Supporting Information

**Cousineau et al. 10.1073/pnas.1207989109**

**SI Text**

**Audio Examples.** The audio examples of stimuli include: (i) the voice minor second (m2) and perfect fifth (P5) chords of Exp. 1 depicted in Fig. 1; and (ii) one example for each of the harmonic (pure tone, harmonic-subset and octave-spaced), inharmonic (jittered, shifted, and 13st-spaced.), beating (dichotic 0.75st), and nonbeating (dichotic 0.75st) stimuli of Exp. 3.

**Roughness for Natural Stimuli Compared with Synthetic Complex Tones.** Beating-based theories of consonance have argued that the amount of roughness found in consonant chords is greater than that found in consonant chords, and ascribe the unpleasant sensation evoked by dissonant chords to the presence of roughness (1–3). One common feature of these classic studies is that they used chords composed of synthetic complex tones with equal amplitude harmonics. Such tones share the harmonic frequency relations of natural musical stimuli, but differ from them in a number of other ways (amplitude and phase of the partials, attack cues, etc.) and do not represent the variability that can be present in natural music listening situations. To investigate whether the relation between dissonance and roughness previously observed in synthetic tones was also present in the natural stimuli used here, we computed a Roughness Index for each of our stimuli, as well as for stimuli generated from synthetic tones. The stimuli were first passed through a bank of gamma-tone filters (4). The envelope of each filter output was extracted via the Hilbert transform, raised to the 0.3 power to simulate compression, and filtered between 20 and 200 Hz, using eighth-order Butterworth filters, to isolate the modulation rates that are standardly thought to contribute to roughness (5, 6). The Roughness Index was then calculated as the total power remaining in each band. Our synthetic tones consisted of the first six harmonics (with frequencies F0 to 6*F0) with equal amplitude and added in sine phase, as these stimuli have generally been used in models of dissonance based on beating (e.g., in refs. 1 and 3).

The variations of the Roughness Index obtained for chords composed of synthetic complex tones reproduce rather closely the ratings of such chords obtained from subjects in ref. 3. However, it is apparent that (i) the amount of beating physically present in the stimuli is generally larger for synthetic chords than for those generated from natural stimuli, and (ii) differences between intervals are more pronounced for the synthetic stimuli (Fig. S1). The voice stimuli in particular exhibit much less beating than do the other timbres, with little variation from interval to interval despite the large variation across intervals in perceived pleasantness.

To summarize the differences in beating between intervals heard as consonant and intervals heard as dissonant, we averaged the Roughness Index across the consonant and dissonant intervals that contributed to our behavioral measures of interval preference, for each of the three note timbres. The results shown in Fig. S2 illustrate that the roughness differences between consonant and dissonant intervals are indeed far more pronounced for intervals composed of synthetic tones than they are for those composed of naturally occurring instrument sounds.

In addition to the large differences in roughness between synthetic and natural intervals, there are pronounced differences even between saxophone and voice intervals. Consequently, if pleasantness ratings for chords were based even partly on roughness, we should see a difference between ratings for the two types of natural stimuli. Contrary to this prediction, normal-hearing listener’s ratings are quite similar for saxophone and voice intervals (Fig. S3). This finding is consistent with the idea that roughness has little bearing on their perception of chord pleasantness. For amusics however, ratings for voice intervals are consistently higher than ratings for saxophone intervals. One possible explanation for these results is that control subjects largely ignore roughness, having learned through a lifetime of exposure that roughness in music is not a reliable cue for dissonance. In contrast, amusics, lacking access to the harmonicity cue that we believe underlies dissonance, cannot learn that roughness is not a reliable cue, and thus show some sensitivity to roughness in their pleasantness judgments of chords. The difference in ratings between instrument type despite the lack of differentiation between different intervals likely reflects the fact that the roughness differences between intervals are small compared with the roughness differences between different timbres (Fig. S2).

Importantly, our stimuli were generated from real-world recordings of an instrument and a voice. It thus seems likely that the acoustic factors that influence consonance judgments under natural conditions are present in our stimuli. It is possible that for artificial stimuli with equal amplitude partials, which produce greater differences in roughness, we might have seen some degree of preference even in amusics. However, such stimuli are not representative of typical musical timbres, in which variations in roughness are subtle and inconsistent.

We conclude from these analyses that roughness is not consistently present in consonant chords, that roughness is sometimes present in consonant chords, and that roughness varies more as a function of instrumentation than as a function of whether a chord is consonant or dissonant. The pioneering work of the 1960s that established roughness as a cornerstone of musical aesthetics may thus have been a red herring.

**Correlation Analyses of Preferences for Consonance and Preferences/ Detection of Harmony and Beating.** To further investigate individual differences in chord ratings and relate them to the perception of harmony and beating, we computed correlations between our measures of preference and detection. When the two groups of listeners were combined, the correlation between the preference for consonance and the harmonic detection score was highly significant ($r = 0.76$, $P = 0.0001$), whereas the correlation between the preference for consonance and the beating detection score was not significant ($r = 0.20$, $P = 0.43$). These results are consistent with the results reported in ref. 7, in which preferences for consonance and harmonicity were correlated, whereas preference for consonance and beating were not. When computed within either the control group or the amusic group alone, however, none of these correlations were significant (Fig. S4, *Upper*). This finding is arguably unsurprising given the consistency within groups; in ref. 7 much larger cohorts were used, which overcame the modest individual differences present in normal listeners, most of whom have substantial preferences for consonance, harmonicity, and the absence of beating.

The very high scores obtained in both groups for beating detection mean that a ceiling effect could in principle be responsible for the absence of correlation between this measure and the preference for consonance. However, we also computed correlations between preference for consonance and our preference measures for harmonicity and beating; these correlations are more immune to this problem given that the preferences are not close to the maximum values possible given the rating scale. As expected, when the two groups were pooled together, a significant correlation was again observed between consonance and harmonicity ($r = 0.48$,...
but not between consonance and beating ($r = 0.17, P = 0.46$) (Fig. S4, Lower).

**Dichotic Versions of the Harmonicity Stimuli.** The frequency components of the harmonic stimuli were spaced more than a critical bandwidth apart in frequency, with the intention that the spectral modifications involved in creating the inharmonic stimuli would not alter the degree of beating. However, although beating is greatly reduced when frequency components are more than a critical band apart, there are conditions in which beats are nonetheless present, typically when frequency components are related by ratios that deviate slightly from small integer ratios (8). Dichotic presentation of frequency components is known to greatly reduce even these other forms of beats (9), and we therefore included alternate harmonicity test stimuli in which the even and odd numbered frequency components were played to opposite ears (dichotic presentation). No difference was found between the ratings obtained for diotic and dichotic versions of the harmonicity stimuli, and the ratings were thus combined in the analyses. The ratings and discrimination data for the diotic and dichotic versions of these stimuli are shown below in Figs. S5 and S6.


**Fig. S1.** Mean Roughness Index (averaged over eight different root notes) is plotted against interval size for chords composed of the three types of stimuli described in the text: recorded voice notes (green), recorded saxophone notes (blue), and synthetic complex tone notes (red). Note that the axis is inverted to make it congruent with the pleasantness rating axis in our figures, such that large amounts of roughness are lower on the graph. The variations of the Roughness Index for synthetic complex tone intervals closely reproduce behavioral ratings obtained in prior studies (e.g., ref. 6). For natural sounds however, the Roughness Index is in general smaller, and its variations more subtle, not resembling the rating pattern seen in behavioral data.
Fig. S2. Mean Roughness Index, averaged across the consonant and dissonant intervals that contributed to our behavioral measures of preference, for each of the three note timbres. Note that for consistency with Fig. S4, the roughness axis is again inverted here. Roughness differences between consonant and dissonant intervals are far more pronounced for intervals composed of synthetic complex tones than they are for those composed of naturally occurring instrument sounds. Large differences in overall roughness are also present between the different timbres of the synthetic and natural stimuli, and in fact exceed the difference between consonant and dissonant intervals within any particular timbre.

Fig. S3. Mean pleasantness ratings of chords by the control (Left) and amusic (Right) groups. For both groups, ratings are averaged over voice stimuli (green diamonds) on the one hand, and over saxophone stimuli (blue circles) on the other hand. Note that to reveal the difference between voice and saxophone in the amusics’ ratings, the data in this figure is not z-scored.
**Fig. S4.** Individual preferences for consonance plotted against detection score (Upper) or preference measure (Lower) for harmonicity (Left) and beating (Right). Correlations between the two scores are reported in each panel across groups of listeners (All), and for the amusic (amu) and the control (ctr) groups independently. Across groups, preference for consonance is correlated with both detection and preference for harmonicity, but with neither of those measures for beating.

**Fig. S5.** Preference ratings for the diotic and dichotic versions of the harmonicity stimuli. Preferences averaged over both diotic and dichotic stimuli, as they appear in Fig. 5 (measure 1 of harmonicity preference), are also replotted here ("all"). The fact that the preference is not larger for diotic stimuli indicates that any residual beating contributed little to the ratings of the diotic stimuli. Here and elsewhere, error bars denote one SE about the mean.
Detection score (% correct)

Harm. Diotic
Harm. Dichotic
Beating

above chance
chance level

Fig. S6. Harmonicity and beating discrimination for each individual stimulus contrast used (three different types of inharmonicity × dichotic/diolic presentation for the harmonicity experiment, and two different frequency separations for the beating experiment). Simulations of a randomly responding individual performing 15 repetitions per condition of a 3-AFC task were used to derive a null distribution for task performance. Of the 10,000 obtained values, 95% were situated below the “significantly above chance” level depicted in this figure.

Table S1. Control and amusic group characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amusics (n = 10)</td>
<td>Controls (n = 10)</td>
<td>P value of t test</td>
</tr>
<tr>
<td>Demographic characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>65.6 ± 5.7</td>
<td>65.7 ± 4.2</td>
<td>0.97 (n.s.)</td>
</tr>
<tr>
<td>Sex</td>
<td>6 female</td>
<td>7 female</td>
<td>—</td>
</tr>
<tr>
<td>Education (y)</td>
<td>17.1 ± 3.6</td>
<td>16.4 ± 2.5</td>
<td>0.62 (n.s.)</td>
</tr>
<tr>
<td>Musical education (y)</td>
<td>1.1 ± 0.9</td>
<td>1.8 ± 1.5</td>
<td>0.23 (n.s.)</td>
</tr>
<tr>
<td>Audiogram (dB hearing loss)</td>
<td>20.2 ± 9.6</td>
<td>24.6 ± 13.0</td>
<td>0.40 (n.s.)</td>
</tr>
<tr>
<td>Music discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBEA Melodic tests (% correct)</td>
<td>60.1 ± 8.4</td>
<td>91 ± 6.55</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>MBEA Rhythmic test (% correct)</td>
<td>67.8 ± 9.0</td>
<td>89.8 ± 6.2</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

The audiogram data are mean values of thresholds at 0.5, 1, 2, 3, 4, and 8 kHz in the two ears. The melodic test scores are expressed in percentages of correct responses and obtained on the scale, contour, and interval tests of the Montreal Battery of Evaluation of Amusia (MBEA) (1). The rhythmic test scores are obtained on the rhythm test of the same battery. Values are group means ± SD.


Other Supporting Information Files

Audio S1: Chord-Voice-m2 (WAV)
Audio S2: Chord-Voice-P5 (WAV)
Audio S3: Harm_PureTone (WAV)
Audio S4: Harm_Subset (WAV)
Audio S5: Harm_Jittered (WAV)
Audio S6: Harm_Shifted (WAV)
Audio S7: Beat_Diotic (WAV)
Audio S8: Beat_Jittered (WAV)