

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

Copyright © 2025 the authors

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

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.

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

44

45

JNeurosci Accepted Manuscript

46 **Abstract**

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

67

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

70

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.

JNeurosci Accepted Manuscript

71 Introduction

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

82 Many behavioral experiments demonstrate that certain (often musical) tonal stimuli are
83 perceived as more similar to each other than others. For example, the pitches of adjacent piano
84 keys and keys an octave apart are perceptually more similar than other pairs. This perceptual
85 pattern can be effectively represented by a helix model (Figure 1A), incorporating height and
86 chroma factors, where a shorter distance between two pitches represents a higher perceptual
87 similarity (Shepard, 1982; Ueda & Ohgushi, 1987). Height as a dimension log-linearly
88 corresponds to the fundamental frequency (F_0) of the tone. Two pitches with a smaller height
89 difference are perceived as more similar. Height information is crucial for separating sound
90 sources, such as distinguishing individuals' voices according to their different frequency ranges.

91 Chroma, on the other hand, signifies the cyclic and recurring patterns of perceptual
92 similarities among pitches separated by an octave (doubling of F_0 frequency), a phenomenon
93 commonly referred to as octave equivalence (Révész, 1954; Shepard, 1964). Chroma plays an
94 important role in recognizing acoustic patterns independent of the sound sources. In speech, for
95 example, octave equivalence facilitates human vocal learning by equalizing the vocal range
96 difference across ages and genders (Wagner & Hoeschele, 2022). In music, a melody shifted by
97 an octave is recognized as being similar or even the same (Demany & Armand, 1984; Wright et
98 al., 2000). Although the way one divides an octave interval into pitch classes (e.g., C, D, E or
99 Do, Re, Mi) can vary across cultures, the recurrence of pitch classes per octave appears a
100 commonly shared feature (Burns 1999; but see Jacoby et al., 2019). Octave equivalence cannot
101 be simply explained by low-level acoustic matching, even though half of the harmonic
102 frequencies inevitably overlap between pitches an octave apart (Jacoby et al., 2019; Wagner et
103 al., 2022).

104 Despite important characterizations of the perceptual representation of pitch over many
105 years, the neural representation of pitch remains incompletely characterized. Pitch is primarily
106 represented in the auditory cortex (e.g., Allen et al., 2022; Bendor & Wang, 2005; Pantev et al.,
107 1989; Penagos et al., 2011; Zatorre, 1988; Zatorre et al., 2002). Specifically, the posterior
108 subregion (posterior planum temporale) is associated with height while the anterior subregion
109 (extending from Heschl's gyrus into the planum polare) is associated with chroma (Briley et al.,
110 2013; Moerel et al., 2015; Warren et al., 2003). Although these basic attributes of pitch seem
111 spatially separable in cortex, several neurobiological questions remain unaddressed. The
112 present experiment is designed with two questions in mind, motivated by issues of timing and
113 neural dynamics. When do the representations of height and chroma emerge? Are they formed
114 at the same time or in different processing stages? And how is the neural activity associated
115 with perceptual similarity?

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

123

124 **Methods**

125

126 *Resource Availability*

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

130

131 *Participants*

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

138

139 *Stimuli*

140 Eight signals spanning three octaves in the pitch classes C, E, and G# were used: G#6
141 (415.3 Hz), C7 (523.3 Hz), E7 (659.3 Hz), G#7 (830.6 Hz), C8 (1046.5 Hz), E8 (1318.5 Hz),
142 G#8 (1661.2 Hz), and C9 (2093.0 Hz). The fundamental frequencies of pitches are
143 logarithmically equidistant, with a step corresponding to 1/3 octave (or 4 semitones). Each pitch
144 was presented in three instrumental timbres (flute, violin, and piano), resulting in 24 individually
145 recorded tones, retrieved from the NSynth Dataset (Engel et al., 2017). The duration of each
146 recording was 0.4 seconds, and the sampling rate was 44.1 kHz. Rather than sinusoidally
147 synthesized tones or a single timbre (as has been typical in previous studies), we used tones
148 from various instruments to permit the generalizability of our findings. The amplitude of tones
149 was pseudorandomly set between 67 and 73 dB-SPL for each presentation.

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

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

159

160 *Procedure*

161 Each participant was exposed to 2400 stimuli (8 frequencies x 3 instrumental timbres x
162 100 repetitions per stimulus). The materials were presented in 100 runs, such that each run
163 contained each stimulus, presented in a pseudorandom order, with half of the runs having
164 consecutive tones in the same pitch (regardless of instrument). The inter-onset intervals within a
165 run roved between 0.9 and 1.4 seconds. To ensure that participants were attending to the
166 stimuli, after each run, they were asked to report whether there were any consecutive tones (i.e.
167 stimuli of the same pitch) by making a yes/no button-press response using their right hand. The
168 next run started 1.5-2 seconds after the response. (Although the behavioral data was
169 unavailable due to technical failure, the post-block, offline behavioral responses are not of
170 critical relevance to the goals and interpretation of this study.) In sum, each stimulus was
171 presented 100 times, and each pitch was presented 300 times.

172

173 *MEG Recording and Preprocessing*

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

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

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

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

196 operator was applied to epochs using dSPM method with regularization parameter $\lambda_2 =$
197 $1/9$. (5) The signals in the source-space were morphed and downsampled to the fsaverage ico3
198 template space. This step is to enhance the statistical power and mitigate overfitting of the
199 decoding analysis by reducing the dimensions of the data and lower multicollinearity among the
200 source signals. (6) To also mitigate overfitting, we reduced the dimensionality by only including
201 signals from the bilateral early auditory cortex, auditory association cortex, and frontal lobe
202 (inferior frontal cortex, insular and frontal opercular cortex, orbital and polar frontal cortex,
203 dorsolateral prefrontal cortex), total in 316 vertices (Figure 1F), were used for subsequent
204 decoding analysis, following the HCP-MMP1 atlas (Glasser et al. 2016), as these regions
205 roughly covers the auditory ventral “what” pathway or have been reported to be associated with
206 processing pitch (Hall & Plack, 2009; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000).
207 In addition to auditory cortex, we included these broad frontal regions for several reasons. First,
208 certain frontal areas are commonly involved in pitch processing (e.g., the inferior frontal area;
209 Griffiths et al., 1999; Zatorre et al., 1994), whereas others are less clearly associated (e.g., the
210 dorsolateral prefrontal cortex, which relates to pitch labeling, and the insular cortex, involved in
211 processing pitch patterns; Bermudez & Zatorre, 2005; Wong et al., 2004). Consequently, there
212 is no definitive guidance on which regions should be specifically included or excluded.
213 Additionally, using decoding analyses instead of traditional fMRI contrasts may reveal more
214 distributed and potentially unexpected regions involved in the task. Anyway, our machine
215 learning model was designed to automatically shrink the weights of locations with minimal
216 contribution, effectively excluding those regions.

217

218 *Multivariate Pattern Decoding Analysis*

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

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2$$

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

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

239

240 *Statistical Analyses*

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

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

263 To analyze ROC-AUC time-series data (Figure 2), we used a one-sample cluster-based
264 permutation test (Maris & Oostenveld, 2007), flipping the signs based on whether values were
265 above or below 0.5 within each participant. To conduct the cluster-based permutation test on
266 LMEM (Figure 3A, Figure 4A and Figure 4C), the dependent and independent variables were
267 randomly shuffled within each participant before fitting an LMEM for each iteration. (Additionally,
268 a small uniform-random number between -0.0001 and 0.0001 was added to the ROC-AUC in
269 cases where the LMEM failed to converge, which occurred in a very small number of iterations.)
270 For all these tests, the number of permutations was set to 10k, the clustering threshold was set
271 at an absolute t-value of 2.5 with a minimum width of 20 ms (5 consecutive time samples), and
272 2-tailed cluster-level p-values were reported.

273

274 **Results**

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

279 Orthogonally, in the by-item analysis, we collapsed all participants per pitch-pair ($df = 27$), and
280 the ROC-AUC was also significantly above chance in the delta, theta, and alpha bands (Figure
281 2B). These findings suggest that neural encoding of pitch profiles is predominantly reflected in
282 the lower frequency bands. Therefore, only delta, theta, and alpha bands are included in the
283 subsequent analyses.

284 The ROC-AUC values were in the range of 0.50 to 0.55 (Figure 2). These relatively
285 modest decoding values are likely caused by the diversity in sound types (different instruments)
286 in our experimental design, which introduces additional variability in the neural responses,
287 consequently impacting the classifier's performance. Despite this experiment-inherent
288 challenge, the magnitude of our ROC-AUC aligns with findings from comparable pitch decoding
289 MEG and fMRI studies that used single timbres (e.g., Czoschke et al., 2021; Sankaran et al.,
290 2018, 2020), and our findings have a better generalizability. Importantly, the identification of
291 statistically significant above-chance clusters implies that the signals are sufficiently robust for
292 further analyses.

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

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

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

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

338 To explore which brain regions are implicated in representing pitch height and chroma,
339 we mapped the weights of the classifier to the cortex (Figure 6A). We further averaged the
340 weights within each atlas label and normalized the mean weight as a percentage to equalize the
341 magnitude of the sum of weights across time (Figure 6B), with the baseline level at 8.33%
342 (100% divided by 12 regions). We specifically focused on the time windows representing height
343 (180-230 ms, the peak range of the height effect) and chroma (300-320 ms). First, the weights
344 at the right early auditory region at both height ($t(16) = 3.92, p = 0.001$, Cohen's $d = 0.98$) and
345 chroma ($t(16) = 3.79, p = 0.002$, Cohen's $d = 0.95$) time windows were above baseline, and so
346 were the right insular and frontal operculum for both height ($t(16) = 6.44, p < 0.001$, Cohen's $d =$
347 1.63) and chroma ($t(16) = 6.07, p < 0.001$, Cohen's $d = 1.52$) windows. Second, we tested the
348 difference between windows within each region, and the weight in right orbital and polar frontal
349 cortex was higher in the chroma than the height window ($t(16) = 4.47, p < 0.001$, Cohen's $d =$
350 1.12). Finally, the lateralization analyses (Figure 6C) showed rightward lateralization in multiple
351 regions and time windows, including early auditory cortex during the chroma window ($t(16) =$
352 $3.02, p = 0.008$, Cohen's $d = 0.76$), auditory association cortex during the height window ($t(16) =$
353 $3.20, p = 0.006$, Cohen's $d = 0.80$), and inferior frontal cortex during the height window ($t(16) =$
354 $3.53, p = 0.003$, Cohen's $d = 0.88$). Both height and chroma seem to engage the right
355 hemisphere more than the left and involve shared cortical regions across auditory and frontal
356 areas; however, the weights of the classifier at the right orbital and polar frontal cortex reflected
357 their differences. Note that interpreting the map of the classifier's weights is not straightforward,
358 as a higher weight does not necessarily imply a larger influence on the classification result
359 (Kriegeskorte & Douglas, 2019). Also, our MEG decoding analyses should be considered
360 complementary to previous fMRI studies from an encoding perspective (e.g., Warren et al.,
361 2003), although they are not directly comparable. Nevertheless, we hope these exploratory
362 findings, despite not controlling for multiple comparisons (Bender & Lange, 2001), can
363 contribute to shaping future localization studies on this topic.

364

365 Discussion

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

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

380 Height is arguably the most salient dimension of a pitch representation. The observation
381 that pitch-pairs with larger height differences have lower neural representational similarity
382 (better decoding performance) is consistent with previous behavioral findings that pitches which
383 have larger fundamental frequency differences in logarithmic scale will be perceived more
384 differently (e.g., Ueda & Ohgushi, 1987). In the brain, a larger pitch height difference elicits
385 stronger event-related potential (ERP) responses (e.g., He et al., 2009; Näätänen et al., 2007),
386 and sounds with closer pitch-eliciting frequencies activate spatially adjacent neural ensembles
387 in the tonotopically organized auditory cortex (Allen et al., 2022). The height representation,
388 spanning from 100 to 300 ms post stimulus onset, is well aligned with the latencies of the
389 auditory ERP components associated with pitch perception, such as N1 and the pitch onset
390 response (e.g., de Cheveigné, 2010). It suggests that the height representation is likely formed
391 at a preattentive, automatic processing stage (Regev et al., 2019), and could be more or less
392 universal in listeners (e.g., Jacoby et al., 2019; Kallman, 1982). Consistent with our findings, a
393 recent MEG study analyzed the neural representational similarity of pitches in a Western tonal
394 music context and showed a consistent finding that the broad-band neural dissimilarity at 150-
395 300 ms is associated with coarse-grained pitch height difference (Sankaran et al., 2020). Our
396 fine-grained data further pinpoints that this association is specific to the delta band and
397 describes its log-linear mathematical relationship.

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

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

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

419 An unexpected and intriguing finding is that neural representations are more similar
420 when pitches are one octave apart, reflecting the effect of octave equivalence and aligning with
421 the helix model; however, they appear more dissimilar when two octaves apart, which is
422 inconsistent with the helix model. This finding provides a minor but noteworthy counterexample
423 to the helix model, aligning with empirical behavioral findings suggesting that the octave
424 equivalence effect may be limited to immediately neighboring octaves (e.g., Wagner et al.,
425 2022). Note that, as there were fewer instances of pitch-pairs with larger height differences in
426 our experiment, their estimation precision will inevitably be lower. Therefore, the model
427 estimation at two octaves should be taken with caution. Together, our findings suggest a need
428 to reexamine the helix model across varying numbers of octaves.

429 The height and chroma representations differ in engaging the right orbital and polar
430 frontal cortex. The implication of orbital and polar frontal cortex was less clear in the literature,
431 despite this region is involved in some pitch or music related tasks (e.g., Fasano et al., 2023;
432 Limb, 2006; Zatorre et al., 1996). One fMRI study showed that these regions are associated
433 with tracking Western tonal structure (Janata et al., 2002). This may be driven by the encoding
434 of relationships among pitches (i.e., height and chroma), as these are fundamental elements in
435 forming musical tonality. Other neuroanatomical studies have reported that the orbitofrontal
436 cortex is connected to auditory cortex via the ventral pathway (Rauschecker & Tian, 2000), and
437 it can top-down modulate the sensory coding and delta band activity (Keitel et al., 2017;
438 Mittelstadt & Kanold, 2023; Winkowski et al., 2018). Therefore, this connection could account for
439 the recurrent processing in the delta band for height at 400 ms and reflect the difference height
440 and chroma. Also, grey matter in orbitofrontal cortex is larger among musicians (Fauvel et al.,
441 2014), and its activity can be associated with musical expertise (Bücher et al., 2023), which may
442 relate to the fact that the effect of octave equivalence is stronger among musicians (Allen, 1967;
443 Krumhansl & Shepard, 1979). Together, these findings suggest a testable hypothesis: the
444 perception of music is formed through the auditory-frontal neural network, progressing from mid-
445 level pitch height and chroma representations to high-level tonal structure. This process may
446 involve recurrent processing and can be modulated by musical expertise.

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

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

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

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

464

465

JNeurosci Accepted Manuscript

466 **Figure Legends**

467

468 **Figure 1. Helical pitch representation, experimental design, and neural signal analyses.**

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

485

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

491

492 **Figure 3. Pitch height and chroma explain delta band (0.5-4 Hz) decoding performance.**

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

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

523

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

530

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

537

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

546 **References**

- 547 Abrams, E. B., Marantz, A., & Gwilliams, L. (2024). Dynamics of pitch perception in the auditory
548 cortex. *bioRxiv*, 2024-06. <https://doi.org/10.1101/2024.06.10.598008>
- 549 Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to
550 spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*,
551 367(6481), 1043-1047.
- 552 Allen, D. (1967). Octave discriminability of musical and non-musical subjects. *Psychonomic*
553 *Science*, 7(12), 421-422.
- 554 Allen, E. J., Mesik, J., Kay, K. N., & Oxenham, A. J. (2022). Distinct representations of tonotopy
555 and pitch in human auditory cortex. *Journal of Neuroscience*, 42(3), 416-434.
- 556 Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie
557 temporal prediction accuracy. *Cerebral Cortex*, 25(9), 3077-3085.
- 558 Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex.
559 *Nature*, 436(7054), 1161-1165.
- 560 Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how? *Journal of*
561 *Clinical Epidemiology*, 54(4), 343-349.
- 562 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and
563 powerful approach to multiple testing. *Journal of the Royal statistical society: series B*
564 (Methodological), 57(1), 289-300.
- 565 Bermudez, P., & Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in
566 nonmusicians: implications for absolute pitch. *Journal of Neuroscience*, 25(34), 7718-
567 7723.
- 568 Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception.
569 *Nature Reviews Neuroscience*, 14(10), 693-707.
- 570 Bücher, S., Bernhofs, V., Thieme, A., Christiner, M., & Schneider, P. (2023). Chronology of
571 auditory processing and related co-activation in the orbitofrontal cortex depends on
572 musical expertise. *Frontiers in Neuroscience*, 16, 1041397.
- 573 Burns, E. M. (1999). Intervals, scales, and tuning. In *The psychology of music* (pp. 215-264).
574 Academic Press.
- 575 Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*,
576 304(5679), 1926-1929.
- 577 Chang, A., Bosnyak, D. J., & Trainor, L. J. (2019). Rhythmicity facilitates pitch discrimination:
578 Differential roles of low and high frequency neural oscillations. *NeuroImage*, 198, 31-43.
- 579 Crottaz-Herbette, S., & Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch
580 and topography codes spectra. *Clinical neurophysiology*, 111(10), 1759-1766.
- 581 Czoschke, S., Fischer, C., Bahador, T., Bledowski, C., & Kaiser, J. (2021). Decoding concurrent
582 representations of pitch and location in auditory working memory. *Journal of*
583 *Neuroscience*, 41(21), 4658-4666.

584 de Cheveigné, A. (2010). Pitch perception. The oxford handbook of auditory science: Hearing,
585 3, 71-104.

586 Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data
587 using higher-order statistics and independent component analysis. *Neuroimage*, 34(4),
588 1443-1449.

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

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

594 Fasano, M. C., Cabral, J., Stevner, A., Vuust, P., Cantou, P., Brattico, E., & Kringelbach, M. L.
595 (2023). The early adolescent brain on music: Analysis of functional dynamics reveals
596 engagement of orbitofrontal cortex reward system. *Human Brain Mapping*, 44(2), 429-
597 446.

598 Fauvel, B., Groussard, M., Chételat, G., Fouquet, M., Landeau, B., Eustache, F., ... & Platel, H.
599 (2014). Morphological brain plasticity induced by musical expertise is accompanied by
600 modulation of functional connectivity at rest. *NeuroImage*, 90, 179-188.

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

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

604 Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Van
605 Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*,
606 536(7615), 171-178.

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

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

612 Griffiths, T. D., Johnsruide, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for
613 the analysis of pitch and duration pattern in segmented sound? *Neuroreport*, 10(18),
614 3825-3830.

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

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

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

623 Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and
624 optimizes human listening behavior. *Proceedings of the National Academy of Sciences*,
625 109(49), 20095-20100.

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

629 Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H.
630 (2019). Universal and non-universal features of musical pitch perception revealed by
631 singing. *Current Biology*, 29(19), 3229-3243.

632 Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The
633 cortical topography of tonal structures underlying Western music. *Science*, 298(5601),
634 2167-2170.

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

637 Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment
638 interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, 147,
639 32-42.

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

642 Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B.
643 (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus.
644 *Cerebral Cortex*, 13(7), 765-772.

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

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

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

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

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

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

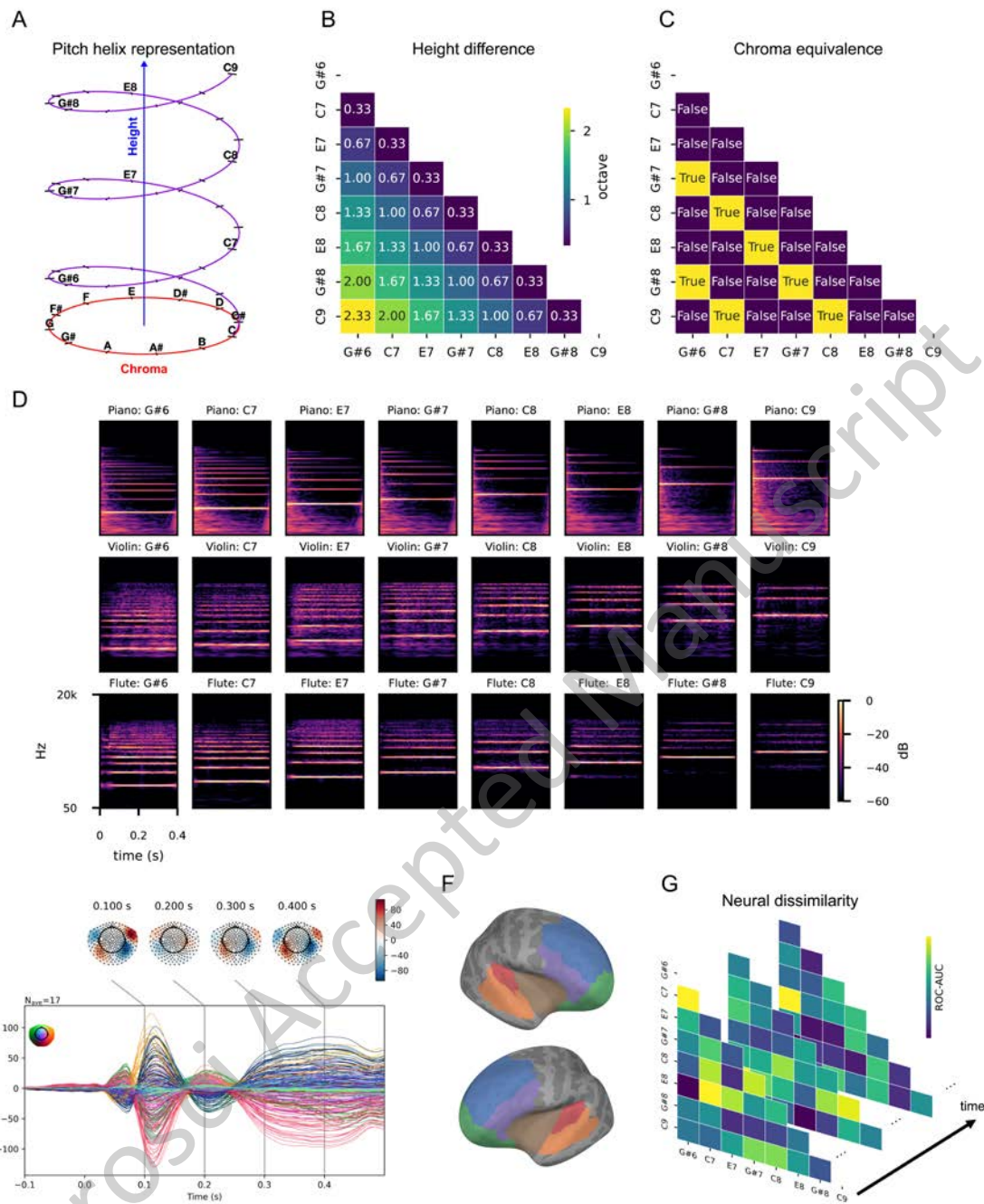
659 Mittelstadt, J. K., & Kanold, P. O. (2023). Orbitofrontal cortex conveys stimulus and task
660 information to the auditory cortex. *Current Biology*, 33(19), 4160-4173.

- 661 Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention.
662 Proceedings of the National Academy of Sciences, 114(42), E8913-E8921.
- 663 Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in
664 basic research of central auditory processing: a review. Clinical Neurophysiology,
665 118(12), 2544-2590.
- 666 Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from
667 generalized linear mixed-effects models. Methods in Ecology and Evolution, 4(2), 133-
668 142.
- 669 Oxenham, A. J. (2023). Questions and controversies surrounding the perception and neural
670 coding of pitch. Frontiers in Neuroscience, 16, 1074752.
- 671 Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the
672 auditory cortex: pitch versus frequency representation. Science, 246(4929), 486-488.
- 673 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay,
674 É. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning
675 Research, 12, 2825-2830.
- 676 Peretz, I. (2016). Neurobiology of congenital amusia. Trends in Cognitive Sciences, 20(11),
677 857-867.
- 678 Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience
679 in nonprimary human auditory cortex revealed with functional magnetic resonance
680 imaging. Journal of Neuroscience, 24(30), 6810-6815.
- 681 Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clinical
682 Neurophysiology, 118(10), 2128-2148.
- 683 Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman
684 primates illuminate human speech processing. Nature Neuroscience, 12(6), 718-724.
- 685 Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and
686 "where" in auditory cortex. Proceedings of the National Academy of Sciences, 97(22),
687 11800-11806.
- 688 Regev, T. I., Nelken, I., & Deouell, L. Y. (2019). Evidence for linear but not helical automatic
689 representation of pitch in the human auditory system. Journal of Cognitive
690 Neuroscience, 31(5), 669-685.
- 691 Révész, G. (1954). Introduction to the psychology of music. University of Oklahoma Press.
- 692 Sankaran, N., Carlson, T. A., & Thompson, W. F. (2020). The rapid emergence of musical pitch
693 structure in human cortex. Journal of Neuroscience, 40(10), 2108-2118.
- 694 Sankaran, N., Thompson, W. F., Carlile, S., & Carlson, T. A. (2018). Decoding the dynamic
695 representation of musical pitch from human brain activity. Scientific reports, 8(1), 839.
- 696 Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with
697 python. In Proceedings of the 9th Python in Science Conference (Vol. 57, No. 61, pp.
698 10-25080).

- 699 Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The journal of the acoustical*
700 *society of America*, 36(12), 2346-2353.
- 701 Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch.
702 *Psychological Review*, 89(4), 305.
- 703 Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in
704 delta-band EEG and is dissociable from the cortical tracking of other acoustic and
705 phonetic features. *European Journal of Neuroscience*, 50(11), 3831-3842.
- 706 Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance
707 inflation factor and other multicollinearity diagnostics from typical regression results.
708 *Basic and Applied Social Psychology*, 39(2), 81-90.
- 709 Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a
710 multidimensional scaling technique. *The Journal of the Acoustical Society of America*,
711 82(4), 1193-1200.
- 712 Wagner, B., & Hoeschele, M. (2022). The Links Between Pitch, Timbre, Musicality, and Social
713 Bonding From Cross-Species Research. *Comparative Cognition & Behavior Reviews*,
714 17.
- 715 Wagner, B., Sturdy, C. B., Weisman, R. G., & Hoeschele, M. (2022). Pitch chroma information is
716 processed in addition to pitch height information with more than two pitch-range
717 categories. *Attention, Perception, & Psychophysics*, 84(5), 1757-1771.
- 718 Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch
719 chroma and pitch height in the human brain. *Proceedings of the National Academy of*
720 *Sciences*, 100(17), 10038-10042.
- 721 Winkowski, D. E., Nagode, D. A., Donaldson, K. J., Yin, P., Shamma, S. A., Fritz, J. B., &
722 Kanold, P. O. (2018). Orbitofrontal cortex neurons respond to sound and activate
723 primary auditory cortex neurons. *Cerebral Cortex*, 28(3), 868-879.
- 724 Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in
725 pitch pattern perception: the effect of linguistic contexts. *Journal of Neuroscience*,
726 24(41), 9153-9160.
- 727 Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiwirth, J. J. (2000). Music perception
728 and octave generalization in rhesus monkeys. *Journal of Experimental Psychology:*
729 *General*, 129(3), 291.
- 730 Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The*
731 *Journal of the Acoustical Society of America*, 84(2), 566-572.
- 732 Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex:
733 music and speech. *Trends in Cognitive Sciences*, 6(1), 37-46.
- 734 Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic
735 perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908-1919.

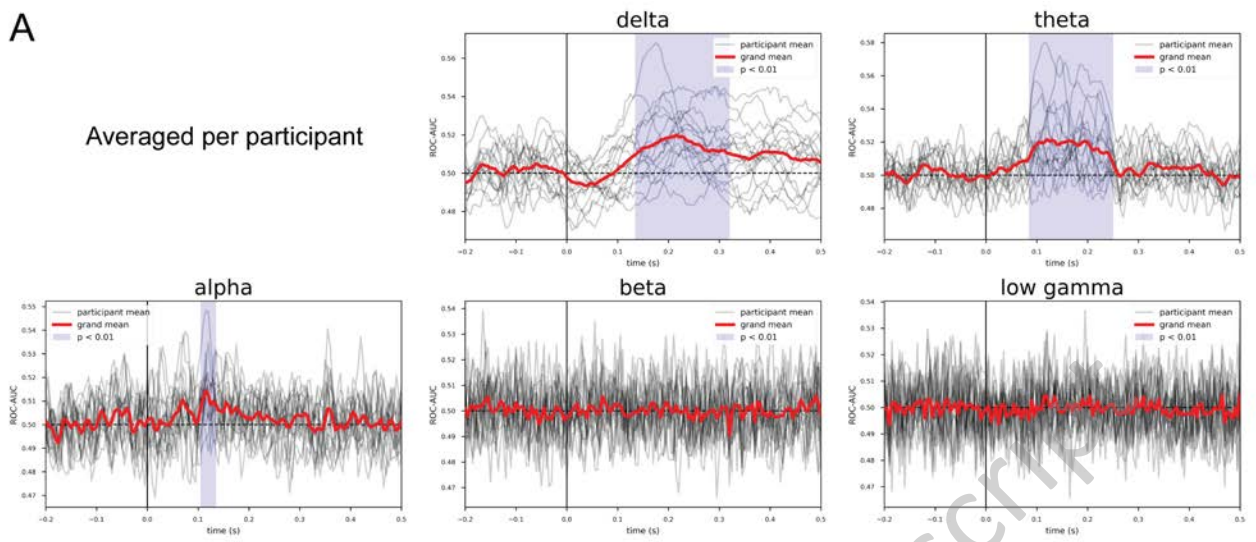
736 Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the
737 mind's ear: a PET investigation of musical imagery and perception. *Journal of cognitive*
738 *neuroscience*, 8(1), 29-46.

JNeurosci Accepted Manuscript

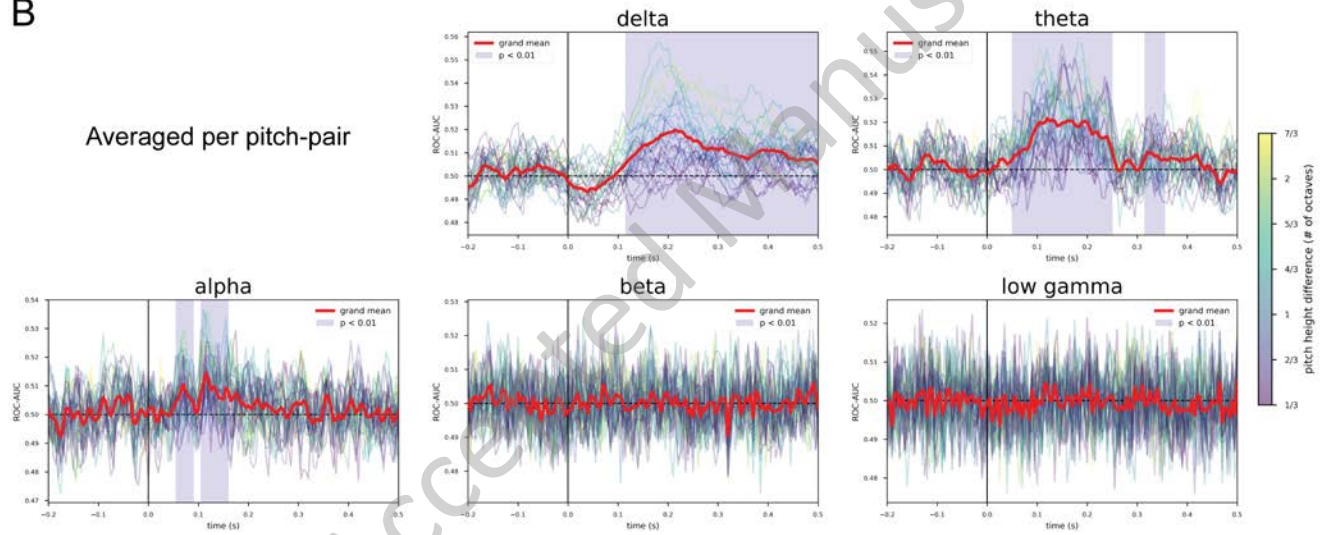


A

Averaged per participant

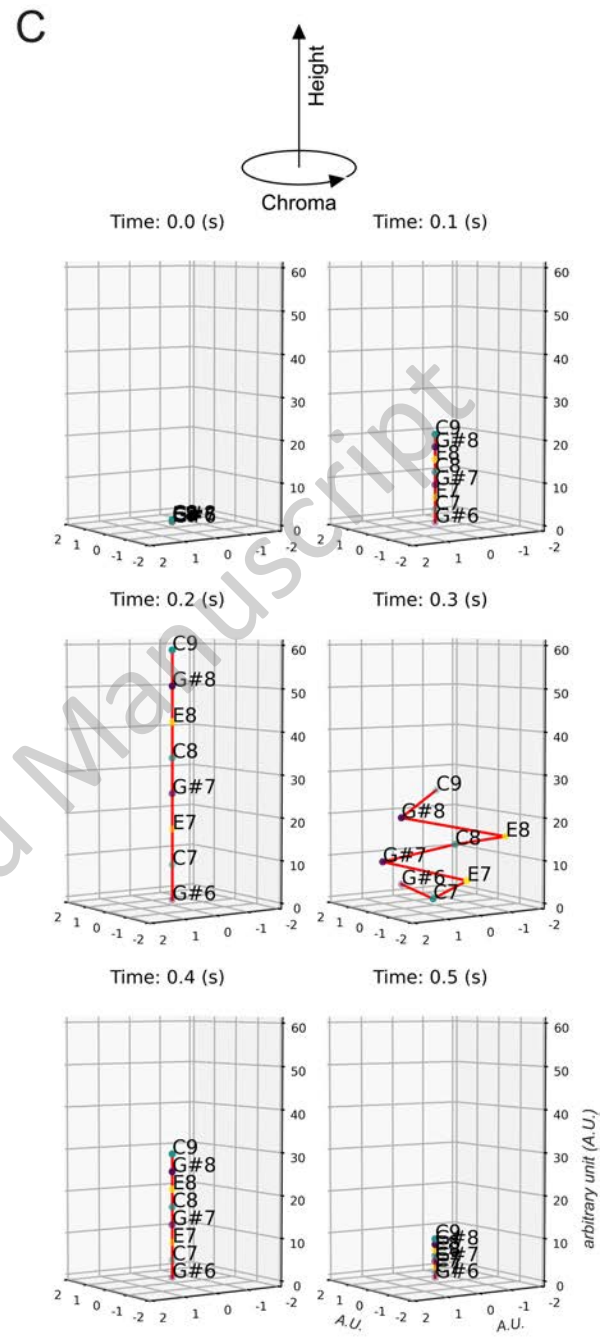
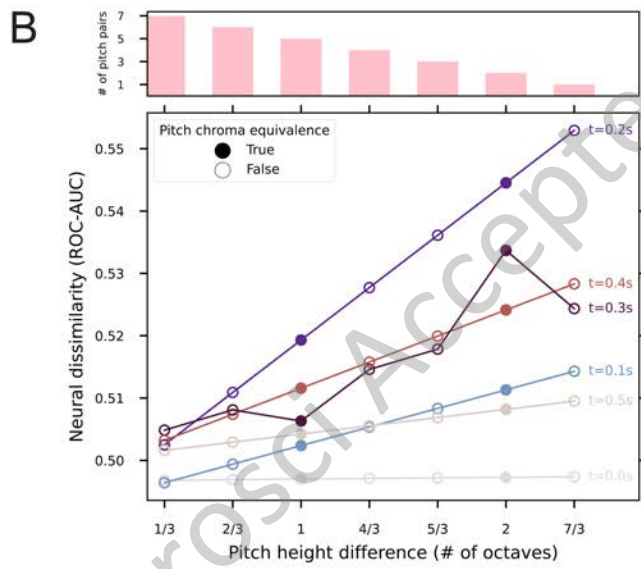
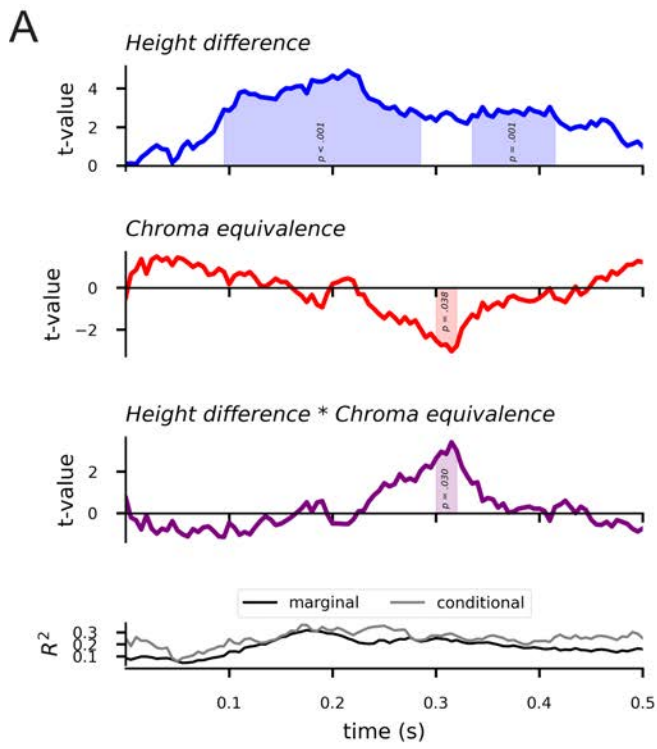
**B**

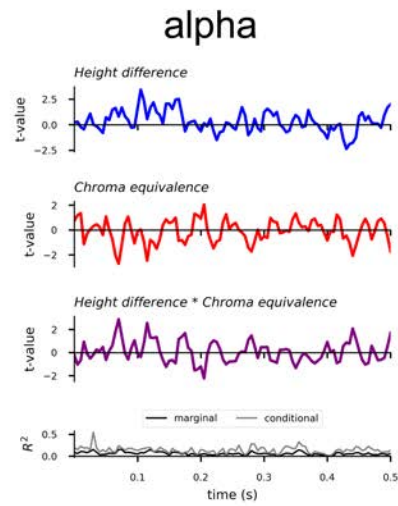
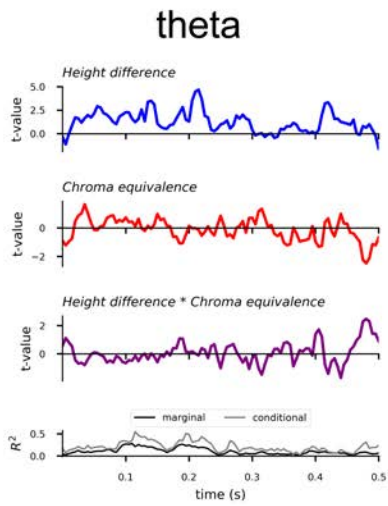
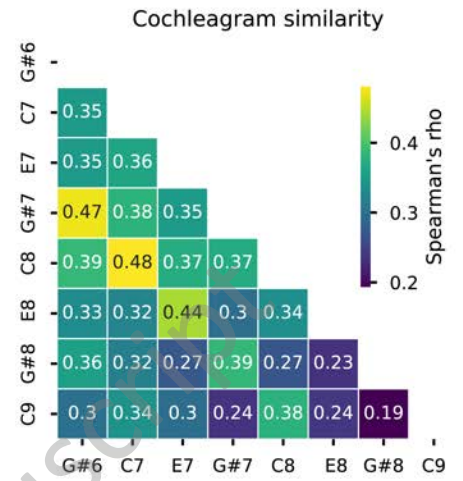
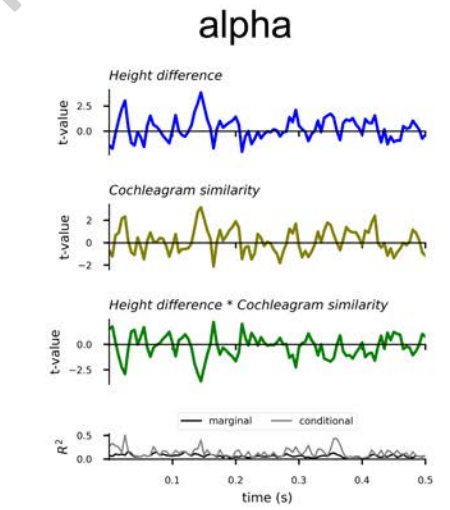
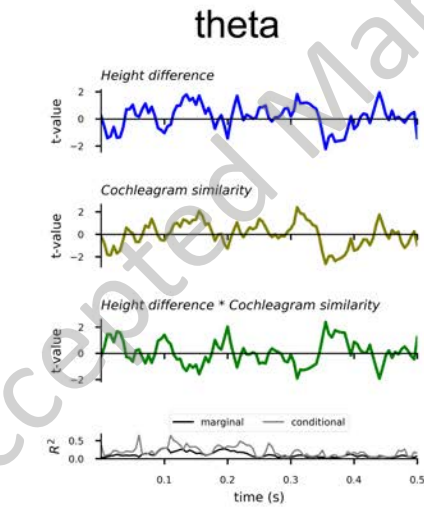
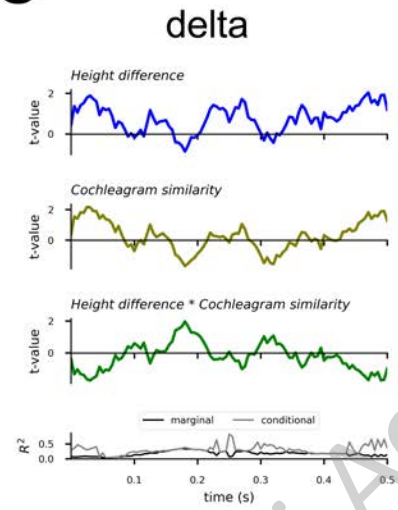
Averaged per pitch-pair



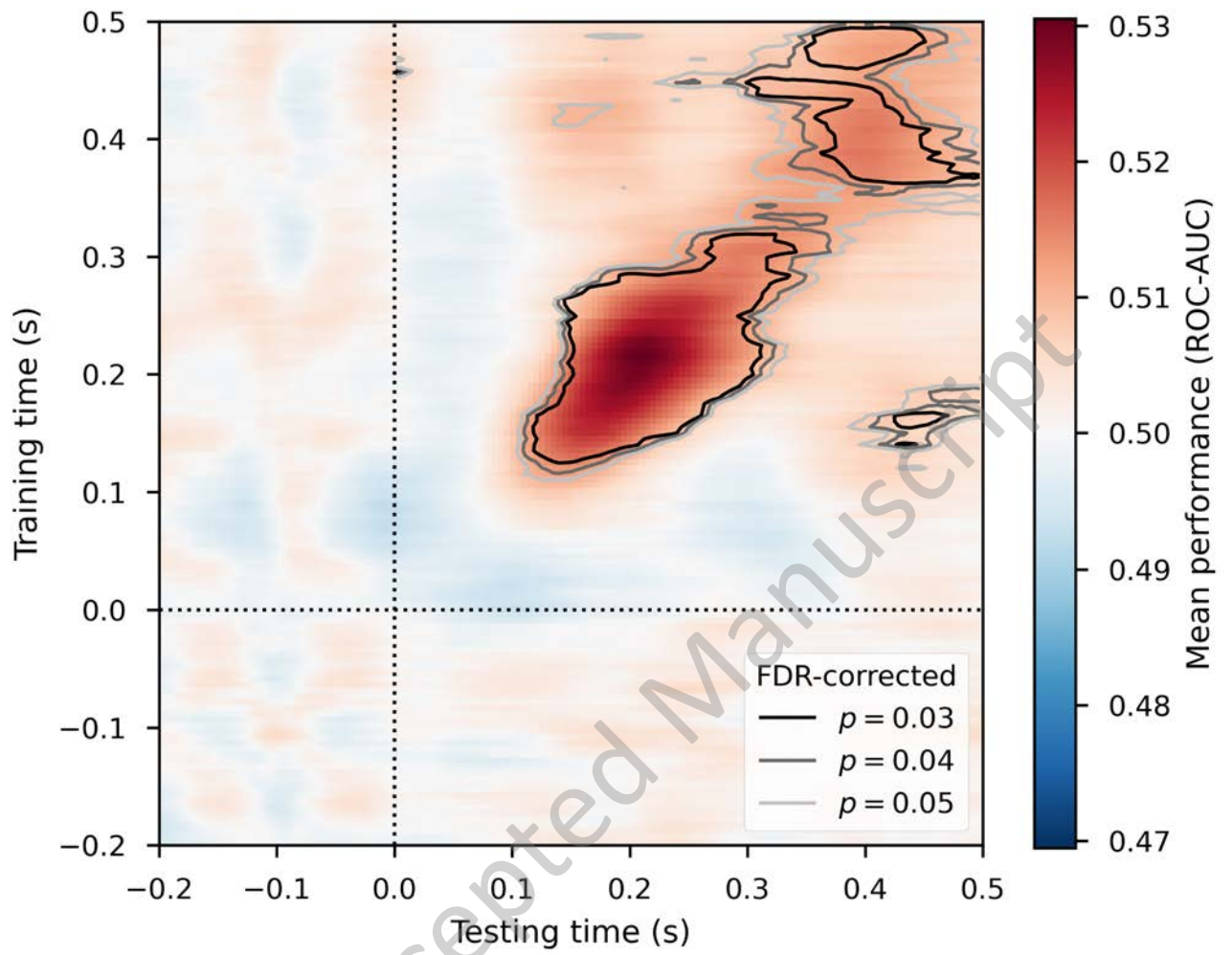
JNeurosci

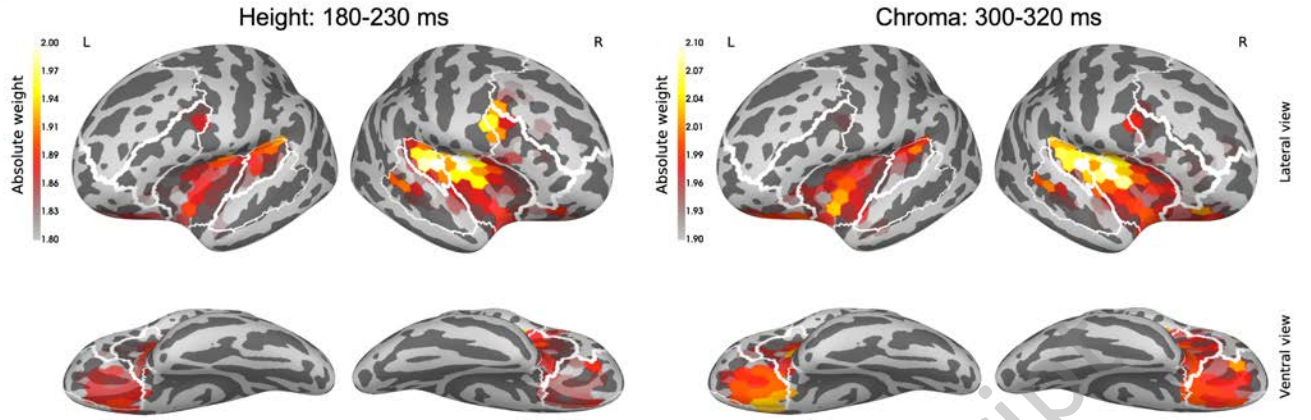
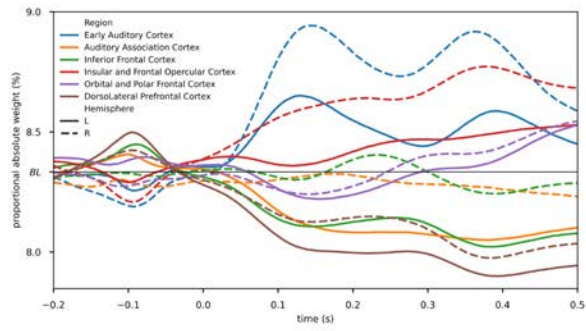
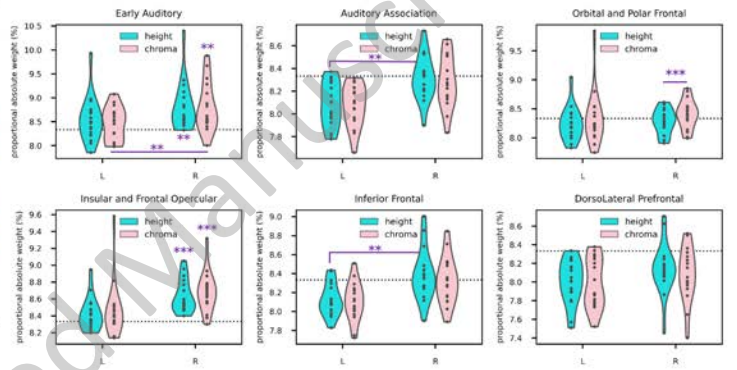
ACCEPTED MANUSCRIPT



A**B****C**

JNeurosci Accepted Manuscript



A**B****C**

JNeurosci Accepted Manuscript