Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing[a)]

Gavin M. Bidelman[b)]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907

Michael G. Heinz[c)]
Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907

Human listeners prefer consonant over dissonant musical intervals and the perceived contrast between these classes is reduced with cochlear hearing loss. Population-level activity of normal and impaired model auditory-nerve (AN) fibers was examined to determine (1) if peripheral auditory neurons exhibit correlates of consonance and dissonance and (2) if the reduced perceptual difference between these qualities observed for hearing-impaired listeners can be explained by impaired AN responses. In addition, acoustical correlates of consonance-dissonance were also explored including periodicity and roughness. Among the chromatic pitch combinations of music, consonant intervals/chords yielded more robust neural pitch-salience magnitudes (determined by harmonicity/periodicity) than dissonant intervals/chords. In addition, AN pitch-salience magnitudes correctly predicted the ordering of hierarchical pitch and chordal sonorities described by Western music theory. Cochlear hearing impairment compressed pitch salience estimates between consonant and dissonant pitch relationships. The reduction in contrast of neural responses following cochlear hearing loss may explain the inability of hearing-impaired listeners to distinguish musical qualia as clearly as normal-hearing individuals. Of the neural and acoustic correlates explored, AN pitch salience was the best predictor of behavioral data. Results ultimately show that basic pitch relationships governing music are already present in initial stages of neural processing at the AN level.
© 2011 Acoustical Society of America. [DOI: 10.1121/1.3605559]

I. INTRODUCTION

In Western tonal music, the octave is divided into 12 equally spaced pitch classes (i.e., semitones). These elements can be further arranged into 7 tone subsets to construct the diatonic major/minor scales that define tonality and musical key. Music theory and composition stipulate that the pitch combinations (i.e., intervals) formed by these scale-tones carry different weight, or importance, within a musical framework (Aldwell and Schachter, 2003). That is, musical pitch intervals follow a hierarchical organization in accordance with their functional role (Krumhansl, 1990). Intervals associated with stability and finality are regarded as consonant while those associated with instability (i.e., requiring resolution) are regarded as dissonant. Given their anchor-like function in musical contexts, it is perhaps unsurprising that consonant pitch relationships occur more frequently in tonal music than dissonant relationships (Budge, 1943; Vos and Troost, 1989). Ultimately, it is the interaction between consonance and dissonance which conveys musical tension and establishes the structural foundations of melody and har-

mony, the fundamental building blocks of Western tonal music (Rameau, 1722/1971; Krumhansl, 1990).

Music cognition literature distinguishes these strictly *musical* definitions from those used to describe the psychological attributes of musical pitch. The term tonal- or sensory-consonance-dissonance refers to the perceptual quality of two simultaneous tones or chords presented in isolation (Krumhansl, 1990) and is distinct from consonance arising from contextual or cognitive influences (Dowling and Harwood, 1986). Perceptually, consonant pitch relationships are described as sounding more pleasant, euphonious, and beautiful than dissonant combinations which sound unpleasant, discordant, or rough (Plomp and Levelt, 1965). Listeners prefer consonant intervals to their dissonant counterparts (Kameoka and Kuriyagawa, 1969a,b; Dowling and Harwood, 1986) and assign them higher status in hierarchical ranking (Krumhansl, 1990; Schwartz *et al.*, 2003)—a fact true even for non-musicians (van de Geer *et al.*, 1962; Tufts *et al.*, 2005; Bidelman and Krishnan, 2009). This perceptual bias for consonant pitch combinations emerges early in life, well before an infant is exposed to the stylistic norms of culturally specific music (Trehub and Hannon, 2006). Indeed, evidence from animal studies indicates that even non-human species (e.g., sparrows and Japanese monkeys) discriminate consonant from dissonant pitch relationships (Izumi, 2000; Watanabe *et al.*, 2005; Brooks and Cook, 2010) and some even show musical preferences similar to human listeners

---

(e.g., Bach preferred over Schoenberg) (Sugimoto *et al.*, 2010; but see McDermott and Hauser, 2004; 2007). The fact that these preferences can exist in the absence of long-term enculturation or music training and may not be restricted solely to humans suggests that certain perceptual attributes of musical pitch may be rooted in fundamental processing and constraints of the auditory system (McDermott and Hauser, 2005; Trehub and Hannon, 2006).

Early explanations of consonance-dissonance focused on the underlying acoustic properties of musical intervals. It was recognized as early as the ancient Greeks, and later by Galileo, that pleasant sounding (i.e., consonant) musical intervals were formed when two vibrating entities were combined whose frequencies formed simple integer ratios (e.g., 3:2 = perfect 5th, 2:1 = octave). In contrast, "harsh" or "discordant" (i.e., dissonant) intervals were created by combining tones with complex ratios (e.g., 16:15 = minor 2nd).[1] By these purely mathematical standards, consonant intervals were regarded as divine acoustic relationships superior to their dissonant counterparts and as a result, were heavily exploited by early composers (for a historic account see Tenney, 1988). Indeed, the most important pitch relationships in music, including the major chord, can be derived directly from the first few components of the harmonic series (Gill and Purves, 2009). Though intimately linked, explanations of consonance-dissonance based purely on these physical constructs (e.g., frequency ratios) are, in and of themselves, insufficient in describing all of the cognitive aspects of musical pitch (Cook and Fujisawa, 2006; Bidelman and Krishnan, 2009). Indeed, it is possible for an interval to be esthetically dissonant while mathematically consonant, or vice versa (Cazden, 1958, p. 205). For example, tones combined at simple ratios (traditionally considered consonant), can be judged to be dissonant when their frequency constituents are inharmonic/stretched (Slaymaker, 1970) or when occurring in an inappropriate (i.e., incongruent) musical context (Dowling and Harwood, 1986).

Helmholtz (1877/1954) offered a psychophysical explanation for sensory consonance-dissonance by observing that when adjacent harmonics in complex tones interfere they create the perception of "roughness" or "beating," percepts closely related to the perceived dissonance of tones (Terhardt, 1974). Consonance, on the other hand, occurs in the absence of beating, when low-order harmonics are spaced sufficiently far apart so as not to interact. Empirical studies suggest this phenomenon is related to cochlear mechanics and the critical-band hypothesis (Plomp and Levelt, 1965). This theory postulates that the overall consonance-dissonance of a musical interval depends on the total interaction of frequency components within single auditory filters. Pitches of consonant dyads have fewer partials which pass through the same critical-bands and therefore yield more pleasant percepts than dissonant intervals whose partials compete within individual channels.

While within-channel interactions may produce some dissonant percepts, modern empirical evidence indicates that the resulting beating/roughness plays only a minor role in the perception of consonance-dissonance and is subsidiary to the harmonicity of an interval (McDermott *et al.*, 2010). In addition, the perception of consonance and dissonance does not rely on monaural interaction (i.e., roughness/beating) alone and can be elicited when pitches are separated between ears, i.e., presented dichotically (e.g., Bidelman and Krishnan, 2009; McDermott *et al.*, 2010). In these cases, properties of acoustics (e.g., beating) and peripheral cochlear filtering mechanisms (e.g., critical band) are inadequate in explaining sensory consonance-dissonance because each ear processes a perfectly periodic, singular tone. In such conditions, consonance must instead be computed centrally by deriving information from the combined neural signals relayed from both cochleae (Bidelman and Krishnan, 2009).

Converging evidence suggests that consonance may be reflected in intrinsic neural processing. Using far-field recorded event-related potentials (Brattico *et al.*, 2006; Krohn *et al.*, 2007; Itoh *et al.*, 2010) and functional imaging (Foss *et al.*, 2007; Minati *et al.*, 2009), neural correlates of consonance, dissonance, and musical scale pitch hierarchy have been identified at cortical and recently subcortical (Bidelman and Krishnan, 2009, 2011) levels in humans. Though such studies often contain unavoidable confounds (e.g., potential long-term enculturation, learned category effects), these reports demonstrate that brain activity is especially sensitive to the pitch relationships found in music even in the absence of training and furthermore, is enhanced when processing consonant relative to dissonant intervals. Animal studies corroborate these findings revealing that single-unit response properties in auditory nerve (AN) (Tramo *et al.*, 2001), inferior colliculus (IC) (McKinney *et al.*, 2001), and primary auditory cortex (A1) (Fishman *et al.*, 2001) show differential sensitivity to consonant and dissonant pitch relationships. Together, these studies offer evidence for a physiological basis for musical consonance-dissonance. However, with limited recording time, stimuli, and small sample sizes, the conclusions of these neurophysiological studies are often restricted. As of yet, no single-unit study has examined the possible differential neural encoding across a complete continuum of musical (and nonmusical) pitch intervals.

Little is known regarding how sensorineural hearing loss (SNHL) alters the complex perception of musical pitch relationships. Reports indicate that even with assistive devices (e.g., hearing aids or cochlear implants), hearing-impaired listeners have abnormal perception of music (Chasin, 2003). Behavioral studies show that SNHL impairs identification (Arehart and Burns, 1999), discrimination (Moore and Carlyon, 2005), and perceptual salience of pitched stimuli (Leek and Summers, 2001). Recently, Tufts *et al.* (2005) examined the effects of moderate SNHL on the perceived sensory dissonance of pure tone and complex dyads (i.e., two-note musical intervals). Results showed that individuals with hearing loss failed to distinguish the relative dissonance between intervals as well as normal-hearing (NH) listeners. That is, hearing impairment (HI) resulted in a reduction in the perceptual contrast between pleasant (i.e., consonant) and unpleasant (i.e., dissonant) sounding pitch relationships. The authors attributed this loss of "musical contrast" to the observed reduction in impaired listeners' peripheral frequency selectivity (level effects were controlled for) as

measured via notched-noise masking. Broadened auditory filters are generally associated with dysfunctional cochlear outer hair cells (OHCs) whose functional integrity is required to produce many of the nonlinearities present in normal hearing, one of which is level-dependent tuning with sharp frequency selectivity at low sound levels (e.g., Liberman and Dodds, 1984; Patuzzi *et al.*, 1989). Ultimately, the results of Tufts *et al.* (2005) imply that individuals with hearing loss may not fully experience the variations in musical tension supplied by consonance−dissonance and that this impairment may be a consequence of damage to the OHC subsystem.

The aims of the present study were threefold: (1) examine population-level responses of AN fibers to determine whether basic temporal firing properties of peripheral neurons exhibit correlates of consonance, dissonance, and the hierarchy of musical intervals/chords; (2) determine if the loss of perceptual "musical contrast" between consonant and dissonant pitch relationships with hearing impairment can be explained by a reduction in neural information at the level of the AN; (3) assess the relative ability of the most prominent theories of consonance−dissonance (including acoustic periodicity and roughness) to explain the perceptual judgments of musical pitch relationships.

II. METHODS

A. Auditory-nerve model

Spike-train data from a computational model of the cat AN (Zilany *et al.*, 2009) was used to evaluate whether correlates of consonance, dissonance, and the hierarchy of musical pitch intervals/chords exist even at the earliest stage of neural processing along the auditory pathway. This phenomenological model represents the latest extension of a well established model that has been rigorously tested against physiological AN responses to both simple and complex stimuli, including tones, broadband noise, and speech-like sounds (Bruce *et al.*, 2003; Zilany and Bruce, 2006, 2007). The model incorporates several important properties observed in the auditory system including, cochlear filtering, level-dependent gain (i.e., compression) and bandwidth control, as well as two-tone suppression. The current generation of the model introduced power-law dynamics and long-term adaptation to the synapse between the inner hair cell and auditory nerve fiber (Zilany *et al.*, 2009). These additions have improved temporal encoding allowing the model to more accurately predict results from animal data including responses to amplitude modulation (i.e., envelope encoding) and forward masking (Zilany *et al.*, 2009). Model threshold tuning curves have been well fit to the CF-dependent variation in threshold and bandwidth for high-spontaneous rate (SR) fibers in normal-hearing cats (Miller *et al.*, 1997). The stochastic nature of AN responses is accounted for by a modified nonhomogenous Poisson process, which includes effects of both absolute and relative refractory periods and captures the major stochastic properties of AN responses (e.g., Young and Barta, 1986). For background and intricate details of the model, the reader is referred to Zilany and Bruce (2007) and Zilany *et al.* (2009).

B. Impaired AN model

SNHL was introduced into the model by altering the control path's C_{IHC} and C_{OHC} scaling coefficients representing inner and outer hair cell functional integrity, respectively (Zilany and Bruce, 2007). Both coefficients range from 0 to 1, where 1 simulates normal hair cell function and 0 indicates complete hair cell dysfunction (Bruce *et al.*, 2003). Lowering C_{IHC} elevates fiber response thresholds without affecting frequency selectivity, consistent with physiologic reports from impaired animal data (Liberman and Dodds, 1984). In contrast, altering C_{OHC} causes both a decrease in model fiber gain (i.e., elevated absolute threshold) and an increase in bandwidth (i.e., broadened tuning curve) (e.g., Liberman and Dodds, 1984; Bruce *et al.*, 2003). Coefficients were chosen based on the default values given by the "fitaudiogram" MATLAB script provided by Zilany and Bruce (2006, 2007) using audiometric data reported for hearing-impaired listeners in Tufts *et al.* (2005) (i.e., flat, moderate, SNHL; pure-tone average (PTA) $\approx$ 45 dB HL). This setting produced the desired threshold shifts by attributing the total hearing loss in dB (HL_{total}) to two-thirds OHC and one-third IHC impairment at each CF (i.e., $HL_{OHC} = \frac{2}{3} HL_{total}$; $HL_{IHC} = \frac{1}{3} HL_{total}$). This etiology is consistent with the effects of noise induced hearing loss in cats (Bruce *et al.*, 2003; Zilany and Bruce, 2006, 2007) and estimated OHC dysfunction in hearing-impaired humans (Plack *et al.*, 2004). Figure 1 shows the audiograms used in simulating AN responses in normal and hearing-impaired conditions. Other than the addition of impairment, NH and HI model predictions were obtained with the same analysis techniques as described in the sections which follow.

C. Stimuli

Musical dyads (i.e., two-note intervals) were constructed to match those found in similar studies on consonance−dissonance (Kameoka and Kuriyagawa, 1969b; Tufts *et al.*, 2005). Individual notes were synthesized using a
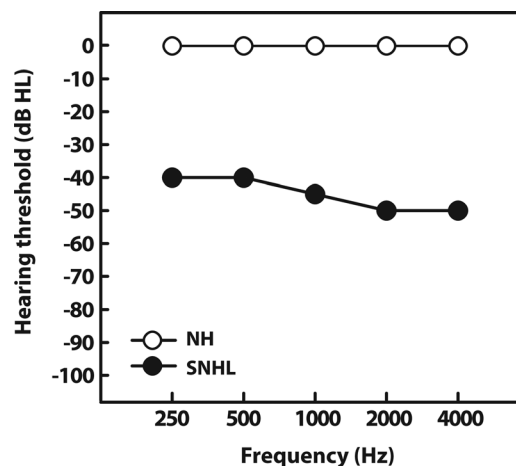
FIG. 1. Audiograms for normal-hearing and hearing-impaired conditions. The impaired audiogram was modeled after data reported by Tufts *et al.* (2005) who measured musical interval consonance rankings in subjects with a flat, moderate, sensorineural hearing loss (SNHL). Pure-tone averages (PTAs) for normal and impaired models are 0 and 45 dB HL, respectively.

tone-complex consisting of six harmonics with equal amplitudes added in cosine phase. However, similar results were obtained using decaying amplitude and random phase components (which are likely more representative of those produced by natural instruments, data not shown). For every dyad, the lower of the two pitches 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 (and nonmusical) intervals within the range of an octave. A total of 220 different dyads were generated by systematically increasing the frequency separation between the lower and higher tones in 1 Hz increments. Thus, separation between notes in the dyad continuum ranged from the musical unison (i.e., $f_{0 \text{ lower}} = f_{0 \text{ higher}} = 220$ Hz) to the perfect octave (i.e., $f_{0 \text{ lower}} = 220$ Hz, $f_{0 \text{ higher}} = 440$ Hz). A subset of this continuum includes the 12 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. Though we report results only for the register between 220–440 Hz, similar results were obtained in the octave above (e.g., 440–880 Hz) and below (e.g., 110–220 Hz) that used presently. Stimulus waveforms were 100 ms in duration including a 10 ms rise−fall time applied at both the onset and offset in order to reduce both spectral splatter in the stimuli and onset components in the responses.

To extend results based on simple musical intervals, we also examined model responses to isolated chords. Chords are comprised of at least three pitches but like musical dyads, listeners rank triads (i.e., three-note chords) according to their degree of "stability," "sonority," or "consonance" (Roberts, 1986; Cook and Fujisawa, 2006). Thus, we aim to determine if perceptual chordal stability ratings (i.e., ∼ consonance) could be predicted from AN response properties. Three pitches were presented simultaneously to the model whose f_0s corresponded to four common chords in Western music (equal temperament): major (220, 277, 330 Hz), minor (220, 261, 330 Hz), diminished (220, 261, 311 Hz), and augmented (220, 277, 349 Hz) triads.

D. Presentation levels

Stimulus level was defined as the overall RMS level of the entire interval/chord (in dB SPL). Presentation levels ranged from 50–70 dB SPL in 5 dB increments (only a subset of these levels are reported here). Because impaired results were produced with a ∼45 dB (PTA) hearing loss, NH predictions were also obtained at a very low intensity (25 dB SPL) to equate sensation levels (SLs) with HI results obtained at 70 dB SPL (i.e., 70 dB SPL – 45 dB HL = 25 dB SL). This control has also been implemented in perceptual experiments studying hearing-impaired listeners response to musical intervals (Tufts et al., 2005) and is necessary to ensure that any differences between normal and impaired results cannot simply be attributed to a reduction in audibility. Level effects were only explored with neural pitch salience—a measure of harmonicity/fusion (described below)—the neural analog of the primary behavioral correlate of consonance-dissonance (McDermott et al., 2010).

E. Neural pitch salience computed via periodic sieve template analysis of AN spike data

A block diagram of the various steps in analyzing AN spike data is illustrated in Fig. 2. To quantify information contained in AN responses that may lead to perceptually salient aspects of musical pitch, a temporal analysis scheme was adopted in order to examine the periodic information contained in the aggregate distribution of AN activity (Cariani and Delgutte, 1996). An ensemble of 70 high-SR (>50 spikes/s) auditory nerve fibers was simulated with characteristic frequencies (CFs) spaced logarithmically between 80–16 000 Hz. Poststimulus time histograms (PSTHs) were first constructed using 100 repetitions of each stimulus (0.1 ms bins) to quantify the neural discharge pattern of each fiber response (Fig. 2, "PSTH"). Only the 15–100 ms steady state portion of the PSTH was analyzed further in order to minimize effects of the onset response and rapid adaptation. While the exclusion of these early response components made little difference to predicted pitch salience, we chose to exclude them because onset responses occur regardless of the eliciting stimulus (e.g., clicks, noise, etc.) and are not directly related to stimulus pitch, per se. The autocorrelation function (ACF) of each PSTH—similar to an all-order interspike interval histogram (ISIH)—was computed from each CF's PSTH, representing the dominant pitch periodicities present in the neural response (Cariani and Delgutte, 1996). Individual ACFs were then weighted with a decaying exponential based on the individual fiber's CF (Fig. 2, "ACF" and 'Weight'): $\tau = 30$ ms (CF ≤ 100 Hz), $\tau = 16$ ms (100 < CF ≤ 440), $\tau = 12$ ms (440 < CF ≤ 880), $\tau = 10$ ms (880 < CF ≤ 1320), $\tau = 9$ ms (CF > 1320) (Cariani, 2004). Weighting gives greater precedence to the shorter pitch intervals an autocorrelation analyzer would have at its disposal (Cedolin and Delgutte, 2005) and accounts for the perceptual lower limit of musical pitch (∼30 Hz) (Pressnitzer et al., 2001). It has been proposed that CF-dependent weighting may emerge naturally as the result of the inherent frequency dependence of peripheral filtering (Bernstein and Oxenham, 2005). Given the inverse relationship between filter bandwidth and the temporal extent of its impulse response, narrower filters associated with lower CFs will yield a wider range of lags over which its channel energy is correlated with itself (Bernstein and Oxenham, 2005). Thus, in theory, lower CFs operate with relatively longer time constants (e.g., 30 ms) than higher CFs (e.g., 9 ms).

Weighted fiber-wise ACFs were then summed to obtain a population-level ACF of the entire neural ensemble (ACF$_{\text{pop}}$). The summary ACF$_{\text{pop}}$ contains information regarding all possible stimulus periodicities present in the neural response. To estimate the neural pitch salience of each musical interval, each ACF$_{\text{pop}}$ was analyzed using "periodic template" analysis. A series of periodic sieves were applied to each ACF$_{\text{pop}}$ in order to quantify the neural activity at a given
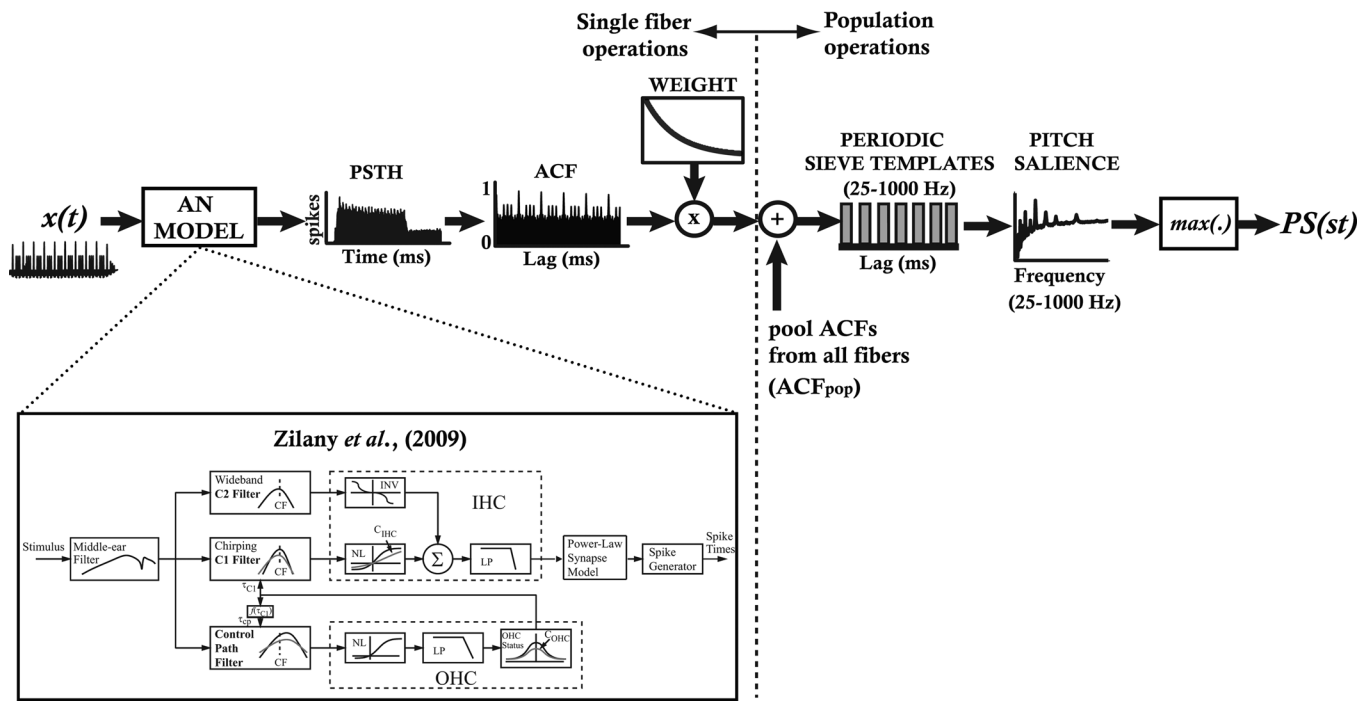
FIG. 2. Procedure for computing neural pitch salience from AN responses to musical intervals. Single-fiber operations vs population-level analyses are separated by the vertical dotted line. Stimulus time waveforms [$x(t)$ = two note pitch interval] were presented to a computational model of the AN (Zilany *et al.*, 2009) containing 70 model fibers (CFs: 80−16 000 Hz). From the PSTH, the time-weighted autocorrelation function (ACF) was constructed for each fiber. Individual fiber ACFs were then summed to create a pooled, population-level ACF (ACF_{pop}). The ACF_{pop} was then passed through a series of periodic sieve templates. Each sieve template represents a single pitch (f_0) and the magnitude of its output represents a measure of neural pitch salience at that f_0. Analyzing the outputs across all possible pitch sieve templates ($f_0 = 25-1000$ Hz) results in a running salience curve for a particular stimulus ("Pitch salience"). The peak magnitude of this function was taken as an estimate of neural pitch salience for a given interval (*PS(st)*, where *st* represents the separation of the two notes in semitones). Inset figure showing AN model architecture adapted from Zilany and Bruce (2006), with permission from The Acoustical Society of America.

pitch period and its integer related multiples (Cedolin and Delgutte, 2005; Larsen *et al.*, 2008; Bidelman and Krishnan, 2009). This periodic sieve analysis is essentially a time-domain equivalent to the classic pattern recognition models of pitch in which a "central pitch processor" matches harmonic information contained in the stimulus to an internal template in order to compute the heard pitch (Goldstein, 1973; Terhardt *et al.*, 1982). Each sieve template (representing a single pitch) was composed of 100 μs wide bins situated at the fundamental pitch period ($T = 1/f_0$) and its multiples (i.e., …$T/$ 2, T, $2T$, …, nT), for all $nT < 50$ ms. All sieve templates with f_0 between 25–1000 Hz (2 Hz steps) were used in the analysis (Fig. 2, "Periodic sieve templates"). The salience for a given pitch was estimated by dividing the mean density of spike intervals falling within the sieve bins by the mean density of activity in the whole interval distribution. ACF_{pop} activity falling within sieve "windows" adds to the total pitch salience while information falling outside the "windows" reduces the total pitch salience (Cariani and Delgutte, 1996; Cedolin and Delgutte, 2005).

The output of all sieves was then plotted as a function of f_0 to construct a running neural pitch salience curve. This curve represents the relative strength of all possible pitches present in the AN response that may be associated with different perceived pitches (Fig. 2, "Pitch salience"). The pitch (i.e., f_0) yielding maximum salience was taken as an estimate of a unitary pitch percept.[2] The peak magnitude at this frequency was then recorded for all 220 dyads tested. Even

though only 12 of these pitch combinations are actually found in Western music, this fine resolution allowed for the computation of a continuous function of neural pitch salience over the range of an entire octave. Note that only one metric is used to characterize the neural representation of intervals containing two pitches. Though not without limitations, the use of a single salience metric has been used to describe the perceptual phenomenon whereby listeners often hear musical intervals as being merged or fused into a single unitary percept (e.g., "pitch fusion" or "pitch unity") (DeWitt and Crowder, 1987; Ebeling, 2008). Comparing how neural pitch salience changes across stimuli allows for the direct contrast in AN encoding not only between actual interval relationships found in Western music practice but also those not recognized by traditional musical systems (i.e., nonmusical pitch combinations).

Example ACF_{pop} (cf. ISIH) responses and their corresponding running pitch salience curves (i.e., output of the periodic sieve analyzer) are shown for NH and HI in Fig. 3, A and B respectively. NH model ACF_{pop} and running salience curves bear striking resemblance to those obtained from NH animals using similar stimuli (i.e., two-tone complexes; Tramo *et al.*, 2001; Larsen *et al.*, 2008). HI ACF_{pop} distributions show similar peak locations and magnitudes to those of NH but with the addition of elevated background energy unrelated to the fundamental pitch period or its harmonics. Thus, contrast between energy at harmonically related pitch periods versus the surrounding background is
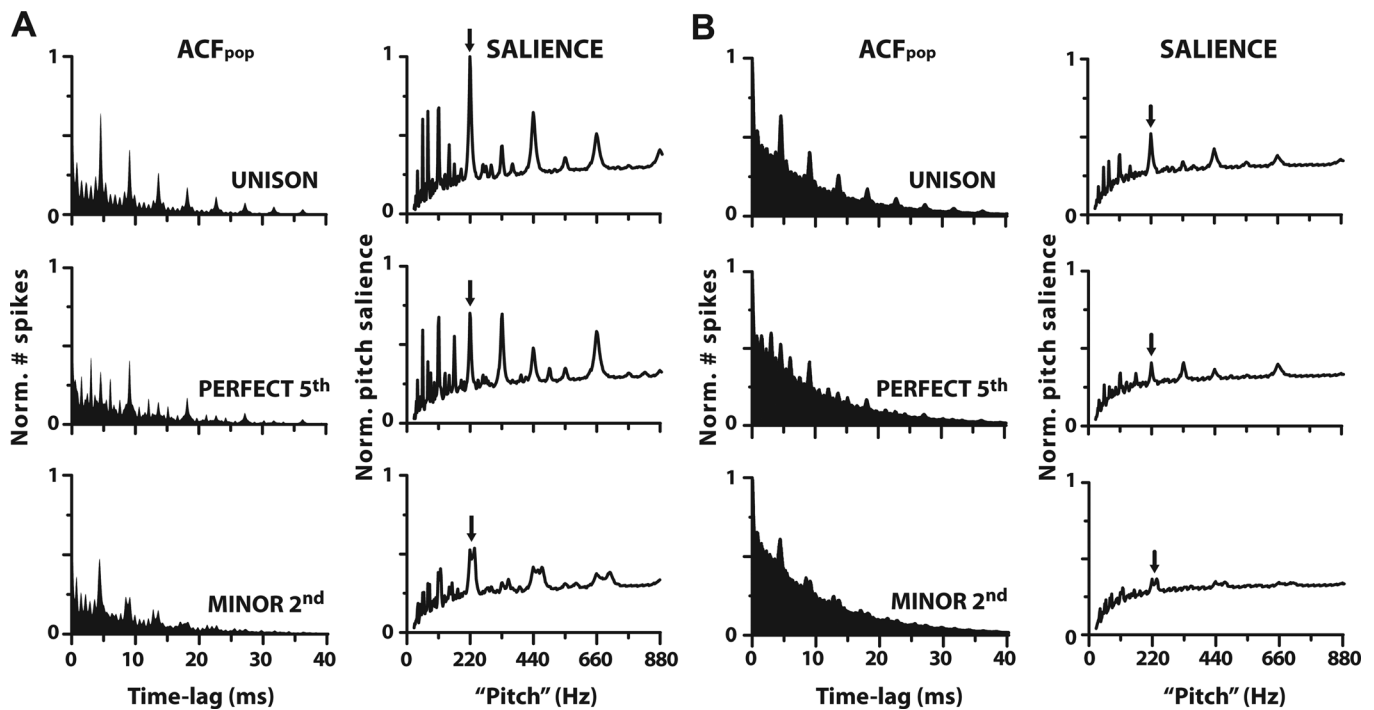
FIG. 3. Pooled autocorrelation functions (cf. ISIH) (left columns) and running pitch salience (i.e., output of periodic sieve analyzer) (right columns) computed for three musical intervals for normal and impaired hearing, A and B, respectively. Pooled ACFs (see Fig. 2, "ACF_pop") quantify periodic activity within AN responses and show clearer, more periodic energy at the fundamental pitch period and its integer related multiples for consonant (e.g., unison, perfect 5th) than dissonant (e.g., minor 2nd) pitch intervals. Running pitch salience curves computed from each ACF_pop quantify the salience of all possible pitches contained in AN responses. Their peak magnitude (arrows) represents a singular measure of salience for the eliciting musical interval and consequently represents a single point in Figs. 4 and 5.

reduced with hearing impairment. Unfortunately, the absence of impaired physiological data in the extant literature precludes the direct comparison between model and animal data for HI results.

F. Neural roughness/beating analysis of AN responses

In addition to measures of salience (i.e., "neural harmonicity") computed via periodic sieve analyses, roughness/beating was computed from AN responses to assess the degree to which this correlate explains perceptual consonance-dissonance judgments (Helmholtz, 1877/1954; Plomp and Levelt, 1965; Terhardt, 1974). Roughness was calculated using the model described by Sethares (1993) which was improved by Vassilakis (2005) to include the effects of register and waveform amplitude fluctuations described in the perceptual literature (Plomp and Levelt, 1965; Terhardt, 1974; Tufts and Molis, 2007). In this model, roughness is computed between any two sinusoids by considering both their frequency and amplitude relationship to one another. Consider frequencies f_1, f_2, with amplitudes A_1, A_2. We define $f_{min} = \min(f_1, f_2)$, $f_{max} = \max(f_1, f_2)$, $A_{min} = \min(A_1, A_2)$, and $A_{max} = \max(A_1, A_2)$. According to Vassilakis (2005, p.141), the roughness (R) between these partials is given by

$$R = X^{0.1}(Y^{3.11})Z \qquad (1)$$

where $X = A_{min} * A_{max}$, $Y = 2A_{min}/(A_{min} + A_{max})$, $Z = e^{-b_1 s(f_{max}-f_{min})} - e^{-b_2 s(f_{max}-f_{min})}$, with parameters $b_1 = 3.5$,

$b_2 = 5.75$, $s = 0.24/(s_1 f_{min} + s_2)$, $s_1 = 0.0207$, $s_2 = 18.96$ chosen to fit empirical data on roughness and musical interval perception (e.g., Plomp and Levelt, 1965; Kameoka and Kuriyagawa, 1969a,b; Vassilakis, 2005). The X term in Eq. (1) represents the dependence of roughness on intensity (amplitude of the added sinusoids), the Y term, the dependence of roughness on the degree of amplitude fluctuation in the signal, and Z, the dependence of roughness/beating on the frequency separation of the two components (Vassilakis, 2005). Total roughness for a complex tone is then computed by summing the individual roughness from all possible (unique) two-tone pairs in the signal. Component magnitudes were first extracted from the spectrum of the pooled PSTH at frequency bins corresponding to locations of harmonics in the input stimulus. The "neural roughness" for each dyad was then computed as the summed contribution of roughness for all pairs of stimulus-related harmonics encoded in AN.

G. Acoustical analysis of periodicity and roughness

In addition to analyzing neural responses to musical intervals and chords, the acoustic properties of these stimuli were analyzed using identical metrics. "Acoustic periodicity" was extracted from stimulus waveforms using the same periodic sieve technique as applied to the neural responses using the average time constant across CFs ($\tau = 15.4$ ms) to weight the stimulus ACF. Similarly, "acoustic roughness" was computed from the spectrum of each stimulus waveform using the same roughness model applied to AN responses (i.e., Vassilakis, 2005). By examining these acoustic
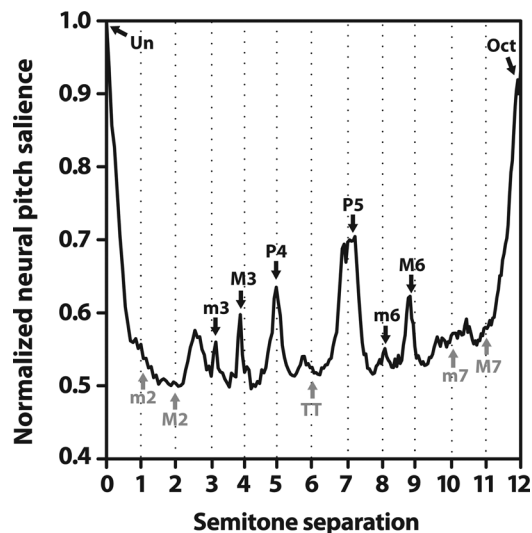
FIG. 4. AN responses correctly predict perceptual attributes of consonance, dissonance, and the hierarchical ordering of musical pitch for normal hearing. Neural pitch salience is shown as a function of the number of semitones separating the interval's lower and higher pitch over the span of an octave (i.e., 12 semitones). The pitch classes recognized by the equal tempered Western music system (i.e., the 12 semitones of the chromatic scale) are demarcated by the dotted lines and labeled along the curve. Consonant musical intervals (black) tend to fall on or near peaks in neural pitch salience whereas dissonant intervals (gray) tend to fall within trough regions, indicating more robust encoding for the former. However, even among intervals common to a single class (e.g., all consonant intervals), AN responses show differential encoding resulting in the hierarchical arrangement of pitch typically described by Western music theory (i.e., Un > Oct > P5, > P4, etc.). All values are normalized to the maximum of the curve which was the unison.

correlates in conjunction with neural correlates we elucidate the relative importance of each of the primary theories postulated in consonance-dissonance perception.

III. RESULTS

A. AN neural pitch salience reveals differential encoding of musical intervals

Neural pitch salience as a function of the number of semitones separating the interval's lower and higher tones is shown in Fig. 4. Pitch combinations recognized by the Western music system (i.e., the 12 semitones of the equal tempered chromatic scale demarcated by dotted lines)[3] tend to fall on or near peaks in the function in the case of consonant intervals, or within trough regions in the case of dissonant musical intervals.[4] The relatively larger magnitudes for consonant over dissonant musical intervals indicate more robust representation for the former (e.g., compare P5 to the nearby TT). Interestingly, among the intervals common to a single class (e.g., all consonant intervals: Un, m3, M3, P4, P5, m6, M6, Oct) AN responses show differential encoding in that pitch salience magnitudes are graded resulting in the hierarchical arrangement of pitch typically described by Western music theory (i.e., Un > Oct > P5, > P4, etc.). In addition, intervals with larger neural pitch salience (e.g., Un, Oct, P5) are also the pitch combinations which tend to produce higher behavioral consonance ratings, i.e., are more pleasant sound-
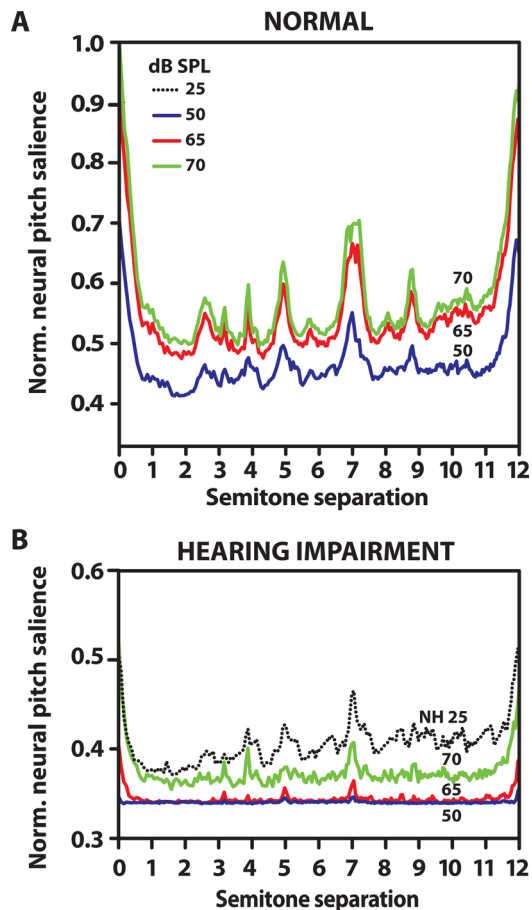

FIG. 5. (Color online) Normal-hearing (A) and hearing-impaired (B) estimates of neural pitch salience as a function of level. Little change is seen in the NH "consonance curve" with decreasing stimulus presentation level. Level effects are more pronounced in the case of HI where consonant peaks diminish with decreasing intensity. Even after equating sensation levels, NH responses still show a greater contrast between consonant peaks and dissonant troughs than HI responses indicating that the reduced contrast seen with HI cannot simply be explained in terms of elevated hearing thresholds. For ease of SL comparison, NH at 25 dB SPL (dotted line) is plotted along with the HI curves in B (i.e., NH at 25 dB SPL and HI at 70 dB SPL are each ~25 dB SL). All values have been normalized to the maximum of the NH 70 dB SPL curve, the unison.

ing to listeners (Plomp and Levelt, 1965; Kameoka and Kuriyagawa, 1969b; Krumhansl, 1990).

B. Hearing impairment reduces contrast in neural pitch salience between musical intervals

Neural pitch salience computed for NH (panel A) and HI (panel B) AN responses are shown in Fig. 5 for several presentation levels (note the difference in ordinate scale for the HI panel). For normal hearing, pitch salience remains relatively robust despite decreases in stimulus intensity. That is, the contrast between encoding of consonant (peaks) and dissonant (troughs) intervals remains relatively distinct and invariant to changing SPL. In comparison, level effects are more pronounced in the case of HI where consonant peaks diminish with decreasing intensity. It is possible that the more compressed "peakedness" (i.e., reduced peak to trough ratio) in the HI condition results from the reduction in audibility due to hearing loss. However, even after equating
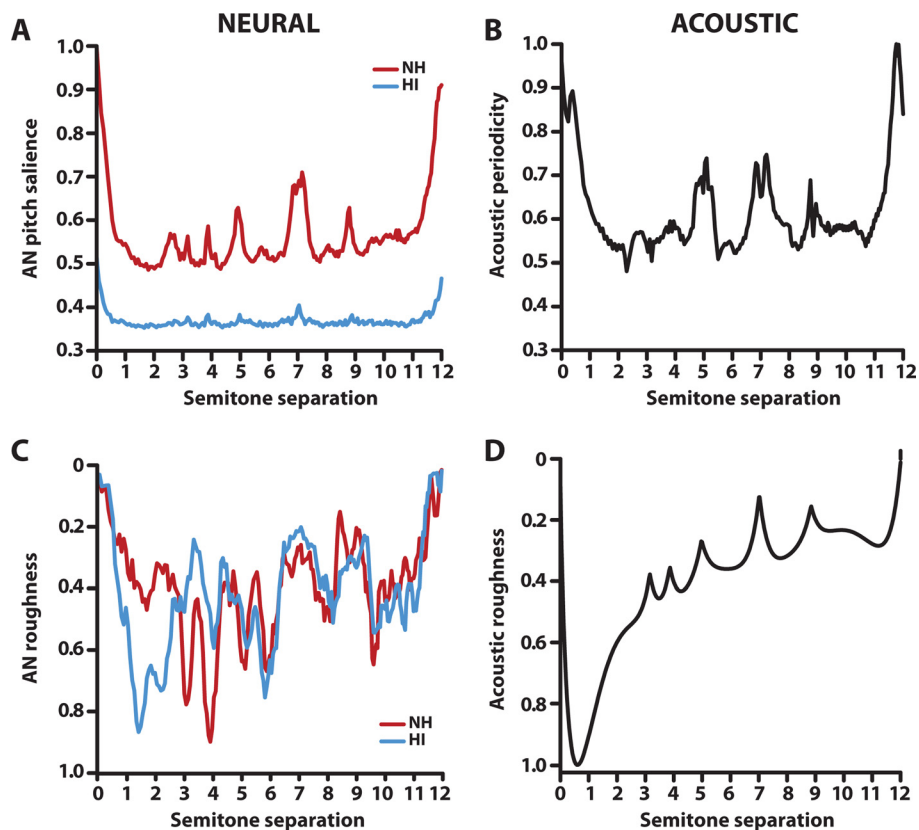
FIG. 6. (Color online) Continuous plots of acoustic and neural correlates of musical interval perception. All panels reflect a presentation level of 70 dB SPL. Ticks along the abscissa demarcate intervals of the equal tempered chromatic scale. Neural pitch salience (A) measures the neural harmonicity/periodicity of dyads as represented in AN responses (same as Fig. 5), and is shown for both normal-hearing (NH) and hearing-impaired (HI) conditions. Similarly, periodic sieve analysis applied to the acoustic stimuli quantifies the degree of periodicity contained in the raw input waveforms (B). Consonant intervals generally evoke more salient, harmonic neural representations and contain higher degrees of acoustic periodicity than adjacent dissonant intervals. Neural (C) and acoustic (D) roughness quantify the degree of amplitude fluctuation/beating produced by partials measured from the pooled PSTH and the acoustic waveform, respectively. Dissonant intervals contain a greater number of closely spaced partials which produce more roughness/beating than consonant intervals in both the neural and acoustic domain. See Fig. 4 for interval labels.

sensation levels, NH responses still show a greater contrast between consonant peaks and dissonant troughs than HI responses (e.g., compare NH at 25 dB SPL and HI at 70 dB SPL: both 25 dB SL). In other words, the reduced contrast seen with impairment cannot simply be explained in terms of elevated hearing thresholds. To quantify the difference in contrast between NH and HI, the "peakedness" was measured at each interval by computing the standard deviation of the salience curve within one-semitone bins centered at each chromatic interval (excluding the unison and octave). Results of a paired samples t-test showed that the NH curve was much more "peaked" than the HI curve at both equal SPL [$t_{10} = 5.02$, $p < 0.001$] and SL [$t_{10} = 2.75$, $p = 0.0205$].

C. Examination of alternate correlates of consonance−dissonance: Acoustics and neural roughness

Continuous functions of neural pitch salience (as in Fig. 5), acoustic periodicity, neural roughness, and acoustic roughness are shown in Fig. 6. Qualitatively, neural and acoustic functions show similar shapes for both periodicity/harmonicity (A vs B) and roughness/beating (C vs D). That is, consonant intervals generally contain higher degrees of periodicity and subsequently evoke more salient, harmonic neural representations than adjacent dissonant intervals. Yet, individual chromatic intervals seem better represented by AN pitch salience than by pure acoustic periodicity in that neural responses show graded, hierarchical magnitudes across intervals (e.g., P5 > P4) which is not generally observed in the raw acoustic waveforms (e.g., P5 = P4). For measures of roughness, neural responses only grossly mimic

the pattern observed acoustically. In general, dissonant intervals (e.g., 1 semitone, m2) contain a greater number of closely spaced partials which produce more roughness/beating than consonant intervals (e.g., 7 semitones, P5) in both the neural and acoustic domain, consistent with perceptual data (e.g., Terhardt, 1974). Qualitatively, HI neural roughness more closely parallels acoustic roughness than it does for NH, especially for intervals of 1–2 (m2, M2) and 3−4 (m3, M3) semitones.

D. Correlations between neural/acoustic measures and behavioral rankings of intervals

To assess the correspondence between acoustic and neural correlates and perceptual rankings of musical intervals, 13 values were extracted from the continuous curves (Fig. 6) at locations corresponding to the 13 semitones (including the unison) of the chromatic scale (e.g., each abscissa tick) and then regressed against the corresponding behavioral consonance scores from NH and HI listeners reported by Tufts et al. (2005). These results are shown in Fig. 7.

Both neural (A) and acoustic (B) periodicity (cf. harmonicity/fusion) show positive correlations with behavioral data. The most consonant musical intervals (e.g., Un, Oct, P5) contain higher degrees of periodicity in their raw waveforms than dissonant intervals (e.g., m2, M2, M7). Similar trends are seen for neural pitch salience. While the exact ordering of intervals differs slightly between normal and impaired conditions, in both cases, the consonant intervals tend to produce higher "neural rankings" than dissonant intervals (e.g., M7, TT, m2) and likewise are also judged more pleasant sounding by listeners. While there is an
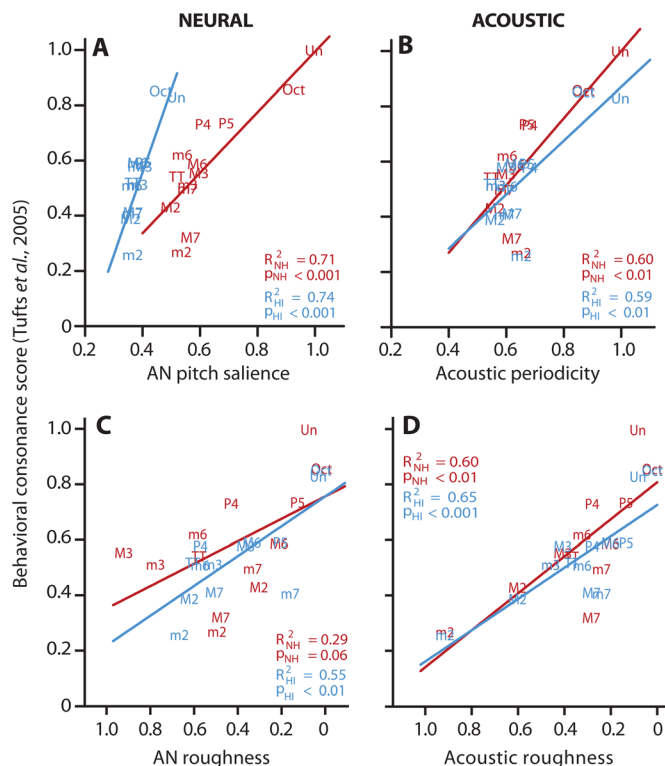
FIG. 7. (Color online) Correlations between neural/acoustic correlates and behavioral consonance scores of equal tempered chromatic intervals for normal and impaired hearing. Both AN pitch salience (A) and acoustic waveform periodicity (B) show positive correlations with behavioral consonance judgments. That is, consonant intervals, judged more pleasant by listeners, are both more periodic and elicit larger neural pitch salience than dissonant intervals. Neural and acoustic roughness (C and D) are negatively correlated with perceptual data (note reversed abscissa) indicating that intervals deemed dissonant contain a larger degree of roughness/beating than consonant intervals. The explanatory power (R^2) of each correlate reveals its strength in predicting the perceptual data: AN neural roughness < (acoustic periodicity ≈ acoustic roughness) < AN neural pitch salience (i.e., harmonicity). Of the neural measures, only AN pitch salience produces the correct ordering and systematic clustering of consonant and dissonant intervals, e.g., maximal separation of the unison (most consonant interval) from the minor 2nd (most dissonant interval). Perceptual data reproduced from Tufts et al. (2005).

overall similarity between normal and impaired orderings, across intervals, neural rankings appear more compressed with HI than in NH (especially for intervals other than the Un, Oct, and P5) consistent with the compressed nature of perceptual responses reported for hearing-impaired listeners (Tufts et al., 2005). Though neural rank orders are derived from responses at the level of the auditory nerve, they show close agreement to rankings stipulated by Western music theory as well as those obtained from human listeners in many psychophysical studies on consonance and dissonance (see Fig. 6 in Schwartz et al., 2003). The significant relationships between both neural pitch salience ($R^2_{NH} = 0.71$; $R^2_{HI} = 0.74$) and acoustic periodicity ($R^2_{NH} = 0.60$; $R^2_{HI} = 0.59$) with behavioral data suggests that one can correctly predict the ordering of perceptual consonance ratings from either of these measures.

In contrast to measures of acoustic periodicity and neural salience, neural and acoustic roughness (C-D) are negatively correlated with perceptual data indicating that
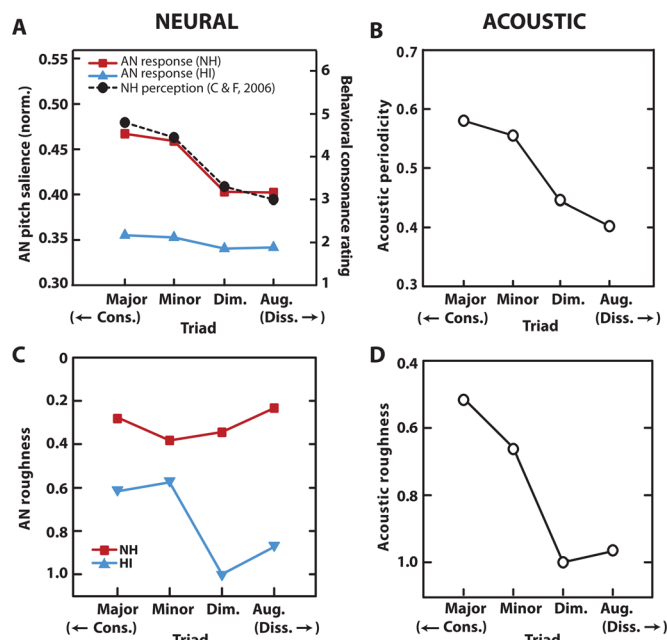

FIG. 8. (Color online) Acoustic and neural correlates of behavioral chordal sonority ratings. Presentation level was 70 dB SPL. Neural pitch salience (A) derived from NH AN responses (squares) show close correspondence to perceptual ratings of chords reported for nonmusician listeners (Cook and Fujisawa, 2006; black circles). Salience values have been normalized with respect to the NH unison presented at 70 dB SPL. Similar to the dyad results, HI estimates for chords (triangles) indicate that the overall differences between triad qualities are muted with hearing loss. Roughness computed from AN (C) shows that only HI responses contain meaningful correlates of harmony perception; NH neural roughness does not predict the ordering of behavioral chordal ratings. In contrast, both acoustic periodicity (B) and roughness (D) provide correlates of chord perception and are inversely related; consonant triads contain larger degrees of periodicity and relatively less roughness than dissonant triads.

intervals deemed dissonant contain a larger degree of roughness/beating than consonant intervals in both their raw waveforms and AN representations. Compared to acoustic roughness which shows relatively close correspondence with NH and HI perceptual data ($R^2_{NH} = 0.60$; $R^2_{HI} = 0.65$), AN roughness is a much poorer predictor of the behavioral data ($R^2_{NH} = 0.29$; $R^2_{HI} = 0.55$). Across hearing configurations, neural roughness is a better predictor of behavioral data for HI listeners than it is for NH.

The explanatory power of these four correlates reveals their strength in predicting perceptual response to musical intervals. While all four are able to predict behavioral scores to some degree, ordered by their R^2, we find AN neural roughness < (acoustic periodicity ≈ acoustic roughness) < AN neural pitch salience. Thus, all acoustic and roughness measures are subsidiary to neural pitch salience (i.e., degree of harmonicity) in their ability to explain perceptual judgments of musical intervals (cf. McDermott et al., 2010).

E. Neural and acoustic correlates of chordal sonority

Neural pitch salience for the four triads computed from normal (squares) and impaired (triangles) responses are shown in Fig. 8(A). As with the dyads, the sieve analysis

identified 220 Hz as the most salient pitch for each triad. For comparison, behavioral ratings of chords reported by Cook and Fujisawa (2006) for normal hearing, non-musician listeners are also shown (circles).

Qualitatively, the pitch salience derived from normal AN responses mimics the perceptual ratings of chords reported by listeners (i.e., major > minor ≫ diminished > augmented). The high correspondence between AN neural pitch salience and perceptual chordal sonority ratings indicates that as with musical intervals, behavioral preferences for certain chords (e.g., major/minor) is predicted from basic AN response properties. Similar to results with single pitch intervals, hearing-impaired estimates indicate that the overall differences between triad qualities are muted, i.e., SNHL may reduce perceptual contrasts between musical chord types. To our knowledge, there are no published studies examining chordal ratings in HI listeners so a direct comparison between impaired model predictions and behavioral results cannot be made.

As with two-tone dyads, acoustic periodicity [Fig. 8(B)], neural roughness [Fig. 8(C)], and acoustic roughness [Fig. 8(D)] make qualitatively similar predictions for the perceptual ordering of chordal triads. The acoustic waveforms for consonant chords (i.e., major and minor) are more periodic than their dissonant counterparts (i.e., diminished and augmented). In addition, the two dissonant triads contain higher degrees of both acoustic and neural roughness/beating than consonant triads (at least for HI model responses), consistent with the unpleasant percept generated by these chords (Roberts, 1986; McDermott et al., 2010; Bidelman and Krishnan, 2011). As with dyads, AN neural roughness does a relatively poor job in predicting behavioral chordal sonority ratings for NH listeners.

IV. DISCUSSION

A. AN responses predict behavioral consonance−dissonance and the hierarchical ordering of musical intervals

Examining temporal response properties, we found that neural phase-locked activity in the AN appears to contain adequate information relevant to the perceptual attributes of musical consonance−dissonance. Pitch combinations defined musically as being consonant are also preferred by listeners for their pleasant sounding qualities (Dowling and Harwood, 1986). Here, we have shown that these same intervals (and chords) seem to elicit differential representations at the level of the AN. Overall, we found the magnitude of neural pitch salience elicited by consonant intervals and chords was larger than that generated by dissonant relationships (Fig. 4). These findings are consistent with previous results obtained from single-unit recordings in live animals (McKinney et al., 2001; Tramo et al., 2001) and brainstem responses recorded in humans (Bidelman and Krishnan, 2009, 2011), which illustrate preferential encoding of consonant over dissonant pitch relationships.

Fundamental to musical structure is the idea that scale tones are graded in terms of their functional importance. Of particular interest here is the similar graded nature of neural activity we observe in AN responses. Musical pitch relationships are not encoded in a strictly binary manner (i.e., consonant versus dissonant) but rather, seem to be processed differentially based on their degree of perceptual consonance (e.g., Kameoka and Kuriyagawa, 1969a; Krumhansl, 1990). This is evident by the fact that even within a given class (e.g., all consonant dyads) intervals elicit graded levels of pitch salience (Figs. 4 and 5). Indeed, we also find that AN responses predict the rank ordering of musical intervals reported by listeners (e.g., compare Fig. 7, present study, to Fig. 6, Schwartz et al., 2003). Taken together, our results suggest that the distribution of temporal firing patterns at a subcortical level contains adequate information to, at least in part, explain the degree of perceptual pleasantness of musical units (cf. Tramo et al., 2001; Bidelman and Krishnan, 2009).

The fact that we observe correlates of consonance−dissonance in cat model responses suggests that these effects are independent of musical training (for perceptual effects of music experience, see McDermott et al., 2010), long-term enculturation, and memory/cognitive capacity. Basic sensory encoding of consonance−dissonance then, may be mediated by domain-general pitch mechanisms not specific to humans (Trehub and Hannon, 2006). It is interesting to note that intervals and chords deemed more pleasant sounding by listeners are also more prevalent in tonal composition (Vos and Troost, 1989). A neurobiological predisposition for these simpler, consonant relations may be one reason why such pitch combinations have been favored by composers and listeners for centuries (Burns, 1999).

B. Effects of hearing impairment on AN representation of musical pitch

Similar patterns were found with hearing loss albeit much more muted in nature. Consistent with the behavioral data of Tufts et al. (2005), hearing impairment did not drastically change the neural rank ordering of musical intervals (Fig. 7). In other words, our data do not indicate that consonant intervals suddenly become dissonant with hearing loss. Rather, impairment seems to act only as a negative blurring effect. HI pitch salience curves mimicked NH profiles but were significantly reduced in terms of their "peakedness" (Fig. 5). The compression of the HI consonance curve suggests that impaired cochlear processing minimizes the contrast between neural representations of consonance and dissonance. Behaviorally, listeners with moderate SNHL have difficulty distinguishing the esthetics of intervals (i.e., consonance versus dissonance) as well as NH listeners (Tufts et al., 2005). Insomuch as our salience metric represents true peripheral encoding of musical pitch, the reduction in neural contrast between consonant and dissonant pitch relationships we find in AN responses may explain the loss of perceptual contrast observed for HI listeners (Tufts et al., 2005).

Tufts et al. (2005) posit that this loss of perceptual musical contrast may be related to the fact that HI listeners often experience a reduced sense of pitch salience (Leek and Summers, 2001). Reduced salience may ultimately lessen the degree of fusion of intervals (i.e., how unitary they

sound; DeWitt and Crowder, 1987) to the point that they no longer contrast with adjacent dissonant intervals. Indeed, our neural salience metric (~harmonicity) is closely related to the psychophysical phenomenon of "pitch fusion" or "tonal affinity" which measures the degree to which two pitches resemble a single tone (DeWitt and Crowder, 1987) or harmonic series (Gill and Purves, 2009; McDermott *et al.*, 2010).[5] It is possible then, that the reduced neural pitch salience estimates for the impaired AN manifest perceptually as either a reduction in the overall salience of an interval or a weakened sense of fusion. Both of these distortions may hinder the ability of HI listeners to fully appreciate the variations between musical consonance and dissonance (Tufts *et al.*, 2005).

The fact that this muted contrast was observed despite controlling for audibility (see Fig. 5 at equal dB SL) suggests this effect resulted from characteristics of the impairment other than purely elevated hearing thresholds (though both are likely at play). For this reason, hearing aids alone are unlikely to restore the neurophysiological and perceptual contrasts for music enjoyed by normal hearing individuals. One possible explanation for the reduced pitch salience magnitudes we observed in HI responses could be a reduction in the available temporal information due to weakened AN phase-locking (Arehart and Burns, 1999). However, with few exceptions (e.g., Woolf *et al.*, 1981) the majority of physiological evidence does not support this explanation (e.g., Harrison and Evans, 1979; Miller *et al.*, 1997; Heinz *et al.*, 2010; Kale and Heinz, 2010). A more likely candidate is the reduction in frequency selectivity that accompanies hearing loss (Liberman and Dodds, 1984). Pitch templates akin to the periodic sieve used here may arise due to the combined effects of basic cochlear filtering and general neuronal periodicity detection mechanisms (cf. Shamma and Klein, 2000; Ebeling, 2008). With the relatively broader peripheral filtering from HI, more harmonics of the stimulus may interact within single auditory filters, ultimately reducing the number of independent information channels along the cochlear partition. Interacting harmonic constituents may have the effect of "jittering" the output of such a temporal based mechanism thereby reducing the correspondence with any one particular pitch template, i.e., fewer coincidences between harmonically related pitch periods (e.g., Ebeling, 2008). Indeed, we find reduced output of the sieve analyzer with hearing impairment [i.e., lower neural pitch salience; Fig. 5(B)] indicating that neural encoding becomes less "template-like" with broadened tuning.

C. Potential limitations and comparisons with previous work

In contrast to the neural pitch salience metric used in the present work, classic models of pitch salience typically operate on the acoustic waveform and estimate a fundamental bass by considering the common denominator (residue F0) between the pitches in a chord/dyad (e.g., Terhardt *et al.*, 1982). Often, this estimated bass is lower than the physical frequencies in the stimulus itself. Though fundamentally different in implementation, our neural salience curves do show

energy at subharmonics (e.g., Fig. 3). Yet, in contrast we find these peaks are much lower than the maximum measured at the lowest F0 of each dyad/triad (220 Hz). Using the same analysis technique (i.e., periodic sieve templates) applied to animal AN recordings, Larsen *et al.* (2008) similarly reported higher salience for the lower pitch of a two-tone complex [their Fig. 5(D)] and not, say, a subharmonic (cf. fundamental bass). The fact that model (present work) and neural (Larsen *et al.*, 2008) data do not predict lower subharmonics, as in Terhardt's acoustic model, is likely a consequence of the exponential functions used to weight fiber ACFs. These functions were employed to give greater precedence to the shorter pitch intervals (Cedolin and Delgutte, 2005) and account for the lower limit of musical pitch (Pressnitzer *et al.*, 2001). One consequence of this weighting then, is that it is unlikely a subharmonic would be the estimated dominant pitch.

Yet, even when applying the same salience metric to both acoustic and neural data, we find that neural salience correlates better with perception than raw acoustical pitch salience (Fig. 7). We speculate that the improved correspondence of the neural data may result from inherent redundancies provided by the peripheral auditory system. Though waveform periodicity is computed via the same harmonic sieve technique, this is arguably only a single-channel operation. In comparison, salience derived from neural responses includes information from 70 fibers each with a CF-dependent weighting function. As such, pitch salience computed neurally is a multi-channel operation. The fact that numerous fibers may respond to multiple harmonics of the stimulus may mean that certain periodicities are reinforced relative to what is contained in the acoustic waveform alone and hence show better connection to percept.

While subtleties may exist between human and cat peripheral processing, e.g., auditory filter bandwidths and absolute audibility curves, it is unlikely these discrepancies would alter our results and conclusions. The effect of filter bandwidth emerges as changes in absolute pitch salience; the estimated pitches (e.g., the peak locations in Fig. 3), remain the same. This is evident by the fact that normal and impaired results (which represent differences in auditory filter bandwidths) produce similar rank orderings of the dyads, albeit with different amounts of "contrast" between intervals (Figs. 5 and 7). If in fact human filter bandwidths are indeed 2-3 times sharper than other mammals (Shera *et al.*, 2002), our estimates of pitch salience reported presently for cat AN responses may underestimate the salience, and hence the true degree of consonance−dissonance represented in human AN. While absolute hearing thresholds between human and cat also differ, the use of relative levels to compute audiometric thresholds (dB HL; Fig. 1) ensures that our results obtained in the cat model can be compared to psychophysical data reported for normal-hearing and hearing-impaired humans (Tufts *et al.*, 2005).

One potential limitation of our approach was the sole use of high-SR fibers. Fiber thresholds of the present model (Zilany *et al.*, 2009) were designed to match the lowest recorded AN fiber thresholds at each CF (Miller *et al.*, 1997; Zilany and Bruce, 2007). Only high-SR model fibers have

been validated against animal data. Using similar stimuli (i.e., two-tone complexes) and analysis procedures (i.e., periodic sieve), Larsen *et al.* (2008) have shown stronger representation of pitch salience in low- and medium-SR fibers as compared to high-SR fibers. Thus, the use of only high-SR fibers may also contribute to an underestimate of absolute pitch-salience magnitudes. The fact that we still observe relative differences between intervals and types of hearing configuration (Fig. 5) suggests that inclusion of low- and medium-SR fibers would only magnify what is already captured in high-SR responses.

In addition, though AN neural responses generally correctly predict listeners' rank ordering of either intervals (Tufts *et al.*, 2005) or triads (Roberts, 1986; Cook and Fujisawa, 2006), Figs. 7 and 8 respectively, there are subtle nuances which conflict with the perceptual literature. For example, neural pitch salience incorrectly predicts the major 2nd (two semitones) to be the most dissonant dyad despite the fact that listeners consistently rate the minor 2nd (one semitone) as most dissonant (see Fig. 6 in Schwartz *et al.*, 2003). Such discrepancies, though small, indicate that neural responses may not accurately reflect the *absolute* percepts across these musical stimuli. Furthermore, McDermott *et al.*, (2010) found that human listeners, on average, rate the major chord as more pleasant than any two-tone interval.[6] At odds with this behavioral result, we find that AN neural pitch salience for the major triad ($\sim$0.47) is lower than the salience of any two-tone interval [compare Fig. 6(A) to 8(A)]. Interestingly, measures of acoustic periodicity also predict similar results, i.e., triads < dyads [compare Fig. 6(B) to 8(B)]. The fact that chords produce lower neural salience (and acoustic periodicity) compared to any of the intervals is a direct consequence of the periodic sieve, a technique which measures the degree to which multiple pitches match a single harmonic series (cf. Stumpf, 1890; Ebeling, 2008). Triads contain multiple (three) and more complex periodicities than any individual two-tone interval. Thus, the measured AN salience (cf. neural harmonicity) is, *ipso facto,* lower for chords than intervals. The fact that listeners nevertheless rate the major chord higher than any dyad (e.g., McDermott *et al,* 2010) may reflect a "cognitive override" to this sensory representation we find in AN. Indeed, in addition to "consonance," the major triad often invokes a connotation (perhaps subconsciously) of positive valence which may influence or enhance its perceived pleasantness (Cook and Fujisawa, 2006).

D. Is consonance−dissonance an acoustical, psychophysical, or neural phenomenon?

Theories of musical consonance−dissonance have ranged from purely acoustical to purely cognitive explanations. Yet, despite their important role in both written and heard music there has not, as of yet, been an entirely satisfactory explanation for the foundations of these musical attributes. One may ask whether consonance-dissonance distinctions originate based on physical properties of the acoustic signal (e.g., frequency ratios), psychophysical mechanisms related to cochlear frequency analysis (e.g., critical band), neural synchronization/periodicity encoding, cognitive constructs (e.g., categorical perception), or a combination of these factors.

Acoustically, tones of consonant pitch combinations share more harmonics than those of dissonant relationships and contain a higher degree of periodicity. As such, one long held theory of consonance−dissonance relates to the raw acoustics of the input waveform (Rameau, 1722/1971; Gill and Purves, 2009; McDermott *et al.*, 2010). Tones with common partials would tend to interact less on the basilar membrane as low-order harmonic constituents pass through distinct auditory filters (cf. critical bands). Together, these features would tend to optimize perceptual fusion and harmonicity (Gill and Purves, 2009; McDermott *et al.*, 2010), minimize roughness and beating (Terhardt, 1974), and lead to more agreeable sounding intervals (i.e., consonance). While these acoustical and biomechanical parameters may provide necessary information for distinguishing consonance and dissonance, they neglect the mechanisms by which this information is transmitted along the auditory pathway from cochlea to cortex.

Our results offer evidence that important aspects of musical pitch percepts may emerge (or are at least represented) in temporal-based, neurophysiologic processing. We found that consonant pitch intervals generate more periodic, harmonically related phase-locked activity than dissonant pitch intervals (Figs. 4 and 5). For consonant dyads, interspike intervals within population AN activity occur at precise, harmonically related pitch periods thereby producing maximal periodicity in their neural representation. Dissonant intervals on the other hand produce multiple, more irregular neural periodicities (Fig. 3). The neural synchronization supplied by consonance is likely to be more compatible with musical pitch extraction templates (e.g., Boomsliter and Creel, 1961; Ebeling, 2008; Lots and Stone, 2008) as it provides a more robust, unambiguous cue for pitch.

While neural pitch salience seems to be the best predictor of behavioral data (explaining 70%−75% of the variance, Fig. 7), examination of the alternate theories of consonance−dissonance revealed other (though weaker) correlates in both acoustic periodicity and acoustic/neural roughness. We observed that waveform periodicity and acoustic roughness each predict $\sim$60%−65% of the variance measured in behavioral interval judgments (Tufts *et al.*, 2005). Weaker relationships exist between NH and HI neural roughness measures and their corresponding behavioral data: roughness in NH AN responses explained only $\sim$30% of behavioral ratings (marginally significant) while HI roughness predicts 55% of HI listener's perceptual responses.[7] The increased neural roughness in HI relative to NH responses is consistent with the increase in perceived roughness experienced by HI listeners for lower modulation rates (Tufts and Molis, 2007), a potential side effect of the increased representation of stimulus envelope following sensorineural hearing loss (Moore *et al.*, 1996; Kale and Heinz, 2010). The fact that acoustic periodicity and roughness (and to a lesser degree neural roughness) can account for even part of the variance in behavioral data suggests that these characteristics provide important and useful information to generate consonance-

dissonance percepts. However, the neural representation of pitch salience (cf. harmonicity/tonal fusion) clearly contributes and improves upon these representations as it shows a closer correspondence with the perceptual data. Consistent with recent psychophysical (McDermott *et al.,* 2010) and neurophysiological (Itoh *et al.,* 2010) evidence, the relative inability of neural roughness/beating to explain behavioral ratings suggests it may not be as important a factor in sensory consonance-dissonance as conventionally thought (e.g., Helmholtz, 1877/1954; Plomp and Levelt, 1965; Terhardt, 1974). Indeed, lesion studies indicate a dissociation between roughness and the perception of dissonance as one percept can be impaired independently of the other (Tramo *et al.,* 2001). Taken together, converging evidence between this and other studies suggests (neural) harmonicity is the dictating correlate of behavioral consonance-dissonance judgments (present study; McDermott *et al.,* 2010).

While the acoustic signature of music may supply ample information to initiate consonance−dissonance, it is clear that we observe a transformation of this information to the neural domain at a subcortical level (cf. Bidelman and Krishnan, 2009). As in language, brain networks engaged during music likely involve a series of computations applied to the neural representation at different stages of processing (e.g., Bidelman *et al.,* 2011). We argue that higher-level abstract representations of musical pitch structure are first initiated in acoustics (Gill and Purves, 2009; McDermott *et al.,* 2010). Physical periodicity is then transformed to musically relevant neural periodicity very early along the auditory pathway (AN; Tramo *et al.,* 2001; present study), transmitted, and further processed in subsequently higher levels in the auditory brainstem (McKinney *et al.,* 2001; Bidelman and Krishnan, 2009, 2011). Eventually, this information ultimately feeds the complex cortical architecture responsible for generating (Fishman *et al.,* 2001) and controlling (Dowling and Harwood, 1986) musical percepts.

V. CONCLUSIONS

Responses to consonant and dissonant musical pitch relationships were simulated using a computational auditory-nerve model of normal and impaired hearing. The major findings and conclusions are summarized as follows:

1. At the level of AN, consonant musical intervals/chords elicit stronger neural pitch salience than dissonant intervals/chords.
2. The ordering of musical intervals according to the magnitude of their neural pitch salience tightly paralleled the hierarchical ordering of pitch intervals/chords described by Western music theory suggesting that information relevant to these perceptually salient aspects of music may already exist even at the most peripheral stage of neural processing. In addition, the close correspondence to behavioral consonance rankings suggests that a listener's judgment of pleasant- or unpleasant-sounding pitch intervals may, at least in part, emerge in low-level sensory processing.
3. Similar (but muted) results were found with cochlear hearing impairment. Hearing loss produced a marked reduction in contrast in AN representation between consonant and dissonant pitches providing a possible physiological explanation for the inability of HI listeners to fully differentiate these two attributes behaviorally (Tufts *et al.,* 2005).
4. Neural pitch salience (i.e., harmonicity) predicts behavioral interval/chord preferences better than other correlates of musical consonance-dissonance including acoustic periodicity, acoustic roughness/beating, and neural roughness/beating.
5. Results were obtained from a model of cat AN so the effects we observe are likely independent of musical training, long-term enculturation, and memory/cognitive capacity. Sensory consonance-dissonance may be mediated by general cochlear and peripheral neural mechanisms basic to the auditory system.
6. It is possible that the choice of intervals and tuning used in compositional practice may have matured based on the general processing and constraints of the mammalian auditory system.

[1]The long held theory that the ear prefers simple ratios is no longer tenable when dealing with modern musical tuning systems. The ratio of the consonant perfect 5th (equal temperament), for example, is 442:295, hardly a small integer relationship.

[2]With few exceptions, the global maximum of the salience function was located at 220 Hz, the root of the dyad/chord (see "Salience" curves, Fig. 3). For some intervals, maximal salience occurred at frequencies other than the root. In the major 3rd and 6th, for instance, the sieve analysis predicted the higher of the two pitches as most salient, 277 Hz and 370 Hz, respectively. Pilot simulations indicated that tracking either changes in salience at 220 Hz or the global maximum of the function produced indistinguishable results from those shown in Figs. 4 and 5; the measured salience magnitudes in either case were nearly identical. Thus, we chose to track maximum salience because it does not assume, *a priori,* knowledge of the stimulus f_0s.

[3]We report data only for equal-tempered intervals since modern Western music is based exclusively on this system of tuning. However, in reality, pitch relationships from any tuning system can be extracted from the continuous neural pitch salience curve of Fig. 4. Although pleasantness ratings for intervals in equal and perfect temperament are similar for most listeners, we find that AN neural pitch salience does reveal subtle nuances in responses elicited by various temperaments (data not shown). Although their ordering remains unaltered, intervals of just intonation tend to produce slightly larger neural salience than their equal tempered counterparts, especially for the thirds and sixths. These intervals differ the most between the two systems. No differences arise for the unison, fifth, and octave whose frequency relationships are nearly identical in both temperaments.

[4]An obvious peak occurs at 2.5 semitones, an interval which does not occur in music (i.e., it straddles two notes on the piano). While the cause of this anomalous peak is not clear, in supplemental simulations we noted that it

did not occur in higher octaves (e.g., 440 – 880 Hz) suggesting that this particular peak has some dependence on register. Consequently, it may also depend on the critical bandwidth (Plomp and Levelt, 1965) which is known to change with frequency (cf. register).

[5]Harmonicity is likely closely related to the principle of tonal fusion (Stumpf, 1890). In the case of consonance (e.g., perfect 5th), more partials coincide between pitches and consequently may produce more fused, agreeable percepts than dissonant pitch relationships (e.g., minor 2nd) which share few, if any harmonics (Ebeling, 2008). Support for the harmonicity hypothesis stems from experiments examining inharmonic tone complexes which show that consonance is obtained when tones share coincident partials and does not necessarily depend on the ratio of their fundamental frequencies (Slaymaker, 1970).

[6]The results of McDermott et al. (2010) are based on a paradigm in which listeners rated isolated musical sounds while maximizing their judgments along a rating scale. When the ratings were later compared with one another (across all stimuli), the major chord was consistently rated more pleasant than any of the dyads. No study to date has contrasted pleasantness judgments of chords and intervals directly, e.g., as in a paired-comparison task (e.g., Tufts et al., 2005; Bidelman and Krishnan, 2009). Such work would unequivocally demonstrate how listeners' rank dyads compared to triads and provide a link between interval (Kameoka and Kuriyagawa, 1969a,b) and chord (Roberts, 1986) perception.

[7]McKinney et al. (2001) describe a roughness metric applied to responses recorded from cat IC neurons quantified as the rate fluctuation of the neural discharge patterns (i.e., standard deviation of the PSTH). Initial attempts to apply this metric to AN PSTHs in the present study produced spurious results: roughness poorly predicted behavioral data—contrary to the mild correspondence reported in the psychoacoustic literature (e.g., Terhardt, 1974). Furthermore, restricting the AN PSTH analysis to lower modulations ($\sim$30$-$150 Hz) that dominate roughness percepts (Terhardt, 1974; McKinney et al., 2001, p. 2) yielded marginally significant *positive* correlations ($R^2_{NH} = 0.32$, $p=0.04$; $R^2_{HI} = 0.26$, $p=0.07$), which we found indecipherable given that roughness is typically *negatively* correlated with consonance ratings (i.e., more roughness corresponds to less consonance). The inability of these particular metrics to describe correlates of consonance-dissonance in AN as compared to IC responses (McKinney et al., 2001) can be attributed to at least two important differences. First, IC correlates of dissonance were primarily observed for roughness extracted from *onset units* which were phase-locked to the difference (i.e., beat) frequency of two-tone intervals; roughness from *sustained units* (which are closer to AN response characteristics) showed little correlation with perceptual ratings. Second, IC neurons show inherent bandpass selectivity to low modulation frequencies (e.g., $\sim$30$-$150 Hz), whereas AN fibers show lowpass modulation selectivity (Joris et al., 2004) and thus do not "highlight," per se, the important modulations which dominate roughness perception. The fact that restricting AN PSTH modulations to 30$-$150 Hz did not produce the expected negative correlation with consonance scores suggests that this second factor may not be as significant as the first (i.e., differences in unit type).

Aldwell, E., and Schachter, C. (2003). *Harmony & Voice Leading* (Thomson/Schirmer, Belmont, CA), pp. 1–656.

Arehart, K. H., and Burns, E. M. (1999). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss," J. Acoust. Soc. Am. 106, 993–997.

Bernstein, J. G., and Oxenham, A. J. (2005). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. 117, 3816–3831.

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). "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., and 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., and Krishnan, A. (2011). "Brainstem correlates of behavioral and compositional preferences of musical harmony," NeuroReport 22, 212–216.

Boomsliter, P., and Creel, W. (1961). "The long pattern hypothesis in harmony and hearing," J. Music Theory 5, 2–31.

Brattico, E., Tervaniemi, M., Naatanen, R., and Peretz, I. (2006). "Musical scale properties are automatically processed in the human auditory cortex," Brain Res. 1117, 162–174.

Brooks, D. I., and Cook, R. G. (2010). "Chord discrimination by pigeons," Music Percept. 27, 183–196.

Bruce, I. C., Sachs, M. B., and Young, E. D. (2003). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am. 113, 369–388.

Budge, H. (1943). *A Study of Chord Frequencies* (Teachers College, Columbia University, NY), pp. 1–82.

Burns, E. M. (1999). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, San Diego), pp. 215–264.

Cariani, P. A. (2004). "A temporal model for pitch multiplicity and tonal consonance," in *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8)*, edited by S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (SMPC, Evanston, IL), pp. 310–314.

Cariani, P. A., and Delgutte, B. (1996). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. 76, 1698–1716.

Cazden, N. (1958). "Musical intervals and simple number ratios," J. Res. Music Educ. 7, 197–220.

Cedolin, L., and Delgutte, B. (2005). "Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve," J. Neurophysiol. 94, 347–362.

Chasin, M. (2003). "Music and hearing aids," Hear. J. 56, 36–41.

Cook, N. D., and Fujisawa, T. X. (2006). "The psychophysics of harmony perception: Harmony is a three-tone phenomenon," Empirical Musicol. Rev. 1, 1–21.

DeWitt, L. A., and Crowder, R. G. (1987). "Tonal fusion of consonant musical intervals: the oomph in Stumpf," Percept. Psychophys. 41, 73–84.

Dowling, J., and Harwood, D. L. (1986). *Music Cognition* (Academic, San Diego), pp. 1–239.

Ebeling, M. (2008). "Neuronal periodicity detection as a basis for the perception of consonance: a mathematical model of tonal fusion," J. Acoust. Soc. Am. 124, 2320–2329.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., and Steinschneider, M. (2001). "Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans," J. Neurophysiol. 86, 2761–2788.

Foss, A. H., Altschuler, E. L., and James, K. H. (2007). "Neural correlates of the Pythagorean ratio rules," NeuroReport 18, 1521–1525.

Gill, K. Z., and Purves, D. (2009). "A biological rationale for musical scales," PLoS One 4, 1–9.

Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. 54, 1496–1516.

Harrison, R. V., and Evans, E. F. (1979). "Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig," Arch. Otolaryngol. 224, 71–78.

Heinz, M. G., Swaminathan, J., Boley, J. D., and Kale, S. (2010). "Across-fiber coding of temporal fine-structure: Effects of noise-induced hearing loss on auditory-nerve responses," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 621–630.

Helmholtz, H. (1877/1954). *On the Sensations of Tone* (Ellis AJ, translator) (Dover, New York). (Original work published 1877), pp. 1–859.

Itoh, K., Suwazono, S., and 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.

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). "Neural processing of amplitude-modulated sounds," Physiol. Rev. 84, 541–577.

Kale, S., and Heinz, M. G. (2010). "Envelope coding in auditory nerve fibers following noise-induced hearing loss," J. Assoc. Res. Otolarylngol. 11, 657–673.

Kameoka, A., and Kuriyagawa, M. (1969a). "Consonance theory part I: consonance of dyads," J. Acoust. Soc. Am. 45, 1451–1459.

Kameoka, A., and Kuriyagawa, M. (1969b). "Consonance theory part II: consonance of complex tones and its calculation method," J. Acoust. Soc. Am. 45, 1460–1469.

Krohn, K. I., Brattico, E., Valimaki, V., and Tervaniemi, M. (2007). "Neural representations of the hierarchical scale pitch structure," Music Percept. 24, 281–296.

Krumhansl, C. L. (**1990**). *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York), pp. 1–307.

Larsen, E., Cedolin, L., and Delgutte, B. (**2008**). "Pitch representations in the auditory nerve: two sconcurrent complex tones," J. Neurophysiol. **100**, 1301–1319.

Leek, M. R., and Summers, V. (**2001**). "Pitch strength and pitch dominance of iterated rippled noises in hearing–impaired listeners," J. Acoust. Soc. Am. **109**, 2944–2954.

Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Lots, I. S., and Stone, L. (**2008**). "Perception of musical consonance and dissonance: an outcome of neural synchronization," J. R. Soc., Interface **5**, 1429–1434.

McDermott, J., and Hauser, M. (**2004**). "Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate," Cognition **94**, B11–21.

McDermott, J., and Hauser, M. D. (**2005**). "The origins of music: Innateness, uniqueness, and evolution," Music Percept. **23**, 29–59.

McDermott, J., and Hauser, M. D. (**2007**). "Nonhuman primates prefer slow tempos but dislike music overall," Cognition **104**, 654–668.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). "Individual differences reveal the basis of consonance," Curr. Biol. **20**, 1035–1041.

McKinney, M. F., Tramo, M. J., and Delgutte, B. (**2001**). "Neural correlates of the dissonance of musical intervals in the inferior colliculus," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 83–89.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., Loveday, C., and Bruzzone, M. G. (**2009**). "Functional MRI/event-related potential study of sensory consonance and dissonance in musicians and nonmusicians," NeuroReport **20**, 87–92.

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Springer Handbook of Auditory Research: Pitch Perception*, edited by C. J. Plack, A. N. Popper, and R. Fay (Springer, Berlin), pp. 234–277.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (**1996**). "Effect of loudness recruitment on the perception of amplitude modulation," J. Acoust. Soc. Am. **100**, 481–489.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). "Outer hair cell receptor current and sensorineural hearing loss," Hear. Res. **42**, 47–72.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Plomp, R., and Levelt, W. J. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Rameau, J.-P. (**1722/1971**). *Treatise on Harmony* (P. Gossett, Trans.) (Dover, New York). (Original work published 1722), pp. 1–444.

Roberts, L. (**1986**). "Consonant judgments of musical chords by musicians and untrained listeners," Acustica **62**, 163–171.

Schwartz, D. A., Howe, C. Q., and Purves, D. (**2003**). "The statistical structure of human speech sounds predicts musical universals," J. Neurosci. **23**, 7160–7168.

Sethares, W. A. (**1993**). "Local consonance and the relationship between timbre and scale," J. Acoust. Soc. Am. **94**, 1218–1228.

Shamma, S., and Klein, D. (**2000**). "The case of the missing pitch templates: how harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–2644.

Shera, C. A., Guinan, J. J. Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U. S. A. **99**, 3318–3323.

Slaymaker, F. H. (**1970**). "Chords from tones having stretched partials," J. Acoust. Soc. Am. **47**, 1569–1571.

Stumpf, C. (**1890**). *Tonpsychologie (Tone psychology)* (Hirzel, Leipzig), pp. 1–446.

Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriyama, Y., Takeshita, H., Nakamura, T., and Hashiya, K. (**2010**). "Preference for consonant music over dissonant music by an infant chimpanzee," Primates **51**, 7–12.

Tenney, J. (**1988**). *A History of 'Consonance' and 'Dissonance'* (Excelsior, New York), pp. 1–117.

Terhardt, E. (**1974**). "On the perception of period sound fluctuations (roughness)," Acustica **20**, 215–224.

Terhardt, E., Stoll, G., and Seewann, M. (**1982**). "Algorithm for the extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**, 679–687.

Tramo, M. J., Cariani, P. A., Delgutte, B., and Braida, L. D. (**2001**). "Neurobiological foundations for the theory of harmony in western tonal music," Ann. N. Y. Acad. Sci. **930**, 92–116.

Trehub, S. E., and Hannon, E. E. (**2006**). "Infant music perception: domain-general or domain-specific mechanisms?," Cognition **100**, 73–99.

Tufts, J. B., and Molis, M. R. (**2007**). "Perception of roughness by listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **121**, EL161–167.

Tufts, J. B., Molis, M. R., and Leek, M. R. (**2005**). "Perception of dissonance by people with normal hearing and sensorineural hearing loss," J. Acoust. Soc. Am. **118**, 955–967.

van de Geer, J. P., Levelt, W. J. M., and Plomp, R. (**1962**). "The connotation of musical consonance," Acta Psychol. **20**, 308–319.

Vassilakis, P. N. (**2005**). "Auditory roughness as a means of musical expression," in *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology*, edited by R. A. Kendall and R. W. H. Savage (UCLA Ethnomusicology Publications, Los Angeles), pp. 119–144.

Vos, P. G., and Troost, J. M. (**1989**). "Ascending and descending melodic intervals: Statistical findings and their perceptual relevance," Music Percept. **6**, 383–396.

Watanabe, S., Uozumi, M., and Tanaka, N. (**2005**). "Discrimination of consonance and dissonance in Java sparrows," Behav. Process. **70**, 203–208.

Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of cochlear outer hair cells," Hear. Res. **4**, 335–346.

Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.

Zilany, M. S., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.

Zilany, M. S., and Bruce, I. C. (**2007**). "Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats," J. Acoust. Soc. Am. **122**, 402–417.

Zilany, M. S., Bruce, I. C., Nelson, P. C., and Carney, L. H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics," J. Acoust. Soc. Am. **126**, 2390–2412.