analyses. However, the LMEM did not reveal any significant findings in any frequency bands (Figure 4C), consistent with previous studies indicating that octave equivalence cannot be explained by acoustic similarity (Jacoby et al., 2019; Wagner et al., 2022). Note that models incorporating cochleagram similarity should be interpreted with caution due to intrinsic multicollinearity among the independent variables (see *Statistical Analyses*), which likely lead to imprecise parameter estimates and thus fail to replicate the height effect observed in Figure 3A.

We next conducted a temporal generalization analysis with the same cross-validation method to examine whether the two significant height difference clusters in Figure 3A represent similar pitch information. The rationale is that if a classifier trained at one time point exhibited above-chance performance at another time point, neural responses at these two time points represent similar information. We used a one-sided t-test to examine which paired train-test time points were above chance (Figure 5), and false discovery rate (FDR) was used for controlling

for multiple comparisons (alpha = 0.05) (Benjamini & Hochberg, 1995). We did not use a cluster-based permutation test here to control for multiple comparisons because the performance on the diagonal line (training time equals testing time) will be certainly higher than other locations. The analyses showed that, unsurprisingly, there was a significant region on the diagonal line, representing each classifier performing best at the trained time point with some generalizability to temporally adjacent points. More interestingly, the classifier trained on the approximately 140-190 ms window could explain the neural responses in the 400-500 ms window, with this effect being stronger than that of a cluster in the diagonally symmetrical position (late window explaining the early window). Although the effects were both weaker than the effect on the diagonal line, and the time window was not perfectly aligned with the clusters shown in Figure 3A, this finding invites the interpretation that the delta band neural activity reflects similar pitch height information peak at two distinct latencies. Moreover, activity in the early window is more robust for training a classifier that generalizes to the late window than vice versa.

To explore which brain regions are implicated in representing pitch height and chroma, we mapped the weights of the classifier to the cortex (Figure 6A). We further averaged the weights within each atlas label and normalized the mean weight as a percentage to equalize the magnitude of the sum of weights across time (Figure 6B), with the baseline level at 8.33% (100% divided by 12 regions). We specifically focused on the time windows representing height (180-230 ms, the peak range of the height effect) and chroma (300-320 ms). First, the weights at the right early auditory region at both height ($t(16) = 3.92$, $p = 0.001$, Cohen's $d = 0.98$) and chroma ($t(16) = 3.79$, $p = 0.002$, Cohen's $d = 0.95$) time windows were above baseline, and so were the right insular and frontal operculum for both height ($t(16) = 6.44$, $p < 0.001$, Cohen's $d = 1.63$) and chroma ($t(16) = 6.07$, $p < 0.001$, Cohen's $d = 1.52$) windows. Second, we tested the difference between windows within each region, and the weight in right orbital and polar frontal cortex was higher in the chroma than the height window ($t(16) = 4.47$, $p < 0.001$, Cohen's $d = 1.12$). Finally, the lateralization analyses (Figure 6C) showed rightward lateralization in multiple regions and time windows, including early auditory cortex during the chroma window ($t(16) = 3.02$, $p = 0.008$, Cohen's $d = 0.76$), auditory association cortex during the height window ($t(16) = 3.20$, $p = 0.006$, Cohen's $d = 0.80$), and inferior frontal cortex during the height window ($t(16) = 3.53$, $p = 0.003$, Cohen's $d = 0.88$). Both height and chroma seem to engage the right hemisphere more than the left and involve shared cortical regions across auditory and frontal areas; however, the weights of the classifier at the right orbital and polar frontal cortex reflected their differences. Note that interpreting the map of the classifier's weights is not straightforward, as a higher weight does not necessarily imply a larger influence on the classification result (Kriegeskorte & Douglas, 2019). Also, our MEG decoding analyses should be considered complementary to previous fMRI studies from an encoding perspective (e.g., Warren et al., 2003), although they are not directly comparable. Nevertheless, we hope these exploratory findings, despite not controlling for multiple comparisons (Bender & Lange, 2001), can contribute to shaping future localization studies on this topic.

Discussion

The study applied machine learning-based decoding analyses to MEG data, revealing the neural dynamics of pitch representation. Three novel findings are reported. First, the low frequency (0.5-13 Hz) neural activity measured in the auditory and frontal cortex represents

pitch profiles differently, with the delta band (0.5-4 Hz) specifically encoding the relative relationships among the pitches (height difference and chroma equivalence). Second, height and chroma representations are processed at different stages, with height preceding chroma. The height representation emerges as early as 100 ms, is sustained for 200 ms, and reappears at approximately 400 ms. The chroma representation arises later, at about 300 ms post stimulus onset and for a brief duration, and its neural representations are consistent with the pitch helix model below two octaves (Figure 1A & Figure 3C). Third, exploratory analyses showed that the neural representations of height and chroma are both right lateralized, convergent with many reports in the literature (e.g., Abrams et al., 2024; Albouy et al., 2020; Zatorre et al., 2002), while the orbital and polar frontal cortex appear to be involved in height and chroma representation in different ways.

Height is arguably the most salient dimension of a pitch representation. The observation that pitch-pairs with larger height differences have lower neural representational similarity (better decoding performance) is consistent with previous behavioral findings that pitches which have larger fundamental frequency differences in logarithmic scale will be perceived more differently (e.g., Ueda & Ohgushi, 1987). In the brain, a larger pitch height difference elicits stronger event-related potential (ERP) responses (e.g., He et al., 2009; Näätänen et al., 2007), and sounds with closer pitch-eliciting frequencies activate spatially adjacent neural ensembles in the tonotopically organized auditory cortex (Allen et al., 2022). The height representation, spanning from 100 to 300 ms post stimulus onset, is well aligned with the latencies of the auditory ERP components associated with pitch perception, such as N1 and the pitch onset response (e.g., de Cheveigné, 2010). It suggests that the height representation is likely formed at a preattentive, automatic processing stage (Regev et al., 2019), and could be more or less universal in listeners (e.g., Jacoby et al., 2019; Kallman, 1982). Consistent with our findings, a recent MEG study analyzed the neural representational similarity of pitches in a Western tonal music context and showed a consistent finding that the broad-band neural dissimilarity at 150-300 ms is associated with coarse-grained pitch height difference (Sankaran et al., 2020). Our fine-grained data further pinpoints that this association is specific to the delta band and describes its log-linear mathematical relationship.

The reappearance of a pitch height representation at approximately 400 ms may reflect a feedback or recurrent communication between the auditory and high-level frontal regions and possibly be associated with the awareness of pitch height. Recurrent processing is often associated with the conscious awareness of the sensory information (Förster et al., 2020). Consistent with this hypothesis, individuals with congenital amusia, a lifelong deficit in musical pitch and its awareness, feature atypical recurrent processing in the right frontotemporal network (Peretz, 2016). Targeted studies are needed to investigate the neural mechanism and the perceptual function of this reappearance.

The chroma representation of pitch, in our study, is elicited at a later latency and is only visible in these analyses for a short duration. This is consistent with the behavioral studies that show that the perceptual effects of chroma are less robust, as they can only be observed in specific task designs which require using chroma information and/or engaging attention (e.g., Hoeschele et al., 2012; Kallman, 1982; Regev et al., 2019; Wagner et al., 2022). Yet our findings demonstrate that the neural representation of pitch chroma can automatically occur without a task that explicitly taxes chroma. Note that the weaker observable behavioral or neural effects of chroma should not be interpreted as indicating that chroma is perceptually less

important, as this remains untested. Interestingly, the latency of the chroma effect overlaps with the latency of the P3a, an ERP component associated with attentional processing of sensory input (Polich, 2007). Together, these properties suggest a hypothesis that chroma information is likely implicit and only accessible at the attentional processing stage for a limited time, which can account for the difficulty of observing its perceptual effects.

An unexpected and intriguing finding is that neural representations are more similar when pitches are one octave apart, reflecting the effect of octave equivalence and aligning with the helix model; however, they appear more dissimilar when two octaves apart, which is inconsistent with the helix model. This finding provides a minor but noteworthy counterexample to the helix model, aligning with empirical behavioral findings suggesting that the octave equivalence effect may be limited to immediately neighboring octaves (e.g., Wagner et al., 2022). Note that, as there were fewer instances of pitch-pairs with larger height differences in our experiment, their estimation precision will inevitably be lower. Therefore, the model estimation at two octaves should be taken with caution. Together, our findings suggest a need to reexamine the helix model across varying numbers of octaves.

The height and chroma representations differ in engaging the right orbital and polar frontal cortex. The implication of orbital and polar frontal cortex was less clear in the literature, despite this region is involved in some pitch or music related tasks (e.g., Fasano et al., 2023; Limb, 2006; Zatorre et al., 1996). One fMRI study showed that these regions are associated with tracking Western tonal structure (Janata et al., 2002). This may be driven by the encoding of relationships among pitches (i.e., height and chroma), as these are fundamental elements in forming musical tonality. Other neuroanatomical studies have reported that the orbitofrontal cortex is connected to auditory cortex via the ventral pathway (Rauschecker & Tian, 2000), and it can top-down modulate the sensory coding and delta band activity (Keitel et al., 2017; Mittelstadt & Kanold, 2023; Winkowski et al., 2018). Therefore, this connection could account for the recurrent processing in the delta band for height at 400 ms and reflect the difference height and chroma. Also, grey matter in orbitofrontal cortex is larger among musicians (Fauvel et al., 2014), and its activity can be associated with musical expertise (Bücher et al., 2023), which may relate to the fact that the effect of octave equivalence is stronger among musicians (Allen, 1967; Krumhansl & Shepard, 1979). Together, these findings suggest a testable hypothesis: the perception of music is formed through the auditory-frontal neural network, progressing from mid-level pitch height and chroma representations to high-level tonal structure. This process may involve recurrent processing and can be modulated by musical expertise.

The specific involvement of the delta band suggests that the formation of a pitch representation could engage a large cortical network. As a rule of thumb, more widespread neural networks tend to recruit lower-frequency neural activity (Buzsaki & Draguhn, 2004). While the recent studies on the auditory delta band activity predominantly focused on its function on temporal processing and entrainment (e.g., Arnal et al., 2015; Chang et al., 2019; Henry & Obleser, 2012), it reflects the long-range communication of auditory regions with motor and frontal regions (e.g., Gourévitch et al., 2020; Keitel et al., 2017; Morillon & Baillet, 2017). On the other hand, one study showed that delta band activity is associated with processing of pitch in the context of speech (Teoh et al., 2019). Different studies are needed to understand the function of the delta band in non-temporal aspects of auditory perception, including pitch.

The decoding performance in the theta and alpha bands was above baseline (Figure 2) but was not explained by height and chroma (Figure 4A). While this null effect could be due to

insufficient statistical power, it is also possible that these bands encode individual pitch profiles rather than pairwise relationships (e.g., height differences or chroma equivalence), which remains to be investigated.

Together, the combined MEG recording and decoding approach reveals the neural dynamics of pitch representations, which are not monolithic or undecomposed.

Figure Legends

Figure 1. Helical pitch representation, experimental design, and neural signal analyses. (A) The helix model of pitch representation includes the vertical height (log-linear) and horizontal chroma (circular) factors. We use musical notation to provide a clear and intuitive understanding of pitch height and chroma differences. (B) The matrix of pairwise height difference among pitches used, in the unit of octave. (C) The matrix to test whether paired pitches are equivalent in chroma. (D) Cochleagrams of the piano, violin and flute tones across 8 pitches. (E) Event-related fields and topographies of neural responses to the tones (onset time at 0 s), recorded by MEG sensors, were grand-averaged across trials and participants. The waveforms are bandpass filtered 0.1-40 Hz for visualization (but not for analyses). (F) We extract the source signals from early auditory cortex (red), auditory association cortex (orange), insular & frontal opercular cortex (brown), inferior frontal cortex (purple), orbital & polar frontal cortex (green), and dorsolateral prefrontal cortex (blue) for decoding analyses. (G) Within each participant, we train a linear Ridge classifier at each time point of the source neural responses for each pitch-pair. The decoding performances were quantified by ROC-AUC, where a larger value represents more separable neural responses, indicating greater neural dissimilarity. The decoding analysis returns a time series of neural dissimilarities of pitch pairs for each participant (Figure 2). The goal is to explain these dissimilarities by pitch height and chroma (Figure 3).

Figure 2. Decoding performance in each neural frequency band. Each plot summarizes ROC-AUC timeseries either (A) averaged per participant or (B) averaged per pitch-pair. The curves in (B) are color-coded according to the pitch height difference. The temporal clusters which were significantly above chance are marked by shading. Horizontal dashed line depicts chance performance (ROC-AUC = 0.5).

Figure 3. Pitch height and chroma explain delta band (0.5-4 Hz) decoding performance. (A) An LMEM model was used to explain the ROC-AUC at each time point with pitch height and chroma factors. The fitted coefficients were converted to t-values. The significant temporal clusters are marked by shading, and the cluster-level p-values are indicated. Pitch height difference positively predicts ROC-AUC in two time windows (blue), suggesting that neural responses are more dissimilar among pitches with larger pitch height differences. Chroma equivalence negatively predicts ROC-AUC in the 300-320 ms time window (red), suggesting that neural responses are more similar among chromatically equivalent pitches (same pitch class). Height difference and chroma equivalence have an interaction effect in the 300-320 ms time window (purple). The marginal and conditional R2 values were reported at each time point as goodness-of-fit metrics (black and grey). (B) Neural dissimilarity as a function of pitch-pair is predicted by the fitted LMEM model at each time point (t). The height difference factor was included at all time points, while the chroma equivalence and interaction factors were only included at their significant time point (0.3 s). At 0 s, neural dissimilarities of pitch pairs were not associated with height difference. Neural dissimilarity was positively associated with height difference, with the linear slope increasing from 0 to 0.2 s and then decreasing from 0.4 to 0.5 s. At 0.3 s, the chroma equivalence and height difference factors jointly influenced neural dissimilarities: the neural responses were more similar when pitch pairs were one octave apart

and less similar when they were two octaves apart. Note that, as there were fewer instances of pitch pairs with larger height differences (upper panel), their estimation precision will inevitably be lower. (C) The 3D visualization illustrates neural dissimilarities (distances) among pitches. The upper panel illustrates the interpretation of the dimensions of the 3D visualization. The data is identical to that in (B) but presented using an alternative visualization method. Pairwise ROC-AUC values, derived from the fitted LMEM model at each time point, were mapped into a three-dimensional Euclidean scale (A.U.: arbitrary unit, with axes oriented to resemble the helix model), where pitches that are closer in distance exhibit more similar neural responses. At time points 0.1 s, 0.2 s, 0.4 s, and 0.5 s, the neural responses to pitches can be represented in a one-dimensional line, resembling the height dimension, with the maximum separation occurring around 0.2 s. At 0.3 s, the neural responses to pitches roughly formed a three-dimensional helix-like shape, resembling both the height and chroma dimensions of the pitch helix model (Figure 1A).

Figure 4. Additional LMEM analyses and cochleagram similarity. (A) Explaining theta (4-8 Hz) and alpha (8-13 Hz) band decoding performances by a LMEM with pitch height and chroma factors. The format is as in Figure 3A. There was no significant cluster. (B) The matrix of pairwise cochleagram similarity among pitches across timbres. (C) Explaining decoding performance at each frequency band by an LMEM with pitch height difference and cochleagram similarity variables. The format is the same as Figure 3A. There was no significant cluster.

Figure 5. Temporal generalization analysis. The decoder trained at each time point (y-axis) was applied to all other time points (x-axis). Black and gray contours indicate training-testing time pairs that are significantly above chance at various FDR-corrected thresholds. In addition to the high decoding performance on the diagonal, where training time equals to testing time, the model trained in the approximately 140-190 ms window could explain the neural responses in the 400-500 ms window, suggesting similar neural representations in these two windows.

Figure 6. Cortical maps and time series of the classifier weights. (A) The cortical maps of the mean absolute weights of the classifier at 180-230 ms and 300-320 ms time windows, corresponding to the temporal clusters of pitch height and chroma of Figure 3A, respectively. (B) The time series of the mean proportional absolute weight by cortical regions, with the black horizontal line representing the baseline (BL). (C) The distribution of mean proportional absolute weight of time windows within each cortical region, with the black dashed line representing the baseline. L: left hemisphere; R: right hemisphere; **: $p < 0.01$; ***: $p < 0.001$, uncorrected for multiple comparisons.

References

Abrams, E. B., Marantz, A., & Gwilliams, L. (2024). Dynamics of pitch perception in the auditory cortex. bioRxiv, 2024-06. https://doi.org/10.1101/2024.06.10.598008

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science, 367(6481), 1043-1047.

Allen, D. (1967). Octave discriminability of musical and non-musical subjects. Psychonomic Science, 7(12), 421-422.

Allen, E. J., Mesik, J., Kay, K. N., & Oxenham, A. J. (2022). Distinct representations of tonotopy and pitch in human auditory cortex. Journal of Neuroscience, 42(3), 416-434.

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. Cerebral Cortex, 25(9), 3077-3085.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. Nature, 436(7054), 1161-1165.

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how? Journal of Clinical Epidemiology, 54(4), 343-349.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Bermudez, P., & Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in nonmusicians: implications for absolute pitch. Journal of Neuroscience, 25(34), 7718-7723.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. Nature Reviews Neuroscience, 14(10), 693-707.

Bücher, S., Bernhofs, V., Thieme, A., Christiner, M., & Schneider, P. (2023). Chronology of auditory processing and related co-activation in the orbitofrontal cortex depends on musical expertise. Frontiers in Neuroscience, 16, 1041397.

Burns, E. M. (1999). Intervals, scales, and tuning. In The psychology of music (pp. 215-264). Academic Press.

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. science, 304(5679), 1926-1929.

Chang, A., Bosnyak, D. J., & Trainor, L. J. (2019). Rhythmicity facilitates pitch discrimination: Differential roles of low and high frequency neural oscillations. NeuroImage, 198, 31-43.

Crottaz-Herbette, S., & Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. Clinical neurophysiology, 111(10), 1759-1766.

Czoschke, S., Fischer, C., Bahador, T., Bledowski, C., & Kaiser, J. (2021). Decoding concurrent representations of pitch and location in auditory working memory. Journal of Neuroscience, 41(21), 4658-4666.

de Cheveigné, A. (2010). Pitch perception. The oxford handbook of auditory science: Hearing, 3, 71-104.

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage, 34(4), 1443-1449.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. The journal of the Acoustical Society of America, 76(1), 57-66.

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., & Simonyan, K. (2017). Neural audio synthesis of musical notes with wavenet autoencoders. In International Conference on Machine Learning (pp. 1068-1077). PMLR.

Fasano, M. C., Cabral, J., Stevner, A., Vuust, P., Cantou, P., Brattico, E., & Kringelbach, M. L. (2023). The early adolescent brain on music: Analysis of functional dynamics reveals engagement of orbitofrontal cortex reward system. Human Brain Mapping, 44(2), 429-446.

Fauvel, B., Groussard, M., Chételat, G., Fouquet, M., Landeau, B., Eustache, F., ... & Platel, H. (2014). Morphological brain plasticity induced by musical expertise is accompanied by modulation of functional connectivity at rest. NeuroImage, 90, 179-188.

Fischl, B. (2012). FreeSurfer. Neuroimage, 62(2), 774-781.

Förster, J., Koivisto, M., & Revonsuo, A. (2020). ERP and MEG correlates of visual consciousness: The second decade. Consciousness and cognition, 80, 102917.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. Nature, 536(7615), 171-178.

Gourévitch, B., Martin, C., Postal, O., & Eggermont, J. J. (2020). Oscillations in the auditory system and their possible role. Neuroscience & Biobehavioral Reviews, 113, 507-528.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 267.

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? Neuroreport, 10(18), 3825-3830.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. Cerebral Cortex, 19(3), 576-585.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ, 6, e4794.

He, C., Hotson, L., & Trainor, L. J. (2009). Maturation of cortical mismatch responses to occasional pitch change in early infancy: effects of presentation rate and magnitude of change. Neuropsychologia, 47(1), 218-229.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proceedings of the National Academy of Sciences, 109(49), 20095-20100.

Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. Attention, Perception, & Psychophysics, 74, 1742-1760.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. Current Biology, 29(19), 3229-3243.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. Science, 298(5601), 2167-2170.

Kallman, H. J. (1982). Octave equivalence as measured by similarity ratings. Perception & Psychophysics, 32(1), 37-49.

Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. NeuroImage, 147, 32-42.

Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. Current Opinion in Neurobiology, 55, 167-179.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cerebral Cortex, 13(7), 765-772.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. Journal of Experimental Psychology: Human Perception and Performance, 5(4), 579.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. The Journal of the Acoustical Society of America, 23(1_Supplement), 147-147.

Limb, C. J. (2006). Structural and functional neural correlates of music perception. The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology: An Official Publication of the American Association of Anatomists, 288(4), 435-446.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods, 164(1), 177-190.

McDonald, G. C. (2009). Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 93-100.

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. The Journal of the Acoustical Society of America, 102(3), 1811-1820.

Mittelstadt, J. K., & Kanold, P. O. (2023). Orbitofrontal cortex conveys stimulus and task information to the auditory cortex. Current Biology, 33(19), 4160-4173.

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences, 114(42), E8913-E8921.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical Neurophysiology, 118(12), 2544-2590.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods in Ecology and Evolution, 4(2), 133-142.

Oxenham, A. J. (2023). Questions and controversies surrounding the perception and neural coding of pitch. Frontiers in Neuroscience, 16, 1074752.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: pitch versus frequency representation. Science, 246(4929), 486-488.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830.

Peretz, I. (2016). Neurobiology of congenital amusia. Trends in Cognitive Sciences, 20(11), 857-867.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. Journal of Neuroscience, 24(30), 6810-6815.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128-2148.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience, 12(6), 718-724.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proceedings of the National Academy of Sciences, 97(22), 11800-11806.

Regev, T. I., Nelken, I., & Deouell, L. Y. (2019). Evidence for linear but not helical automatic representation of pitch in the human auditory system. Journal of Cognitive Neuroscience, 31(5), 669-685.

Révész, G. (1954). Introduction to the psychology of music. University of Oklahoma Press.

Sankaran, N., Carlson, T. A., & Thompson, W. F. (2020). The rapid emergence of musical pitch structure in human cortex. Journal of Neuroscience, 40(10), 2108-2118.

Sankaran, N., Thompson, W. F., Carlile, S., & Carlson, T. A. (2018). Decoding the dynamic representation of musical pitch from human brain activity. Scientific reports, 8(1), 839.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference (Vol. 57, No. 61, pp. 10-25080).

Shepard, R. N. (1964). Circularity in judgments of relative pitch. The journal of the acoustical society of America, 36(12), 2346-2353.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. Psychological Review, 89(4), 305.

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. European Journal of Neuroscience, 50(11), 3831-3842.

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic and Applied Social Psychology, 39(2), 81-90.

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. The Journal of the Acoustical Society of America, 82(4), 1193-1200.

Wagner, B., & Hoeschele, M. (2022). The Links Between Pitch, Timbre, Musicality, and Social Bonding From Cross-Species Research. Comparative Cognition & Behavior Reviews, 17.

Wagner, B., Sturdy, C. B., Weisman, R. G., & Hoeschele, M. (2022). Pitch chroma information is processed in addition to pitch height information with more than two pitch-range categories. Attention, Perception, & Psychophysics, 84(5), 1757-1771.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. Proceedings of the National Academy of Sciences, 100(17), 10038-10042.

Winkowski, D. E., Nagode, D. A., Donaldson, K. J., Yin, P., Shamma, S. A., Fritz, J. B., & Kanold, P. O. (2018). Orbitofrontal cortex neurons respond to sound and activate primary auditory cortex neurons. Cerebral Cortex, 28(3), 868-879.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. Journal of Neuroscience, 24(41), 9153-9160.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. Journal of Experimental Psychology: General, 129(3), 291.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. The Journal of the Acoustical Society of America, 84(2), 566-572.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences, 6(1), 37-46.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. Journal of Neuroscience, 14(4), 1908-1919.

736 Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the
737 mind's ear: a PET investigation of musical imagery and perception. Journal of cognitive
738 neuroscience, 8(1), 29-46.