



Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex



Gavin M. Bidelman^{a,b,*}, Jeremy Grall^c

^a Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

^b School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA

^c Rudi E. Scheidt School of Music, University of Memphis, Memphis, TN, USA

ARTICLE INFO

Article history:

Accepted 5 July 2014

Available online 12 July 2014

Keywords:

Auditory event-related potentials (ERPs)

Musical tonality

Pitch-onset response (POR)

Pitch perception

Neural organization

ABSTRACT

Pitch relationships in music are characterized by their degree of consonance, a hierarchical perceptual quality that distinguishes how pleasant musical chords/intervals sound to the ear. The origins of consonance have been debated since the ancient Greeks. To elucidate the neurobiological mechanisms underlying these musical fundamentals, we recorded neuroelectric brain activity while participants listened passively to various chromatic musical intervals (simultaneously sounding pitches) varying in their perceptual pleasantness (i.e., consonance/dissonance). Dichotic presentation eliminated acoustic and peripheral contributions that often confound explanations of consonance. We found that neural representations for pitch in early human auditory cortex code perceptual features of musical consonance and follow a hierarchical organization according to music-theoretic principles. These neural correlates emerge pre-attentively within ~150 ms after the onset of pitch, are segregated topographically in superior temporal gyrus with a rightward hemispheric bias, and closely mirror listeners' behavioral valence preferences for the chromatic tone combinations inherent to music. A perceptual-based organization implies that parallel to the phonetic code for speech, elements of music are mapped within early cerebral structures according to higher-order, perceptual principles and the rules of Western harmony rather than simple acoustic attributes.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Western music practice stipulates that constituent tones which form its scales and tonal structure carry different weight, or importance, within a musical framework (Aldwell and Schachter, 2003; Rameau, 1722/1971). Consequently, different pitch combinations (e.g., intervals, chords) follow a hierarchical organization in accordance with their functional role in musical composition (Krumhansl, 1990). Tonal pairs associated with stability and finality are regarded as consonant while those associated with instability (i.e., requiring resolution) are deemed dissonant. Given their anchor-like function in musical contexts, consonant relationships tend to occur more frequently in tonal music (Budge, 1943; Vos and Troost, 1989), are preferred by listeners relative to their dissonant counterparts (Bidelman and Krishnan, 2009; Kameoka and Kuriyagawa, 1969b; McDermott et al., 2010; Plomp and Levelt, 1965; Schwartz et al., 2003), and consequently carry higher status in both music-theoretic and perceptual-cognitive ranking (Aldwell and Schachter, 2003; Itoh et al., 2010; Krumhansl, 1990; Malmberg, 1918; Rameau, 1722/1971). It is the hierarchical organization and ebb

and flow between these perceptual-cognitive attributes which produce the sense of musical key and pitch structure intrinsic to Western tonal music (Rameau, 1722/1971).

Given its fundamental role in music, the origins of musical consonance have enjoyed a rich history of explanations developed over many centuries. These accounts have ranged from the purely mathematical and acoustical factors noted by ancient Greek scholars (e.g., frequency ratios of two vibrating bodies) (Galilei, 1638/1963; Gill and Purves, 2009) to psycho-physiological properties of the peripheral auditory system (e.g., cochlear filtering and auditory masking) (Helmholtz, 1877/1954; Plomp and Levelt, 1965) [for review, see Bidelman (2013)]. Modern neuroimaging studies have challenged many of these long-held beliefs by demonstrating that when acoustic or peripheral factors are controlled (which typically covary with perceived consonance), brain activity provides a sufficient basis to account for the fundamental organizational principles of tonal music (Bidelman, 2013; Bidelman and Krishnan, 2009; Itoh et al., 2010). Both human (scalp-recorded potentials) and animal (intracellular) electrophysiological recordings reveal preferential encoding of consonant relative to dissonant pitch relationships (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009, 2011; Tramo et al., 2001). Moreover, listeners' behavioral preferences for intervallic and chordal harmonies (i.e., simultaneously sounding pitches) are well predicted based on underlying

* Corresponding author at: School of Communication Sciences and Disorders, University of Memphis, 807 Jefferson Ave., Memphis, TN, 38105, USA. Fax: +1 901 525 1282.

E-mail address: g.bidelman@memphis.edu (G.M. Bidelman).

sensory pitch representations along the early auditory pathway (rostral brainstem and auditory nerve) (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009). These studies demonstrate that nascent features of tonal structure and musical pitch hierarchy are present in the earliest (pre-attentive) stages of the auditory system in neurocomputations operating below conscious awareness.

Multichannel event related potentials (ERPs) provide direct assays of neuronal activity and thus, the potential to further clarify the nature of cerebral mechanisms engaged in processing musical pitch relations. Prominent ERP components elicited by auditory stimuli emerge within a few hundred milliseconds following stimulus onset (e.g., the classic P1–N1–P2–N2) and are thought to reflect initial sound processing in early auditory cortical fields within the supratemporal plane (Picton et al., 1999; Scherg et al., 1989). Employing these obligatory responses, neural correlates of musical consonance/dissonance have been identified in the latency of the P2–N2 complex, ~200–300 ms after the onset of sound (Habibi et al., 2013; Itoh et al., 2010). However, these components are both generated and modulated by a wide range of stimuli and largely reflect the encoding of energy onset (Näätänen and Picton, 1987) in addition to any one specific acoustic feature (e.g., musical pitch).

To disentangle overlapping obligatory onset responses from those specific to pitch, Krumbholz et al. (2003) designed a novel stimulus paradigm in which a continuous sound is constructed from a segment of noise followed by a segment of iterated rippled noise (IRN). IRN is created by delaying broadband noise and adding it back onto itself matched in intensity and overall spectral profile. This delay-and-add process yields a noisy pitch percept corresponding to the reciprocal of the time delay (d), whose perceptual salience scales with the number of iterations (n) (Yost, 1996).¹ The recursion process produces temporal regularity in the noise and a sinusoidal ripple in its long-term power spectrum yielding a harmonically rich sound. Yet, IRN lacks a prominent temporal envelope and does not produce typical cochlear place cues (e.g., prominent peaks in the excitation pattern) associated with most pitch bearing signals (see Fig. 1 in Krumbholz et al., 2003). Concatenating IRN with a preceding noise yields a salient perceptual change at the transition; the percept smoothly changes from a “hiss” (noise segment) to a musical note with distinct pitch (IRN segment).

This stimulus design offers two important advantages for studying musical pitch: (1) the segments differ only in a single acoustic parameter (periodicity) and thus isolate the percept of interest, i.e., pitch; (2) the segmentation (i.e., noise → IRN) allows for the temporal separation of the onset of pitch from the onset of the stimulus energy as a whole thereby disentangling the response to these two events. Neuromagnetic (MEG) responses to these stimuli show clear onset components, reflecting an obligatory response to the initial noise segment and further response deflections following the initiation of pitch (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003). This so-called pitch onset response (POR) is thought to reflect cortical activity to pitch-bearing information.² This is suggested by its response properties and neuroanatomical generators. Studies demonstrate a strong dependence of POR latency and magnitude on specific features of pitch (e.g., salience, fundamental

frequency). An earlier, more robust POR is generated by salient pitch percepts whereas a smaller, delayed response is evoked by weaker pitch percepts (Krumbholz et al., 2003; Seither-Preisler et al., 2006). Strikingly similar responses are produced by either monaurally or binaurally (e.g., Huggins pitch) generated pitch, suggesting that even disparate pitch percepts converge into a common cortical representation reflected by the POR (Chait et al., 2006; Hertrich et al., 2005). Finally, source analyses (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003) corroborated by human depth electrode recordings (Schonwiesner and Zatorre, 2008) indicate that POR generators are localized to the anterolateral portion of Heschl's gyrus (HG), the putative site of pitch processing and pitch sensitive neurons (Bendor and Wang, 2005; Griffiths et al., 1998; Johnsrude et al., 2000; Penagos et al., 2004; Zatorre, 1988). Given both its sensitivity and consistency across a number of studies, the POR offers an ideal window for studying early cortical representations of musical pitch relationships. To our knowledge, this is the first study to systematically examine the effects of musical pitch attributes (consonance/dissonance) on the neuroelectric POR response.

To this end, we recorded cortical POR responses while participants listened passively to dyadic (two-note) musical intervals varying in their degree of consonance/dissonance. In previous attempts to decipher their neural basis, it is often unclear whether the alleged brain correlates reflect an underlying percept or merely changes in neural activity that covary with stimulus acoustics or byproducts of phenomena propagated from the auditory periphery [e.g., cochlear beating, roughness] (Bidelman, 2013). Our stimulus paradigm included two controls to safeguard against such confounds. First, intervals were presented to listeners dichotically (one note to each ear). While dichotic presentation does not alter musical consonance percepts, critically, it removes acoustic and peripheral factors (e.g., beating/roughness) that often cloud interpretation of its origins (Bidelman, 2013; McDermott et al., 2010). Additionally, we adopted the novel noise → pitch stimulus paradigm introduced above to temporally offset musical pitch intervals from the overall onset of acoustic stimulus energy. These stimuli repress obligatory neural activity that dominates auditory evoked responses and isolates a pitch-specific cortical activity (Krumbholz et al., 2003). Under the hypothesis that chromatic intervals are represented in the brain according to music-theoretic and/or perceptual principles, we expected the magnitude of early cortical activity to follow hierarchical relations of musical pitch structure and predict listeners' subsequent behavioral consonance judgments. Source reconstruction of neural responses probed whether the spatial encoding of music is based on feature based scheme following the perceptual construct of consonance. The latter would indicate that musical pitch intervals are mapped according to their perceptual rather than rudimentary acoustic features, paralleling the abstract phonetic organization observed for speech (Gutschalk and Uppenkamp, 2011; Scharinger et al., 2011; Shestakova et al., 2004).

Methods

Participants

Nine, normal-hearing adults participated in the experiment (3 females). All participants were strongly right-handed as measured by the Edinburgh Handedness Inventory (laterality: $97.2 \pm 6.2\%$) (Oldfield, 1971), had a similar level of education ($\mu \pm \sigma$: 20.7 ± 1.95 years), and were monolingual speakers of American English. We recruited listeners with musical training (18.9 ± 7.1 years) given their larger auditory evoked potentials (Bidelman et al., 2011b,c; Itoh et al., 2010; Zendel and Alain, 2014), and hence, better response signal-to-noise ratio necessary for accurate source localization. It should be noted, however, that consonance ratings are largely independent of musical training; both musicians and nonmusicians show similar behavioral preferences (Bidelman and Krishnan, 2009, 2011; McDermott et al., 2010) and cortical ERPs to musical pitch intervals (Itoh et al.,

¹ The manipulation of pitch via IRN periodicity is fundamentally different than the effects of simply adding noise to a pitched signal (e.g., complex tone). The former alters the internal representation of the pitch itself while the latter primarily alters detectability (Gockel et al., 2006).

² Recent debate in the neuroimaging literature questions the use of IRN in studying the neural correlates of pitch. fMRI studies argue that the cortical response to IRN may not be driven to pitch *per se*, but rather other complex acoustic features inherent to IRN (e.g., spectrotemporal flux; Barker et al., 2011; Hall and Plack, 2009). Yet, our recent EEG study confirms that the cortical POR response to IRN indexes changes in pitch rather than other latent acoustic features of IRN, e.g., slow spectrotemporal flux (Krishnan et al., 2012). The ability of neuroelectric (EEG) but not hemodynamic (fMRI) responses to reflect the perceptual dimension of pitch was also confirmed in a recent study directly comparing the two techniques using the same pitch stimuli; BOLD signals (an indirect measure of neural activity) were shown to reflect only slow, non-pitch-specific stimulus fluctuations whereas EEG/MEG coded stimulus features specific to pitch operating at a millisecond resolution (Steinmann and Gutschalk, 2012).

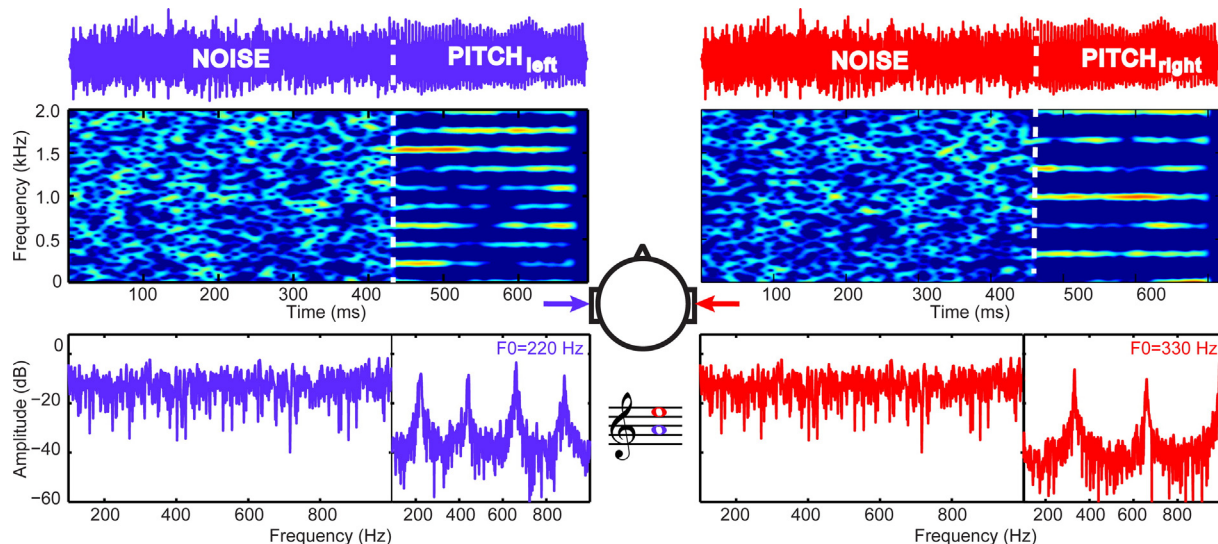


Fig. 1. Temporal and spectral stimulus characteristics. Time waveform (top row) and spectrogram (middle row) of the acoustic stimuli. The dotted line demarcates the transition from random noise to iterated rippled noise (IRN) that contained a clear pitch (450 ms after stimulus onset). (bottom row) Frequency spectra (FFTs) computed within each of the two stimulus segments. Note the absence of acoustic periodicity in the precursor noise but periodicity in the latter pitch segment as shown in the clear bands of energy at the fundamental frequencies (220/330 Hz) and their harmonics. Musical dyads were created from two instances of IRN stimuli where the lower and higher tone of each interval were presented dichotically (i.e., one note to each ear). In this example, the perfect 5th, the left ear heard only the lower pitch ($F_0 = 220$ Hz) whereas the right ear heard only the higher pitch ($F_0 = 330$ Hz).

2010), only magnified compared to their nonmusician peers. Subjects reported no history of hearing, neurological, or psychiatric disorders. Each gave written informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board.

Stimuli

Thirteen stimuli were created corresponding to the 13 chromatic dyads (i.e., two-tone intervals) within the octave (unisons to octave, semitone spacing). Each 700 ms stimulus consisted of two consecutive segments: a noise precursor (which contained no sensation of pitch), contiguous with a following segment containing the pitch interval (Fig. 1). The noise precursor consisted of a 450 ms segment of Gaussian noise filtered between 80 and 4000 Hz. The pitch interval portion was 250 ms in duration and was created using IRN. Here, $n = 64$ iterations were used to create IRN with salient pitch (Bidelman et al., 2011a; Krishnan et al., 2012).³ We varied the IRN delay (d) to produce the various pitches of the Western musical scale. The pitch segment was filtered between 80 and 4000 Hz to match the bandwidth of the preceding noise segment. The two segments were then equated in RMS amplitude and crossfaded (5 ms \cos^2 ramps) with overlapping ramps. This ensured that at the transition from noise to pitch, the envelope of the composite stimulus remained flat (see Fig. 1). This also ensured that there were no perceptible discontinuities in intensity or

³ For studying consonance, complex pitches (e.g., IRN) are desired as they (i) produce rich harmonic spectra that closely mimic that of natural instruments and (ii) they require the auditory system to compute a percept based on the combined spectral information across frequency channels. However, complex spectra contain multiple harmonics which, unless controlled experimentally (e.g., with dichotic presentation: current study; Bidelman and Krishnan, 2009; McDermott et al., 2010) interact in the auditory periphery. These interactions can introduce additional cues (e.g., beating/roughness) which reinforce and covary with consonance percepts but are not generated by central brain processing. Peripheral factors can be partially alleviated by using pure tones. However, behavioral studies indicate that the pattern of consonance for pure tone intervals differs dramatically from that obtained with complex tones (Cariani, 2004; Kameoka and Kuriyagawa, 1969a, b). The consonance of pure tones is largely monotonic, is dependent on whether or not the two frequencies fall within the same auditory filter, and is invariant outside of a single critical band (Itoh et al., 2010; Plomp and Levelt, 1965). In contrast, the consonance of complex tones is non-monotonic, depends on the total harmonicity between all the spectral components of the two tone complexes, and requires cross-frequency comparisons performed by the central auditory nervous system (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009; McDermott et al., 2010).

other cues (e.g., clicks) between noise and pitch segments and the total acoustic energy remained constant over its duration. The resulting percept was thus a smooth transition from noise to pitch. Critically, this stimulus segmentation allowed for the temporal separation of the onset of pitch from the onset of the stimulus as a whole thereby allowing us to examine the encoding of musical pitch without the overlap of obligatory onset responses (Gutschalk et al., 2004; Krumbholz et al., 2003).

Dyads were created by combining multiple instances of these IRN pitches and presenting them dichotically between ears; pitched segments sounded simultaneously and thus created the perception of a musical interval. That is, each ear (LE/RE) heard a noise \rightarrow pitch stimulus where the pitch segment was selected to be a different note along the chromatic musical scale and $\text{pitch}_{LE} \neq \text{pitch}_{RE}$ (except unison). Dichotic presentation minimized peripheral confounds (e.g., cochlear nonlinearities, roughness/beating: Bidelman and Heinz, 2011; Itoh et al., 2010; McDermott et al., 2010) which result from combining multiple tones during monaural stimulation. It also guaranteed that brain responses would reflect the central extraction of pitch after binaural integration (Bidelman and Krishnan, 2009) and that the resulting neural correlates were not simply driven by stimulus acoustics. For example, the ratio between an interval's fundamental frequencies and acoustic periodicity strength typically covaries with consonance judgments; these cues are unavailable when intervals are presented dichotically (for review, see Bidelman, 2013). Though each ear hears only a singular tone, the sensation of musical consonance is maintained via binaural integration (Bidelman and Krishnan, 2009; Houtsma and Goldstein, 1972). Dichotic presentation is also justified based on recent studies which demonstrate that neuromagnetic POR responses are largely indistinguishable when elicited by either monaural or binaural pitch percepts (Chait et al., 2006; Hertrich et al., 2005). Following our previous reports (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009), for every dyad, the lower of the two pitches – presented to the left ear – was fixed with a fundamental frequency (F_0) of 220 Hz (A3 on the Western music scale) while the upper F_0 was varied to produce different musical intervals. F_0 separation was varied systematically (semitone steps) to yield the 13 equal tempered pitch intervals recognized in Western music: unison (Un, $F_{0\text{higher}} = 220$ Hz), minor 2nd (m2, 233 Hz), major 2nd (M2, 247 Hz), minor 3rd (m3, 262 Hz), major 3rd (M3, 277 Hz), perfect 4th (P4, 293 Hz), tritone (TT, 311 Hz), perfect 5th (P5, 330 Hz), minor 6th (m6, 349 Hz), major 6th (M6, 370 Hz),

minor 7th (m7, 391 Hz), major 7th (M7, 415 Hz), octave (Oct, 440 Hz), where $F_{0\text{lower}}$ was always 220 Hz.

Electrophysiological recordings

Data acquisition and preprocessing

Participants reclined comfortably in an electro-acoustically shielded booth to facilitate recording of neurophysiologic responses. They were instructed to relax and refrain from extraneous body movement (to minimize myogenic artifacts), ignore the sounds they hear (to divert attention to the auditory stimuli), and were allowed to watch a muted subtitled movie to maintain a calm yet wakeful state. Stimulus presentation was controlled by a MATLAB (The MathWorks) routed to a TDT RP2 interface (Tucker–Davis Technologies) and delivered binaurally at an intensity of 82 dB SPL through insert earphones (ER-2A; Etymotic Research) using fixed, rarefaction polarity. Stimulus intensity was calibrated using the entire stimulus played through a Larson–Davis SPL meter (Model LxT) and measured in a 2-cc coupler (IEC 60126). Left and right ear channels were calibrated separately. Listeners heard 200 exemplars of each of the 13 musical interval stimuli presented at a repetition rate of 0.68/s (ISI = 750 ms). Stimuli were presented in separate blocks with one interval per block. The order of blocks was randomized within and across participants.

Neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10–20 locations around the scalp (Oostenveld and Praamstra, 2001). EEGs were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC–200 Hz. Responses were then stored to disk for offline analysis. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. During online acquisition, all electrodes were referenced to an additional sensor placed ~1 cm posterior to Cz. However, data were re-referenced offline to a common average reference. Contact impedances were maintained below 10 k Ω throughout the duration of the experiment.

Subsequent preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom routines coded in MATLAB. Data visualization and scalp topographies were computed using EEG/ERPLAB (Delorme and Makeig, 2004) (<http://www.erpinfo.org/erplab>). Prior to artifact correction, excessively noisy channels were interpolated and paroxysmal segments (i.e., >500 μ V) were automatically discarded. Ocular artifacts (saccades and blink artifacts) were then corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom et al., 2004). The PCA decomposition provided a set of independent components which best explained the topography of the blink/saccadic artifacts. The scalp projection of the first two PCA loadings was subtracted from the continuous EEG traces to nullify ocular contamination in the final ERPs. Cleaned EEGs were then epoched (–200–1100 ms), baseline-corrected to the pre-stimulus period, digitally filtered (1–30 Hz; zero-phase filters), and subsequently averaged in the time domain to obtain ERPs for each stimulus condition per participant.

ERP response analysis

For the frequency range of stimuli used in this study (220–440 Hz), the POR wave appears as a frontally-distributed negativity ~125 ms after the onset of pitch periodicity (Krumbholz et al., 2003; Seither-Preisler et al., 2006). Thus, to quantify the response and ensure consistency with our previous report (Krishnan et al., 2012), we measured POR magnitude at the Fpz electrode, computed as voltage difference between the maximum and minimum deflections within the 500–620 ms time window after the time-locking stimulus onset (i.e., arrows Fig. 3A). This search window encapsulated the onset and offset of periodicity in our stimuli and thus, the expected latency of the cortical POR response to pitch. Quantifying the response based on a peak-to-peak magnitude is preferable in that (i) it acknowledges that evoked responses are often a complex rather than a single isolated peak and (ii)

helps remove potential biases in peak measurement (e.g., low-frequency drift; baseline inconsistencies) that often distort amplitude measures based on single peak quantification (Luck, 2005, p. 231). The time at which the maximum negativity occurred was taken as a measure of response latency (i.e., absolute latency: ~525 ms; relative latency from onset of pitch: ~125 ms). This latency range is consistent with previous studies which indicate that the relative latency of the pitch-specific POR component in relation to the onset of periodicity (Gutschalk et al., 2004; Krumbholz et al., 2003).

Source waveform analysis

We estimated the location, strength, and orientation of the most likely intracerebral generators underlying the neural encoding of musical consonant and dissonant pitch relationships using source dipole analysis (Picton et al., 1999; Scherg et al., 1989). Localization was first performed on the group averaged scalp-recorded potentials using a realistic, boundary element volume conductor model (Fuchs et al., 1998a, 2002) standardized to the MNI brain (Mazziotta et al., 1995). Two symmetrical mirrored, rotating dipoles were fit in each hemisphere using seed points in bilateral superior temporal gyri. To constrain the inverse solution, the minimum distance between the two dipoles was set at ≥ 50 mm. Paired dipole models were fit in an iterative manner within the time window between 525 and 650 ms after the time-locking stimulus (Fuchs et al., 1998b). Fit solutions accounted for $\geq 80\%$ of the variance in voltage distribution recorded at the scalp. This analysis window was guided by determining the latency of the peak POR response from the group averaged ERPs (see Fig. 3). Using identical parameters, dipoles were then fit for each individual subject and condition. Source strengths (magnitude of source time courses) were extracted from each dipole pair per subject and condition. The peak response between 525 and 650 ms (absolute latency) was taken as a singular measure of hemispheric activation for each stimulus. In the current study, dipole fits were used to describe the location, strength, and orientation of the most likely sources underlying the neural encoding of musical pitch. Dipole locations, reflecting the mean “center of gravity” of neural activity, were visualized by projecting the stereotaxic coordinates of each solution onto the standardized MNI brain (Mazziotta et al., 1995).

To test whether or not consonant and dissonant interval classes where spatially segregated in cerebral cortex, we compared the Euclidean distance between responses to each interval class. For each participant, the (x, y, z) coordinates obtained from source dipole fits were averaged separately to determine the average spatial location for all consonant vs. dissonant intervals. Our interval groupings were based on both well-established music-theoretic principles (Aldwell and Schachter, 2003) and psychological classifications (Schwartz et al., 2003) of the chromatic intervals (see Fig. 2). Euclidean distance was computed as $ED = \sqrt{(x_D - x_C)^2 + (y_D - y_C)^2 + (z_D - z_C)^2}$, where (x_D, y_D, z_D) and (x_C, y_C, z_C) are the mean coordinates of all dissonant and all consonant dyads, respectively. A *t*-test was then used to determine if the mean “center of gravity” of neural activation between consonant and dissonant classes differed spatially on the cortical surface.

Behavioral consonance ratings.

Subjective ratings of consonance and dissonance were measured using a paired comparison task (Bidelman and Krishnan, 2009). The 13 chromatic dyads were presented dichotically (as in the ERP experiment) to each participant at an intensity of ~70 dB SPL through circumaural headphones (Sennheiser HD 280). Only the pitch portion of the stimuli was used in the behavioral task; the precursor noise used in the ERP experiment was removed. In each trial, listeners heard a pair of musical intervals, played consecutively, and were asked to select the interval they thought sounded more consonant (i.e., pleasant, beautiful, euphonious) from the pair (Bidelman and Krishnan, 2009;

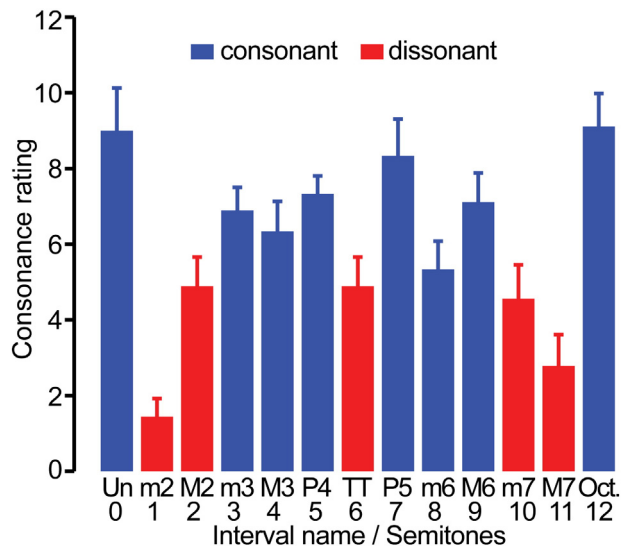


Fig. 2. Behavioral consonance ratings of musical intervals. Consonance judgments for the dichotic presentation of pitch intervals from the chromatic musical scale. Dyads considered consonant according to music theory (Aldwell and Schachter, 2003) (blue: Un, Oct, P5, P4, m3, M3, m6, M6) are preferred over those considered dissonant (red: m2, M2, TT, m7, M7). Perceptual rankings are not strictly binary (i.e., consonant vs. dissonant) but show a hierarchical arrangement as described by Western music theory. Error bars = 1 s.e.m.

Plomp and Levelt, 1965). Limiting the amount of time participants had to hear each stimulus (250 ms) forced them to respond based on the sensory percept rather than basing their decision on esthetic judgment (e.g., Binder et al., 2004)—which may be influenced by their musical knowledge and/or personal musical tastes. The order of the two intervals in a pair was randomly assigned and all possible pairings were presented to each listener. In total, participants heard $\binom{13}{2} = 78$ interval pairs such that each musical interval was contrasted with every other interval. A consonance rating for each dyad was then computed by counting the number of times it was selected relative to the total number of possible comparisons. Participants completed the behavioral task after electrophysiological recordings. The entire experimental session including electrophysiological and behavioral response collection took ~2 h to complete.

Results

Behavioral rankings of musical pitch relationships.

Mean behavioral consonance ratings for the 13 intervals of the chromatic music scale are shown in Fig. 2. Repeated measures (rm)ANOVA revealed a strong main effect of interval width on perceptual judgments [$F_{12, 96} = 7.75, p < 0.0001$]. Subjects generally selected consonant intervals (e.g., unison (Un), octave (Oct), perfect 5th/4th (P5/P4), major/minor 3rds (M3/m3), etc.) more frequently than dissonant intervals (e.g., minor/major 2nds (m2/M2), tritone (TT), minor/major 7ths (m7/M7), etc.), suggesting that the former was judged more pleasant sounding than the latter. Notably, listeners' ratings were not strictly binary (i.e., all consonant > all dissonant), but rather, followed a hierarchical ordering (i.e., Un > Oct > P5 > P4 > m3/M3 > ... > M7 > m2). Intervallic rankings observed here are consistent with both historic (Kameoka and Kuriyagawa, 1969b; Malmberg, 1918; Plomp and Levelt, 1965; Stumpf, 1989) and modern (Bidelman and Krishnan, 2009; McDermott et al., 2010; Schwartz et al., 2003) reports of musical interval ratings and the notion that musical pitch intervals/

chords are perceived in a hierarchical fashion (Bidelman, 2013; Krumhansl, 1990; Schwartz et al., 2003).

Neuroelectric responses to musical intervals.

While recording electrical brain activity, participants listened passively to the musical intervals blocked randomly in presentation. Time traces and corresponding scalp topography of the cortical neuroelectric responses are shown in Fig. 3. A clear P1–N1–P2 onset complex is observed in the 50–200 ms time window reflecting the obligatory brain response to onset of acoustic stimulus energy. Following the transition from a noise precursor (0–450 ms) to the initiation of musical pitch (450 ms), a prominent frontally distributed negativity emerged at the ~560 ms time mark corresponding to the pitch-specific POR response. The scalp topography and inversion of wave polarity between temporal and frontal electrode sites are consistent with bilateral neural generators in lateral aspects of the Sylvian fissure, i.e., Heschl's gyrus (Krumholz et al., 2003; Schonwiesner and Zatorre, 2008) (Fig. 3B). Importantly, the peak-to-peak magnitude of the POR complex was modulated by the specific musical interval [Fig. 4A; $F_{12, 96} = 2.87, p = 0.002$]. No reliable stimulus effect was observed in POR latency [$F_{12, 96} = 1.33, p = 0.22$].

Brain-behavior correspondences.

Critically, we found that neurophysiological response magnitudes were strongly associated with behavioral ratings [Pearson's correlation: $r = 0.52, p = 0.03$] suggesting that early cortical processing can predict an individual's perceptual judgments of musical pitch relationships (Fig. 4B, left panel). Consonant musical intervals judged more pleasant by listeners also yielded more robust neurophysiological responses in comparison to dissonant intervals, as evident by the clustering of the two categories. Intervals deemed most consonant according to music practice (e.g., Un, Oct, P5) are separated maximally in distance from those deemed highly dissonant (e.g., m2, TT, M7) in neural-perceptual space. Similarly, while POR latency did not show a reliable stimulus effect, it was negatively correlated with behavioral ratings [$r = -0.51, p = 0.03$]; consonant intervals deemed more pleasant sounding by listeners occurred earlier than dissonant intervals with more prolonged latencies (Fig. 4B, right panel). Collectively, these results suggest that stronger and earlier POR responses are associated with a higher degree of perceptual consonance.

Source generators of musical consonant and dissonant pitch relations

Grand average dipole source locations are shown for the 13 chromatic intervals in Fig. 5A. POR responses were largely circumscribed to more lateral portions of HG, consistent with electromagnetic POR responses (Gutschalk et al., 2004; Schonwiesner and Zatorre, 2008) and fMRI foci (Griffiths et al., 1998; Patterson et al., 2002) recorded in response to single IRN pitches (Fig. 5B). Yet, we found that dichotic two-tone intervals were mapped topographically according to their binary music classification, i.e., a topographic separation between consonant vs. dissonant intervals. Consonant intervals evoked activity clustered toward the anterolateral portion of HG whereas dissonant intervals clustered posteromedially. Across subjects, the Euclidean distance between consonant and dissonant interval clusters was significant [$t_8 = 3.35, p = 0.014$] (Fig. 5C). The clear separation in the “center of gravity” of neural activity between interval classes suggests that intervals are mapped not via simple acoustics but based on the perceptual attributes of music (i.e., consonance/dissonance classification).

Source waveforms extracted from the left and right hemisphere generators are shown in Fig. 5D. To test for a potential differential encoding scheme for consonance between hemispheres, source waveforms were collapsed within each of the two interval classes. We then conducted a fully crossed rmANOVA with factors hemisphere (2 levels: left vs. right)

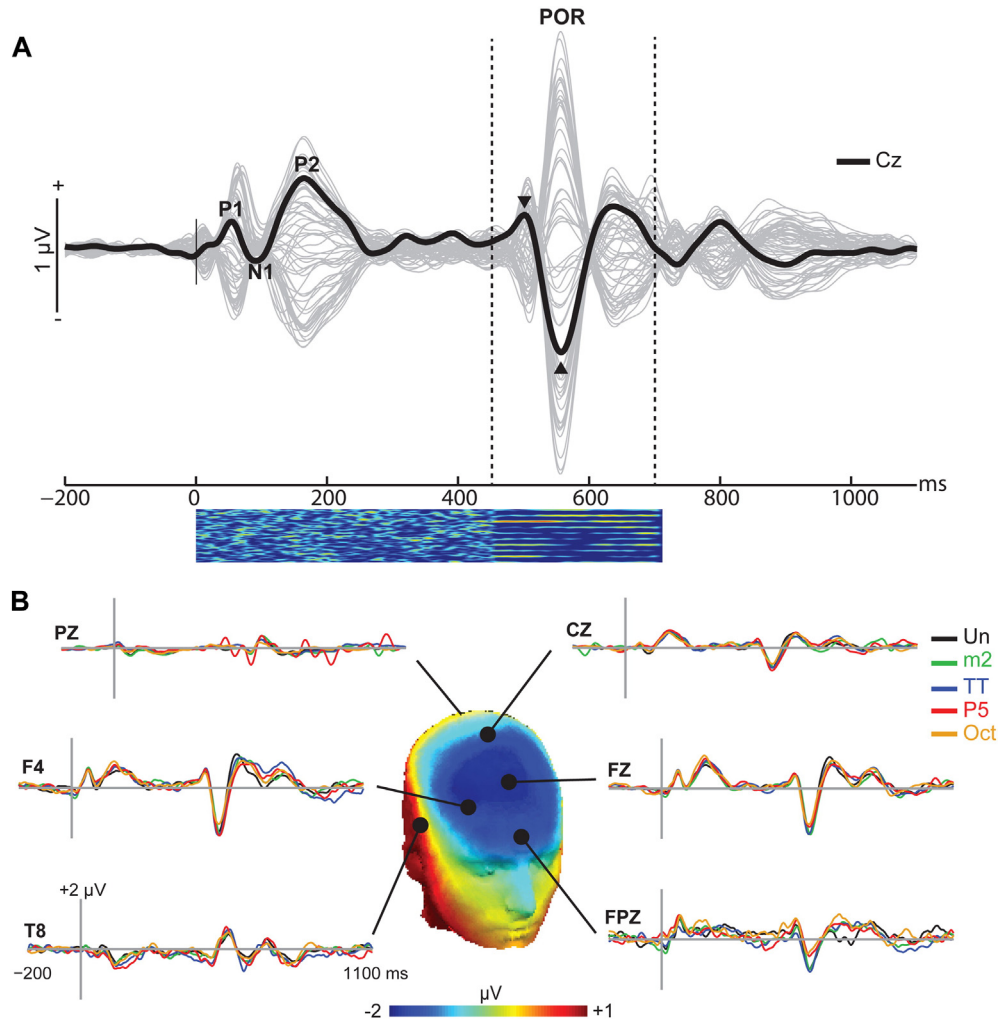


Fig. 3. Cortical evoked responses to musical intervals. (A) Grand average butterfly plot of ERPs in response to the unison interval. Thin gray traces show individual sensor waveforms; the thick black line denotes the response recorded at the vertex (Cz electrode). Arrows demarcate peaks used to quantify POR magnitude. For reference, a stimulus spectrogram is shown below the timescale. Note the difference between early obligatory components (P1, N1, P2), elicited by the onset of acoustic stimulus energy, and the POR wave (~560 ms), reflecting the neural encoding of musical pitch. (B) Scalp distribution and grand averaged cortical ERPs across various consonant and dissonant pitch relationships. The scalp topography depicts the response to the unison. Note the polarity inversion between temporal and frontal electrode sites, consistent with bilateral neural generators in the superior temporal plane (Krumbholz et al., 2003).

and interval class (2 levels: consonant vs. dissonant) to assess hemispheric laterality in the encoding of musical pitch relations. This analysis yielded a significant hemisphere \times interval class interaction [$F_{1, 8} = 9.16, p = 0.016$]. Post-hoc contrasts revealed that this effect was due to a significant change in right hemisphere source strength between interval classes [i.e., consonance > dissonance: $t_8 = 1.96, p = 0.04$] whereas the left hemisphere did not distinguish consonant from dissonant pitch categories [$t_8 = 0.35, p = 0.36$]. Complementing these functional asymmetries, correlations between neuronal source strength and behavioral data revealed that listeners' perceptual judgments were predicted from right [$r = 0.51, p = 0.037$] but not left [$r = -0.33, p = 0.26$] hemisphere activity (insets, Fig. 5D). Together, these results suggest that in right auditory cortex, neural activity elicited by musical pitch relationships (i) is spatially segregated according to a binary perceptual music classification (consonant vs. dissonant) and (ii) covaries in strength of activation according to listeners' perceptual judgments.

Discussion

There are notable commonalities among many of the music systems of the world including the division of the octave into specific scale steps and the use of a stable reference pitch to establish key structure. It has been argued that culturally-specific music is simply an elaboration of

only a few of these universal traits (Carterette and Kendall, 1999), one of which is the preference for consonance (Fritz et al., 2009). The perceptual bias for consonance is observed early in life well before human infants are exposed to culturally-specific music (Trainor et al., 2002) and has been observed even in avian species and non-human primates (Izumi, 2000; Sugimoto et al., 2010; Watanabe et al., 2005). It is plausible that the fundamental organizational principles of tonal music normally regarded as abstract cognitive traits (Krumhansl, 1990) are actually rooted in early and relatively automatic auditory processing as observed in this and previous studies (Bidelman, 2013; Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009; Trainor et al., 2002; Tramo et al., 2001).

Early cortical processing reveals brain basis for musical consonance and dissonance

In this regard, our results demonstrate that pitch relationships are mapped in auditory cortical brain areas according to the foundational principles governing musical harmony and tonality. Auditory brain activity emerging ~120 ms after the initiation of musical pitch was more robust and earlier when elicited by consonant than dissonant pitch relationships. More critically, we found that neural magnitudes mirrored listeners' perceptual judgments as neurophysiological responses (both

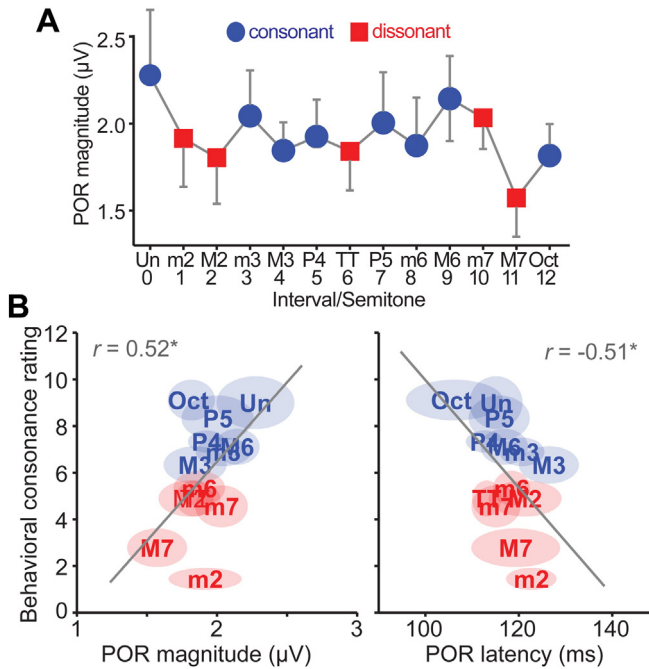


Fig. 4. Cortical ERPs show differential encoding of musical pitch intervals and predict listeners' behavioral consonance judgments. (A) Mean POR magnitudes are shown as a function of semitone width of each musical interval. As with behavioral responses (cf. Fig. 2), consonant pitch relationships (e.g., Un, Oct, P5, P4) elicit more robust and earlier neuroelectric responses than their dissonant counterparts (e.g., M2, TT, M7) and also show hierarchical encoding across the octave. (B) Brain behavioral correlations. Across listeners, consonant intervals elicit larger and earlier cortical evoked POR magnitudes than dissonant intervals and are subsequently judged more pleasant sounding. Note the systematic clustering of consonant and dissonant intervals and the maximal separation of the unison and octave (most consonant intervals) from the minor 2nd (most dissonant interval). Filled ellipses denote 1 s.e.m. in either the behavioral or neural dimension, respectively. POR latencies are expressed relative to the onset of pitch (see Fig. 1, ~450 ms). * $p < 0.05$.

sensor and particularly source activity) directly predicted the degree of pleasantness a listener associated with a particular pitch interval (Figs. 4B and 5D). Moreover, right (but not left) auditory cortical sources revealed an encoding scheme where interval classes were spatially clustered and distinguished according to their perceptual musical class (i.e., consonant vs. dissonant). These neural correlates were observed in the absence of peripheral (i.e., cochlear nonlinearities) or acoustic influences (e.g., frequency ratios); dichotic presentation ensured that each ear heard only a singular, undistorted tone free of these acoustic interactions. These findings reaffirm the notion that consonance perception does not require monaural interactions *per se*, e.g., roughness/beating (Bidelman and Krishnan, 2009; Houtsma and Goldstein, 1972; McDermott et al., 2010). Nevertheless, we found that cortical brain activity closely paralleled behavioral judgments for musical pitch relations. These results indicate that while acoustic and peripheral factors (e.g., cochlear distortion products) may contribute in everyday music listening, central neurobiological processing, i.e., how the auditory nervous system fires, provides a sufficient basis to account for consonance percepts (Bidelman, 2013; Tramo et al., 2001). Collectively, our findings imply that automatic, pre-attentive brain processing drives the perceptual response to music and determines which type of pitch combinations sound most agreeable or discordant to the ear (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009, 2011).

Our study was limited to examining the neural encoding of two-tone pitch combinations played simultaneously (i.e., harmonically) and thus, aligns with theories on "tonal- or sensory-consonance/dissonance" (Krumhansl, 1990). Harmonic pitch intervals presented in isolation do not involve contextual or cognitive influences which may also drive consonance ratings in realistic music textures (Dowling and Harwood,

1986). Though they evoked qualitatively similar perceptual ratings (Bidelman and Krishnan, 2009; Krumhansl and Shepard, 1979), listeners are generally more tolerant of melodic (i.e., sequentially presented notes) than harmonic dissonance. Moreover, while consonance ratings are largely based on pitch, judgments can vary across instrumental timbre; more muted ratings occur for vocal vs. complex tone stimuli (e.g., McDermott et al., 2010). In some sense, the harmonic nature of our stimuli coupled with our use of complex pitches with artificial timbre (IRN) probably represents somewhat of an extreme or optimal case of dissonance processing. It is conceivable that we may have observed slightly different brain-behavioral correlations with melodically presented intervals and/or different stimulus timbres. Future studies should clarify whether or not the neural correlates of consonance/dissonance observed here similarly hold for other musical pitch stimuli, presentation modes (i.e., melodic vs. harmonic intervals), and contexts (e.g., tones in isolation vs. tones set within a melody).

Nevertheless, our findings offer compelling evidence that the perceptual attributes of musical preferences (at least concurrently sounding pitch combinations) are rooted in early neurobiological processing at or before the level of primary auditory cortex (Bidelman, 2013). Consonant intervals deemed more pleasurable sounding evoked more robust and earlier neural responses than their dissonant counterparts and closely predicted listeners' behavioral ratings (Fig. 4). However, given that participants in the current study had formal musical training, it remains possible that at least some of our findings result from the long-term exposure of our listeners to the pitch combination of Western music. While musical training might amplify the perceptual contrast between consonance and dissonance (McDermott et al., 2010), musically naïve listeners show similar interval/chordal preferences to trained musicians (Bidelman and Krishnan, 2009, 2011; Roberts, 1986). Thus, we have no reason to believe that our results would not hold for nonmusicians. Nonetheless, it remains possible that experience-dependent reorganization resulting from musical training may create internalized "templates" for the chromatic intervals and partially explain the functional organization reflected in our ERP data (cf. Itoh et al., 2010). Importantly however, we observed that these strong brain correlates in the absence of attentional engagement and overt musical task. That musical scale properties are automatically extracted in early cortical processing is supported by the fact that pitch incongruities (e.g., detection of out-of-tune notes) are registered in auditory cortex prior to the intervention of focused attention (Brattico et al., 2006). As such, the neurobiological basis underlying certain perceptual attributes of music might emerge in rather primitive, pre-attentive brain processing (see also, Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009).

Neural generators of the POR response

Source dipole analysis of the POR demonstrated recruitment of anterolateral portions of Heschl's gyrus (Fig. 5B). These findings are consistent with pitch-specific brain areas identified via fMRI (Griffiths et al., 1998; Norman-Haignere et al., 2013; Patterson et al., 2002) and electromagnetic POR responses evoked by single IRN tones (Gutschalk et al., 2004; Schonwiesner and Zatorre, 2008) but extend these results to dichotic, tone-tone pitch intervals. While the POR most probably reflects neural activity from the primary auditory cortices (current study; Krumbholz et al., 2003), it remains possible that other source generators also contribute to the sensor-level responses. This is evident by the fact that at the sensor-level (i.e., scalp-recorded responses), adjacent stimuli (e.g., Un, m2) tended to produce similar POR response magnitudes (Fig. 3) and were largely driven by the differential encoding between the Un and M7, intervals maximally separated in acoustic space (Fig. 4A). It is possible that we may have observed a stronger modulation of the scalp-recorded potential (i.e., sensor-level POR) if both pitches of an interval were presented to the same ear. Monaural presentation produces more salient consonance percepts due to additional within-channel cues (e.g., beating, roughness) not present in dichotic

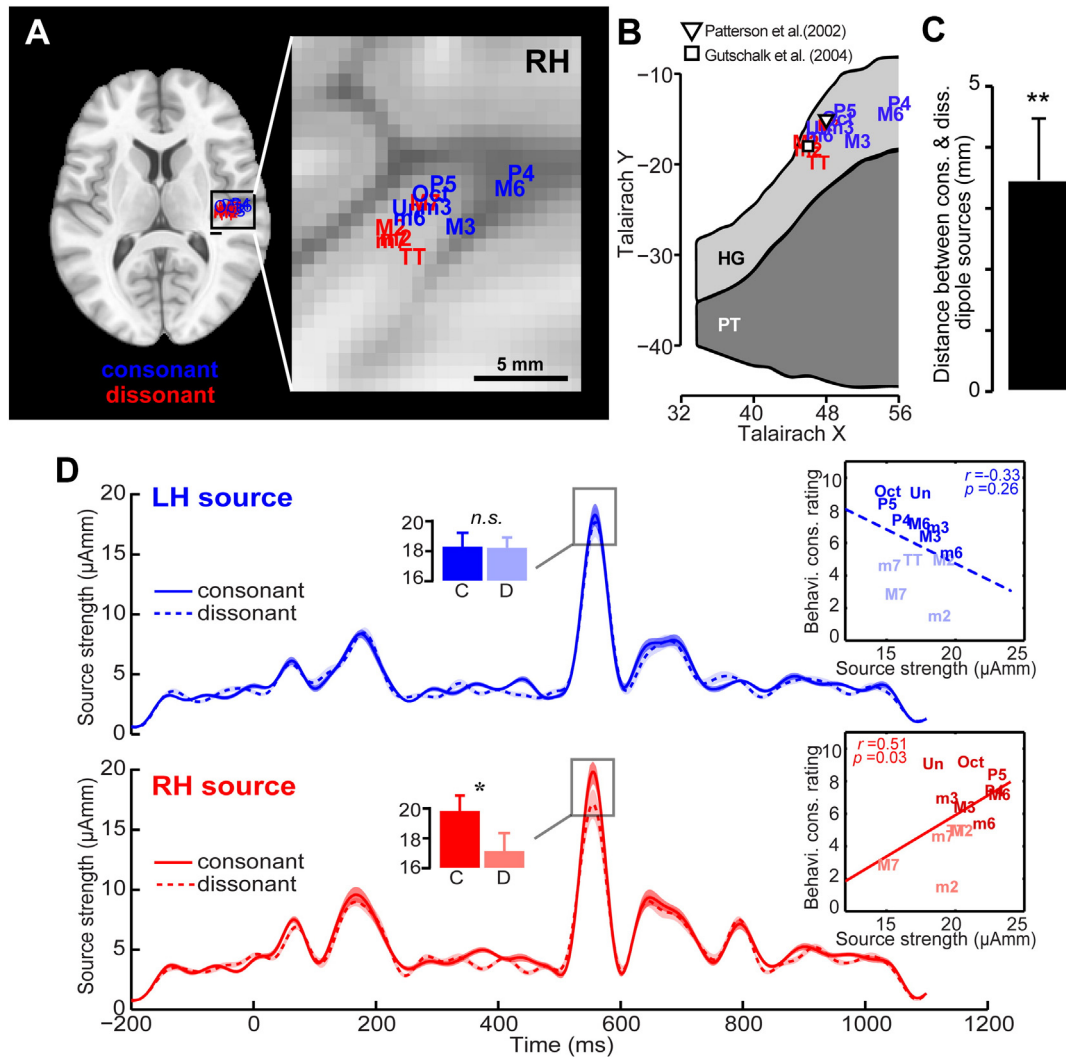


Fig. 5. Musical consonance and dissonance are segregated topographically in superior temporal gyrus. (A) Average dipole locations across listeners for the 13 chromatic intervals. Source foci are plotted in stereotaxic coordinates projected onto a single horizontal slice ($z = +11$ mm) of the MNI brain (Mazziotta et al., 1995). Scale bars = 5 mm. Consonant intervals (blue) tend to evoke activity clustered toward the anterolateral portion of Heschl's gyrus whereas dissonant intervals (red) cluster posteromedially. (B) Dipole locations projected onto a schematized axial plane through Heschl's gyrus (HG) and planum temporale (PT) with sulcal landmarks described by Leonard et al. (1998) based on 53 normal brains. Dipole locations for POR pitch responses described by Gutschalk et al. (2004) (\square) and mean fMRI activation to IRN pitch stimuli reported in Patterson et al. (2002) (∇) are shown for comparison. (C) As measured by Euclidean distance between consonant and dissonant clusters, the clear separation in the "center of gravity" of neural activity between interval classes suggests that intervals are mapped based on their abstract perceptual classification (consonance vs. dissonance) rather than tonotopy. (D) Source waveforms extracted from left and right hemisphere dipoles. Pooling intervals within classes (inset bar plots), source strength for consonant intervals (solid lines) is stronger than dissonant intervals (dotted lines) in the right but not left hemisphere (i.e., RH: consonance > dissonance; LH: consonance = dissonance). Pooled across listeners, the degree of perceptual consonance across intervals is predicted by right but not left hemisphere source activity. * $p < 0.05$, ** $p < 0.01$.

listening (Bidelman, 2013; McDermott et al., 2010). These cues are known to enhance neurophysiological encoding of consonance and dissonance (Bidelman, 2013; Bidelman and Heinz, 2011) and would presumably produce stronger modulations in the POR wave. Future studies are needed to compare PORs to musical consonance under different modes of stimulus presentation.

Nevertheless, our source analysis demonstrates that a dipole pair in bilateral auditory cortices is largely adequate in accounting for the majority of the response and correlation with behavior. Nevertheless, the volume conducted nature of ERPs cannot preclude the possibility that additional brain areas contribute to our scalp-record data. Indeed, it is well-known that distal non-auditory regions including prefrontal cortex can modulate even the earliest auditory cortical ERPs (e.g., P1, N1) (Chao and Knight, 1997; Knight et al., 1999). Moreover, neuroimaging data has revealed the engagement of distinct paralimbic and frontal neocortical regions (e.g., medial prefrontal cortex) that modulate hemodynamic brain activity as a function of musical dissonance (e.g., Blood et al., 1999). Thus, while we find that auditory cortical sources are

sufficient in describing the perceived pleasantness of musical pitch relationships (Fig. 5), it remains possible that non-auditory regions play a role in establishing these neural correlates.

Spatial organization for the perceptual dimension of musical consonance

For behaviorally-relevant signals like speech, constituent units (e.g., vowel phonemes) are multidimensional. Consequently, speech sounds are mapped in cortex according to abstract and invariant perceptual categories (i.e., phonetic classes) rather than their simple acoustic attributes (Scharinger et al., 2011; Shestakova et al., 2004). Our source analysis aimed to identify whether a similarly abstract functional organization exists for the fundamental constituents of music (i.e., consonant and dissonant pitch intervals) either in a spatial or functional code.

Intriguingly, we found that the chromatic pitch intervals were topographically segregated along auditory cortex in accordance with their perceptual classification. This implies that musical pitch relations are mapped cortically along a perceptual dimension (i.e., consonance vs.

dissonance) rather than low-level acoustic features (e.g., frequency separation between interval tones). Source reconstruction demonstrated mean activation to the chromatic scale intervals localized to Heschl's gyrus situated within the superior temporal plane (Fig. 5). Critically, these cortical regions distinguished interval classes according to their music-theoretic categories whereby consonant intervals clustered anterolateral to dissonant intervals (Figs. 5A and B), an effect most prominent in right auditory cortex. The clear separation in the “center of gravity” of neural activity (i.e., Euclidean distance) between consonant and dissonant classes suggests that chromatic intervals are mapped not exclusively based on stimulus acoustics but more provocatively, perceptual attributes relevant to musical harmony. Such an engrained neural architecture for the music-theoretic and perceptual principles of tonal relations might account for listeners' implicit knowledge of the structure and rules of Western harmony that develops early in life and without formal musical training (Ettlinger et al., 2011; Koelsch et al., 2000; Schellenberg et al., 2005).

However, our stimuli were generated using IRN which may shift the spatial bias of auditory cortical activation away from what would be evoked using more traditional pitch stimuli (e.g., tone complexes). Indeed, Butler and Trainor (2012) have recently shown that source location estimates of the earlier cortical ERPs (P1, N1) vary depending on the type of evoking pitch stimulus (i.e., complex tones vs. IRN) despite having the same F0 pitch. This suggests that the absolute anatomical locations of neural responses might be biased in the current study given our use of IRN. This would not, however, account for the relative spatial segregation we still find between consonance and dissonant interval categories. Interestingly, Butler and Trainor (2012) also reported that the spatial orientations of the later auditory ERPs (150 ms after stimulus onset) were invariant to the type of pitch evoking stimuli; IRN and tone complexes were elicited by similar generators in this later time window. This suggests that slightly later processing might reflect an integrated and common pitch percept, akin to the abstract neural organization observed for speech (Bidelman et al., 2013, in press; Chang et al., 2010). Inasmuch as our neural recordings reveal a similar level of analysis (they have similar latency: ~150 ms), the POR might also reflect brain correlates for the integrated perception of musical pitch relations. This is further supported by the fact that in the current study, PORs were evoked under dichotic listening but were nonetheless modulated according to a perceptual rather than acoustic dimension of music (i.e., consonance–dissonance).

In this regard, our data corroborate recent neuroimaging studies which have described similar abstract perceptual organization for pitch attributes in auditory cortex. Indeed, fMRI studies have demonstrated that pitch “chroma” – a perceptual dimension of pitch reflecting the cyclical nature of single tones within the octave – recruits cortical regions proximal but anterior to primary auditory cortex (Briley et al., 2013; Warren et al., 2003). Moreover, recent comprehensive fMRI mappings of the cortical regions selective to pitch percepts suggest that pitch-specific responses are located in a stereotyped region in anterolateral auditory cortex (Griffiths et al., 1998; Norman-Haignere et al., 2013; Penagos et al., 2004) but also recruit non-primary regions anterior to HG (Norman-Haignere et al., 2013). Our data agree with these fMRI studies and further implicate anterolateral HG in processing (i) the pitch combinations of music and (ii) their perceptual, rather than purely acoustic dimensions (Fig. 5). However, our analyses reveal that the consonance of pitch intervals seems to be coded largely in primary auditory cortex, and does not necessarily encroach into pitch-specific areas identified in non-primary auditory areas (cf. Norman-Haignere et al., 2013). Discrepancies between studies are likely the result of differences in the spatial resolution of neuroimaging approaches (fMRI: Norman-Haignere et al., 2013) vs. ERPs (ERP: current study; Gutschalk et al., 2004), our use of multiple simultaneous pitches, and dichotic vs. monaural presentation. Dichotic presentation generally produces weaker consonance percepts than when two pitches are presented to the same ear (McDermott et al., 2010). Differences between fMRI

and ERPs notwithstanding, it is conceivable that the relatively weaker pitch percepts used in the current study (dichotic intervals) recruit a more restricted expanse of pitch-specific neurons than pitch activity evoked under monaural stimulation (cf. Norman-Haignere et al., 2013).

Musical consonance is determined by the strength of early cortical activity in right hemisphere

In contrast to the binary spatial organization for musical pitch relationships, the strength of neuronal source activation not only distinguished perceptual music classes (Fig. 5C) but closely mirrored listeners' behavioral judgments (Fig. 5D). These findings indicate that degree of pleasantness experienced in response to musical pitch relations is determined not by the location *per se*, but rather, the strength of activity along the cortical surface. The closer correspondence of right compared to left hemisphere in predicting perceptual attributes of musical consonance is consistent with the well-known hemispheric asymmetry in fine spectral analysis and the specialization of right hemisphere in music information processing (Patterson et al., 2002; Zatorre and Salimpoor, 2013; Zatorre et al., 2002). Our results provide important evidence that early auditory cortical encoding (~150 ms after the onset of pitch), traditionally regarded as only sensory processing of sound features (e.g., intensity, frequency, duration) (Picton, 2010), reflects more than an echo of the external acoustic input. Rather, our data provide insight that even early auditory brain mechanisms reflect a higher-order coding and abstraction of attributes that renders intended auditory behaviors (Bidelman et al., 2013; Chang et al., 2010; Eulitz and Lahiri, 2004).

What might be the behavioral consequence of a neural organization for musical pitch relations? Recent functional imaging work demonstrates that the esthetic reward and pleasure involved in hearing music rests critically on an interaction between mesolimbic reward centers of the brain and early perceptual analyses performed by the auditory cortices (Salimpoor et al., 2013). We propose that a categorical organization for pitch combinations based on perceptual consonance may facilitate efficient predictive coding which drives listener's harmonic expectations and the resulting pleasurable response to music. By establishing an early perceptual-based configuration in auditory cortices, the awareness of musical tonality would then be guided by an automatic predictive coding process whereby sensory input is compared online to internal harmonic predictions established by our consonance-based feature map (cf. van Wassenhove et al., 2005). Given the preponderance of consonant tones in music practice (Budge, 1943; Vos and Troost, 1989), statistical learning would tend to strengthen these maps over repeated exposure to the tonal relationships found in a music system (Saffran et al., 1999). Like speech mapping, this reinforcement need not be entirely explicit (Saffran et al., 1999); listeners can acquire implicit knowledge of the hierarchical rules that typify Western harmony without formal music exposure or training (Koelsch et al., 2000; Schellenberg et al., 2005). It is conceivable that these affinities arise or might be built from the functional neural architecture and predisposition for consonance observed here.

More generally, our results have broader implications for understanding the neural organization of important human cognitive traits. Though our study pertained to the neurophysiological organization of musical pitch units, our data dovetail recent neuroimaging in the domain of language which shows parallel organizational principles for the phonemic units of speech. Electrical brain activity elicited during linguistic tasks reveals that speech sounds are mapped topographically in primary auditory cortex and adjacent areas according to phonetic rather than acoustic rules (Gutschalk and Uppenkamp, 2011; Scharinger et al., 2011; Shestakova et al., 2004). Neural representations for speech are thus organized to facilitate processing of discrete and invariant perceptual categories rather than continuous features of the acoustic signal. Remarkably, brain mechanisms supporting this higher-

dimensional mapping appear to operate very early in the neural chronometry of speech processing (Bidelman et al., 2013, in press; Chang et al., 2010). Our findings reveal a similarly abstract and highly efficient organization for music operating within a few hundred milliseconds after the onset of sound. That the fundamental building blocks of both linguistic and musical systems are automatically mapped according to perceptual categories in early auditory cortical centers highlights an isomorphism in the cerebral organization of these two cognitive systems and reveals new commonalities between the music and language domains.

Acknowledgments

We thank Stefanie Hutka and Dr. Chris Smalt for comments on earlier versions of the article. This work was supported in part by a grant from the GRAMMY® Foundation awarded to G.M.B.

References

- Aldwell, E., Schachter, C., 2003. *Harmony & Voice Leading*. Thomson/Schirmer, Boston.
- Barker, D., Plack, C.J., Hall, D.A., 2011. Human auditory cortical responses to pitch and to pitch strength. *Neuroreport* 22, 111–115.
- Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Bidelman, G.M., 2013. The role of the auditory brainstem in processing musically-relevant pitch. *Front. Psychol.* 4, 1–13.
- Bidelman, G.M., Heinz, M.G., 2011. Auditory-nerve responses predict pitch attributes related to musical consonance–dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.* 130, 1488–1502.
- Bidelman, G.M., Krishnan, A., 2009. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171.
- Bidelman, G.M., Krishnan, A., 2011. Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport* 22, 212–216.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011a. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011b. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn.* 77, 1–10.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011c. Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neurosci. Lett.* 503, 203–207.
- Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212.
- Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 1–12. <http://dx.doi.org/10.1111/ejn.12627> (in press).
- Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301.
- Blood, A.J., Zatorre, R.J., Bermudez, P., Evans, A., 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2, 382–387.
- Brattico, E., Tervaniemi, M., Naatanen, R., Peretz, I., 2006. Musical scale properties are automatically processed in the human auditory cortex. *Brain Res.* 1117, 162–174.
- Briley, P.M., Breakey, C., Krumbholz, K., 2013. Evidence for pitch chroma mapping in human auditory cortex. *Cereb. Cortex* 23, 2601–2610.
- Budge, H., 1943. *A Study of Chord Frequencies*. Teachers College Columbia University, New York.
- Butler, B.E., Trainor, L.J., 2012. Sequencing the cortical processing of pitch-evoking stimuli using EEG analysis and source estimation. *Front. Psychol.* 3, 180.
- Cariani, P.A., 2004. A temporal model for pitch multiplicity and tonal consonance. In: Lipscomb, S.D., Ashley, R., Gjerdingen, R.O., Webster, P. (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8)*. SMPC, Evanston, IL, pp. 310–314.
- Carterette, E.C., Kendall, R.A., 1999. Comparative music perception and cognition. In: Deutsch, D. (Ed.), *The Psychology of Music*. Academic Press, San Diego, pp. 725–791.
- Chait, M., Poeppel, D., Simon, J.Z., 2006. Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cereb. Cortex* 16, 835–848.
- Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432.
- Chao, L.L., Knight, R.T., 1997. Prefrontal deficits in attention and inhibitory control with aging. *Cereb. Cortex* 7, 63–69.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21.
- Dowling, J., Harwood, D.L., 1986. *Music Cognition*. Academic Press, San Diego.
- Ettlinger, M., Margulis, E.H., Wong, P.C.M., 2011. Implicit memory in music and language. *Front. Psychol.* 2.
- Eulitz, C., Lahiri, A., 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., Koelsch, S., 2009. Universal recognition of three basic emotions in music. *Curr. Biol.* 19, 573–576.
- Fuchs, M., Drenckhahn, R., Wischmann, H.-A., Wagner, M., 1998a. An improved boundary element method for realistic volume-conductor modeling. *IEEE Trans. Biomed. Eng.* 45, 980–997.
- Fuchs, M., Wagner, M., Wischmann, H.-A., Kohler, T., Theissen, A., Drenckhahn, R., Buchner, H., 1998b. Improving source reconstructions by combining bioelectric and biomagnetic data. *Electroencephalogr. Clin. Neurophysiol.* 107, 93–111.
- Fuchs, M., Kastner, J., Wagner, M., Hawes, S., Ebersole, J., 2002. A standardized boundary element method volume conductor model. *Clin. Neurophysiol.* 113, 702–712.
- Galilei, G., 1638/1963. *Discorsi e dimostrazioni matematiche interno à due nuove scienze attenenti alla mecanica ed i movimenti locali* [Dialogues concerning Two New Sciences] (H. Crew & A. de Salvio, Trans.). McGraw-Hill Book Co., Inc., New York (Original work published 1638).
- Gill, K.Z., Purves, D., 2009. A biological rationale for musical scales. *PLoS One* 4, 1–9.
- Gockel, H., Moore, B.C., Plack, C.J., Carlyon, R.P., 2006. Effect of noise on the detectability and fundamental frequency discrimination of complex tones. *J. Acoust. Soc. Am.* 120, 957–965.
- Griffiths, T.D., Buechel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.* 1, 422–427.
- Gutschalk, A., Uppenkamp, S., 2011. Sustained responses for pitch and vowels map to similar sites in human auditory cortex. *Neuroimage* 56, 1578–1587.
- Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S., Scherg, M., 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *Neuroimage* 15, 207–216.
- Gutschalk, A., Patterson, R.D., Scherg, M., Uppenkamp, S., Rupp, A., 2004. Temporal dynamics of pitch in human auditory cortex. *Neuroimage* 22, 755–766.
- Habibi, A., Wirantana, V., Starr, A., 2013. Cortical activity during perception of musical pitch: comparing musicians and nonmusicians. *Music. Percept.* 30, 463–479.
- Hall, D.A., Plack, C.J., 2009. Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576–585.
- Helmholtz, H., 1877/1954. *On the Sensations of Tone* (Ellis AJ, translator). Dover, New York (Original work published 1877).
- Hertrich, I., Mathiak, K., Menning, H., Lutzenberger, W., Ackermann, H., 2005. MEG responses to rippled noise and Huggins pitch reveal similar cortical representations. *Neuroreport* 16, 193–196.
- Houtsma, A.J., Goldstein, J.L., 1972. The central origin of the pitch of complex tones: evidence from musical interval recognition. *J. Acoust. Soc. Am.* 51, 520–529.
- Itoh, K., Suwazono, S., Nakada, T., 2010. Central auditory processing of noncontextual consonance in music: an evoked potential study. *J. Acoust. Soc. Am.* 128, 3781.
- Izumi, A., 2000. Japanese monkeys perceive sensory consonance of chords. *J. Acoust. Soc. Am.* 108, 3073–3078.
- Johnsrude, I.S., Penhune, V.B., Zatorre, R.J., 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155–163.
- Kameoka, A., Kuriyagawa, M., 1969a. Consonance theory part I: consonance of dyads. *J. Acoust. Soc. Am.* 45, 1451–1459.
- Kameoka, A., Kuriyagawa, M., 1969b. Consonance theory part II: consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.* 45, 1460–1469.
- Knight, R.T., Richard Staines, W., Swick, D., Chao, L.L., 1999. Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychol.* 101, 159–178.
- Koelsch, S., Gunter, T., Friederici, A.D., Schroger, E., 2000. Brain indices of music processing: “nonmusicians” are musical. *J. Cogn. Neurosci.* 12, 520–541.
- Krishnan, A., Bidelman, G.M., Smalt, C.J., Ananthakrishnan, S., Gandour, J.T., 2012. Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia* 50, 2849–2859.
- Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lutkenhoner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl’s gyrus. *Cereb. Cortex* 13, 765–772.
- Krumhansl, C.L., 1990. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York.
- Krumhansl, C.L., Shepard, R.N., 1979. Quantification of the hierarchy of tonal functions within a diatonic context. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 579–594.
- Leonard, C.M., Puranik, C., Kuldau, J.M., Lombardino, L.J., 1998. Normal variation in the frequency and location of human auditory cortex landmarks Heschl’s Gyrus: where is it? *Cereb. Cortex* 8, 397–406.
- Luck, S., 2005. *An Introduction to the Event-Related Potential Technique*. MIT Press.
- Malmberg, C.F., 1918. The perception of consonance and dissonance. *Psychol. Monogr.* 25, 93–133.
- Mazziotta, J.C., Toga, A.W., Evans, A., Lancaster, J.L., Fox, P.T., 1995. A probabilistic atlas of the human brain: theory and rational for its development. *Neuroimage* 2, 89–101.
- McDermott, J.H., Lehr, A.J., Oxenham, A.J., 2010. Individual differences reveal the basis of consonance. *Curr. Biol.* 20, 1035–1041.
- Nääätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Norman-Haignere, S., Kanwisher, N., McDermott, J., 2013. Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci.* 33, 19451–19469.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719.

- Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776.
- Penagos, H., Melcher, J.R., Oxenham, A.J., 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* 24, 6810–6815.
- Picton, T.W., 2010. *Human Auditory Evoked Potentials*, 1st ed. Plural Publishing, San Diego.
- Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., Trujillo, N.J., 1999. Intracerebral sources of human auditory-evoked potentials. *Audiol. Neuro Otol.* 4, 64–79.
- Plomp, R., Levelt, W.J., 1965. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38, 548–560.
- Rameau, J.-P., 1722/1971. *Treatise on Harmony* (P. Gossett, Trans.). Dover Publications, Inc., New York (Original work published 1722).
- Roberts, L., 1986. Consonant judgments of musical chords by musicians and untrained listeners. *Acustica* 62, 163–171.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Salimpoor, V.N., van den Bosch, I., Kovacevic, N., McIntosh, A.R., Dagher, A., Zatorre, R.J., 2013. Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219.
- Scharinger, M., Idsardi, W.J., Poe, S., 2011. A comprehensive three-dimensional cortical map of vowel space. *J. Cogn. Neurosci.* 23, 3972–3982.
- Schellenberg, E.G., Bigand, E., Poulin-Charronnat, B., Garnier, C., Stevens, C., 2005. Children's implicit knowledge of harmony in Western music. *Dev. Sci.* 8, 551–566.
- Scherg, M., Vajsar, J., Picton, T.W., 1989. A source analysis of the late human auditory evoked potentials. *J. Cogn. Neurosci.* 1, 336–355.
- Schonwiesner, M., Zatorre, R.J., 2008. Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Exp. Brain Res.* 187, 97–105.
- Schwartz, D.A., Howe, C.Q., Purves, D., 2003. The statistical structure of human speech sounds predicts musical universals. *J. Neurosci.* 23, 7160–7168.
- Seither-Preisler, A., Patterson, R.D., Krumbholz, K., Seither, S., Lutkenhoner, B., 2006. From noise to pitch: transient and sustained responses of the auditory evoked field. *Hear. Res.* 218, 50–63.
- Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V., Huotilainen, M., 2004. Orderly cortical representation of vowel categories presented by multiple exemplars. *Brain Res. Cogn. Brain Res.* 21, 342–350.
- Steinmann, I., Gutschalk, A., 2012. Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. *J. Neurophysiol.* 107, 3458–3467.
- Stumpf, C., 1989. Konsonanz und Dissonanz. *Beitrage Ak Musikwiss* 1, 91–107.
- Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T., Hashiya, K., 2010. Preference for consonant music over dissonant music by an infant chimpanzee. *Primates* 51, 7–12.
- Trainor, L., Tsang, C., Cheung, V., 2002. Preference for sensory consonance in 2- and 4-month-old infants. *Music. Percept.* 20, 187.
- Tramo, M.J., Cariani, P.A., Delgutte, B., Braid, L.D., 2001. Neurobiological foundations for the theory of harmony in western tonal music. *Ann. N. Y. Acad. Sci.* 930, 92–116.
- van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1181–1186.
- Vos, P.G., Troost, J.M., 1989. Ascending and descending melodic intervals: statistical findings and their perceptual relevance. *Music. Percept.* 6, 383–396.
- Wallstrom, G.L., Kass, R.E., Miller, A., Cohn, J.F., Fox, N.A., 2004. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *Int. J. Psychophysiol.* 53, 105–119.
- Warren, J.D., Uppenkamp, S., Patterson, R.D., Griffiths, T.D., 2003. Separating pitch chroma and pitch height in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10038–10042.
- Watanabe, S., Uozumi, M., Tanaka, N., 2005. Discrimination of consonance and dissonance in Java sparrows. *Behav. Process.* 70, 203–208.
- Yost, W.A., 1996. Pitch of iterated rippled noise. *J. Acoust. Soc. Am.* 100, 511–518.
- Zatorre, R.J., 1988. Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566–572.
- Zatorre, R.J., Salimpoor, V.N., 2013. From perception to pleasure: music and its neural substrates. *Proc. Natl. Acad. Sci. U. S. A.* 110 (Suppl. 2), 10430–10437.
- Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
- Zendel, B.R., Alain, C., 2014. Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiol. Aging* 35, 55–63.