american physiological society
JNP JOURNAL OF NEUROPHYSIOLOGY

J Neurophysiol 130: 291–302, 2023.
First published June 28, 2023; doi:10.1152/jn.00042.2023

RESEARCH ARTICLE

Sensory Processing

Neurophysiological time course of timbre-induced music-like perception

Alejandra E. Santoyo,[1]* Mariel G. Gonzales,[1]* Zunaira J. Iqbal,[1]* Kristina C. Backer,[1,2] Ramesh Balasubramaniam,[1,2] Heather Bortfeld,[1,2,3] and Antoine J. Shahin[1,2]

[1]Department of Cognitive and Information Sciences, University of California, Merced, California, United States; [2]Health Sciences Research Institute, University of California, Merced, California, United States; and [3]Department of Psychology, University of California, Merced, California, United States

Abstract

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 (HG), ascending laterally to nonprimary 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 nonmusical sound-streams. The nonmusical 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 nonmusical. Musical processing exhibited right dominant α power enhancement, followed by a lateralized increase in θ phase-locking and spectral power. The θ phase-locking was stronger in musicians than in nonmusicians. 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 & NOTEWORTHY EEG induced by streams of pitchless noise segments varying in timbre were classified as music-like and exhibited a right-lateralized hierarchy in processing similar to pitched melodic processing. This study provides evidence 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 as in cochlear-implant listeners, as well as the role of nonpitched sounds in the induction of music-like perceptual states.

EEG; melody processing; oscillatory activity; pitch processing; timbre processing

INTRODUCTION

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

Musical perception is typically attributed to pitch, timbre, and loudness contours—that is, the pattern of changes in an auditory dimension (e.g., pitch, timbre, and loudness) over time (1, 2). A key perceptual attribute of timbre is brightness

---

*A. E. Santoyo, M. G. Gonzales, and Z. J. Iqbal contributed equally to this study.
Correspondence: A. J. Shahin (ashahin@ucmerced.edu).
Submitted 26 January 2023 / Revised 26 June 2023 / Accepted 26 June 2023

that is correlated to a sound's spectral centroid—the center of mass of any given stream of sound. Higher spectral centroids typically elicit the perception of brighter tones. Previous studies have drawn connections between the mechanisms among pitch, timbre, and amplitude contour discrimination. McDermott et al. (2) used a modified melody recognition task to investigate if listeners can discriminate between transpositions of brightness and loudness contours as easily as they can with pitch contours. They showed that while pitch is more accurately encoded than timbre and loudness, contours in loudness and brightness are also useful, though not to the same degree as pitch, in the recognition of familiar melodies. Similarly, Cousineau et al. (3) used varied pitch, brightness, and loudness sequences at different lengths (i.e., 1, 2, and 4 pure tones in the sequence) and tasked participants with indicating if varied-length pairs were the same or different. As sequence length increased, discrimination accuracy decreased significantly for the loudness sequences whereas discrimination accuracy decreased only marginally for pitch and timbre. Furthermore, it has been found that musicians perform significantly better at short-term recognition of variable-pitch sequences than nonmusicians, though there was no significant difference in performance between musicians and nonmusicians in discriminating constant-pitch sequences (4). A recent study by Graves et al. (1) on congenital amusia, a neurodevelopmental disorder in which individuals exhibit a deficit in musical perception, examined individuals with amusia performing a short-term memory task in which they had to extract pitch, brightness, and loudness contours using novel melodies. They found that the individuals with amusia displayed greater impairment in discriminating between pitch- and brightness-based melodies compared with loudness-based melodies, suggesting that pitch and timbre may be processed by a common locus in the brain.

Commonalities between pitch and timbre processing are further elucidated by pitch and timbral brightness-induced judgments of Shepard tones. Classic Shepard tones are generated by combining shifting tones with superimposed sine waves while maintaining a fixed spectral envelope, thus, producing the illusory perception of a sound with perpetually increasing or decreasing pitch. Siedenburg (5) compared the classic Shepard tone with a "pseudo-Shepard tone," created by varying the spectral envelope while maintaining the spectral fine structure (tone). This manipulation induced the illusory perception of shifts in timbral brightness analogous to the shifts in pitch induced by the classic Shepard tone. In a follow-up study (6), the same Shepard tone variations associated with pitch (shifting spectral fine structure, SFS) and timbral brightness (shifting spectral envelope, SE) were again compared in conditions where both SFS and SE shifts were applied simultaneously in either a synergistic or antagonistic manner (same or opposite directions). Behavioral results showed that the two direction changes were very rarely perceived concurrently and the stronger cue did not correlate with the component shifted (SFS or SE) but instead with the nature of the sound itself. That is, when the sound was harmonic, more weight was put on SFS, and when the sound was inharmonic, more weight was put on SE (6). These two studies promote a unitary account of pitch and timbre perception.

Behavioral accounts have been key in guiding neurophysiological research on musical encoding, which has demonstrated a hierarchy of sound processing. Simple sounds are largely processed in the brain stem and Heschl's gyrus (HG) (7, 8), but as the sound becomes more complex (e.g., richer in pitch dynamics), processing extends to anterolateral regions of the AC—superior and inferior temporal gyri (9–12). Patterson et al. (13) sought to understand the hierarchy of pitch processing by experimenting with regular-interval (RI) sounds, which are created by overlapping regularized time intervals of broadband noise (14). Manipulating the temporal regularity of the noise segments affects perception, with a stronger noise (no-pitch) perception as temporal regularity decreases and emergence of pitch-like perception as temporal regularity increases (13, 14). When RIs of several unique temporal intervals ("pitches") are sequentially combined in a stream, they sound melodic-like.

Patterson et al. (13) tracked neural engagement using functional magnetic resonance imaging (fMRI), while people processed no-pitch, fixed-pitch, and melody-like RIs. They observed the emergence of a hierarchical pattern of cortical networks. As RIs became more complex and melodic-like, areas of activation extended beyond HG and planum temporale (PT) to include the superior temporal gyrus (STG) and planum polare (PP) (13, see also Ref. 15). Moreover, right hemisphere activation was more robust relative to left in STG and PP, but not in HG and PT. Similarly, Hyde et al. (10) found that fixed pitch processing occurred bilaterally in HG, whereas melody processing activated regions in secondary auditory cortices, including PT and STG. Most noteworthy was their finding that the right PT exhibited a linear increase in activity corresponding to parametric increases in pitch distance; increases in activation in the left hemisphere were more tempered. Taken together, these studies indicate that, as sound increases in pitch complexity (e.g., is perceived as increasingly musical), there is a corresponding right-lateralized shift to higher-level processing within the hierarchy.

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

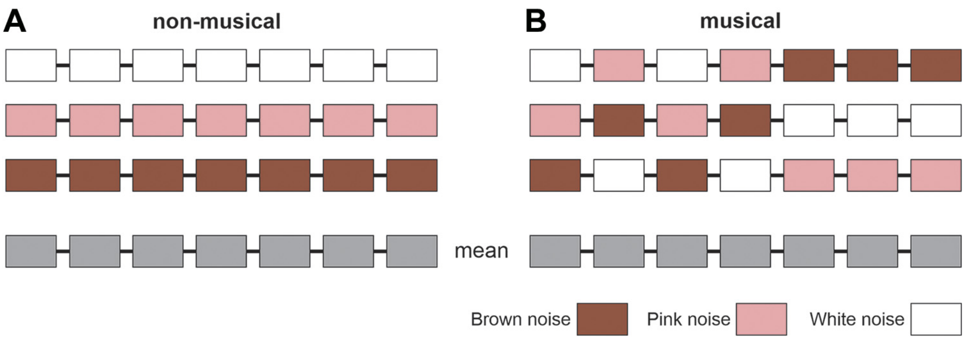

Figure 1. Experimental design. *A*: colored streams depicting white, pink, and brown noise streams of the nonmusical condition. Each stream was comprised of seven segments of the same type of noise. *B*: colored streams depicting white, pink, and brown noise streams of the musical condition. Each stream was comprised of seven segments of all three types of noise combined in a different order. Importantly, the mean across the three streams within each condition was the same across the two conditions.

perception of music on pitch cues. The only distinction between the two conditions was that the inclusion of all three noise types within each musical stream produced variations in timbre (i.e., spectral envelopes) as a function of time. We hypothesized that: *1*) stronger α desynchronization (suppression) should be observed for the musical than the nonmusical condition in lateral temporal channels, indexing stronger engagement of nonprimary AC networks during musical processing; *2*) α desynchronization should be more pronounced in the right hemisphere; *3*) the aforementioned two predictions should be more pronounced in musicians than nonmusicians, as musical encoding should be stronger in musicians (16–18), leading to a more robust lateralization and hierarchical shift.

METHODS

Participants

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

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

Stimuli

Two-hundred millisecond brown, pink, and white noise-bursts were created using Adobe Audition (stimuli are available at https://doi.org/10.6084/m9.figshare.21430614.v2). In total, there were seven noise bursts within each stream with 37.5 ms of silence between bursts, lasting a total duration of 1,625 ms. Each nonmusical stream ($n = 3$) comprised just one of the three types of noise (brown, pink, or white), and musical streams ($n = 3$) were created by combining all three noise types into three unique patterns (see Fig. 1 and *Experimental Design and Statistical Analyses* for details).

Procedure

EEG was recorded as participants were seated in a sound-attenuated booth ∼1 m away from a 27-in. monitor. Auditory stimuli were presented at ∼70 dBA sound pressure level (SPL) through two loudspeakers located to the left and the right of the monitor at a −/+45° angle, creating the perception that the sound came from the center (0°). The study consisted of 10 blocks of 60 trials each, with a single sound stream presented per trial. The nonmusical and musical stimuli were presented in a random order within each block. No training with the sounds took place before study participation. Participants were instructed to listen to the stimuli while fixating their gaze on a cross in the center of the monitor and indicate whether they perceived each stimulus as nonmusical or musical. A visual cue to respond was displayed immediately following the offset of each trial. Participants indicated their responses via keyboard presses, "n" for nonmusical and "m" for musical; they used their right hand to respond. No feedback was given regarding correct/incorrect responses. Presentation software (NeuroBehavioral Systems Inc., Berkeley, CA) was used to present the stimuli and record responses. EEG was recorded using a high-impedance 64-channel Biosemi System (Active Two system, 10–20 Ag-AgCl electrode, with Common Mode Sense and Driven Right Leg passive electrodes serving as grounds, A/D rate: 1,024 Hz).

Data Analysis

Behavior.

Individual response logfiles were parsed using in-house MATLAB (The MathWorks Inc., Natick, MA) code, producing reaction times (RTs) and classification proportions (musical, nonmusical). For each stimulus condition (musical, nonmusical), we calculated the proportion of musical responses relative to the total stimulus condition trials. There was no need to compute the nonmusical proportions because the total proportions of musical and nonmusical responses across the two conditions were equal to 1. RT was measured from onset of sound until a button press response.

EEG.

EEG analyses were conducted using ERPLAB (19), EEGLAB (20), and FieldTrip (21) toolboxes and in-house MATLAB code. Processing steps included: *1*) Individual continuous EEG blocks were downsampled to 512 Hz and concatenated into one continuous data set for each subject. *2*) Each continuous data set was epoched (segmented) from −1.25 to 2.5 s around the onset of each noise stream and baselined to the entire epoch (mean removed). *3*) Independent component analysis (ICA) was conducted on each individual data set, producing 64 ICA components. *4*) ICA components that reflected ocular artifacts were removed (mean of 2 per subject) and bad channels (maximum of 2) were interpolated using EEGLAB's spherical interpolation. *5*) Individual datasets were average referenced. Datasets were further processed in two ways: to produce files suitable for auditory-evoked potentials (AEPs) and files suitable for Oscillatory activity.

AEPs. Individual files were filtered between 0.1 and 30 Hz (zero-phase Butterworth, fourth order), re-epoched around acoustic onsets from −0.2 to 2 s, linearly detrended to remove the slow shift, and re-baselined to the prestimulus period (−0.2 to 0 s). Then, trials with amplitude shifts greater than ±150 μV at any channel were removed. The mean number of remaining trials for the musical and nonmusical conditions was 279 for each condition. Because initial analyses revealed an overwhelming musical response for the musical stimuli and nonmusical response for nonmusical stimuli (see Fig. 2), we segregated the conditions based on stimulus type and not percept type. Doing so maintained an equal number of trials and equal acoustic characteristics between the two conditions. Maintaining an equal number of trials is essential for the analysis of intertrial phase coherence (ITPC), as ITPC values are influenced by trial number (22). Finally, trials within each stimulus condition (musical, nonmusical), were averaged, creating separate AEP files for each condition.

Oscillatory activity. Individual files with amplitude shifts larger than 200 μV at any channel were rejected and trials were separated into two conditions (musical and nonmusical). The remaining mean number of trials for the musical and nonmusical conditions were 271 and 270, respectively. Event-related spectral perturbation (ERSP, spectral power) and ITPC spectrograms were generated for each condition, channel, and participant using the timef.m function of the EEGLAB toolbox. We examined activity for the period of −1,250 to 2,500 ms relative to noise-stream onset for the 2.7–50 Hz frequency range. The analysis used a sinusoidal wavelet-based discrete Fourier transform (DFT) of the time-domain signal (750-ms Hanning window, 12 ms and 1 Hz steps) with two cycles at the lowest frequency, increasing linearly to 37 cycles at the highest frequency. For the ERSP analysis, poststimulus activity was baselined to the preacoustic onset in the period spanning from −1,250 to −375 ms.

Experimental Design and Statistical Analyses

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

The statistical analyses comparing reaction time (RT) or classification between the musical and nonmusical conditions and the interaction with group (musicians, nonmusicians) were based on analysis of variance (ANOVA). *t* Tests (corrected for multiple comparisons using the Bonferroni test) were further used to compare classification values relative to chance level. Nonresponses were excluded (only for behavioral analysis).

EEG activity (AEP amplitude, spectral power, or phase-locking) differences between conditions were examined using the nonparametric cluster-based permutation test (CBPT; 23) as implemented in the FieldTrip toolbox (21). The CBPT takes into account the activity at all time-points and channels. For the AEP amplitude data, the CBPT organizes significantly different EEG amplitudes (based on dependent-samples *t* tests between the musical and nonmusical conditions done at each data sample, $\alpha = 0.05$) into continuous clusters according to the significant samples' spatial adjacency and temporal adjacency. The *t* statistics at each time-channel sample within each cluster were then summed. To

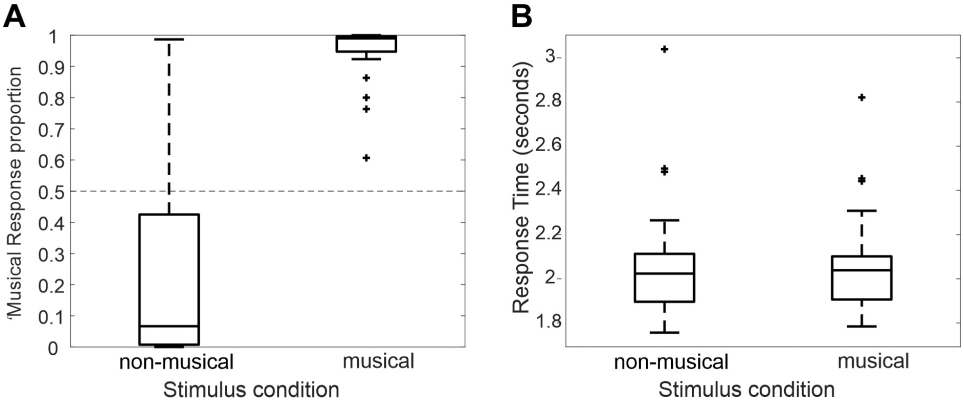

Figure 2. *A*: musical response proportions for the musical and nonmusical stimulus conditions. *B*: reaction time (RT) for musical and nonmusical stimulus conditions.

create the null distribution, these steps were repeated for each resampling of the data (2,000 permutations; Monte Carlo simulation). For each permutation, the largest cluster t statistic sum was recorded. Finally, the summed t statistics for each cluster in the real data were compared against this null distribution of maximal cluster t statistic sums, using a cluster-level α value of 0.05. Any clusters that surpass this α value are considered to contain activity that is significantly different between the musical and nonmusical conditions.

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

For post hoc analysis, and to test for interaction of oscillatory activity and group (musicians vs. nonmusicians), we used ANOVAs. Correlation results were based on Pearson correlation coefficient (PCC) method, corrected for multiple comparisons using the Bonferroni test.

RESULTS

Behavior

Figure 2*A* depicts the classification proportions of the musical response observed for the musical and nonmusical stimulus conditions. We limited our ANOVA to the musical classification for the two conditions. An ANOVA with the variables group and condition (musical, nonmusical) revealed that there was only a main effect of condition ($F_{1,21} = 87.3$, $P = 0.0001$; $\eta^2 = 0.81$) with no other main effect or interaction between the variables ($F < 1$). The condition effect was due to a strong labeling of the musical stimulus as musical and a weak labeling of the nonmusical stimulus as musical.

The musical classification within each stimulus condition was further subjected to a one-sample t test against chance level (50%). Both tests revealed that the musical classification was above chance within the musical condition ($t_{22} = 2{,}403$, $P = 0.00001$; mean musical = 0.944, 95% CI = 0.90–0.99) and below chance within the nonmusical condition ($t_{22} = 739$, $P = 0.000001$; mean nonmusical = 0.23, 95% CI = 0.091–0.37). Although these effects seem obvious, this labeling reinforces the notion that irregular noise patterns (i.e., irregular timbre patterns) can be perceived as musical despite participants' lack of familiarity with the stimuli. Moreover, the lack of a group effect demonstrates that there was no evidence of a musicianship advantage in identifying a musical stimulus as musical and vice versa for the nonmusical stimuli.

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

EEG

AEPs.

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

Our AEP results can be summarized as follows: First, the waveforms exhibited eight distinct transient (steady state) potentials: an initial onset AEP response (typical P1-N1-P2 morphology), followed by six peaks and lastly followed by an offset response (typical P1-N1-P2 morphology but with attenuated amplitude). These AEPs clearly reflect the onsets/offsets of the seven noise bursts, i.e., noise-stream envelope. Second, the musical and nonmusical AEP waveforms begin to diverge around 325 ms, whereby the musical waveform weakens in amplitude (becomes closer to zero for both the positive and negative clusters). The latency of divergence is ∼80–90 ms after the onset of the second noise burst of the noise stream; the onset of the second noise burst is depicted by a red dotted vertical line within the plots shown in Fig. 3. This is the first instance when the musical stream can be perceived as irregular relative to the nonmusical stream because the second noise-burst in the musical streams has a different timbre than that of the first noise-burst. The divergence between the waveforms of the two conditions continues until after 1,400 ms. However, toward the end of this divergence (∼1,200–1,300 ms), there is a short period in which the musical AEP's amplitude becomes greater (moves away from 0 µV) than the nonmusical AEP amplitude (which moves toward 0 µV). This short period occurs ∼250 ms after the onset of the fifth noise burst (depicted by the blue dotted vertical line in Fig. 3) in the sound stream. This is also around the same time as the onset of the sixth noise-burst in the sound stream. Third, the topography (frontocentral and temporo-occipital) of these AEP differences is reflective of sources within and surrounding HG. In short, differences in musical and nonmusical AEPs can be observed following the onset of the first irregularity in the noise stream. Musical AEPs are continually diminished until after the onset of the 5th/6th segment(s) in the noise stream in which they eventually rebound.

Musicianship. Post hoc ANOVA on the mean AEP values (between 350 ms and 1,350 ms) at the frontocentral channels and within the time-points reaching significance did not reveal an interaction between condition and group ($F_{1,21} = 3.1$, $P = 0.09$). Thus, musicianship was not a factor in AEP differences. Despite the lack of significance, it is worth noting that musicians tended to exhibit weaker AEPs than nonmusicians especially for the musical condition. If we contend that AEPs reflect low-level auditory processing, then this would suggest

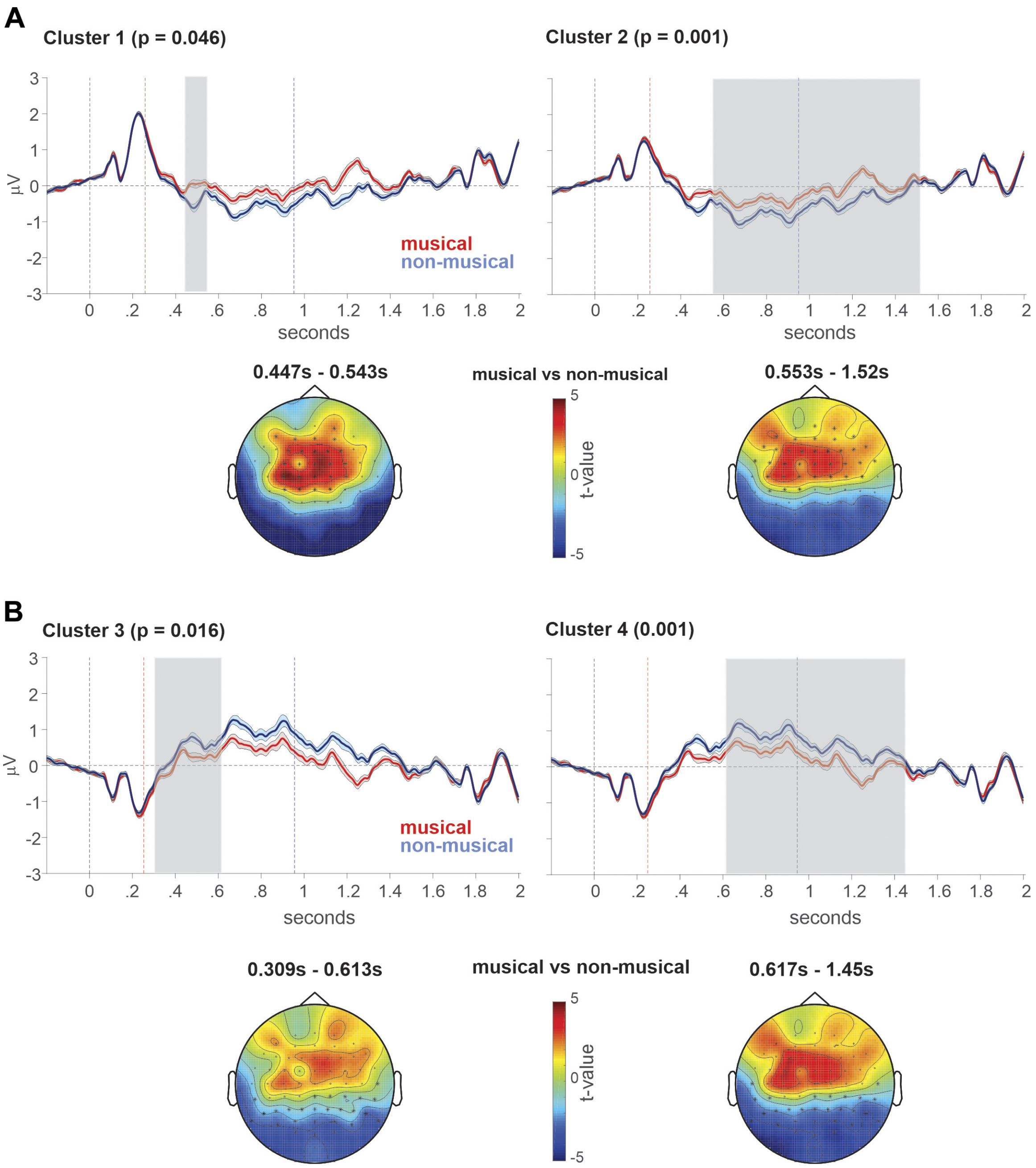

Figure 3. Auditory-evoked potential (AEP) waveforms of significant clusters [*n* =4, 2 positive (*A*) and 2 negative (*B*)]. The gray shaded areas represent the significant time windows distinguishing the musical and nonmusical AEPs. The negative and positive clusters represent the opposite poles of the same neural sources. Below are the corresponding *t* value topographies reflecting differences between the musical from nonmusical conditions. The topographies are consistent with sources that are mainly originating within and/or surrounding Heschl's gyrus (HG). The black dotted vertical line indexes sound stream onset (0 ms). The red dotted vertical line indicates the onset of the second noise segment in the stream. The blue dotted vertical line indicates the onset of the fifth segment in the stream.

that musicians shift their processing to higher auditory levels during processing more so than nonmusicians.

Oscillatory activity.

Spectral power. The CBPT revealed one positive cluster ($P = 0.026$), indicating significantly greater spectral power for the musical condition than nonmusical condition, which was observed in all 64 channels. Figure 4 (*top*) reveals the *t* value spectrograms of spectral power distinguishing musical versus nonmusical for a subset of left, mid, and right channels. Figure 4 (*bottom*) shows the *t* value topographies of α (9–15 Hz, ∼650–1,350 ms) and θ (3–8 Hz ∼1,150–1,650 ms) oscillatory activities within the window of significance for these frequencies. Our findings demonstrate the following: First, both θ and α activities were greater for the musical condition than the nonmusical condition. Second, α was concentrated more in temporal sites, especially in the right

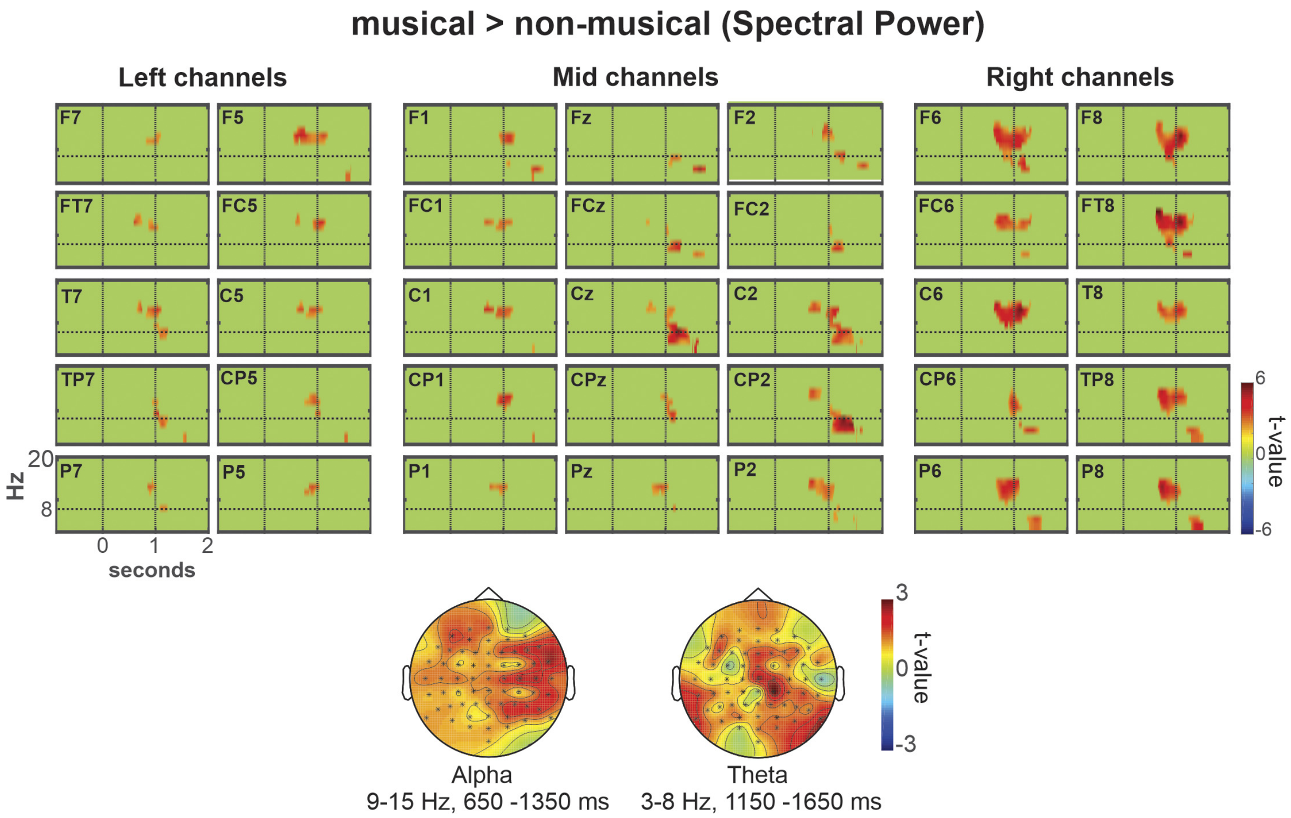

Figure 4. *Top*: *t* value spectrograms indexing significant spectral power differences between the musical and nonmusical conditions. *Bottom*: *t* value topographies reflecting significant spectral power (θ and α) differences between the musical and nonmusical conditions.

hemisphere. Third, θ was concentrated frontocentrally and temporoparietally, with right hemispheric dominance that was mainly evident in the low-θ (3–4 Hz) spectral power. Theta's topography is consistent with generators within or surrounding HG. Fourth, the different topographies of θ and α strongly suggest that θ and α have different neural origins. Fifth, while θ and α overlapped in time, α (maximally exhibited at temporal sites) began earlier and θ (maximally exhibited at frontocentral and temporal sites) ended later.

Phase-locking. As in the spectral power analysis, the phase-locking analysis produced one scalp-wide positive cluster ($P = 0.001$), indicating that the musical condition resulted in stronger phase-locking than the nonmusical condition. Figure 5 (*top*) depicts the *t* value spectrograms distinguishing phase-locking of musical versus nonmusical for a subset of left, mid, and right channels. There was scalp-wide activity of low θ (3–4 Hz) occurring around 1,130–1,440 ms. The *t* value topography of this activity—lateral central and bilateral temporoparietal—may suggest auditory generators that are more lateral from what we expect from HG generators (as reported earlier). Thus, the origins of this θ phase-locking activity is uncertain; they more likely reflect generators outside of HG, e.g., within STG.

Correlations. We examined the degree of correlation between θ and α 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 θ and α power regardless of condition (averaged across the musical and nonmusical conditions ($r = 0.69$; $P = 0.0006$)). Separately, correlation for either condition was highly significant (musical: $r = 0.67$, $P = 0.001$; nonmusical: $r = 0.7$, $P = 0.0005$).

Musicianship. Post hoc ANOVA on the θ and α 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 θ 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 θ phase-locking occurring in musicians for the musical versus nonmusical condition ($P = 0.000001$; Bonferroni test) relative to nonmusicians who showed a weaker but still significant difference in phase-locking for the musical versus nonmusical condition ($P = 0.0036$; Bonferroni test) (Fig. 7).

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

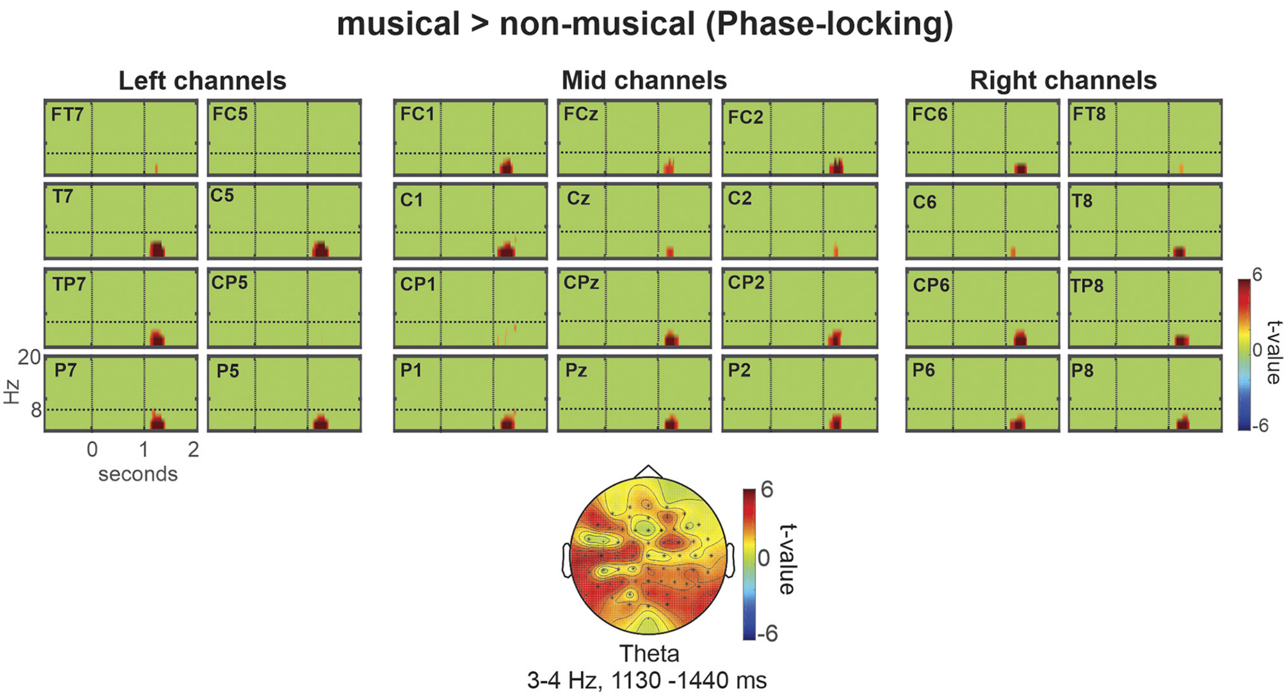

Figure 5. *Top*: *t* value spectrograms indexing significant phase-locking differences between the musical and nonmusical conditions. *Bottom*: *t* value topography reflecting significant low-θ phase-locking differences between the musical and nonmusical conditions.

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

Recent work shows that humans can experience musicality of irregular streams of sounds where pitch is not explicitly varied and, in some cases, not present, as in beatboxing (24). Herein, we use a common definition of musicality "the quality of having a pleasant sound; melodiousness" (Oxford Languages, https://languages.oup.com/google-dictionary-en/). However, for more nuanced definition see Ref. 25. Because beatboxing involves sequences of sounds that sometimes lack pitch (e.g., unvoiced percussion sounds), beatboxers are thus expected to be more attuned to streams of sound that are irregular but contain no pitch or no variations in pitch. Interestingly, recent work has shown that beatboxers produce different responses to acoustic variations in pitched sounds, compared with musicians trained in traditional musical instruments (26), suggesting the existence of unique individual differences in processing sounds and music that is dependent on an individual's musical experience. Here, we ask whether sounds that do not involve pitch but produce timbre irregularities can elicit similar brain response in "traditional"

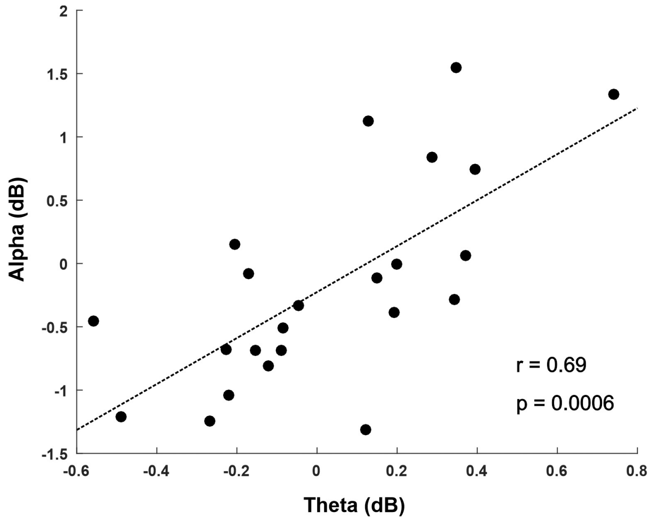

Figure 6. Scatterplot of α power as a function of θ power, observed within the significant windows, distinguishing the musical and nonmusical conditions.

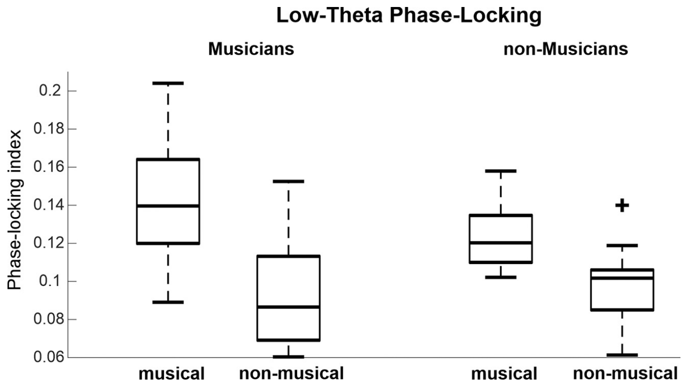

Figure 7. θ Phase-locking, observed within the significant windows, distinguishing the musical and nonmusical conditions in musicians and nonmusicians.

musical sounds that do contain pitch. Our results show that the processing of irregular pitchless noise streams perceived as musical produces an increase in α power with right hemisphere dominance, followed by an increase in θ phase-locking and spectral power. Because EEG does not allow definitive localization of θ and α neural generators, our interpretation emphasizes the time course of neural activity while limiting information on neural origins to basic scientific deduction.

The θ spectral power scalp topographies exhibited frontocentral activity with posterior temporo-occipital activity that resembles processing/activity typically generated within and surrounding HG, as with AEPs (Fig. 3). HG generators, as opposed to higher level (e.g., nonprimary auditory cortex or nPAC) generators, are particularly involved in processing simple sound features such as sound onsets and offsets (27, 28), leading to a stronger temporal alignment of the AC activity with the noise-stream envelope. Thus, θ activity, including AEPs, more likely reflects the engagement of low-level auditory networks (e.g., HG) indexing basic sound features. Second, the right-lateralized α enhancement (i.e., synchronization) is consistent with a gyral origin, such as within STG in nPAC. We know that α enhancement is associated with neural inhibition, whereas α suppression (desynchronization) indicates neural excitation (29–37, 55). We interpret this α enhancement as reflecting inhibitory mechanisms in nPAC.

At first glance, the α result seems to contradict our hypothesis, i.e., we expected excitation of nPAC during musical processing. However, a well-established account linking α enhancement and item retention in working memory is well suited to explain our findings. Prior studies have demonstrated that an increase in item retention in working memory coincides with an increase in α power (e.g., see Refs. 29, 38–41). For example, Jensen et al. (29) used a modified Sternberg task, in which individuals were required to retain 2, 4, or 6 consonants presented simultaneously on a screen, and judge whether a subsequent probe matched one of the items in working memory. During the retention period, α power (9–12 Hz) systematically increased with an increase of the number of items retained in working memory. Jensen and coworkers interpreted this α increase as signifying the brain's effort to inhibit encoding of irrelevant information within the neural population associated with retaining items in working memory. Furthermore, Andrillon et al. (42) demonstrated that repeated exposure to noise patterns leads to the perception of acoustic uniqueness over time as implicit memory is formed in listeners tasked with detecting sequence repetitions. This neural adaptation was marked by the formation of AEPs (N1-P2) time-locked to when noise patterns were later recognized behaviorally. Likewise, the aforementioned AEPs were also elicited when listeners were distracted by an auditory task that prevented them from attending to the noise sequences. These findings demonstrate that short-term neural plasticity can give rise to perceptual sensitivity to subtle acoustic cues in pitchless sounds with repeated exposure with or without focused attention. Similarly, we posit that the current noise sequential encoding is reflected in auditory short-term memory. Our sound streams consisted of increasing numbers of items (7 items maximum) unfolding over time, requiring serial maintenance in auditory short-term memory for a subsequent judgment of whether the sum of the items (noise bursts) constituted a musical sequence.

An interesting outcome of the current study is the group distinction. Despite the musical noise stream being an unconventional form of music, i.e., comprised of noise, musical streams evoked stronger low-θ phase-locking than nonmusical toward the end of the stream (following identification of the noise stream as a musical), especially in musicians. We know from previous accounts that the degree of θ phase-locking reflects musicality (43, 44). In their thoughtful study, Vanden Bosch der Nederlanden et al. (44) demonstrated that neural entrainment to syllable rate in linguistic utterances in speech and song led to better θ phase-locking to phrasal and syllabic information during sung speech. Thus, it is reasonable to conclude that phase-locking, indexing musicality, would be more prominent in musicians than in nonmusicians. Also, work from our laboratory and others demonstrated enhancement of P1, N1, and P2 AEPs (evoked by music sounds) in musicians relative to nonmusicians (16–18, 45–51). The N1-P2 AEPs are phase-locked in the θ band, whereas the P1 is reflected in higher frequency phase-locking, such as in the β (15–30 Hz) and γ (>30 Hz) evoked response (46, 52). However, a puzzling aspect of the current θ phase-locking is its strong lateral topography. The topography is not consistent with typical HG generators (e.g., fronto-central as in AEPs); rather they are more consistent with gyral (higher-level) nonprimary auditory generators. A

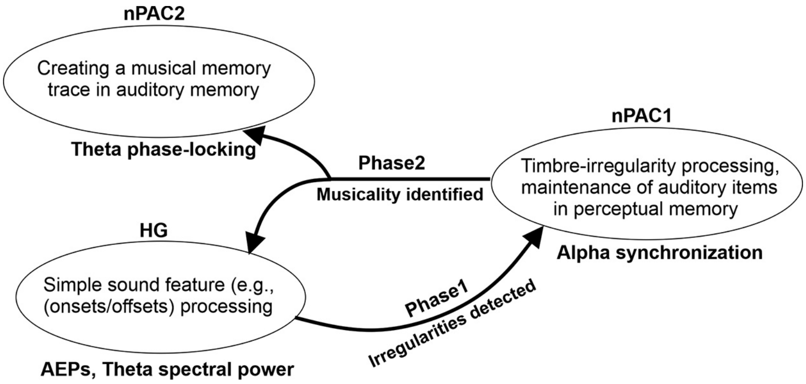

Figure 8. Theoretical framework conveying the temporal dynamics between low- and high-level auditory networks as a function of stimulus complexity and perceptual outcome. AEP, auditory-evoked potential; HG, Heschl's gyrus; nPAC, nonprimary auditory cortex.

potential explanation is that the θ phase-locking represents an encoding phase in auditory short-term memory. Evidence suggests that θ phase-locking is associated with memory encoding, especially across inferior frontal and hippocampal networks (53), including during sequence learning (54), as in the current study. Thus, we may amend our interpretation to state that the α enhancement (inhibition of irrelevant items in auditory memory) in nPAC, is followed by phase-locking in auditory memory networks, reflecting the encoding of the musical trace once the stream is identified as "music-like."

Given these findings, we propose the following theoretical framework (Fig. 8) regarding the neural time-course of melodic processing, which is largely facilitated by the EEG's superior temporal resolution. Initially, both the musical and nonmusical streams activate the auditory cortex, including primary (i.e., HG) and nonprimary networks (e.g., STG). Following the second item in the stream, irregularity in the musical stream is detected (∼ 250 ms), whereupon processing is shifted toward nPAC. As a result, we observe a reduction in the AEP amplitude of musical condition relative to the nonmusical condition following the onset of the second item, because processing is shifted laterally to nPAC. In nPAC, maintenance or retention of items (individual noise-bursts) begins, leading to enhancement of α during the retention period (showing significant activity between ∼ 650 ms and 1,350 ms). Once the decision is made, i.e., musical is identified, memory retention ceases, resetting θ phase as the memory trace is encoded in auditory memory, followed by a shift in processing back to HG. During this time period (1,130–1,650 ms), a θ power increase, coinciding with a brief AEP rebound, occurs for musical relative to nonmusical stimuli, signaling a processing shift back to lower levels (e.g., HG). The strong correlation observed between the earlier α and later θ powers further support the premise of this framework. Specifically, once nPAC concludes processing the musical stream (e.g., the stream was identified as musical), it signals HG to resume processing of subsequent incoming streams.

Limitations

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

Conclusions

The current findings extend our understanding of the mechanisms underlying music perception by targeting less common forms of music, pitchless music-like sequences. Despite this deviation from conventional music identity, the neural mechanisms identified herein support a similar dynamic observed with conventional music. Our results have implications for understanding the evolution of sound perception and music processing. This is particularly relevant in understanding music processing as it relates to the growing population of individuals with auditory prosthetics (e.g., cochlear implants), who have a degraded sense of pitch, and the introduction or reintroduction of more basic forms of music (e.g., beatboxing), that historically remained latent.

DATA AVAILABILITY

Stimuli and data can be accessed at https://doi.org/10.6084/m9.figshare.21430614.v2.

ACKNOWLEDGMENTS

We thank Dr. Laurel Trainor for valuable input on the study.

GRANTS

This work was supported by the National Institute on Deafness and Other Communications Disorders Grant R01 DC-013543 (to A.J.S.).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

A.E.S. and A.J.S. conceived and designed research; A.E.S., M.G.G., Z.J.I., K.C.B., and R.B. performed experiments; A.E.S., M.G.G., Z.J.I., K.C.B., R.B., and A.J.S. analyzed data; A.E.S. and A.J.S. interpreted results of experiments; A.E.S., M.G.G., Z.J.I., K.C.B., R.B., and A.J.S. prepared figures; A.E.S., M.G.G., Z.J.I., K.C.B., R.B., and A.J.S. drafted manuscript; A.E.S., K.C.B., R.B., H.B., and A.J.S. edited and revised manuscript; A.E.S., M.G.G., Z.J.I., K.C.B., R.B., H.B., and A.J.S. approved final version of manuscript.

REFERENCES

1. **Graves JE**, **Pralus A**, **Fornoni L**, **Oxenham AJ**, **Caclin A**, **Tillmann B.** Short-and long-term memory for pitch and non-pitch contours: insights from congenital amusia. *Brain Cogn* 136: 103614, 2019. doi:10.1016/j.bandc.2019.103614.
2. **McDermott JH**, **Lehr AJ**, **Oxenham AJ.** Is relative pitch specific to pitch? *Psychol Sci* 19: 1263–1271, 2008. doi:10.1111/j.1467-9280.2008.02235.x.
3. **Cousineau M**, **Carcagno S**, **Demany L**, **Pressnitzer D.** What is a melody? On the relationship between pitch and brightness of timbre. *Front Syst Neurosci* 7: 127, 2013. doi:10.3389/fnsys.2013.00127.
4. **Siedenburg K**, **McAdams S.** Short-term recognition of timbre sequences: music training, pitch variability, and timbral similarity. *Music Perception* 36: 24–39, 2018. doi:10.1525/mp.2018.36.1.24.
5. **Siedenburg K.** Timbral Shepard-illusion reveals perceptual ambiguity and con- text sensitivity of brightness perception. *J Acoust Soc Am* 143: EL93–EL98, 2018. doi:10.1121/1.5022983.
6. **Siedenburg K**, **Graves J**, **Pressnitzer D.** A unitary model of auditory frequency change perception. *PLoS Comput Biol* 19: e1010307, 2023. doi:10.1371/journal.pcbi.1010307.

7. **Griffiths TD**, **Uppenkamp S**, **Johnsrude I**, **Josephs O**, **Patterson RD.** Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci* 4: 633–637, 2001. doi:10.1038/88459.
8. **Griffiths TD**, **Hall DA.** Mapping pitch representation in neural ensembles with fMRI. *J Neurosci* 32: 13343–13347, 2012. doi:10.1523/JNEUROSCI.3813-12.2012.
9. **Hackett TA**, **Preuss TM**, **Kaas JH.** Architectonic Identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441: 197–222, 2001. doi:10.1002/cne.1407.
10. **Hyde KL**, **Peretz I**, **Zatorre RJ.** Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46: 632–639, 2008. doi:10.1016/j.neuropsychologia.2007.09.004.
11. **Tsai CG**, **Li CW.** Is it speech or song? Effect of melody priming on pitch perception of modified Mandarin speech. *Brain Sci* 9: 286, 2019. doi:10.3390/brainsci9100286.
12. **Zatorre RJ**, **Belin P**, **Penhune VB.** Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6: 37–46, 2002. doi:10.1016/s1364-6613(00)01816-7.
13. **Patterson RD**, **Uppenkamp S**, **Johnsrude IS**, **Griffiths TD.** The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36: 767–776, 2002. doi:10.1016/s0896-6273(02)01060-7.
14. **Yost WA**, **Patterson R**, **Sheft S.** The role of the envelope in processing iterated rippled noise. *J Acoust Soc Am* 104: 2349–2361, 1998. doi:10.1121/1.423746.
15. **Puschmann S**, **Uppenkamp S**, **Kollmeier B**, **Thiel CM.** Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage* 49: 1641–1649, 2010. doi:10.1016/j.neuroimage.2009.09.045.
16. **Schneider P**, **Sluming V**, **Roberts N**, **Bleeck S**, **Rupp A.** Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann NY Acad Sci* 1060: 387–394, 2005. doi:10.1196/annals.1360.033.
17. **Schneider P**, **Groß C**, **Bernhofs V**, **Christiner M**, **Benner J**, **Turker S**, **Zeidler BM**, **Seither-Preisler A.** Short-term plasticity of neuro-auditory processing induced by musical active listening training. *Ann NY Acad Sci* 1517: 176–190, 2022. doi:10.1111/nyas.14899.
18. **Shahin A**, **Bosnyak DJ**, **Trainor LJ**, **Roberts LE.** Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 23: 5545–5552, 2003. doi:10.1523/JNEUROSCI.23-13-05545.2003.
19. **Lopez-Calderon J**, **Luck SJ.** ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci* 8: 213, 2014. doi:10.3389/fnhum.2014.00213.
20. **Delorme A**, **Makeig S.** EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134: 9–21, 2004. doi:10.1016/j.jneumeth.2003.10.009.
21. **Oostenveld R**, **Fries P**, **Maris E**, **Schoffelen JM.** FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011: 156869, 2011. doi:10.1155/2011/156869.
22. **Fisher NI.** *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University Press, 1993.
23. **Maris E**, **Oostenveld R.** Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164: 177–190, 2007. doi:10.1016/j.jneumeth.2007.03.024.
24. **Proctor M**, **Bresch E**, **Byrd D**, **Nayak K**, **Narayanan S.** Paralinguistic mechanisms of production in human "beatboxing": a real-time magnetic resonance imaging study. *J Acoust Soc Am* 133: 1043–1054, 2013. doi:10.1121/1.4773865.
25. **Honing H.** On the biological basis of musicality. *Ann NY Acad Sci* 1423: 51–56, 2018. doi:10.1111/nyas.13638.
26. **Krishnan S**, **Lima CF**, **Evans S**, **Chen S**, **Guldner S**, **Yeff H**, **Manly T**, **Scott SK.** Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cereb Cortex* 28: 4063–4079, 2018. doi:10.1093/cercor/bhy208.
27. **Heinrich A**, **Carlyon RP**, **Davis MH**, **Johnsrude IS.** Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci* 20: 1737–1752, 2008. doi:10.1162/jocn.2008.20069.
28. **Shahin AJ**, **Bishop CW**, **Miller LM.** Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage* 44: 1133–1143, 2009a. doi:10.1016/j.neuroimage.2008.09.045.
29. **Jensen O**, **Gelfand J**, **Kounios J**, **Lisman JE.** Oscillations in the α band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb Cortex* 12: 877–882, 2002. doi:10.1093/cercor/12.8.877.
30. **Kerlin JR**, **Shahin AJ**, **Miller LM.** Attentional gain control of ongoing cortical speech representations in a "cocktail party. *J Neurosci* 30: 620–628, 2010. doi:10.1523/JNEUROSCI.3631-09.2010.
31. **Mazaheri A**, **Picton TW.** EEG spectral dynamics during discrimination of auditory and visual targets. *Brain Res Cogn Brain Res* 24: 81–96, 2005. doi:10.1016/j.cogbrainres.2004.12.013.
32. **Obleser J**, **Weisz N.** Suppressed α oscillations predict intelligibility of speech and its acoustic details. *Cereb Cortex* 22: 2466–2477, 2012. doi:10.1093/cercor/bhr325.
33. **Pfurtscheller G**, **Stancák A Jr**, **Neuper C.** Event-related synchronization (ERS) in the α band—an electrophysiological correlate of cortical idling: a review. *Int J Psychophysiol* 24: 39–46, 1996. doi:10.1016/s0167-8760(96)00066-9.
34. **Rohenkohl G**, **Nobre AC.** α Oscillations related to anticipatory attention follow temporal expectations. *J Neurosci* 31: 14076–14084, 2011. doi:10.1523/JNEUROSCI.3387-11.2011.
35. **Sauseng P**, **Klimesch W**, **Gerloff C**, **Hummel FC.** Spontaneous locally restricted EEG α activity determines cortical excitability in the motor cortex. *Neuropsychologia* 47: 284–288, 2009. doi:10.1016/j.neuropsychologia.2008.07.021.
36. **Shahin AJ**, **Picton TW**, **Miller LM.** Brain oscillations during semantic evaluation of speech. *Brain Cogn* 70: 259–266, 2009. doi:10.1016/j.bandc.2009.02.008.
37. **Weisz N**, **Hartmann T**, **Müller N**, **Lorenz I**, **Obleser J.** Alpha rhythms in audition: cognitive and clinical perspectives. *Front Psychol* 2: 73, 2011. doi:10.3389/fpsyg.2011.00073.
38. **Klimesch W.** EEG α and θ oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 29: 169–195, 1999. doi:10.1016/s0165-0173(98)00056-3.
39. **Obleser J**, **Wöstmann M**, **Hellbernd N**, **Wilsch A**, **Maess B.** Adverse listening conditions and memory load drive a common α oscillatory network. *J Neurosci* 32: 12376–12383, 2012. doi:10.1523/JNEUROSCI.4908-11.2012.
40. **Sauseng P**, **Klimesch W**, **Stadler W**, **Schabus M**, **Doppelmayr M**, **Hanslmayr S**, **Gruber WR**, **Birbaumer N.** A shift of visual spatial attention is selectively associated with human EEG α activity. *Eur J Neurosci* 22: 2917–2926, 2005. doi:10.1111/j.1460-9568.2005.04482.x.
41. **Wianda E**, **Ross B.** The roles of α oscillation in working memory retention. *Brain Behav* 9: e01263, 2019. doi:10.1002/brb3.1263.
42. **Andrillon T**, **Kouider S**, **Agus T**, **Pressnitzer D.** Perceptual learning of acoustic noise generates memory-evoked potentials. *Curr Biol* 25: 2823–2829, 2015. doi:10.1016/j.cub.2015.09.027.
43. **Shahin AJ**, **Trainor LJ**, **Roberts LE**, **Backer KC**, **Miller LM.** Development of auditory phase-locked activity for music sounds. *J Neurophysiol* 103: 218–229, 2010. doi:10.1152/jn.00402.2009.
44. **Vanden Bosch der Nederlanden CM**, **Joanisse MF**, **Grahn JA.** Music as a scaffold for listening to speech: better neural phase-locking to song than speech. *NeuroImage* 214: 116767, 2020. doi:10.1016/j.neuroimage.2020.116767.
45. **Baumann S**, **Meyer M**, **Jäncke L.** Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J Cogn Neurosci* 20: 2238–2249, 2008. doi:10.1162/jocn.2008.20157.
46. **Doelling KB**, **Poeppel D.** Cortical entrainment to music and its modulation by expertise. *Proc Natl Acad Sci USA* 112: E6233–E6242, 2015. doi:10.1073/pnas.1508431112.
47. **Pantev C**, **Oostenveld R**, **Engelien A**, **Ross B**, **Roberts LE**, **Hoke M.** Increased auditory cortical representation in musicians. *Nature* 392: 811–814, 1998. doi:10.1038/33918.
48. **Pantev C**, **Roberts LE**, **Schulz M**, **Engelien A**, **Ross B.** Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12: 169–174, 2001. doi:10.1097/00001756-200101220-00041.
49. **Sanju HK**, **Kumar P.** Enhanced auditory evoked potentials in musicians: a review of recent findings. *J Otol* 11: 63–72, 2016. doi:10.1016/j.joto.2016.04.002.
50. **Shahin A**, **Roberts LE**, **Trainor LJ.** Enhancement of auditory cortical development by musical experience in children. *Neuroreport* 15: 1917–1921, 2004. doi:10.1097/00001756-200408260-00017.
51. **Shahin A**, **Roberts LE**, **Pantev C**, **Trainor LJ**, **Ross B.** Modulation of P2 auditory-evoked responses by the spectral complexity of

musical sounds. *Neuroreport* 16: 1781–1785, 2005. doi:10.1097/01.wnr.0000185017.29316.63.

52. **Shahin AJ**, **Roberts LE**, **Chau W**, **Trainor LJ**, **Miller LM.** Music training leads to the development of timbre-specific γ band activity. *NeuroImage* 41: 113–122, 2008. doi:10.1016/j.neuroimage.2008.01.067.
53. **Jensen O.** Reading the hippocampal code by theta phase-locking. *Trends Cogn Sci* 9: 551–553, 2005. doi:10.1016/j.tics.2005.10.003.
54. **Reddy L**, **Self MW**, **Zoefel B**, **Poncet M**, **Possel JK**, **Peters JC**, **Baayen JC**, **Idema S**, **VanRullen R**, **Roelfsema PR.** Theta-phase dependent neuronal coding during sequence learning in human single neurons. *Nat Commun* 12: 4839, 2021. doi:10.1038/s41467-021-25150-0.
55. **Jensen O**, **Mazaheri A.** Shaping functional architecture by oscillatory α activity: gating by inhibition. *Front Hum Neurosci* 4: 186, 2010. doi:10.3389/fnhum.2010.00186.