

UC Merced

UC Merced Previously Published Works

Title

Neurophysiological time course of timbre-induced music-like perception

Permalink

<https://escholarship.org/uc/item/2s23d8nx>

Journal

Journal of Neurophysiology, 130(2)

ISSN

0022-3077

Authors

Santoyo, Alejandra E
Gonzales, Mariel G
Iqbal, Zunaira J
[et al.](#)

Publication Date

2023-08-01

DOI

10.1152/jn.00042.2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed

1
2 **Neurophysiological time course of timbre-induced music-like perception**

3
4 Alejandra E. Santoyo^{1ε}, Mariel G. Gonzales^{1ε}, Zunaira J. Iqbal^{1ε}, Kristina C. Backer^{1,2}, Ramesh
5 Balasubramaniam^{1,2}; Heather Bortfeld^{1,2,3}, and Antoine J. Shahin^{1,2*}

6
7 ¹Department of Cognitive and Information Sciences, University of California, Merced.

8 ²Health Science Research Institute, University of California, Merced.

9 ³Department of Psychology, University of California, Merced.

10
11 ε Equal contributions.

12
13
14 **Running title:** Neurophysiology of timbre irregularities

15
16 ***Corresponding author:**

17 Antoine J. Shahin

18 Department of Cognitive and Information Sciences

19 University of California, Merced

20 Email: ashahin@ucmerced.edu

21
22 **Number of pages: 24**

23 **Number of Figures: 8**

24 **Abstract: 250 words**

25 **Introduction: 1334 words**

26 **Discussion: 1686 words**

27
28
29 **Conflict of interest:** The authors declare no conflict of interest.

30
31 **Acknowledgments:** We thank Dr. Laurel Trainor for her valuable input on the study. This work
32 was supported by the National Institute on Deafness and Other Communications Disorders Grant
33 R01 DC-013543 (A.J.S.).

34
35
36
37
38
39
40
41
42 **Keywords:** EEG, Melody processing, Oscillatory activity, Pitch processing, Timbre processing

47 **ABSTRACT**

48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78

Traditionally, pitch variation in a sound stream has been integral to music identity. We attempt to expand music’s definition, by demonstrating that the neural code for musicality is independent of pitch encoding. That is, pitchless sound streams can still induce music-like perception and a neurophysiological hierarchy similar to pitched melodies. Previous work reported that neural processing of sounds with no-pitch, fixed-pitch, and irregular-pitch (melodic) patterns, exhibits a right-lateralized hierarchical shift, with pitchless sounds favorably processed in Heschl’s gyrus, ascending laterally to non-primary auditory areas for fixed-pitch and even more laterally for melodic patterns. The objective of this EEG study was to assess whether sound encoding maintains a similar hierarchical profile when musical perception is driven by timbre irregularities in the absence of pitch changes. Individuals listened to repetitions of three musical and three non-musical sound-streams. The non-musical streams were comprised of seven 200-ms segments of white, pink, or brown noise, separated by silent gaps. Musical streams were created similarly, but with all three noise types combined in a unique order within each stream to induce timbre variations and music-like perception. Subjects classified the sound streams as musical or non-musical. Musical processing exhibited right dominant alpha power enhancement, followed by a lateralized increase in theta phase-locking and spectral power. The theta phase-locking was stronger in musicians than in non-musicians. The lateralization of activity suggests higher-level auditory processing. Our findings validate the existence of a hierarchical shift, traditionally observed with pitched-melodic perception, underscoring that musicality can be achieved with timbre irregularities alone.

New and Noteworthy: EEG induced by streams of pitchless noise segments varying in timbre were perceived as music-like and exhibited a right-lateralized hierarchy in processing similar to pitched melodic processing. This study underscores that the neural-code of musicality is independent of pitch encoding. The results have implications for understanding music processing in individuals with degraded pitch perception, such in cochlear-implant listeners, as well as the role of non-pitched sounds in the induction of music-like perceptual states.

79 **INTRODUCTION**

80 When listening to music, we experience a sequence of notes of varying pitch, rhythm,
81 and timbre, smoothly transitioning across time, and creating a melodic auditory stream. While
82 pitch irregularities, giving rise to melodies, have been a central focus of music identity, it
83 remains unknown whether musical perception and hierarchical encoding associated with musical
84 processing in the auditory cortex (AC) is contingent on the existence of pitch in sound. We argue
85 that this is not necessary. We provide evidence that hierarchical neurophysiological processing
86 associated with melody is also exhibited for music-like pitchless forms of sounds, varying only
87 in timbre.

88 Musical perception is typically attributed to pitch, timbre, and loudness contours—that is,
89 the pattern of changes in an auditory dimension (e.g., pitch, timbre, loudness) over time
90 (McDermott et al., 2008; Graves et al., 2019). A key perceptual attribute of timbre is brightness
91 which is correlated to a sound’s spectral centroid—center of mass of any given stream of sound.
92 Higher spectral centroids typically elicit perception of higher brightness. Previous studies have
93 drawn connections between the mechanisms among pitch, timbre, and amplitude contour
94 discrimination. McDermott and colleagues (2008) used a modified melody recognition task to
95 investigate if listeners can discriminate between transpositions of brightness and loudness
96 contours as easily as they can with pitch contours. They showed, while pitch is more accurately
97 encoded than timbre and loudness, contours in loudness and brightness are also useful, though
98 not to the same degree as pitch, in the recognition of familiar melodies. Similarly, Cousineau and
99 colleagues (2014) used varied pitch, brightness, and loudness sequences at different lengths (i.e.,
100 1, 2, and 4 pure tones in the sequence) and tasked participants with indicating if varied-length
101 pairs were the same or different. As sequence length increased, discrimination accuracy
102 decreased significantly for the loudness sequences while discrimination accuracy decreased only
103 marginally for pitch and timbre. Furthermore, it has been found that musicians perform
104 significantly better at short-term recognition of variable-pitch sequences than non-musicians,
105 though there was no significant difference in performance between musicians and non-musicians
106 in discriminating constant-pitch sequences (Siedenburg & McAdams, 2018). A recent study by
107 Graves and colleagues (2019) on congenital amusia, a neurodevelopmental disorder in which
108 individuals exhibit a deficit in musical perception, examined individuals with amusia performing
109 a short-term memory task in which they had to extract pitch, brightness, and loudness contours

110 using novel melodies. They found that the individuals with amusia displayed greater impairment
111 in discriminating between pitch and brightness-based melodies compared to loudness-based
112 melodies, suggesting that pitch and timbre may be processed by a common locus in the brain.

113 Commonalities between pitch and timbre processing are further elucidated by pitch and
114 timbral brightness-induced judgements of Shepard tones. Classic Shepard tones are generated by
115 combining shifting tones with superimposed sine waves while maintaining a fixed spectral
116 envelope; thus, producing the illusory perception of a sound with perpetually increasing or
117 decreasing pitch. Seidenburg (2018) compared the classic Shepard tone with a “pseudo-Shepard
118 tone”, created by varying the spectral envelope while maintaining the spectral fine structure
119 (tone). This manipulation induced the illusory perception of shifts in timbral brightness
120 analogous to the shifts in pitch induced by the classic Shepard tone. In a follow-up study
121 (Siedenburg et al., 2023), the same Shepard tone variations associated with pitch (shifting
122 spectral fine structure, SFS) and timbral brightness (shifting spectral envelope, SE) were again
123 compared in conditions where both SFS and SE shifts were applied simultaneously in either a
124 synergistic or antagonistic manner (same or opposite directions). Behavioral results showed that
125 the two direction changes were very rarely perceived concurrently, and that the stronger cue did
126 not correlate with the component shifted (SFS or SE), but instead with the nature of the sound
127 itself. That is, when the sound was harmonic, more weight was put on SFS, and when the sound
128 was inharmonic, more weight was put on SE (Seidenburg et al., 2023). These two studies
129 promote a unitary account of pitch and timbre perception.

130 Behavioral accounts have been key in guiding neurophysiological research on musical
131 encoding, which has demonstrated a hierarchy of sound processing. Simple sounds are largely
132 processed in the brain stem and Heschl’s gyrus (HG) (Griffiths et al., 2001; Griffiths & Hall,
133 2012), but as sound becomes more complex (e.g., richer in pitch dynamics), processing extends
134 to anterolateral regions of the AC—superior and inferior temporal gyri (Hackett et al., 2001;
135 Hyde et al., 2008; Tsai & Li., 2019; Zatorre et al., 2002). Patterson and colleagues (2002) sought
136 to understand the hierarchy of pitch processing by experimenting with regular-interval (RI)
137 sounds, which are created by overlapping regularized time intervals of broadband noise (Yost,
138 1998). Manipulating the temporal regularity of the noise segments affects perception, with a
139 stronger noise (no-pitch) perception as temporal regularity decreases and emergence of pitch-like
140 perception as temporal regularity increases (Yost et al., 1996; Patterson et al., 2002). When RIs

141 of several unique temporal intervals (“pitches”) are sequentially combined in a stream, they
142 sound melodic-like.

143 Patterson and colleagues (2002) tracked neural engagement using fMRI, while people
144 processed no-pitch, fixed-pitch, and melody-like RIs. They observed the emergence of a
145 hierarchical pattern of cortical networks. As RIs became more complex and melodic-like, areas
146 of activation extended beyond HG and planum temporale (PT) to include the superior temporal
147 gyrus (STG) and planum polare (PP) (Patterson et al., 2002; see also Puschmann et al., 2010).
148 Moreover, right hemisphere activation was more robust relative to left in STG and PP, but not in
149 HG and PT. Similarly, Hyde et al., (2008) found that fixed pitch processing occurred bilaterally
150 in HG, while melody processing activated regions in secondary auditory cortices, including PT
151 and STG. Most noteworthy was their finding that the right PT exhibited a linear increase in
152 activity corresponding to parametric increases in pitch distance; increases in activation in the left
153 hemisphere were more tempered. Taken together, these studies indicate that, as sound increases
154 in pitch complexity (e.g., is perceived as increasingly musical), there is a corresponding right-
155 lateralized shift to higher-level processing within the hierarchy.

156 The aforementioned studies laid the groundwork for the current study, whereby we
157 attempted to examine whether musical perception based solely on timbre manipulations (without
158 specific pitch changes) would elicit neural activity consistent with that observed with pitch
159 manipulations. The abovementioned behavioral studies hint at similar neural accounts, and the
160 abovementioned fMRI studies, which manipulated pitch, served as our frame of reference. We
161 used EEG to examine whether hierarchical neural shifts in processing take place when timbre
162 differences in pitchless sounds are perceived as music-like. EEG is inferior to fMRI in its ability
163 to spatially locate neural activity, however our objective here was to assess the time-course of
164 these shifts, a quality better assessed with EEG than fMRI. We compared the processing of “non-
165 musical” to “musical” sound streams. We used quotes for musical and non-musical conditions to
166 emphasize that the perceptual effect is music-like; henceforth we drop the quotes. All our non-
167 musical streams consisted of seven 200-ms bursts of either white, pink, or brown noise with
168 37.5-ms silent gaps separating the noise bursts (see Fig. 1). Musical streams were created
169 similarly, with the only difference being that all three types of noise were combined in a unique
170 order within each stream. Since noise bursts were used in both the musical and non-musical
171 conditions, subjects could not base their perception of music on pitch cues. The only distinction

172 between the two conditions was that the inclusion of all three noise types within each musical
173 stream produced variations in timbre (i.e., spectral bandwidths) as a function of time. We
174 hypothesized that: 1) stronger alpha desynchronization (suppression) should be observed for the
175 musical than the non-musical condition in lateral temporal channels, indexing stronger
176 engagement of non-primary AC networks during musical processing; 2) alpha desynchronization
177 should be more pronounced in the right hemisphere; 3) the above two predictions should be more
178 pronounced in musicians than non-musicians, as musical encoding should be stronger in
179 musicians (Shahin et al, 2003, Schneider et al., 2005, 2022), leading to a more robust
180 lateralization and hierarchical shift.

181

182 **METHODS**

183 **Participants**

184 Twenty-five participants were recruited through the University of California, Merced
185 online recruiting database and through recruitment flyers. Prior to commencing the experiment,
186 all participants provided informed consent and completed a general questionnaire assessing their
187 education level, handedness, language and musical backgrounds, hearing status, and neurological
188 history. One participant was excluded due to neurological history and another due to a high
189 percentage of EEG artifacts. Thus, 23 adult participants (9 females; 18 years and older, $M = 25.1$
190 years, $SD = 7.1$; all right-handed except for one ambidextrous) were included in the final
191 analyses. Participants were monetarily compensated after completion of the experiment. The
192 experimental protocol was approved by the Institutional Review Board at UC Merced.

193 To account for musicianship (musical skill), the 23 participants were divided into two
194 groups according to years of formal music training. Musicians ($n = 11$; age 25.18 years) had an
195 average of 7 years of formal music training, while non-musician controls ($n = 12$; age 25.08
196 years) had an average of less than 0.2 years of formal training. Instruments of training among
197 musicians included piano, guitar, violin, flute, cello, saxophone, viola, clarinet, and trumpet.

198

199 **Stimuli**

200 Two-hundred millisecond brown, pink, and white noise-bursts were created using Adobe
201 Audition (stimuli are available at <https://figshare.com/s/d560fd87ee54c1b5e13a>). In total, there
202 were 7 noise-bursts within each stream with 37.5 ms of silence between bursts, lasting a total

203 duration of 1,625 ms. Each non-musical stream ($n = 3$) comprised of just one of the three types
204 of noise (brown, pink, or white), and musical streams ($n = 3$) were created by combining all three
205 noise types into three unique patterns (see Figure 1 and Experimental Design and Statistical
206 analyses section for details).

207 -----
208 Figure 1 here
209 -----
210

211 Procedure

212 EEG was recorded as participants were seated in a sound-attenuated booth
213 approximately 1 meter away from a 27-inch monitor. Auditory stimuli were presented at about
214 70 dBA SPL through 2 loudspeakers located to the left and to the right of the monitor at a $\pm 45^\circ$
215 angle, creating the perception that the sound came from the center (0°). The study consisted of 10
216 blocks of 60 trials each, with a single sound stream presented per trial. The non-musical and
217 musical stimuli were presented in a random order within each block. No training with the sounds
218 took place prior to study participation. Participants were instructed to listen to the stimuli while
219 fixating their gaze on a cross in the center of the monitor and indicate whether they perceived
220 each stimulus as non-musical or musical. A visual cue to respond was displayed immediately
221 following the offset of each trial. Participants indicated their responses via keyboard presses, “n”
222 for non-musical and “m” for musical; they used their right hand to respond. No feedback was
223 given regarding correct/incorrect responses. Presentation software (NeuroBehavioral Systems
224 Inc., Berkeley, CA) was used to present the stimuli and record responses. EEG was recorded
225 using a high-impedance 64-channel Biosemi System (Active Two system, 10–20 Ag-AgCl
226 electrode, with Common Mode Sense and Driven Right Leg passive electrodes serving as
227 grounds, A/D rate: 1,024 Hz).

228 Data Analysis

230 Behavior

231 Individual response logfiles were parsed using in-house MATLAB (The MathWorks Inc.,
232 Natick, MA) code, producing reaction times (RT) and classification proportions (musical, non-
233 musical). For each stimulus condition (musical, non-musical) we calculated the proportion of
234 musical responses relative to the total stimulus condition trials. There was no need to compute

235 the non-musical proportions because the total proportions of musical and non-musical responses
236 across the two conditions was equal to 1. RT was measured from onset of sound until a button
237 press response.

238

239 **EEG**

240 EEG analyses were conducted using ERPLAB (Lopez-Calderon & Luck, 2014),
241 EEGLAB (Delorme & Makeig, 2004), and FieldTrip (Oostenveld et al., 2011) toolboxes and in-
242 house MATLAB code. Processing steps included: 1) Individual continuous EEG blocks were
243 downsampled to 512 Hz and concatenated into one continuous dataset for each subject. 2) Each
244 continuous dataset was epoched (segmented) from -1.25 to 2.5 s around the onset of each noise
245 stream and baselined to the entire epoch (mean removed). 3) Independent component analysis
246 (ICA) was conducted on each individual dataset, producing 64 ICA components. 4) ICA
247 components that reflected ocular artifacts were removed (mean of 2 per subject) and bad
248 channels (maximum of 2) were interpolated using EEGLAB's spherical interpolation. 5)
249 Individual datasets were average referenced. Datasets were further processed in two ways: to
250 produce files suitable for auditory evoked potentials (AEPs) and files suitable for Oscillatory
251 activity.

252

253 **AEPs**

254 Individual files were filtered between 0.1 and 30 Hz (zero-phase Butterworth, fourth
255 order), re-epoched around acoustic onsets from -0.2 to 2 s, linearly detrended to remove the slow
256 shift, and re-baselined to the pre-stimulus period (-0.2 to 0 s). Then, trials with amplitude shifts
257 greater than $\pm 150 \mu\text{V}$ at any channel were removed. The mean number of remaining trials for the
258 musical and non-musical conditions was 279 for each condition. Because initial analyses
259 revealed an overwhelming musical response for the musical stimuli and non-musical response
260 for non-musical stimuli (see Fig. 2), we segregated the conditions based on stimulus type and not
261 percept type. Doing so maintained an equal number of trials and equal acoustic characteristics
262 between the two conditions. Maintaining an equal number of trials is essential for the analysis of
263 inter-trial phase coherence (ITPC), as ITPC values are influenced by trial number (Fisher, 1993).
264 Finally, trials within each stimulus condition (musical, non-musical), were averaged, creating
265 separate AEP files for each condition.

266

267 ***Oscillatory activity***

268 Individual files with amplitude shifts larger than 200 μ V at any channel were rejected and
269 trials were separated into the two conditions (musical and non-musical). The remaining mean
270 number of trials for the musical and non-musical conditions were 271 and 270, respectively.
271 Event-related spectral perturbation (ERSP, spectral power) and ITPC spectrograms were
272 generated for each condition, channel, and participant using the timef.m function of the
273 EEGLAB toolbox. We examined activity for the period of -1250 to 2500 ms relative to noise-
274 stream onset for the 2.7 – 50 Hz frequency range. The analysis used a sinusoidal wavelet-based
275 Discrete Fourier transform (DFT) of the time-domain signal (750-ms Hanning window, 12 ms
276 and 1 Hz steps) with two cycles at the lowest frequency, increasing linearly to 37 cycles at the
277 highest frequency. For the ERSP analysis, post-stimulus activity was baselined to the pre-
278 acoustic onset in the period spanning from -1250 to -375 ms.

279

280 **Experimental Design and Statistical Analyses**

281 The experimental design was as follows: Six sound streams, each containing 7 noise-
282 bursts separated by 37.5 ms silence, were created using three types of noise (white, pink, brown).
283 The control streams (non-musical) consisted of 7 repetitions of one noise type, yielding three
284 unique streams. In the experimental (musical) condition, each of the three streams consisted of
285 all three noise types, uniquely ordered in such a way that the acoustical sum of the three streams
286 within the musical condition was identical to the sum of the three non-musical streams.
287 Subsequent behavioral and EEG analyses were based on the sum, not individual streams. This
288 was deliberate to rule out physical differences between the two conditions (the summed stream is
289 identical between musical and non-musical), while maintaining perceptual influence.

290 The statistical analyses comparing reaction time (RT) or classification between the
291 musical and non-musical conditions and the interaction with group (Musicians, non-Musicians)
292 were based on Analysis of Variance (ANOVA). T-tests (corrected for multiple comparisons
293 using Bonferroni test) were further used to compare classification values relative to chance level.
294 Non-responses were excluded (only for behavioral analysis).

295 EEG activity (AEP amplitude, spectral power, or phase-locking) differences between
296 conditions were examined using the nonparametric cluster-based permutation test (CBPT; Maris

297 & Oostenveld, 2007) as implemented in the FieldTrip toolbox (Oostenveld et al., 2011). The
298 CBPT takes into account the activity at all time-points and channels. For the AEP amplitude
299 data, the CBPT organizes significantly different EEG amplitudes (based on dependent-samples t-
300 tests between the musical and non-musical conditions done at each data sample, $\alpha = 0.05$)
301 into continuous clusters according to the significant samples' spatial adjacency and temporal
302 adjacency. The t-statistics at each time-channel sample within each cluster were then summed.
303 To create the null distribution, these steps were repeated for each resampling of the data (2,000
304 permutations; Monte Carlo simulation). For each permutation, the largest cluster t-statistic sum
305 was recorded. Finally, the summed t-statistics for each cluster in the real data were compared
306 against this null distribution of maximal cluster t-statistic sums, using a cluster-level alpha value
307 of 0.05. Any clusters that surpass this alpha value are considered to contain activity that is
308 significantly different between the musical and non-musical conditions.

309 The CBPT statistical approach was also used on spectral power and phase-locking data.
310 The only difference between the AEP and time-frequency CBPTs is that for the time-frequency
311 data, the cluster formation is conditioned upon not only spatial (channel) and temporal adjacency
312 between the significantly different samples (identified via univariate t-tests), but also the
313 adjacency of the significantly different samples in frequency space. The results of the CBPTs
314 applied to the spectral power and phase-locking spectrograms reveal differences in spectral
315 power or phase-locking values, respectively, between the musical and non-musical conditions.

316 For post hoc analysis, and to test for interaction of oscillatory activity and group
317 (Musicians vs. Non-Musicians), we used ANOVAs. Correlation results were based on Pearson
318 Correlation Coefficient (PCC) method, corrected for multiple comparisons using Bonferroni test.

319

320 **RESULTS**

321 **Behavior**

322 Figure 2A depicts the classification proportions of the musical response observed for the
323 musical and non-musical stimulus conditions. We limited our ANOVA to the musical
324 classification for the two conditions. An ANOVA with the variables group and condition
325 (musical, non-musical) revealed that there was only a main effect of condition ($F_{(1, 21)} = 87.3$, $p =$
326 0.0001 ; $\eta^2 = 0.81$) with no other main effect or interaction between the variables ($F < 1$). The

327 condition effect was due to a strong labeling of the musical stimulus as musical and a weak
328 labeling of the non-musical stimulus as musical.

329 The musical classification within each stimulus condition was further subjected to a one
330 sample t-test against chance level (50%). Both tests revealed that the musical classification was
331 above chance within the musical condition ($t_{(22)} = 2403$, $p = 0.00001$; Mean musical = 0.944,
332 95% CI = 0.90 to 0.99) and below chance within the non-musical condition ($t_{(22)} = 739$, $p =$
333 0.000001; Mean non-musical = 0.23, 95% CI = 0.091 to 0.37). While these effects seem obvious,
334 this labeling reinforces the notion that irregular noise patterns (i.e., irregular timbre patterns) can
335 be perceived as musical despite participants' lack of familiarity with the stimuli. Moreover, the
336 lack of a group effect demonstrates that there was no evidence of a musicianship advantage in
337 identifying a musical stimulus as musical and vice versa for the non-musical stimuli.

338 A similar ANOVA for the RT revealed no main effects or interaction between the
339 variables ($F < 1$). The lack of RT differences between the two conditions and among groups
340 provides evidence that the two conditions were comparable in difficulty, and that musicians and
341 non-musicians did not exhibit differences in cognitive/attentional efforts when labeling the two
342 conditions.

343 -----
344 Figure 2 here
345 -----
346

347 EEG

348 AEPs

349 Figure 3 depicts the AEP results. The CBPT revealed four significant clusters (Two
350 positive clusters, $p = 0.046$ and $p = 0.001$, respectively; two negative clusters, $p = 0.016$ and $p =$
351 0.001, respectively). Figure 3A shows the AEP waveforms for the positive clusters (i.e., musical
352 had a more positive/less negative amplitude than non-musical), and Figure 3B shows the AEP
353 waveforms for the negative clusters (i.e., musical had a more negative/less positive amplitude
354 than non-musical). Below the waveforms are the t-value topographies distinguishing musical and
355 non-musical within the significant windows (gray shaded areas). Clearly, the positive (clusters 1
356 and 2) and negative (clusters 3 and 4) clusters represent the same source(s), given the highly
357 overlapping time windows and near identical scalp topographies. Thus, we will discuss the
358 positive and negative clusters as reflecting a common activity.

359 Our AEP results can be summarized as follows: First, the waveforms exhibited eight
360 distinct transient (steady state) potentials: an initial onset AEP response (typical P1-N1-P2
361 morphology), followed by six peaks and lastly followed by an offset response (typical P1-N1-P2
362 morphology but with attenuated amplitude). These AEPs clearly reflect the onsets/offsets of the
363 seven noise-bursts, i.e., noise-stream envelope. Second, the musical and non-musical AEP
364 waveforms begin to diverge around 325 ms, whereby the musical waveform weakens in
365 amplitude (becomes closer to zero for both the positive and negative clusters). The latency of
366 divergence is about 80-90 ms after the onset of the second noise-burst of the noise-stream; the
367 onset of the second noise burst is depicted by a red dotted vertical line within the plots shown in
368 Fig. 3. This is the first instance when the musical stream can be perceived as irregular relative to
369 the non-musical stream because the second noise-burst in the musical streams has a different
370 timbre than that of the first noise-burst. The divergence between the waveforms of the two
371 conditions continues until after 1400 ms. However, towards the end of this divergence (about
372 1200 – 1300 ms), there is a short period in which the musical AEP's amplitude becomes greater
373 (moves away from 0 μ V) than the non-musical AEP amplitude (which moves toward 0 μ V). This
374 short period occurs about 250 ms after the onset of the fifth noise-burst (depicted by the blue
375 dotted vertical line in Fig. 3) in the sound stream. This is also around the same time as the onset
376 of the sixth noise-burst in the sound stream. Third, the topography (frontocentral and temporo-
377 occipital) of these AEP differences is reflective of sources within and surrounding HG. In short,
378 differences in musical and non-musical AEPs can be observed following the onset of the first
379 irregularity in the noise stream. musical AEPs are continually diminished until after the onset of
380 the 5th/6th segment(s) in the noise stream in which they eventually rebound.

381 -----
382 Figure 3 here
383 -----
384

385 **Musicianship.** Post hoc ANOVA on the mean AEP values (between 350 ms – 1350 ms) at the
386 frontocentral channels and within the time-points reaching significance did not reveal an
387 interaction between condition and group ($F_{(1, 21)} = 3.1, p = 0.09$). Thus, musicianship was not a
388 factor in AEP differences. Despite the lack of significance, it is worth noting that musicians
389 tended to exhibit weaker AEPs than non-musicians especially for the musical condition. If we

390 contend that AEPs reflect low-level auditory processing, then this would suggest that musicians
391 shift their processing to higher auditory levels during processing more so than non-musicians.

392

393 *Oscillatory activity*

394 **Spectral power.** The CBPT revealed one positive cluster ($p = 0.026$), indicating significantly
395 greater spectral power for the musical condition than non-musical condition, which was observed
396 in all 64 channels. Figure 4 (top panels) reveal the t-value spectrograms of spectral power
397 distinguishing musical versus non-musical for a subset of left, mid, and right channels. Bottom
398 panels show the t-value topographies of alpha (9-15 Hz, ~650 ms to ~1350 ms) and theta (3-8 Hz
399 ~1150 ms to ~1650 ms) oscillatory activities within the window of significance for these
400 frequencies. Our findings demonstrate the following: First, both theta and alpha activities were
401 greater for the musical condition than the non-musical condition. Second, alpha was concentrated
402 more in temporal sites, especially in the right hemisphere. Third, theta was concentrated
403 frontocentrally and temporoparietally, with right hemispheric dominance that was mainly evident
404 in the low-theta (3-4 Hz) spectral power. Theta's topography is consistent with generators within
405 or surrounding HG. Fourth, the different topographies of theta and alpha strongly suggest that
406 theta and alpha have different neural origins. Fifth, while theta and alpha overlapped in time,
407 alpha (maximally exhibited at temporal sites) began earlier and theta (maximally exhibited at
408 frontocentral and temporal sites) ended later.

409

410 Figure 4 here
411 -----

412

413 **Phase-locking.** As in the spectral power analysis, the phase-locking analysis produced one scalp-
414 wide positive cluster ($p = 0.001$), indicating that the musical condition resulted in stronger phase-
415 locking than the non-musical condition. Figure 5 (top panels) depict the t-value spectrograms
416 distinguishing phase-locking of musical versus non-musical for a subset of left, mid, and right
417 channels. There was scalp-wide activity of low theta (3- 4 Hz) occurring around 1130 – 1440 ms.
418 The t-value topography of this activity—lateral central and bilateral temporoparietal—may
419 suggest auditory generators that are more lateral from what we expect from HG generators (as
420 reported earlier). Thus, the origins of this theta phase-locking activity is uncertain; they more
421 likely reflect generators outside of HG, e.g., within STG.

422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456

Figure 5 here

Correlations. We examined the degree of correlation between theta and alpha power (means within their significant time windows) to assess a possible link between them. This was an important correlation since both clearly index different neural generators (different topographies). Figure 6 demonstrates the correlation between theta and alpha power regardless of condition (averaged across the musical and non-musical conditions ($r = 0.69$; $p = 0.0006$)). Separately, correlation for either condition was highly significant (musical: $r = 0.67$, $p = 0.001$; non-musical: $r = 0.7$, $p = 0.0005$).

Figure 6 here

Musicianship. Post hoc ANOVA on the theta and alpha spectral power (conducted separately) obtained in the CBPT analyses with the variables group and condition did not reveal a main effect of group or an interaction between the variables ($F < 2$). However, both ANOVAs did show an effect of condition ($F > 5$), thus confirming the CBPT results.

A similar post hoc ANOVA on the theta phase-locking values also did not reveal a main effect of group. However, it revealed a main effect of condition (confirming the CBPT results, $F > 70$) and an interaction between group and condition ($F_{(1, 21)} = 8.9$, $p = 0.007$; $\eta^2 = 0.3$). We attribute the interaction to significantly stronger theta phase-locking occurring in musicians for the musical vs. non-musical condition ($p = 0.000001$; Bonferroni test) relative to non-musicians who showed a weaker but still significant difference in phase-locking for the musical vs. non-musical condition ($p = 0.0036$; Bonferroni test).

Figure 7 here

DISCUSSION

Pitch perception has been a central focus in psychoacoustic research and has been instrumental in forming and testing theories of sound representations, processing, and perception in the auditory system. Unlike with spoken language, research on music processing has traditionally had a strong focus on pitch, the perceptual construct of the fundamental frequency,

457 or f_0 . One of the major questions raised in this paper is whether the experience of music is
458 limited only to sounds that vary in pitch. Indeed, there are no agreed-upon rules concerning the
459 types of sounds that give rise to musical experience. Any sounds with irregular patterns can be
460 considered musical or melodic, as long as an individual perceives them as such.

461 Recent work in music perception shows that humans can experience musicality of
462 irregular streams of sounds where pitch is not explicitly varied and, in some cases, not present, as
463 in beatboxing (Proctor et al, 2013). Because beatboxing involves sequences of sounds that
464 sometimes lack pitch (e.g., unvoiced percussion sounds), beatboxers are thus expected to be
465 more attuned to streams of sound that are irregular but contain no pitch or no variations in pitch.
466 Interestingly, recent work has shown that beatboxers produce different responses to acoustic
467 variations in pitched sounds, compared to musicians trained in traditional musical instruments
468 (Krishnan et al, 2018), suggesting the existence of unique individual differences in processing
469 sounds and music that is dependent on an individual's musical experience. Here we ask whether
470 sounds that do not involve pitch but produce timbre irregularities can elicit similar brain response
471 in "traditional" musical sounds that do contain pitch. Our results show that the processing of
472 irregular pitchless noise streams perceived as musical, produces an increase in alpha power with
473 right hemisphere dominance, followed by an increase in theta phase-locking and spectral power.
474 Because EEG does not allow definitive localization of theta and alpha neural generators, our
475 interpretation emphasizes the time course of neural activity, while limiting information on neural
476 origins to basic scientific deduction.

477 The theta spectral power scalp topographies exhibited fronto-central activity with
478 posterior temporo-occipital activity that resembles processing/activity typically generated within
479 and surrounding HG, as with AEPs (Fig 3). HG generators, as opposed to higher level (e.g., non-
480 primary auditory cortex or nPAC) generators, are particularly involved in processing simple
481 sound features such as sound onsets and offsets (Heinrich et al., 2008; Shahin et al., 2009a),
482 leading to a stronger temporal alignment of the AC activity with the noise-stream envelope.
483 Thus, theta activity, including AEPs, more likely reflects engagement of low-level auditory
484 networks (e.g., HG) indexing basic sound features. Second, the right lateralized alpha
485 enhancement (i.e., synchronization) is consistent with a gyral origin, such as within STG in
486 nPAC. We know that alpha enhancement is associated with neural inhibition, while alpha
487 suppression (desynchronization) indicates neural excitation (Pfurtscheller et al 1996; Jensen et

488 al., 2002; Mazaheri & Picton, 2005; Sauseng et al., 2009; Shahin et al., 2009b; Kerlin et al.,
489 2010; Obleser & Weisz, 2012; Rohenkohl & Nobre, 2011; Weisz et al., 2011). We interpret this
490 alpha enhancement as reflecting inhibitory mechanisms in nPAC.

491 At first glance, the alpha result seems to contradict our hypothesis, i.e., we expected
492 excitation of nPAC during musical processing. However, a well-established account linking
493 alpha enhancement and item retention in working memory is well-suited to explain our findings.
494 Prior studies have demonstrated that an increase in item retention in working memory coincides
495 with an increase in alpha power (e.g., Klimesch et al. 1999; Jensen et al 2002; Sauseng et al.,
496 2005; Obleser et al., 2012; Wianda and Ross 2019). For example, Jensen et al. (2002) used a
497 modified Sternberg task, in which individuals were required to retain 2, 4, or 6 consonants
498 presented simultaneously on a screen, and judge whether a subsequent probe matched one of the
499 items in working memory. During the retention period, alpha power (9-12 Hz) systematically
500 increased with an increase of the number of items retained in working memory. Jensen and
501 colleagues interpreted this alpha increase as signifying the brain's effort to inhibit encoding of
502 irrelevant information within the neural population associated with retaining items in working
503 memory. Furthermore, Andrillon et. al., (2015) demonstrated that repeated exposure to noise
504 patterns lead to the perception of acoustic uniqueness over time as implicit memory is formed in
505 listeners tasked with detecting sequence repetitions. This neural adaptation was marked by the
506 formation of AEPs (N1-P2) time-locked to when noise patterns were later recognized
507 behaviorally. Likewise, the aforementioned AEPs were also elicited when listeners were
508 distracted by an auditory task that prevented them from attending to the noise sequences. These
509 findings demonstrate that short-term neural plasticity can give rise to perceptual sensitivity to
510 subtle acoustic cues in pitchless sounds with repeated exposure with or without focused
511 attention. Similarly, we posit that the current noise sequential encoding is reflected in auditory
512 short-term memory. Our sound streams consisted of increasing numbers of items (7 items
513 maximum) unfolding over time, requiring serial maintenance in auditory short-term memory for
514 a subsequent judgment of whether the sum of the items (noise-bursts) constituted a musical
515 sequence.

516 An interesting outcome of the current study is the group distinction. Despite the musical
517 noise-stream being an unconventional form of music, i.e., comprised of noise, musical streams
518 evoked stronger low-theta phase-locking than non-musical toward the end of the stream

519 (following identification of the noise-stream as a musical), especially in musicians. We know
520 from previous accounts that the degree of theta phase-locking reflects musicality (Shahin et al.,
521 2010; Vanden Bosch der Nederlanden et. al, 2020). In their thoughtful study, Vanden Bosch der
522 Nederlanden et al. (2020) demonstrated that neural entrainment to syllable rate in linguistic
523 utterances in speech and song led to better theta phase-locking to phrasal and syllabic
524 information during sung speech. Thus, it is reasonable to conclude that phase-locking, indexing
525 musicality, would be more prominent in musicians than non-musicians. Also, work from our lab
526 and others demonstrated enhancement of P1, N1 and P2 AEPs (evoked by music sounds) in
527 musicians relative to non-musicians (Pantev et al., 1998, 2001; Shahin et al, 2003, 2004, 2005;
528 Schneider et al., 2005, 2022; Sanju and Kumar 2016; Baumann et al., 2008; Doelling and
529 Poeppel, 2015). The N1-P2 AEPs are phase-locked in the theta band, while the P1 is reflected in
530 higher frequency phase-locking, such as in the beta (15 – 30 Hz) and gamma (>30 Hz) evoked
531 response (Shahin et al., 2008; Doelling and Poeppel, 2015). However, a puzzling aspect of the
532 current theta phase-locking is its strong lateral topography. The topography is not consistent with
533 typical HG generators (e.g., fronto-central as in AEPs); rather they are more consistent with
534 gyral (higher-level) non-primary auditory generators. A potential explanation is that the theta
535 phase-locking represents an encoding phase in auditory short-term memory. Evidence suggests
536 that theta phase-locking is associated with memory encoding, especially across inferior frontal
537 and hippocampal networks (Jensen 2005), including during sequence learning (Reddy et al.,
538 2021), as in the current study. Thus, we may amend our interpretation to state that the alpha
539 enhancement (inhibition of irrelevant items in auditory memory) in nPAC, is followed by phase-
540 locking in auditory memory networks, reflecting the encoding of the musical trace once the
541 stream is identified as “music-like”.

542 Given these findings, we propose the following theoretical framework (Fig. 8) regarding
543 the neural time-course of melodic processing, which is largely facilitated by the EEG’s superior
544 temporal resolution. Initially, both the musical and non-musical streams activate the auditory
545 cortex, including primary (i.e., HG) and non-primary networks (e.g., STG). Following the second
546 item in the stream, irregularity in the musical stream is detected (~ 250 ms), whereupon
547 processing is shifted toward nPAC. As a result, we observe a reduction in the AEP amplitude of
548 musical condition relative to the non-musical condition following the onset of the second item,
549 because processing is shifted laterally to nPAC. In nPAC, maintenance or retention of items

550 (individual noise-bursts) begins, leading to enhancement of alpha during the retention period
551 (showing significant activity between ~ 650 ms and 1350 ms). Once the decision is made, i.e.,
552 musical is identified, memory retention ceases, resetting theta-phase as the memory trace is
553 encoded in auditory memory, followed by a shift in processing back to HG. During this time
554 period (1130 to 1650 ms), a theta power increase, coinciding with a brief AEP rebound, occurs
555 for musical relative to non-musical stimuli, signaling a processing shift back to lower levels (e.g.,
556 HG). The strong correlation observed between the earlier alpha and later theta powers further
557 support the premise of this framework. Specifically, once nPAC concludes processing the
558 musical stream (e.g., the stream was identified as musical), it signals HG to resume processing of
559 subsequent incoming streams.

560 -----
561 Figure 8 here
562 -----

563 **Limitations**

565 A limitation of this study is that we could not make a distinction of whether timbre
566 variation (i.e., spectral variation), musical judgment, or both led to the present EEG results.
567 While we controlled for contributions of the physical (acoustical) attributes of the musical and
568 non-musical conditions, perceptually, it remains possible that timbre-variation judgment, as
569 opposed to musical identification judgement, was the main cause of the neural shifts.
570 Conversely, we cannot say with certainty that timbre-variation was the cause since the task was
571 based on musical identification and not spectral change judgments. The initial expectation for the
572 current design was that we would have enough trials of musical and non-musical identification
573 for each of the single-noise (non-musical) and tri-noise (musical) streams. However, the response
574 was severely lopsided, limiting our ability to compare EEG between different percepts of the
575 same stimulus type.

576 577 **CONCLUSION**

579 The current findings extend our understanding of the mechanisms underlying music
580 perception by targeting less common forms of music, pitchless music-like sequences. Despite
581 this deviation from conventional music identity, the neural mechanisms identified herein support
582 a similar dynamic observed with conventional music. Our results have implications for

583 understanding the evolution of sound perception and music processing. This is particularly
584 relevant in understanding music processing as it relates to the growing population of individuals
585 with auditory prosthetics (e.g., cochlear implants), who have degraded sense of pitch, and the
586 introduction or reintroduction of more basic forms of music (e.g., beat boxing), that historically
587 remained latent.

588

589 REFERENCES

- 590 Andrillon T, Kouider S, Agus T, Pressnitzer D (2015) Perceptual learning of acoustic noise
591 generates memory-evoked potentials. *Current Biology*, 25:1-7.
- 592 Baumann S, Meyer M, Jäncke L (2008) Enhancement of auditory-evoked potentials in musicians
593 reflects an influence of expertise but not selective attention. *Journal of cognitive*
594 *neuroscience* 20(12) 2238–2249.
- 595 Cousineau M, Carcagno S, Demany L, Pressnitzer D (2014) What is a melody? on the
596 relationship between pitch and brightness of timbre. *Frontiers in Systems Neuroscience*,
597 7(127).
- 598 Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG
599 dynamics including independent component analysis. *J Neurosci Methods* 15;134(1):9-21.
- 600 Doelling KB, Poeppel D (2015). Cortical entrainment to music and its modulation by expertise.
601 *Proc Natl Acad Sci U S A*. 112(45) E6233-42.
- 602 Fisher NI (1993) *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University
- 603 Graves JE, Pralus A, Fornoni L, Oxenham AJ, Caclin A, Tillmann B (2019) Short-and long-term
604 memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and*
605 *Cognition*, 136:103614.
- 606 Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the
607 temporal regularity of sound in the human brainstem. *Nature Neuroscience* 4(6) 633–637.
- 608 Griffiths TD, Hall DA (2012) Mapping pitch representation in neural ensembles with fMRI.
609 *Journal of Neuroscience* 32(39) 13343–13347.
- 610 Hackett TA, Preuss TM, Kaas JH (2001) Architectonic Identification of the core region in
611 auditory cortex of macaques, chimpanzees, and humans. *The Journal of Comparative*
612 *Neurology* 441(3) 197–222.

613 Heinrich A, Carlyon RP, Davis MH, Johnsrude IS (2008). Illusory vowels resulting from
614 perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci*
615 20(10) 1737–1752.

616 Hyde KL, Peretz I, Zatorre RJ. Evidence for the role of the right auditory cortex in fine pitch
617 resolution. *Neuropsychologia* 46(2) 632–639

618 Jensen O, Gelfand J, Kounios J, Lisman JE (2002) Oscillations in the alpha band (9–12 Hz)
619 increase with memory load during retention in a short-term memory task. *Cerebral cortex*
620 12(8) 877-882.

621 Jensen O (2005). Reading the hippocampal code by theta phase-locking. *Trends Cogn Sci.* 9(12)
622 551-553.

623 Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity:
624 gating by inhibition. *Frontiers in human neuroscience*, 4, 186.

625 Krishnan S, Lima CF, Evans S, Chen S, Guldner S, Yeff H, Manly T, Scott SK (2018)
626 Beatboxers and guitarists engage sensorimotor regions selectively when listening to the
627 instruments they can play. *Cerebral Cortex* 28(11) 4063–4079.

628 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J*
629 *Neurosci Methods.* *Journal of neuroscience methods* 164(1) 177–190.

630 McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? *Psychological*
631 *Science*, 19(12):1263-1271.

632 Kerlin, JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech
633 representations in a “cocktail party”. *Journal of Neuroscience* 30(2) 620-628.

634 Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance:
635 a review and analysis. *Brain research reviews* 29(2-3) 169-195.

636 Lopez-Calderon J, Luck SJ (2014) ERPLAB: an open-source toolbox for the analysis of event-
637 related potentials. *Frontiers in human neuroscience* 8 213.

638 Mazaheri A, Picton TW (2005) EEG spectral dynamics during discrimination of auditory and
639 visual targets. *Cognitive Brain Research* 24(1) 81-96.

640 Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its
641 acoustic details. *Cerebral cortex* 22(11) 2466-2477.

642 Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening
643 conditions and memory load drive a common α oscillatory network. *The Journal of*
644 *Neuroscience* 32(36), 12376–12383.

645 Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for
646 advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational*
647 *intelligence and neuroscience* 2011 156869.

648 Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M (1998) Increased auditory
649 cortical representation in musicians. *Nature* 392(6678) 811–814.

650 Pantev C, Roberts LE, Schulz M, Engelien A, Ross B (2001) Timbre-specific enhancement of
651 auditory cortical representations in musicians. *Neuroreport* 12(1) 169–174.

652 Patterson RD, Uppenkamp S, Johnsruide IS, Griffiths TD (2002) The processing of temporal
653 pitch and melody information in auditory cortex. *Neuron* 36(4) 767–776.

654 Pfurtscheller G, Stancák Jr A, Neuper C (1996) Event-related synchronization (ERS) in the alpha
655 band--an electrophysiological correlate of cortical idling: a review. *International journal of*
656 *psychophysiology: official journal of the International Organization of Psychophysiology*
657 24(1-2) 39–46.

658 Proctor M, Bresch E, Byrd D, Nayak K, Narayanan S. Paralinguistic mechanisms of production
659 in human "beatboxing": a real-time magnetic resonance imaging study. *J Acoust Soc Am.*
660 133(2):1043-1054

661 Puschmann S, Uppenkamp S, Kollmeier B, Thiel CM (2010) Dichotic pitch activates pitch
662 processing centre in Heschl's gyrus. *Neuroimage* 49(2), 1641-1649.

663 Reddy L, Self MW, Zoefel B, Poncet M, Possel JK, Peters JC, Baayen JC, Idema S, VanRullen
664 R, Roelfsema PR (2021). Theta-phase dependent neuronal coding during sequence learning
665 in human single neurons. *Nat Commun* 12(1) 4839.

666 Rohenkohl G, Nobre AC (2011) Alpha oscillations related to anticipatory attention follow
667 temporal expectations. *Journal of Neuroscience* 31(40) 14076-14084.

668 Sanju HK, Kumar P (2016) Enhanced auditory evoked potentials in musicians: A review of
669 recent findings. *Journal of otology* 11(2) 63–72.

670 Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR,
671 Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human
672 EEG alpha activity. *The European journal of neuroscience* 22(11) 2917–2926.

673 Sauseng P, Klimesch W, Gerloff C, Hummel FC (2009) Spontaneous locally restricted EEG
674 alpha activity determines cortical excitability in the motor cortex. *Neuropsychologia* 47(1)
675 284-288.

676 Schneider P, Sluming V, Roberts N, Bleeck S, Rupp A (2005) Structural, functional, and
677 perceptual differences in Heschl's gyrus and musical instrument preference. *Annals of the*
678 *New York Academy of Sciences* 1060 387–394.

679 Schneider P, Groß C, Bernhofs V, Christiner M, Benner J, Turker S, Zeidler BM, Seither-
680 Preisler A (2022) Short-term plasticity of neuro-auditory processing induced by musical
681 active listening training. *Annals of the New York Academy of Sciences* 1517(1) 176–190.

682 Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c
683 auditory evoked potentials in musicians. *Journal of Neuroscience* 23(13) 5545-5552.

684 Shahin A, Roberts LE, & Trainor LJ (2004) Enhancement of auditory cortical development by
685 musical experience in children. *Neuroreport* 15(12) 1917-1921.

686 Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B (2005) Modulation of P2 auditory-evoked
687 responses by the spectral complexity of musical sounds. *Neuroreport* 16(16) 1781-1785.

688 Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM (2008). Music training leads to the
689 development of timbre-specific gamma band activity. *Neuroimage*. 41(1) 113-22.

690 Shahin AJ, Bishop CW, Miller LM (2009a) Neural mechanisms for illusory filling-in of
691 degraded speech. *NeuroImage* 44(3) 1133–1143.

692 Shahin AJ, Picton TW, Miller LM (2009b) Brain oscillations during semantic evaluation of
693 speech. *Brain and cognition* 70(3) 259-266.

694 Shahin AJ, Trainor LJ, Roberts LE, Backer KC, Miller LM (2010) Development of auditory
695 phase-locked activity for music sounds. *Journal of Neurophysiology* 103(1) 218-229.

696 Siedenburg K (2018) Timbral Shepard-illusion reveals perceptual ambiguity and con- text
697 sensitivity of brightness perception. *Journal of the Acoustical Society of America*, 143(2)
698 EL93-EL98.

699 Siedenburg K, Graves J, Pressnitzer D (2023) A unitary model of auditory frequency change
700 perception. *PLOS Computational Biology*, 19(1):1-30.

701 Siedenburg K, McAdams S (2018) Short-term recognition of timbre sequences: Music training,
702 pitch variability, and timbral similarity. *Music Perception*, 36(1):24-39.

703 Tsai CG, Li CW (2019) Is It Speech or Song? Effect of Melody Priming on Pitch Perception of
704 Modified Mandarin Speech. *Brain sciences* 9(10) 286.

705 Vanden Bosch der Nederlanden C M, Joannisse MF, Grahn JA (2020) Music as a scaffold for
706 listening to speech: Better neural phase-locking to song than speech. *NeuroImage* 214 Article
707 116767.

708 Weisz N, Hartmann T, Müller N, Lorenz I, Obleser J (2011) Alpha rhythms in audition:
709 cognitive and clinical perspectives. *Frontiers in psychology* 2 73.

710 Wianda E, Ross B (2019) The roles of alpha oscillation in working memory retention. *Brain and*
711 *behavior* 9(4) e01263.

712 Yost WA, Patterson R, Sheft S (1998) The role of the envelope in processing iterated rippled
713 noise. *The Journal of the Acoustical Society of America* 104(4) 2349-2361.

714 Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and
715 speech. *Trends in Cognitive Sciences* 6(1) 37–46.

716

717 **Data accessibility**

718 Stimuli and data can be accessed at <https://figshare.com/s/d560fd87ee54c1b5e13a>

719

720 **FIGURE CAPTIONS**

721 **Figure 1: Experimental design.** **A.** Colored streams depicting white, pink and brown noise
722 streams of the non-musical condition. Each stream was comprised of 7 segments of the same
723 type of noise. **B.** Colored streams depicting white, pink and brown noise streams of the musical
724 condition. Each stream was comprised of 7 segments of a mix of all three types of noise.
725 Importantly, the mean across the three streams within each condition was the same across the
726 two conditions.

727 **Figure 2: A.** musical response proportions for the musical and non-musical stimulus conditions.
728 **B.** Reaction time (RT) for musical and non-musical stimulus conditions.

729

730 **Figure 3:** Auditory evoked potential (AEP) waveforms of significant clusters ($n = 4$, 2 positive
731 **(A)** and 2 negative **(B)**). The gray shaded areas represent the significant time windows
732 distinguishing the musical and non-musical AEPs. The negative and positive clusters represent
733 the opposite poles of the same neural sources. Below are the corresponding t-value topographies

734 reflecting differences between the musical from non-musical conditions. The topographies are
735 consistent with sources that are mainly originating within and/or surrounding HG. Black dotted
736 vertical line indexes sound stream onset (0 ms). The red dotted vertical line indicates the onset of
737 the second noise segment in the stream. Blue dotted vertical line indicates the onset of the fifth
738 segment in the stream.

739

740 **Figure 4: Top panels.** T-value spectrograms indexing significant spectral power differences
741 between the musical and non-musical conditions. **Bottom panels.** T-value topographies
742 reflecting significant spectral power (theta and alpha) differences between the musical and non-
743 musical conditions.

744

745 **Figure 5: Top panels.** T-value spectrograms indexing significant phase-locking differences
746 between the musical and non-musical conditions. **Bottom panel.** T-value topography reflecting
747 significant low-theta phase-locking differences between the musical and non-musical conditions.

748

749 **Figure 6:** Scatterplot of alpha power as a function of theta power, observed within the significant
750 windows, distinguishing the musical and non-musical conditions.

751

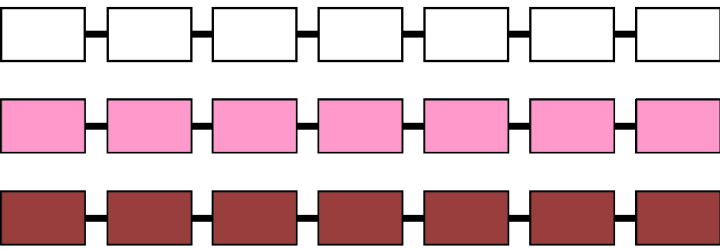
752 **Figure 7:** Theta phase-locking, observed within the significant windows, distinguishing the
753 musical and non-musical conditions in musicians and non-musicians.

754

755 **Figure 8:** Theoretical framework conveying the temporal dynamics between low- and high-level
756 auditory networks as a function of stimulus complexity and perceptual outcome.

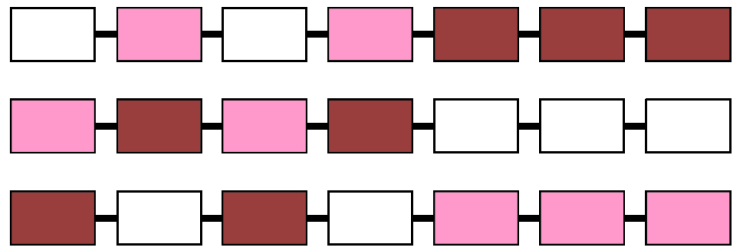
757

No pitch, constant timbre






Perceived non-musical

No pitch, irregular timbre

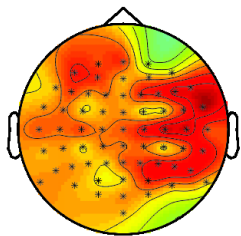


Perceived musical

Brown noise 
Pink noise 
White noise 

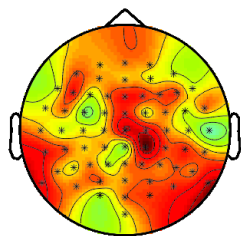
musical vs non-musical (hierarchical processing)

spectral power



Alpha

9-15 Hz, 650 -1350 ms



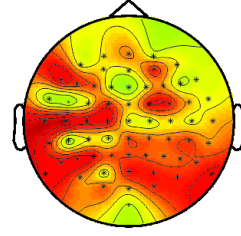
Theta

3-8 Hz, 1150 -1650 ms



t-value

phase-locking



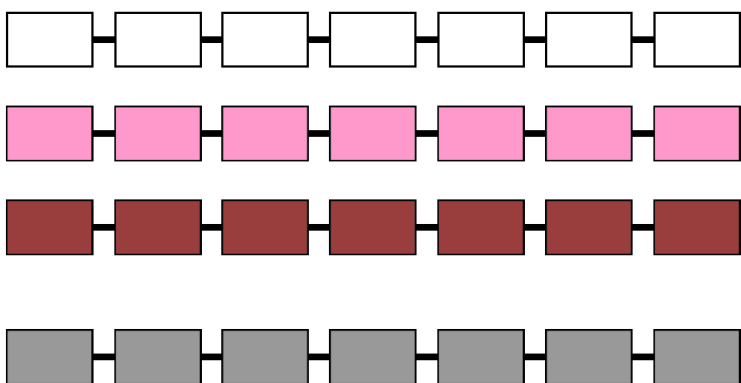
Theta

3-4 Hz, 1130 -1440 ms

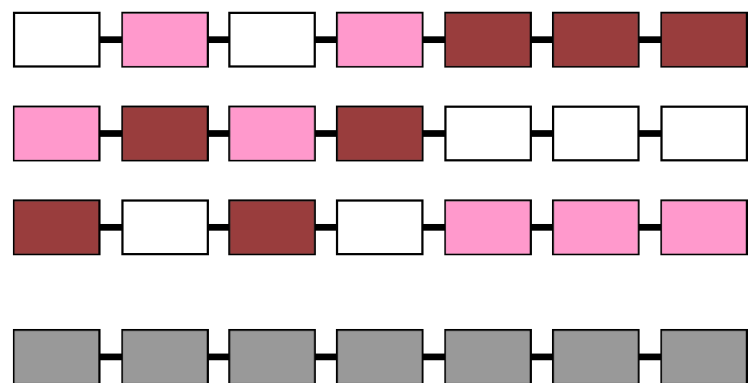


t-value

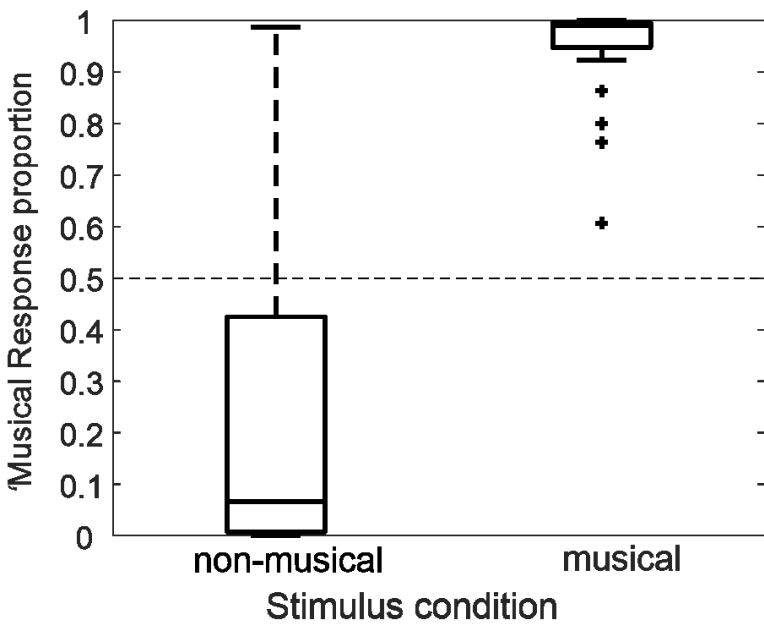
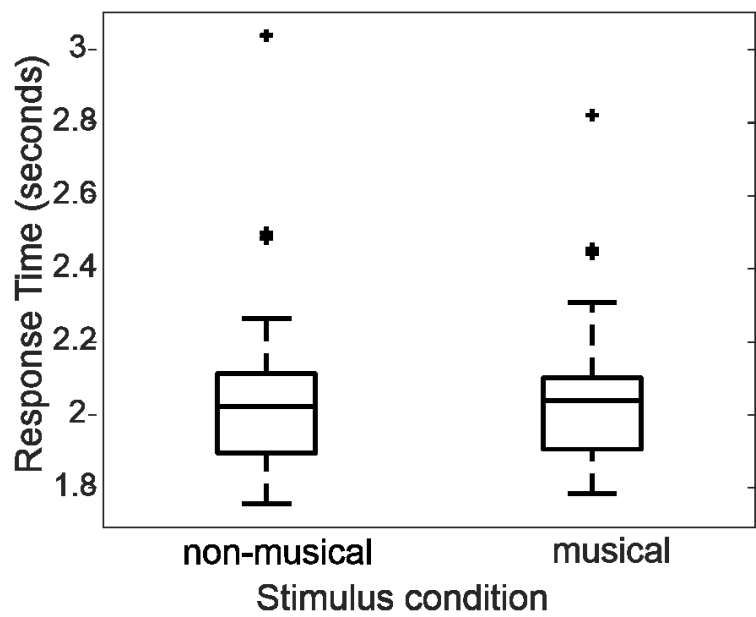
A non-musical

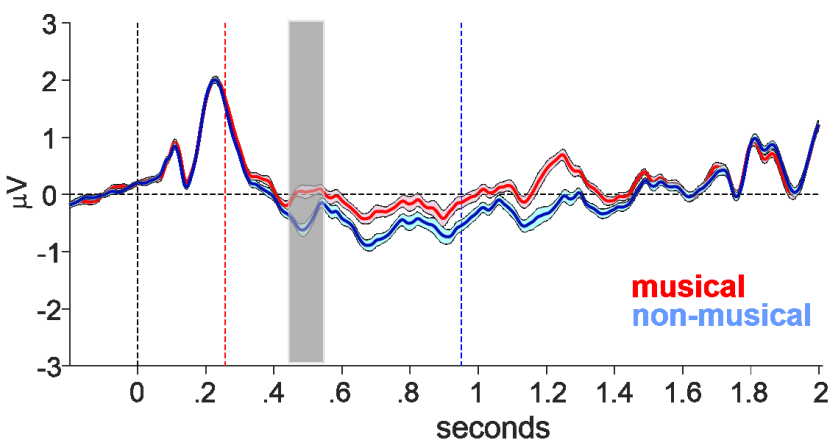
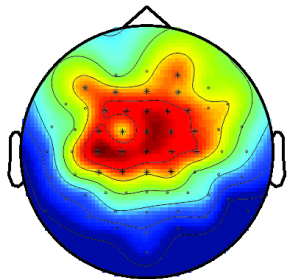
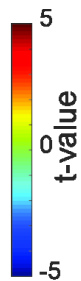
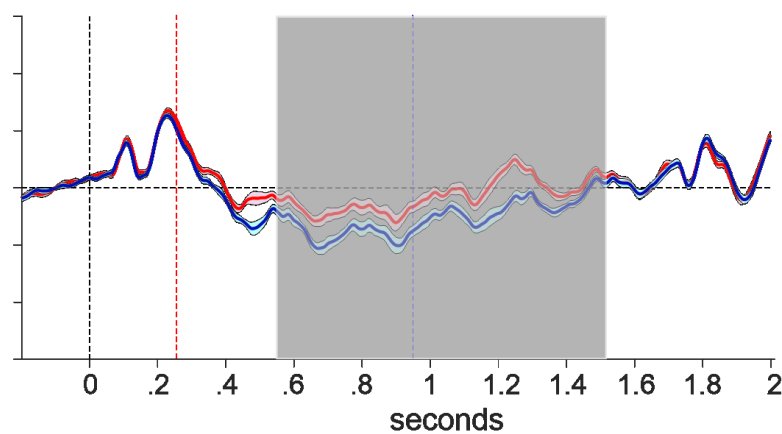
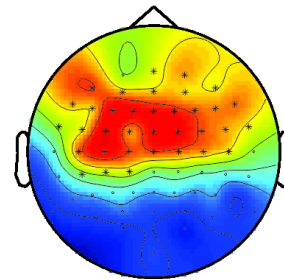
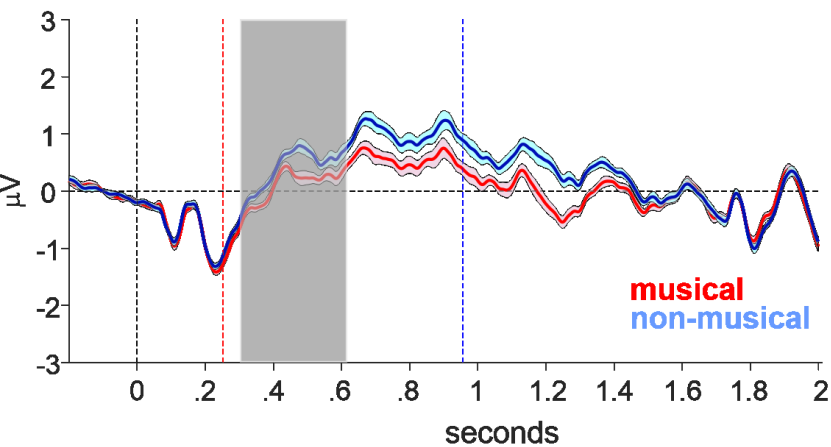
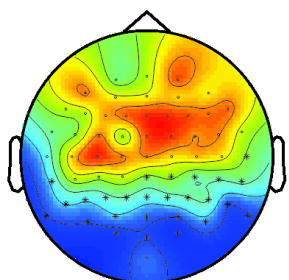
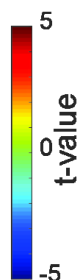
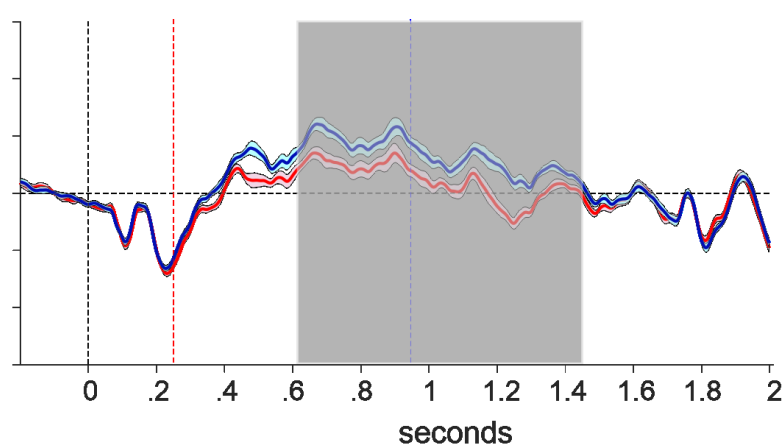
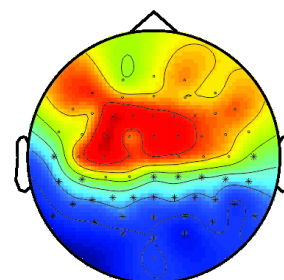


B musical



Brown noise  Pink noise  White noise 

A**B**

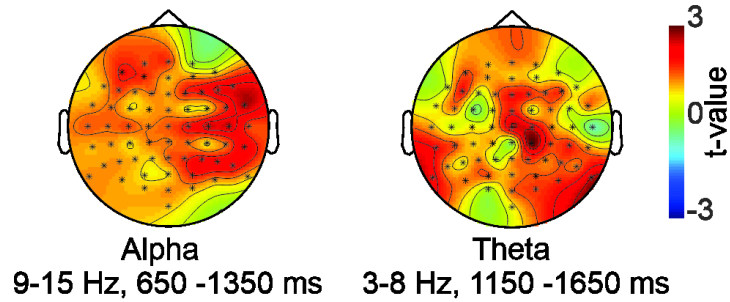
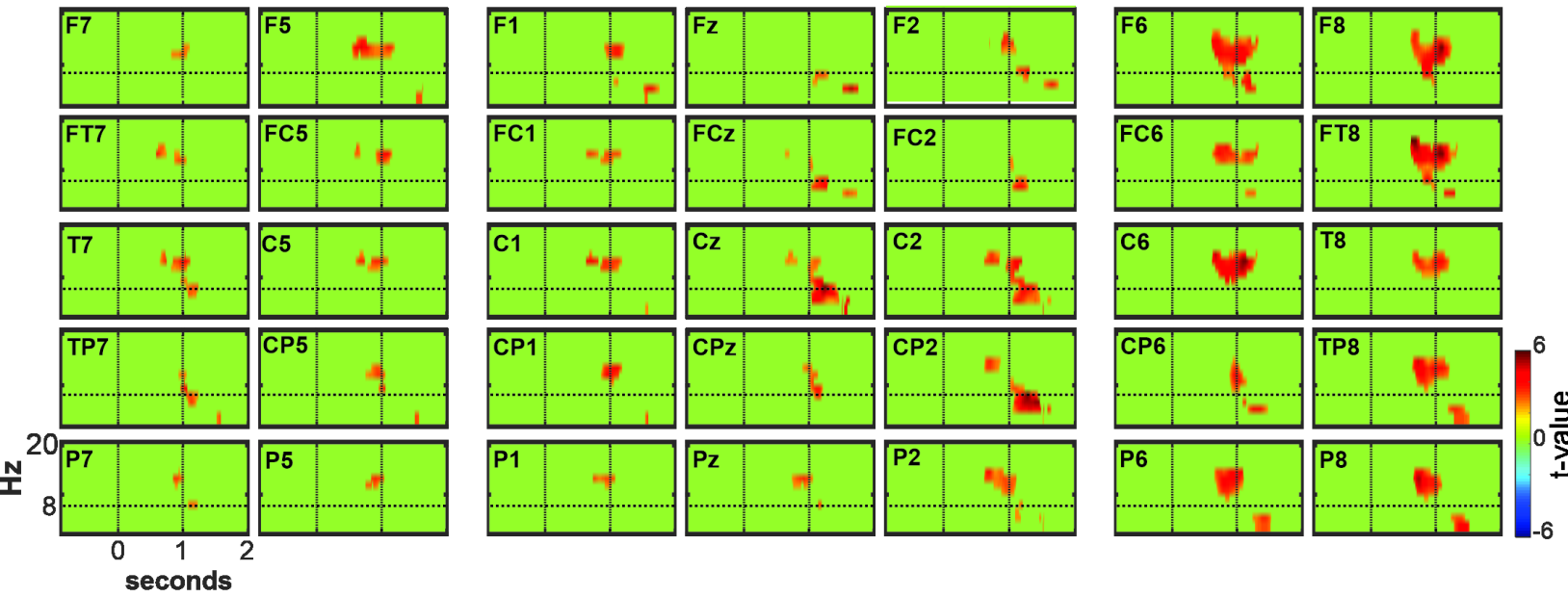
A**Cluster 1 ($p = 0.046$)****0.447s - 0.543s****musical vs non-musical****Cluster 2 ($p = 0.001$)****0.553s - 1.52s****B****Cluster 3 ($p = 0.016$)****0.309s - 0.613s****musical vs non-musical****Cluster 4 (0.001)****0.617s - 1.45s**

musical > non-musical (Spectral Power)

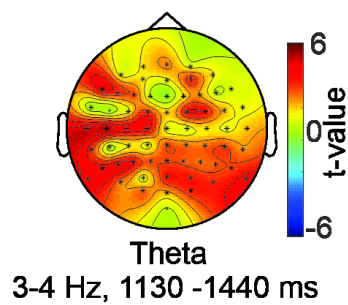
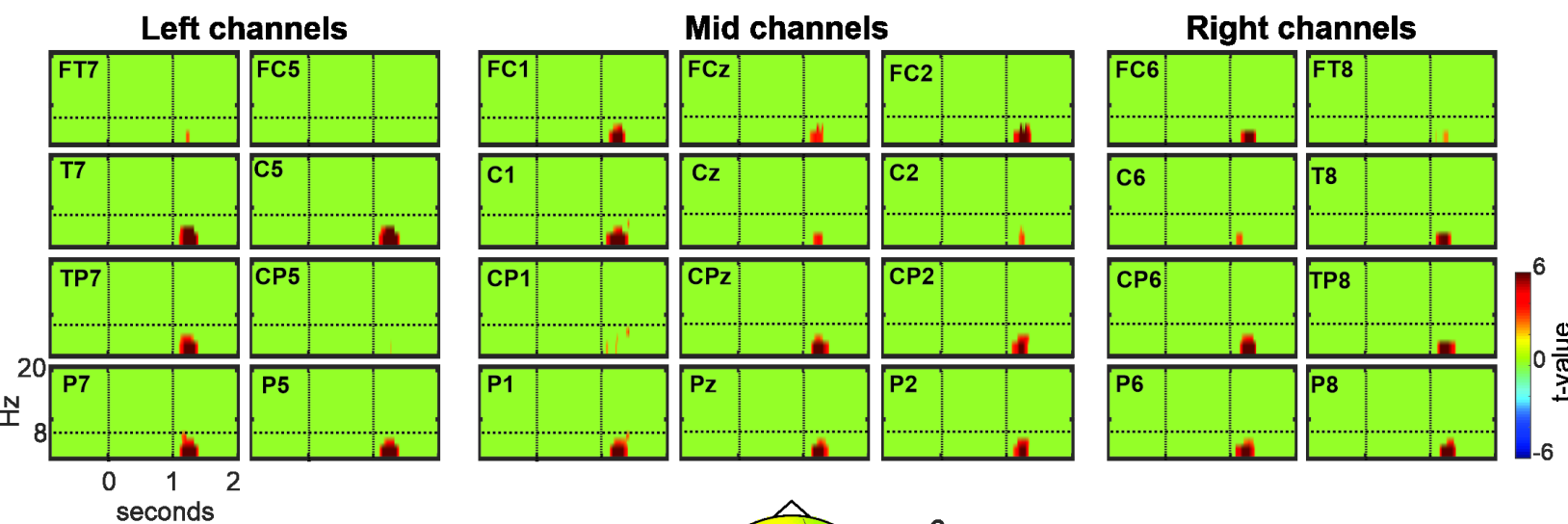
Left channels

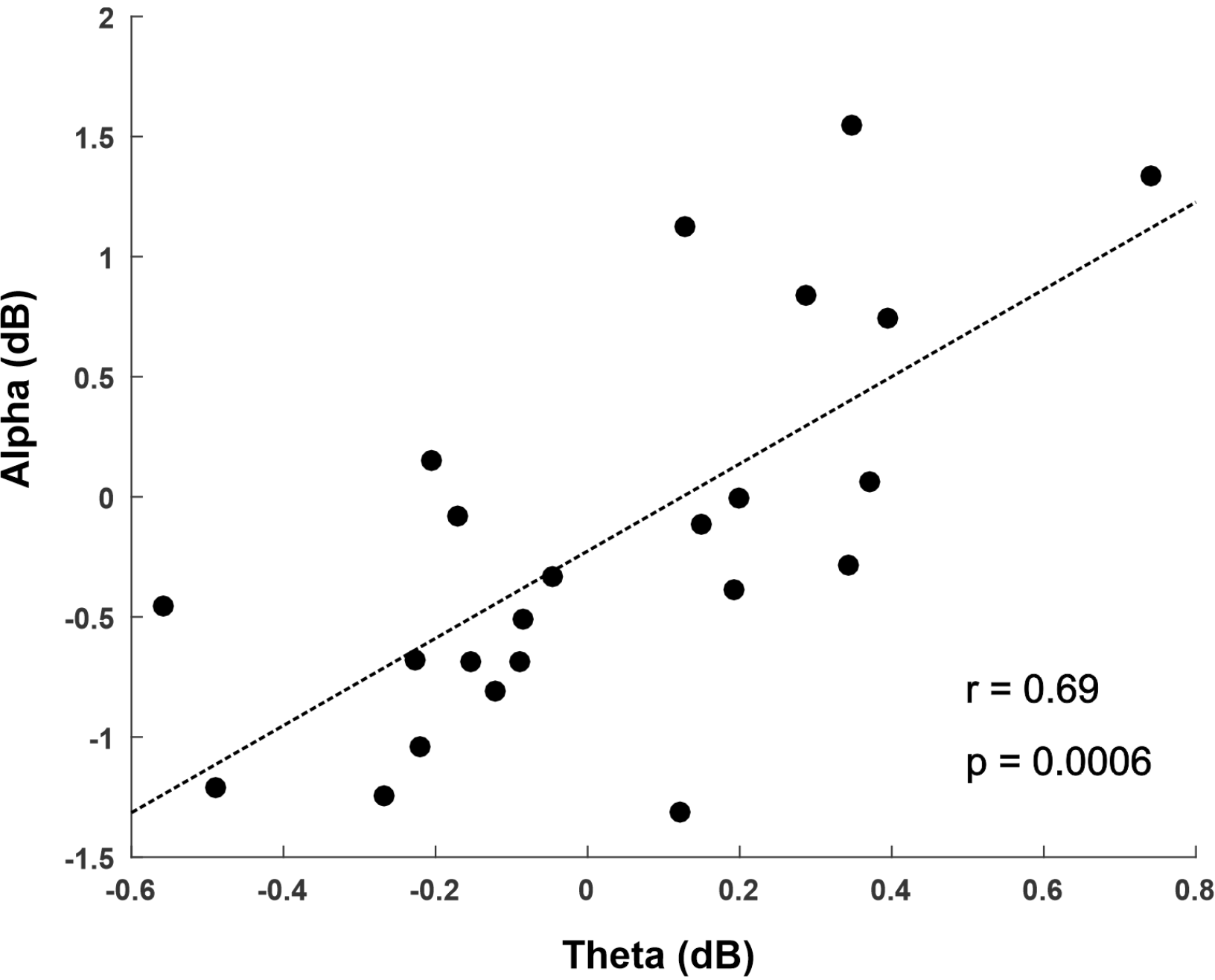
Mid channels

Right channels



musical > non-musical (Phase-locking)





Low-Theta Phase-Locking

Musicians

non-Musicians

