A Neurodynamic Account of Musical Tonality

Edward W. Large, Ji Chul Kim,
Nicole Kristine Flaig
University of Connecticut

Jamshed J. Bharucha
Cooper Union

Carol Lynne Krumhansl
Cornell University

Science since antiquity has asked whether mathematical relationships among acoustic frequencies govern musical relationships. Psychophysics rejected frequency ratio theories, focusing on sensory phenomena predicted by linear analysis of sound. Cognitive psychologists have since focused on long-term exposure to the music of one's culture and short-term sensitivity to statistical regularities. Today evidence is rapidly mounting that oscillatory neurodynamics is an important source of nonlinear auditory responses. This leads us to reevaluate the significance of frequency relationships in the perception of music. Here, we present a dynamical systems analysis of mode-locked neural oscillation that predicts cross-cultural invariances in music perception and cognition. We show that this theoretical framework combines with short- and long-term learning to explain the perception of Hindustani rāgas, not only by encultured Indian listeners but also by Western listeners unfamiliar with the style. These findings demonstrate that intrinsic neurodynamics contribute significantly to the perception of musical structure.

Received: December 5, 2014, accepted April 30, 2015.

Key words: tonality, neurodynamics, oscillation, mode-locking, plasticity

Is musical knowledge mainly acquired through long-term exposure to the music of one's culture? Or do intrinsic properties of neural processing constrain music perception and cognition? What is the role of short-term exposure and rapid statistical learning? While all these likely play a role, it is unknown how these fundamental mechanisms combine to enable the rich cognitive and emotional experience of music. Here we use dynamical systems theory to transform recent findings about nonlinear auditory neural processing into predictions about the perception of musical relationships. We ask whether this theory can explain cross-cultural invariances in the perception of Hindustani classical music, a highly developed style different from the more well-studied Western tonal-harmonic music. The dynamical principles explain fundamental similarities between unfamiliar Western listeners and encultured Indian listeners. These combine with statistical learning and culture-specific knowledge to provide a new model of tonal organization.

Consonance and dissonance are fundamental concepts in the science of music, with a long history of theory and experiment. The earliest observation, dating back at least to Pythagoras, is that small integer ratios, such as 2:1, 3:2, and 4:3, produce more pleasing or consonant musical intervals of the octave, fifth, and fourth—the "perfect consonances"—because they are mathematically pure (Burns, 1999). Pythagoras designed a system for tuning musical instruments based on the perfect consonances (Table 1), and 500 years later Ptolemy proposed several small-integer-ratio tuning systems, known as just intonation (JI; Table 1), still current in musical practice. Three significant non-Western musical traditions—Indian, Chinese, and Arab-Persian—also use intervals that approximate small integer ratios (Burns, 1999). In the eighteenth century, Euler (1739) hypothesized that the mind directly perceives and aesthetically appreciates simple integer frequency ratios.

Helmholtz (1885/1954) observed that purity of mathematical ratios could not explain perceived consonance in equal tempered tuning systems (ET; Table 1), in which small integer ratios are approximated by irrational numbers. Instead, he proposed that as the auditory system performs a linear analysis of complex sounds, proximal harmonics interfere with one another and produce a sensation of roughness, which he equated with dissonance. Small integer ratios yield more consonant musical intervals, he surmised, because they have more harmonics in common and thus fewer harmonics that interfere. When extrapolated to complex tones, the interaction of pure tone components predicts the perception of consonance well (Kameoka & Kuriyagawa, 1969).

Music Perception, volume 33, issue 3, pp. 319–331, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/MP.2016.33.03.319

TABLE 1. *Frequency Ratios Used in Tuning Systems and Farey Ratios Chosen by the Neurodynamic Model*

Interval from C to	Pythagorean tuning		Just intonation		Equal temperament		Farey ratio	
C	1:1	(1.000)	1:1	(1.000)	2^0	(1.000)	1:1	(1.000)
C♯/D♭	$2^8:3^5$	(1.053)	16:15	(1.067)	$2^{1/12}$	(1.059)	16:15	(1.067)
D	$3^2:2^3$	(1.125)	9:8	(1.125)	$2^{2/12}$	(1.122)	9:8	(1.125)
D♯/E♭	$2^5:3^3$	(1.185)	6:5	(1.200)	$2^{3/12}$	(1.189)	6:5	(1.200)
E	$3^4:2^6$	(1.266)	5:4	(1.250)	$2^{4/12}$	(1.260)	5:4	(1.250)
F	$2^2:3$	(1.333)	4:3	(1.333)	$2^{5/12}$	(1.335)	4:3	(1.333)
F♯/G♭	$3^6:2^9$	(1.424)	45:32	(1.406)	$2^{6/12}$	(1.414)	17:12	(1.417)
G	3:2	(1.500)	3:2	(1.500)	$2^{7/12}$	(1.498)	3:2	(1.500)
G♯/A♭	$2^7:3^4$	(1.580)	8:5	(1.600)	$2^{8/12}$	(1.587)	8:5	(1.600)
A	$3^3:2^4$	(1.688)	5:3	(1.667)	$2^{9/12}$	(1.682)	5:3	(1.667)
A♯/B♭	$2^4:3^2$	(1.778)	7:4	(1.750)	$2^{10/12}$	(1.782)	16:9	(1.778)
B	$3^5:2^7$	(1.898)	15:8	(1.875)	$2^{11/12}$	(1.888)	15:8	(1.875)
C (octave)	2:1	(2.000)	2:1	(2.000)	2^1	(2.000)	2:1	(2.000)

Recently, there has been a renewed interest in the neurophysiological basis of consonance. Theories based on the neural processing of pitch relationships have led to the development of concepts such as harmonicity and dynamical stability. Harmonicity is the degree to which the frequency spectrum of two complex tones resembles the harmonic spectrum of the difference tone of the fundamental frequencies (Gill & Purves, 2009; Tramo, Cariani, Delgutte, & Braida, 2001). Dynamical stability is based on the synchronization properties of ensembles of coupled neural oscillators (Shapira Lots & Stone, 2008), which we describe in further detail below. Interestingly, both theories relate consonance and dissonance to simple integer frequency ratios. Integer ratio-based predictions have been shown to account for generalizations about musical scales cross-culturally (Gill & Purves, 2009) and to account for the standard ordering of consonances as described in Western music theory (Shapira Lots & Stone, 2008). Perceptual studies have shown that harmonicity has a greater effect on consonance judgments in individuals with music training (McDermott, Lehr, & Oxenham, 2010), little effect in individuals with congenital amusia (Cousineau, McDermott, & Peretz, 2012), and can be dissociated from roughness (Cousineau et al., 2012; McDermott et al., 2010).

Recent theory and perceptual experiments have focused on the phenomenon of musical tension and resolution, which is distinct from the phenomenon of consonance and dissonance. The experience of tension and resolution, central to both Western and non-Western tonal music, is the outcome of expectations generated in a musical context. In a tonal melody, for example, certain pitches are felt as more stable than others, providing a sense of completion, that expectations have been fulfilled. More stable pitches function as points of organizational focus and engender a sense of resolution. Less stable pitches are felt as points of tension; they function relative to the more stable ones and are heard to point toward or be attracted to them (Bharucha, 1984; Lerdahl, 2001). Stability and attraction are the properties that differentiate musical sound sequences from arbitrary sound sequences and are thought to enable simple sound patterns to carry meaning (Meyer, 1956; Zuckerkandl, 1956).

Empirical studies measure perceived stability and ask how it depends on the musical context. One method that has been used to index tonal stability is the probe-tone method. In probe-tone experiments (e.g., Krumhansl & Kessler, 1982; Krumhansl & Shepard, 1979), listeners rate how well each tone within an octave fits with or completes a tonal context presented before it, which might be a scale, a sequence of chords, or a musical passage. Consistent patterns of ratings, called tone profiles, have been found for a variety of contexts including Western keys, Indian rāgas, Balinese melodies, 20th century music, and sequences specially constructed to have novel tone distributions (for a review, see Krumhansl & Cuddy, 2010). In Western major and minor contexts the tonic is the most stable, followed by the fifth and third scale degrees, then the remaining diatonic scale degrees, and finally the nondiatonic pitches (Krumhansl, 1990; see also Figure 1A).

Stability judgments differ from consonance judgments in a number of ways. Consonance correlates only weakly with stability ratings in major mode contexts and does not account well for minor contexts (Krumhansl, 1990). It does not explain why a chord may sound consonant or dissonant depending on its place in a musical sequence (Johnson-Laird, Kang, & Leong, 2012), and it makes a relatively small contribution to the sense of tension and release experienced when listening to music (Lerdahl, 2001). Moreover, consonance

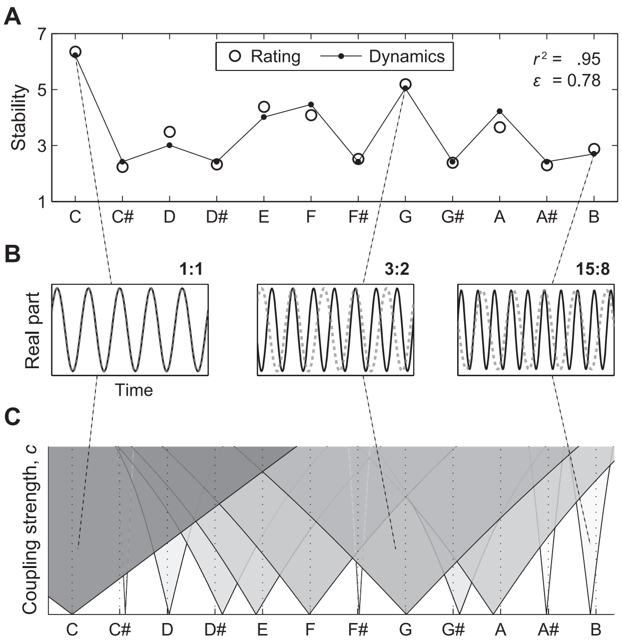

FIGURE 1. Tonal stability and mode-locking dynamics. (A) A tone profile (Krumhansl & Kessler, 1982) summarizing the stability of tones in the Western major mode (circles). Predictions of stability (Equation 2) for the major mode (Large, 2010) based on an analysis of mode-locking neurodynamics (dots). The value of ϵ for the maximum correlation is shown along with the r^2 statistic. (B) Mode-locked oscillations at different integer ratios. The real parts of z_1 and z_2 (Equation 1) are plotted over time. Each inset shows the interaction between an oscillator tuned to the frequency of C (z_1, dashed line) and an oscillator tuned to the frequency of the pitch labeled in the panel A (z_2, solid line). (C) Resonance regions for the mode-locks governing equal-tempered intervals. The horizontal axis represents the natural frequency ratio of two neural oscillators, with the labels indicating the natural frequency of the second oscillator while the first oscillator is fixed at the reference pitch C. The integer ratios chosen for the chromatic intervals, called Farey ratios, are shown in Table 1 (see Method for the procedure of choosing the ratios).

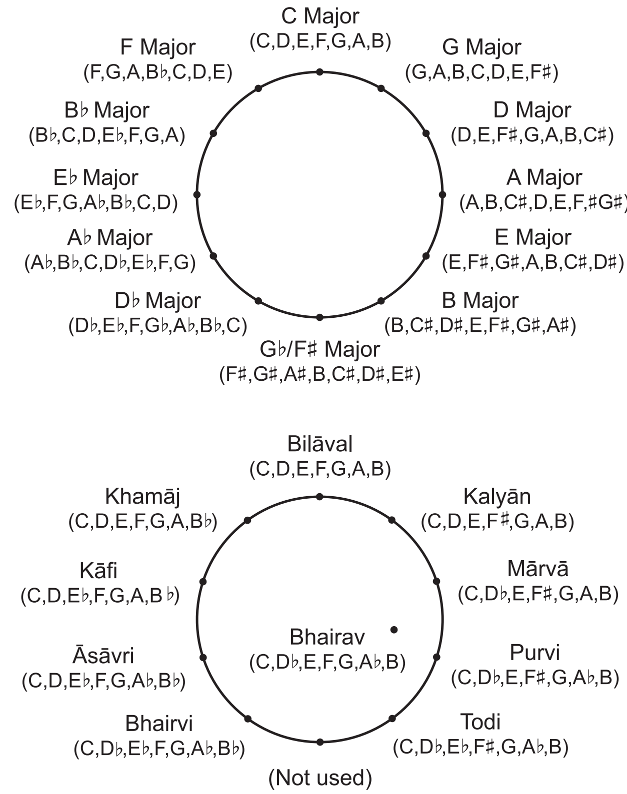

FIGURE 2. Western and Hindustani scale systems. The circle of fifths (top) and the circle of thāts (bottom) are shown with the scale tones listed in parentheses, starting with the reference pitch (tonic for the Western system and Sa for the Hindustani system).

relationships are fixed, implying that musical relationships that depend upon them would be largely invariant historically and cross-culturally.

To ask what aspects of tonal stability may be invariant across cultures, we turn to Hindustani music. The Hindustani tonal system shares some properties with Western music but differs in important respects. Its basic set of tones contains intervals that approximate the same integer-ratio intervals used in Western music. Many of the scales (thāts) are heptatonic, containing seven tones, but this may vary depending on the particular rāga. Whereas Western music is anchored in harmony, Hindustani music is anchored to a drone, consisting of the tonic (Sa) and usually the fifth (Pa). These two tones are fixed in all the scales, but the remaining tones can appear in one of two positions differing by a semitone, creating a richer and more diverse scale system than is used in the West. As a consequence, the intervals between adjacent pitches in the Indian scale system are more varied than in the major and minor scales of Western music. Systematic relationships between a subset of Hindustani scales can be displayed in a circular representation, similar to the layout of keys on the Western circle of fifths (Figure 2; see Jairazbhoy, 1971, for a theoretical account of the circle of thāts). However, comparison shows that although some of the Hindustani scales use the same tones as a Western key, all Hindustani scales shown are different modes with the same tonic (C or Sa) whereas all the Western keys are the same mode with different tonics.

Given such differences, one might assume that perceived stability relationships are quite different in Hindustani and Western music, and that large differences would be found between Indian listeners and Western listeners because of long-term cultural experience. Interestingly, this prediction has turned out to be incorrect. In one experiment (Castellano, Bharucha, & Krumhansl, 1984), the probe-tone method was used to quantify perceived stability relationships in Hindustani music for two groups of listeners: Indian listeners, who were familiar with Hindustani music, and unfamiliar Western listeners. The Indian and Western listeners were matched in music training (which was at a moderate level) so that any differences found could be attributed to cultural experience. Surprisingly, the ratings of Western listeners correlated strongly with those of Indian listeners, both matching the predictions of Hindustani music theory. Thus, the experience of stability did not require long-term exposure to the music of a particular culture; even the relatively short melodies used in this experiment reliably convey stability information to unfamiliar listeners.

One possible account of this surprising result came from examining the musical contexts. It revealed that the relative durations of tones correlated with the stability ratings of both groups of listeners. This suggested that perhaps the Western listeners abstracted stability based on the relative durations of the tones in the short melodic contexts. Music training may have enhanced the ability to extrapolate to an unfamiliar style (Krumhansl & Cuddy, 2010; Oram & Cuddy, 1995), so larger differences might have been found with Western nonmusicians who are less adept at assimilating unfamiliar tone distributions. However, the primary interest here was the effect of exposure to the style not music training per se. Related work found that the distribution of pitches in Western tonal music, their relative frequencies of occurrence and durations, resembles Western tone stability profiles (Krumhansl, 1985). It has since been shown that manipulating statistical regularities over both the short and longer term can affect stability ratings (Loui, Wessel, & Kam, 2010; Oram & Cuddy, 1995). These findings provided evidence that what is now called statistical learning applies not only to language and other domains (Kirkham, Slemmer, & Johnson, 2002; Saffran, 2003), but also to music.

The difficulty with an explanation of tonal stability based solely on the relative duration of tones is that it implies that any given duration distribution could establish a hierarchy of stability equally well regardless of pitch set. This raises some perplexing questions. If any set of pitches would do equally well, why are small integer ratio tuning systems so pervasive (Burns, 1999)? Why is it that, despite the fact that the statistical distributions differ across musical cultures and styles, regularities are also apparent such that consonant intervals are favored (Krumhansl, 2000; Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Krumhansl et al., 2000)? What accounts for the effects of frequency ratio, harmony, and pitch set on stability relationships (Krumhansl & Cuddy, 2010; Loui et al., 2010; Oram & Cuddy, 1995; Schellenberg & Trehub, 1996)? Might the relative duration statistics found in music reflect some underlying effect of acoustic frequency relationships? What is needed is a model that can take into account both frequency ratio and relative duration effects, as well as possible influences of culture-specific knowledge. Here we propose a new model consistent with evidence of nonlinear auditory processing, which predicts that dynamical stability in an oscillatory neural network contributes strongly to the perception of tonal stability.

Nonlinearities in Auditory Processing

Helmholtz's (1885/1954) observations, and most subsequent treatments of frequency ratios and their role in determining perceived musical relationships, have rested heavily on the assumption that the auditory system performs a linear frequency decomposition of incoming acoustic signals. In linear systems, small integer ratios have no special properties per se; therefore the only principle available to explain consonance is that of interference. However, we now know that the auditory nervous system is highly nonlinear, and in nonlinear systems frequency ratio relationships are important determinants of system behavior (Hoppensteadt & Izhikevich, 1997).

Recent evidence about nonlinearities in auditory processing may lead to a better understanding of how frequency ratios constrain the perception of musical relationships. Nonlinear responses to sound are found in the active processes of the cochlea (Robles, Ruggero, & Rich, 1997), and in neural populations of the cochlear nucleus, inferior colliculus, and higher auditory areas (Escabi & Schreiner, 2002; Langner, 2007; Sutter & Schreiner, 1991). Nonlinear spectrotemporal receptive fields (STRFs) have been identified in the inferior colliculus of the cat (Escabi & Schreiner, 2002) and the gerbil (Langner, 2007), and in cat primary auditory cortex (Sutter & Schreiner, 1991). In humans, nonlinear frequency-following responses (FFRs) have been observed in the brainstem using electroencephalography (EEG; Pandya & Krishnan, 2004), and in the auditory cortex using steady-state methods in magnetoencephalography (MEG; Purcell, Ross, Picton, & Pantev, 2007). Highly patterned nonlinear responses to harmonic musical intervals have been measured in the human auditory brainstem response (Lee, Skoe, Kraus, & Ashley, 2009) and have been captured in a nonlinear model of central auditory processing (Large & Almonte, 2012; Lerud, Almonte, Kim, & Large, 2014).

In central auditory circuits, action potentials phase-lock to both the fine time structure and the temporal envelope modulations of auditory stimuli at many different levels, including cochlear nucleus, superior olive, inferior colliculus, thalamus, and the A1 cortical region (Langner, 1992). If phase-locked activity arises from active nonlinear circuits, one might also expect to observe mode-locking, a generalization of phase-locking, in which a periodic stimulus interacts with the intrinsic dynamics of a nonlinear oscillator, causing k cycles of the oscillator to lock to m cycles of the stimulus (where k and m are integers; see Figure 1B). Recently, mode-locking to acoustic signals has been reported in guinea pig cochlear nucleus chopper and onset neurons (Laudanski, Coombes, Palmer, & Sumner, 2010). Mode-locked spiking patterns are often observed in vitro under DC injection (Brumberg & Gutkin, 2007), and mode-locked dynamics can be observed in generic models of neurons (Lee & Kim, 2006), neural populations (Hoppensteadt & Izhikevich, 1996; Large, Almonte, & Velasco, 2010), and auditory brainstem neurodynamics (Lerud et al., 2014).

Neurodynamic Model

We introduce a dynamical systems analysis that translates observations of mode-locking auditory dynamics into hypotheses about the perception of musical relationships. The conceptual framework for tonal stability is as follows. Each tone in a sequence creates a memory trace in an oscillatory network at or near its fundamental frequency. The neuronal oscillations interact with one another, creating complex patterns of mode-locking. In such a network certain frequencies would have greater dynamical stability, and we test whether this dynamical stability predicts perceived tonal stability.

When two oscillatory neural populations with natural frequencies, f_1 and f_2, are coupled, their long-term dynamics are governed by a resonant relationship. If the ratio f_1:f_2 is close enough to an integer ratio k:m, they mode-lock to one another with k cycles of one locking to m cycles of the other (Figure 1B). Mode-locking occurs in regions of parameter space called Arnol'd tongues (Figure 1C; see also Glass, 2001), whose boundaries are determined by stable (attracting) steady-state solutions of Equation 1, below. Two neuronal oscillators whose intrinsic frequencies fall within a resonance region mode-lock by adopting instantaneous frequencies different from their intrinsic frequencies (Video S1[1]). Figure 1C shows that resonance regions based on simpler mode-locks (e.g., 3:2) are wider than those based on higher-order ratios (e.g., 15:8), indicating that mode-locking responses driven by small integer ratios have greater dynamical stability.

Predictions about stability are derived from a dynamical systems analysis of weakly interacting oscillatory neural populations with different natural frequencies. The generic dynamics of such a system is captured by

$$
\begin{aligned}
\tau_1 \frac{dz_1}{dt} &= z_1(\alpha + \mathrm{i}2\pi + \beta_1|z_1|^2 + \epsilon\beta_2|z_1|^4 + \cdots) \\
&\quad + c\epsilon^{(k+m-2)/2} z_2^k \bar{z}_1^{m-1}, \\
\tau_2 \frac{dz_2}{dt} &= z_2(\alpha + \mathrm{i}2\pi + \beta_1|z_2|^2 + \epsilon\beta_2|z_2|^4 + \cdots) \\
&\quad + c\epsilon^{(k+m-2)/2} z_1^m \bar{z}_2^{k-1},
\end{aligned}
\tag{1}
$$

which describes the results of an analysis of two oscillatory populations ($i = 1, 2$) whose frequencies, $f_1 = 1/\tau_1$ and $f_2 = 1/\tau_2$, are close enough to the integer ratio k:m so that they mode-lock in that ratio (for details, see Hoppensteadt & Izhikevich, 1997; Large et al., 2010). Here, z_1 and z_2 are complex-valued (two-dimensional) state variables that represent the amplitude and phase of oscillations. The parameter α controls the bifurcation of autonomous behavior (spontaneous oscillation when $\alpha > 0$, damped oscillation when $\alpha < 0$), the β_n parameters specify nonlinear intrinsic dynamics, c represents synaptic coupling strength between neural oscillators, ϵ is a small number indicating weak interaction, $\bar{z}$ is the complex conjugate of z, and the roman i denotes the imaginary unit.

Predictions of Tonal Stability

Predicting tonal stability requires finding stable steady-state solutions of systems of equations such as Equation 1, where the τ parameters are chosen based on the frequencies of the scale (i.e., $\tau_\mathrm{i} = 1/f_\mathrm{i}$). To find the best

[1] Video S1 can be found in the supplementary materials section at the online version of this paper. The video shows that out of an infinite number of natural resonances within an octave (i.e., between 1:1 and 2:1), the neurodynamic model chooses the strongest resonance that is close enough to each equal-tempered (ET) ratio. For each ET ratio from the unison to the octave, a pure-tone dyad tuned to the ratio is played, two ET-tuned oscillators are set into motion, mode-locking in the chosen integer ratio (governed by Equation 1; the bottom panel shows the real part of oscillations), and the resonance region (Arnol'd tongue) for the mode-locking oscillators is displayed (top panel). For human eyes, oscillations are slowed down so that the lower C becomes 1 Hz.

fitting model using this approach, however, we would need to specify all parameters, namely the α, β_n, and the c, for all oscillators in the system, resulting in excessive degrees of freedom, and necessitating search of a prohibitively large parameter space. Here we present a simpler approach that leads to closed-form predictions with a single degree of freedom. This approach assumes a simplified network in which each oscillator interacts with only one other oscillator, whose frequency is that of the tonic. Such a network can be specified as pairs of equations like Equation 1. Steady-state amplitude for each oscillator in Equation 1 depends on the strength of its input, for instance, $c\ \epsilon^{(k+m-2)/2} z_2^k \bar{z}_1^{m-1}$. The strength is proportional to $\epsilon^{(k+m-2)/2}$, therefore we use this simple formula to approximate stability within a given context.

First, k and m must be determined based on the frequencies of the scale being tested. As Figure 1C shows, there are regions of stability not just *at* integer ratios, but *around* integer ratios. We choose the smallest k and m (i.e., the largest resonance region) for which k/m approximates the actual frequency ratio closely enough (Table 1; see Method). This approach does not require *mathematical purity* of ratios to predict perceptual effects, it could be applied to any set of frequency ratios. Next, steady-state amplitude can be estimated by specifying a single parameter, ϵ. Assuming $\epsilon < 1$ (weak interaction), input strength is highest when $k + m$ is smallest, so mode-locking responses driven by small integer ratios produce higher amplitude responses. Finally, the input to z_i depends on z_i itself. This means that, when k, $m > 1$, an oscillator cannot achieve nonzero amplitude merely by interaction with another oscillator; it requires external input at its own frequency. Thus, only tones in the context sequence enter into the interaction. In sum, we use $\epsilon^{(k+m-2)/2}$ to predict the relative stability of different mode-locked states where the k's and m's are determined based on the frequency ratios of the tones that occur in the scale (see Method), and any tones not sounded in the context sequence are assigned a stability value of 0.

For a complex network of interacting frequencies, computer simulations can be used to numerically integrate the nonlinear differential equations, one for each frequency, producing more detailed predictions. The Appendix contains equations of a complete computer model, and Videos S2 and S3 show animations of oscillator amplitude as a function of time, which predicts stability for specific musical sequences.[2]

[2] Videos S2 and S3 can be found in the supplementary materials section at the online version of this paper.

Two lines of recent research have considered whether dynamical stability could explain aspects of music perception and cognition. One analysis showed that the width of resonance regions in a model of coupled neural oscillators predicts the standard ordering of consonances in Western music theory (Shapira Lots & Stone, 2008). Another study showed that the relative stability of mode-locks could account for tonal stability in Western major and minor modes (Large, 2010). That study considered pair-wise resonant relationships, described above, between each oscillation at each tone frequency and the tonic frequency because the tonic functions as a cognitive reference point in tonal contexts (Krumhansl, 1990). Figure 1C shows the resonance regions chosen to govern equal-tempered intervals ranging from the unison to the major seventh (Video S1 and Table 1; see footnote 1). Figure 1A shows that this model predicted a high proportion of variance in listener judgments of the major mode, and it was higher than any of the psychoacoustic models tested (Krumhansl, 1990); it also generalized to the minor mode ($r^2 = .77$ for $\epsilon = 0.85$), which most of the psychoacoustic models did not. Thus, the theoretical stability of mode-locking interactions in oscillatory neural populations predicts empirically measured tonal stability for Western tonal contexts.

Cross-Cultural Judgments of Tonal Stability

A critical test of this approach is whether it generalizes across musical cultures. Castellano et al.'s (1984) cross-cultural study of the perception of Hindustani music provides precisely the kind of empirical data needed to test this hypothesis. They collected stability judgments for two groups of listeners. One group consisted of eight Indian listeners who were students at Cornell University and had been exposed to Indian music starting at an average age of 12 years, six played an Indian instrument with an average of 4 years of formal training, and all listened to Indian music an average of 3 hr per week and to Western music an average of 10 hr per week. The other group consisted of 8 Western listeners, seven of whom played a Western instrument, averaging 6 years of formal training. They listened to Western music 18 hr per week. They had minimal exposure to Indian music, although one reported having heard some, and none had any formal instruction in Indian music.

The stimuli were synthesized digitally by a DMX-1000 (Digital Music Systems, Inc.) signal-processing computer under the control of a PDP-11/24 computer. They were based on an A4 of 440 Hz. Each tone consisted of a fundamental and 6 higher harmonics with the following amplitude ratios: 1, 1/3, 1, 1/3, 1, 1/3, 1. The amplitude envelope consisted of a 10-ms onset gradient followed by a linear decay to 0, simulating a plucked instrument. Frequency modulation (FM) was used to simulate the sound of the C (Sa) and G (Pa) drone tones, with a ratio of 7:5 between modulator and carrier frequencies and a DMX modulation index of 10. The stimuli also contained glides between notes simulating ornamentation in Indian music.

The sthāyi (the initial phase of a fixed composition which establishes a rāga) and theme (some brief passage incorporating characteristic melodic features) for each of 10 rāgas were taken from Daniélou (1968). There were 10 blocks of trials, played in a different random order for each listener, which had the following format. Each began with two successive presentations of the sthāyi, which averaged 65.8 s, with a range of 37.8 – 93.0 s. After 2 practice trials, 12 trials presented the theme of the rāga followed by a probe tone that was sounded without the drone. The themes' durations averaged 19.4 s, ranging from 10.8 to 29.4 s, and probe tones were 1 s in duration. The probe tones were all 12 tones of the chromatic scale. Listeners were instructed to judge how well each probe tone fit with the preceding theme in a musical sense, on a 7-point scale ranging from *fits poorly* (1) to *fits well* (7). The resulting data for Western and Indian listeners are shown in Table S1, which also shows the durations of the tones in each rāga theme.[3]

Method

The first step was to choose the strongest natural resonance (based on an integer ratio) that approximates each equal-tempered frequency ratio within a specified tolerance. This was done by searching through the sequences of irreducible fractions (called Farey ratios; Hardy & Wright, 1979) between 1 and 2 in increasing order until the required tolerance is met, here chosen to be 1%. The found Farey ratios, shown in Table 1, are identical to the JI ratios except for the tritone (from C to F♯/G♭) and the minor seventh (from C to A♯/B♭).

Next, to compute the stability of each scale tone, we considered pair-wise interactions with the tonic frequency. This enables a simple closed-form equation, based on Equation 1, that gives a good approximation of relative stability. We used $\epsilon^{(k+m-2)/2}$ to predict the relative stability of different mode-locked states where the k's and m's are the numerators and denominators of the Farey ratios obtained above. Any tones not sounded are assigned a stability value of 0. In other words,

$$\text{stability} \approx \begin{cases} \epsilon^{(k+m-2)/2} & \text{if } k/m \text{ is in context,} \\ 0 & \text{if } k/m \text{ not in context.} \end{cases} \quad (2)$$

Because this approximation is based only on the input, it holds for a wide range of parameter values in Equation 1. Numerical simulations confirm that it holds for more complex network interactions as well (see Appendix and Videos S2 and S3; see footnote 2).

Finally, the ϵ parameter was fit to the average of the Indian and Western data for each rāga separately by performing a least squares fit of Equation 2 with $0 < \epsilon < 1$ as the only free parameter. The resulting stability predictions were used, along with other predictors, in step-wise regression analyses of the behavioral ratings from both the Indian and Western listeners.

Results

The ratings for the two groups were correlated with each other for each rāga separately. The correlations between groups averaged $r^2(11) = .87$, ranging from .70 to .93, $p < .01$ for every rāga. The ratings of the two groups were then averaged. As seen in Figure 3, the neurodynamic model predictions correlated significantly with the averaged perceptual ratings for every rāga, mean $r^2(11) = .90$, min = .84, max = .93, all $p < .01$. For comparison, we computed the correlations between the average data and the durations of the tones in each rāga. These correlations, mean $r^2(11) = .84$, min = .77, max = .91, all $p < .01$, were lower, $t(9) = 2.68$, $p = .03$. Thus, the model fit the averaged data better than the durations.

The same analysis, comparing the neurodynamic model with the durations as predictors of the data, was done for the Indian and Western data separately. For the Indian data, the model predictions correlated significantly with the perceptual ratings for every rāga, mean $r^2(11) = .90$, min = .77, max = .98, all $p < .01$. For comparison, we computed the correlations between the Indian data and the durations of the tones in each rāga. For the Indian listeners, these correlations, mean $r^2(11) = .83$, min = .74, max = .94, all $p < .01$, were lower than the correlations with the model, $t(9) = 2.60$, $p = .01$. So, again the model fit the data better than the durations.

For the Western data, the model predictions correlated significantly with the perceptual ratings for every rāga, mean $r^2(11) = .84$, min = .76, max = .94, all

[3] Table S1 can be found in the supplementary materials section at the online version of this paper.

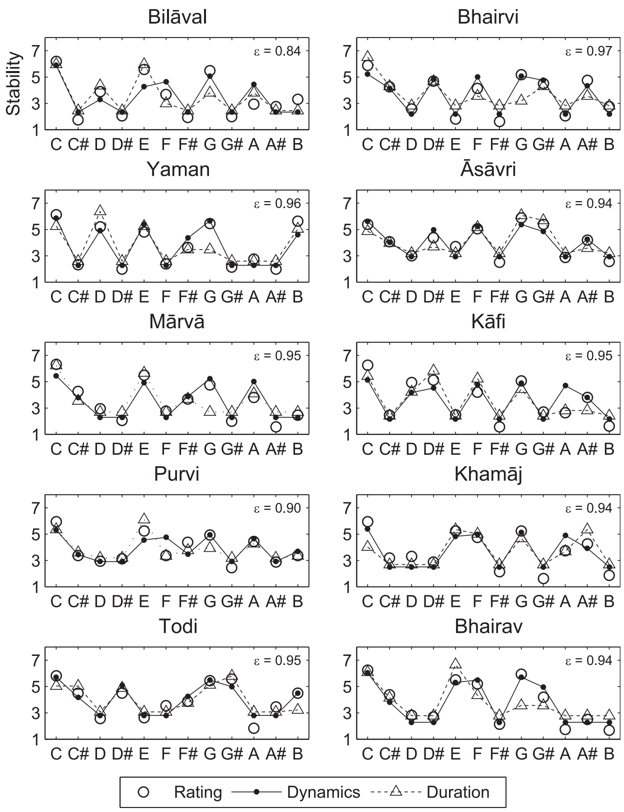

FIGURE 3. The fit of the neurodynamic model predictions and the duration statistics to the average data. Shown in the figure are the tone profiles for 10 North Indian rāgas averaged across Indian and Western listeners (circles), the predictions of stability (Equation 2) for each rāga, based on an analysis of mode-locking dynamics (dots) along with the ϵ value for the maximum correlation, and the predictions based on the relative durations of tones (triangles).

$p < .01$, but were lower than those for the Indian data, $t(9) = 2.60$, $p = .01$. We then computed the correlations between the Western data and the durations of the tones in each rāga. For the Western listeners, these correlations, mean $r^2(11) = .80$, min $= .72$, max $= .88$, all $p < .01$, were not significantly different from the correlations with the model ($p = .17$). Finally, we compared the Indian and Western listeners' correlations between their ratings and tone durations. These were not significantly different from one another ($p = .32$). Thus, for the Indian listeners the model outperformed the durations, but not for the Western listeners, and both groups were equal in the extent to which their data matched the durations.

Next, the listener judgments were combined into a single variable (10 rāgas $\times$ 12 tones) for step-wise regression analyses that included a wide range of factors that might be expected to generalize across cultures. The factors considered were: predictions of the neurodynamic model (Equation 2), tone durations, harmonicity ($[k + m - 1]/[km]$, see Gill & Purves, 2009), resonance region widths (Shapira Lots & Stone, 2008), and six measures of consonance (Krumhansl, 1990, p. 57, Table 3.1). For the neurodynamic model, the parameter ϵ was fit to the average of the Indian and Western data for each rāga separately (see Method).

Because the stability ratings of Indian and Western listeners were highly correlated, the first regression analysis was performed to determine which variables could account for the grand average ratings. Table 2 (top) shows the results. The neurodynamic model entered at step 1 and accounted for a highly significant amount of the variance for all rāgas taken together, R^2 (119) $= .79$, $p < .0001$, as well as separately. Duration came in at step 2 with both dynamics and duration contributing significantly (at $p < .0001$) to fitting the averaged data, R^2 (118) $= .84$, $p < .0001$, suggesting a combination of intrinsic dynamics and tone duration accounts well for these patterns. Of the remaining variables, the width of the resonance region (Shapira Lots & Stone, 2008) was the only one to contribute significantly (at $p = .04$) to the fit of the data. Because the contribution is relatively modest and the resonance region model is closely related to the neurodynamic model conceptually, we did not consider it in the next analysis.

We then asked whether culturally specific variables might explain the remaining differences between the Indian and Western participants (bottom Table 2). We considered several variables, including thāt (scale) membership, drone, the vādi and samvādi, which are tones emphasized in the individual rāga, that might be more salient for the Indian listeners, and the tone profiles of several major and minor keys related to C major and C minor that might be more salient for the Western listeners. For the Indian data, two of these made highly significant contributions over and above the neurodynamic model and tone duration: whether the tone was in the drone ($p = .0002$; the p-values given are those when the variable was first added to the model) and whether the tone was in the thāt ($p < .0001$). The former shows an appreciation of the central role of the drone in anchoring melodies in the style; the latter reflects the concept of scale membership. For the Western data, two factors made significant contributions: the tone profiles of C minor ($p = .005$) and G major ($p = .03$), which reflect the Western concept of major and minor scales. The fit of the full regression models to the respective data is shown in Figures 4 and 5.

TABLE 2. *Results of Step-wise Regressions Predicting the Average Indian and Western Data (Top) and the Indian and Western Data Separately (Bottom)*

	Average Indian and Western data		
Step	Predictor	*p*-value	Cumulative R^2
1	Dynamics	< .0001	.79
2	Duration	< .0001	.84
3	Resonance region width (Lots & Stone)	.04	.85

	Indian data			Western data		
Step	Predictor	*p*-value	Cumulative R^2	Predictor	*p*-value	Cumulative R^2
1	Dynamics	< .0001	.80	Dynamics	< .0001	.69
2	Duration	< .0001	.84	Duration	< .0001	.73
3	Drone	.0002	.86	C Minor	.005	.75
4	Thāt	< .0001	.88	G Major	.03	.76

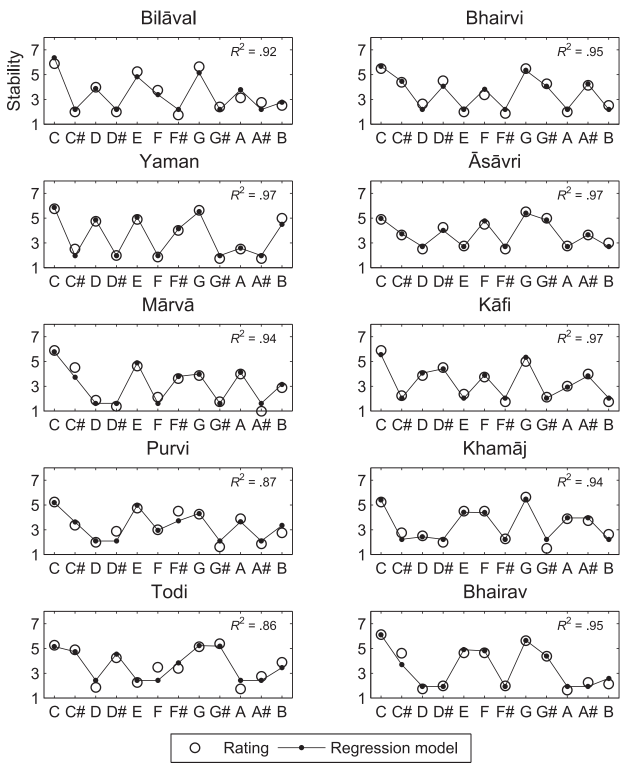

FIGURE 4. The fit of the full Indian regression model (dots) to the Indian probe-tone rating data (circles) for each rāga (see Table 2, bottom left).

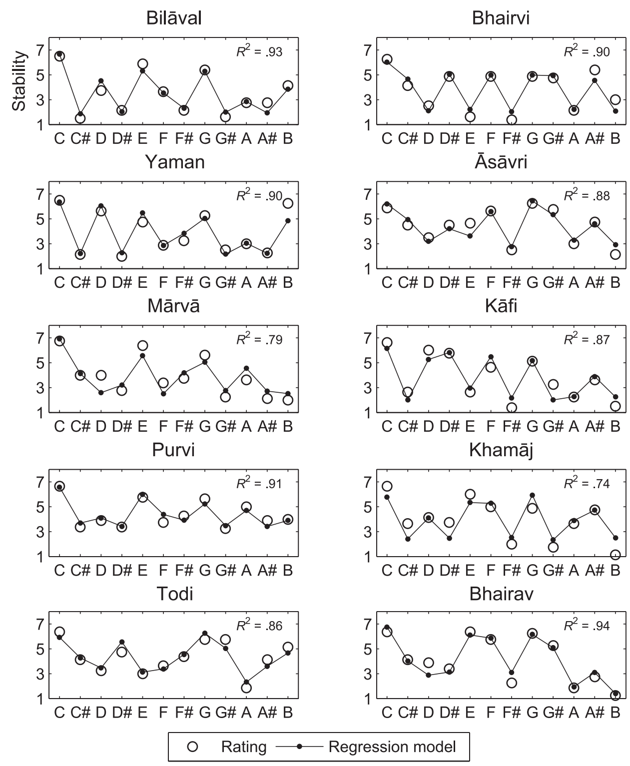

FIGURE 5. The fit of the full Western regression model (dots) to the Western probe-tone rating data (circles) for each rāga (see Table 2, bottom right).

Discussion

Here we have shown how tonal stability, a high-level perception that is central to the experience of tonal music, can be accounted for by a biologically plausible model of neural oscillations in a dynamical system. Previous models have considered only stability relations in Western music as judged by Western listeners. The current study shows that a fully generic model of mode-locking neurodynamics predicts substantive musical invariances, matching empirical observations in a cross-cultural study of tonal stability judged by listeners familiar and unfamiliar with the style of Hindustani music. The simple neurodynamic predictions are improved by taking into account the duration of tones in the context sequences. A third layer of explanation is added by considering the effect of culture-specific knowledge to account for residual differences between the encultured and naïve participants. More realistic models, based on complex numerical simulations that incorporate both mode-locking and synaptic plasticity, are confirming and refining the predictions presented here (see Appendix and Videos S2 and S3; see footnote 2). These findings raise new and interesting possibilities for studying tonal cognition using theoretical models that are consistent with auditory neurophysiology (Hoppensteadt & Izhikevich, 1997; Large & Almonte, 2012; Lerud et al., 2014).

The neurodynamic approach may shed light on a number of persistent mysteries surrounding music cognition and the development of musical languages. The perceptual stability of mode-locked states explains the strong propensity for perfect consonances (2:1, 3:2, 4:3) in the scales of almost all musical cultures, including the sophisticated small integer ratio tuning systems of the European, Chinese, Indian, and Arab-Persian musical systems (Burns, 1999), that began to develop more than 9,000 years ago (Zhang, Harbottle, Wang, & Kong, 1999). Regions of stability around integer ratios can explain the considerable latitude in interval tuning found in different styles, which has been used to discredit simple-ratio theories (Burns, 1999; Helmholtz, 1885/1954), and the well-established finding that intervals are perceived categorically (Burns, 1999; Smith, Nelson, Grohskopf, & Appleton, 1994). Third, the fact that stability properties are intrinsic to neurodynamics explains why even infants have more stable memory traces for the perfect consonances than the dissonances in melodic sequences (Schellenberg & Trehub, 1996). These dynamical principles also make strong neurophysiological predictions and have been shown to predict nonlinear responses to musical intervals measured in the auditory brainstem response (Large & Almonte, 2012; Lee et al., 2009; Lerud et al., 2014). This approach could also shed light on the development of musical regularities, implying that certain pitches occur more frequently because they have greater dynamical stability in underlying neural networks. Future research will explore whether dynamical systems theory can be extended to describe other invariant aspects of musical structure and cognition in terms of underlying neurodynamics.

Author Note

This work was supported by NSF Grant BCS-1027761 to Edward W. Large. The authors wish to thank Stephen McAdams, Peter Pfordresher, and an anonymous reviewer for their comments and suggestions which helped improve this article.

Correspondence concerning this article should be addressed to Edward W. Large, Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020. E-mail: edward.large@uconn.edu

References

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*, 485-518. doi: 10.1016/0010-0285(84)90018-5

Brumberg, J. C., & Gutkin, B. S. (2007). Cortical pyramidal cells as non-linear oscillators: Experiment and spike-generation theory. *Brain Research*, *1171*, 122-137. doi:10.1016/j.brainres.2007.07.028

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deustch (Ed.), *The psychology of music* (2nd ed., pp. 215-264). San Diego, CA: Academic Press.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*, 394-412.

Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences*, *109*, 19858–19863. doi:10.1073/pnas.1207989109

Daniélou, A. (1968). *The rāga-s of Northern Indian music*. London: Barrie & Rockliff the Cresset P.

Escabi, M. A., & Schreiner, C. E. (2002). Nonlinear spectro-temporal sound analysis by neurons in the auditory midbrain. *Journal of Neuroscience*, *22*, 4114-4131.

Euler, L. (1739). *Tentamen Novae Theoriae Musicae* [An attempt at a new theory of music]. St. Petersburg: Typographia Academiae Scientiarum.

Gill, K. Z., & Purves, D. (2009). A biological rationale for musical scales. *PLoS ONE*, *4*(12), e8144. doi:10.1371/journal.pone.0008144

Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature*, *410*, 277-284. doi:10.1038/35065745

Hardy, G. H., & Wright, E. M. (1979). *An introduction to the theory of numbers* (5th ed.). New York: Oxford University Press.

HELMHOLTZ, H. L. F. (1954). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover Publications. (Original revised edition published 1885)

HOPPENSTEADT, F. C., & IZHIKEVICH, E. M. (1996). Synaptic organizations and dynamical properties of weakly connected neural oscillators I: Analysis of a canonical model. *Biological Cybernetics*, *75*, 117-127. doi:10.1007/s004220050279

HOPPENSTEADT, F. C., & IZHIKEVICH, E. M. (1997). *Weakly connected neural networks*. New York: Springer.

JAIRAZBHOY, N. A. (1971). *The rāgs of North Indian music: Their structure and evolution*. London: Faber & Faber.

JOHNSON-LAIRD, P. N., KANG, O. E., & LEONG, Y. C. (2012). On musical dissonance. *Music Perception*, *30*, 19-35. doi:10.1525/mp.2012.30.1.19

KAMEOKA, A., & KURIYAGAWA, M. (1969). Consonance theory part I: Consonance of dyads. *Journal of the Acoustical Society of America*, *45*, 1451-1459. doi:10.1121/1.1911623

KIRKHAM, N. Z., SLEMMER, J. A., & JOHNSON, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35-B42. doi:10.1016/S0010-0277(02)00004-5

KRUMHANSL, C. L. (1985). Perceiving tonal structure in music. *American Scientist*, *73*, 371-378.

KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press. doi:10.1093/acprof: oso/9780195148367.001.0001

KRUMHANSL, C. L. (2000). Tonality induction: A statistical approach applied cross-culturally. *Music Perception*, *17*, 461-479.

KRUMHANSL, C. L., & CUDDY, L. L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research: Music perception* (Vol. 36, pp. 51-87). New York: Springer. doi:10.1007/978-1-4419-6114-3_3

KRUMHANSL, C. L., & KESSLER, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334-368. doi:10.1037/0033-295X.89.4.334

KRUMHANSL, C. L., LOUHIVUORI, J., TOIVIAINEN, P., JÄRVINEN, T., & EEROLA, T. (1999). Melodic expectancy in Finnish folk hymns: Convergence of behavioral, statistical, and computational approaches. *Music Perception*, *17*, 151-196.

KRUMHANSL, C. L., & SHEPARD, R. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346-384. doi:10.1016/0010-0285(79)90016-1

KRUMHANSL, C. L., TOIVANEN, P., EEROLA, T., TOIVIAINEN, P., JÄRVINEN, T., & LOUHIVUORI, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*, 13-58. doi:10.1016/S0010-0277(00)00068-8

LANGNER, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, *60*, 115-142. doi:10.1016/0378-5955(92)90015-F

LANGNER, G. (2007). Temporal processing of periodic signals in the auditory system: Neuronal representation of pitch, timbre, and harmonicity. *Zeitschrift für Audiologie*, *46*, 8-21.

LARGE, E. W. (2010). A dynamical systems approach to musical tonality. In R. Huys & V. Jirsa (Eds.), *Nonlinear dynamics in human behavior* (pp. 193-211). New York: Springer. doi:10.1007/978-3-642-16262-6_9

LARGE, E. W., & ALMONTE, F. V. (2012). Neurodynamics, tonality, and the auditory brainstem response. *Annals of the New York Academy of Sciences*, *1252*, E1-E7. doi:10.1111/j.1749-6632.2012.06594.x

LARGE, E. W., ALMONTE, F., & VELASCO, M. (2010). A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, *239*, 905-911. doi:10.1016/j.physd.2009.11.015

LAUDANSKI, J., COOMBES, S., PALMER, A. R., & SUMNER, C. J. (2010). Mode-locked spike trains in responses of ventral cochlear nucleus chopper and onset neurons to periodic stimuli. *Journal of Neurophysiology*, *103*, 1226-1237. doi:10.1152/Jn.00070.2009

LEE, K. M., SKOE, E., KRAUS, N., & ASHLEY, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, *29*, 5832-5840. doi:10.1523/JNEUROSCI.6133-08.2009

LEE, S. G., & KIM, S. (2006). Bifurcation analysis of mode-locking structure in a Hodgkin-Huxley neuron under sinusoidal current. *Physical Review E*, *73*, 041924. doi:10.1103/PhysRevE.73.041924

LERDAHL, F. (2001). *Tonal pitch space*. New York: Oxford University Press. doi:10.1093/acprof: oso/9780195178296.001.0001

LERUD, K. D., ALMONTE, F. V., KIM, J. C., & LARGE, E. W. (2014). Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hearing Research*, *308*, 41-49. doi:10.1016/j.heares.2013.09.010

LOUI, P., WESSEL, D. L., & KAM, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27, 377-388. doi:10.1525/mp.2010.27.5.377

MCDERMOTT, J. H., LEHR, A. J., & OXENHAM, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*, 1035-1041. doi:10.1016/j.cub.2010.04.019

MEYER, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.

ORAM, N., & CUDDY, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*, 103-118. doi:10.1007/BF00447080

PANDYA, P. K., & KRISHNAN, A. (2004). Human frequency-following response correlates of the distortion product at 2F1-F2. *Journal of the American Academy of Audiology*, *15*, 184-197. doi:10.3766/jaaa.15.3.2

Purcell, D. W., Ross, B., Picton, T. W., & Pantev, C. (2007). Cortical responses to the 2f1-f2 combination tone measured indirectly using magnetoencephalography. *Journal of the Acoustical Society of America, 122*, 992-1003. doi:10.1121/1.2751250

Robles, L., Ruggero, M. A., & Rich, N. C. (1997). Two-tone distortion on the basilar membrane of the chinchilla cochlea. *Journal of Neurophysiology*, *77*, 2385-2399.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110-114. doi:10.1111/1467-8721.0124

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*, 272-277. doi:10.1111/j.1467-9280.1996.tb00373.x

Shapira Lots, I., & Stone, L. (2008). Perception of musical consonance and dissonance: An outcome of neural synchronization. J*ournal of The Royal Society Interface*, *5*, 1429-1434. doi:10.1098/rsif.2008.0143

Smith, J. D., Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, *52*, 23-54. doi:10.1016/0010-0277(94)90003-5

Sutter, M. L., & Schreiner, C. (1991). Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, *65*, 1207-1226.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences, 930*, 92-116. doi:10.1111/j.1749-6632.2001.tb05727.x

Zhang, J., Harbottle, G., Wang, C., & Kong, Z. (1999). Oldest playable musical instruments found at Jiahu early Neolithic site in China. *Nature*, *401*, 366-368. doi:10.1038/43865

Zuckerkandl, V. (1956). *Sound and symbol: Music and the external world* (W. R. Trask, Trans.). Princeton, NJ: Princeton University Press.

Appendix

SIMULATION AND EXTENSION OF NEURODYNAMIC MODEL

Videos S2 and S3 demonstrate more complex simulations that are refining and extending the predictions of the purely theoretical model discussed in the main text (see footnote 2). The simulations involve training an oscillator network capable of Hebbian plasticity with musical sequences and then testing its responses to another set of sequences. Oscillators in this network, and the input to this network, are tuned using equal temperament. We simulated two octaves of oscillators from C3 to C5. The network oscillators operate in a supercritical regime of Hopf bifurcation ($\alpha = 0.002$, $\beta_1 = -2$, $\beta_2 = -4$) where they oscillate spontaneously without external driving. The choice of parameters was somewhat arbitrary, and they were not fit to the data. Initial conditions for all oscillators were $z_i = 0$ for all i.

The model generates memory traces as neural oscillations at the frequencies of the tones in the context sequences; oscillations interact as defined by Equations A1 and A2. The network was stimulated with musical sequences, and the pattern of oscillator amplitudes in memory is interpreted directly as tonal stability (no actual experiments were simulated, so probe tones were not introduced).

Each oscillator has three types of input (Equation A1, below):

1. External input: A melody or a chord progression consisting of a sequence of complex-valued sinusoids. Each oscillator receives input at its own natural frequency through a linear driving term, $x_i(t)$.
2. Network input: 24 two-frequency monomials, one monomial for each of the other oscillators in the simulation (the terms with C_{ij} in Equation A1). Self-connections are not allowed. Each monomial is chosen according to normal form theory, that is, it is the lowest-order monomial that is resonant for the two oscillators in question. Thus, these monomials are the same as those used to make the closed-form predictions in the main text. However, that analysis considered only connections with the tonic oscillator. In the simulation all network frequencies interact.
3. Network input: 275 three-frequency monomials, one monomial for each pair of other network oscillators (the terms with D_{ijk} in Equation A1). Self-connections are not allowed (and there are some other restrictions). Each monomial is the lowest-order resonant monomial for a given set of three frequencies.

The dynamics of each oscillator in the network are described by

$$\tau_i \frac{dz_i}{dt} = z_i\left(\alpha + \mathrm{i}2\pi + \beta_1|z_i|^2 + \frac{\epsilon\beta_2|z_i|^4}{1-\epsilon|z_i|^2}\right) + \sum_{j\neq i} C_{ij}\epsilon^{(k_{ij}+m_{ij}-2)/2} z_j^{k_{ij}} \bar{z}_i^{m_{ij}-1} + \sum_{j\neq i}\sum_{k\neq i} D_{ijk}\epsilon^{(p_{ijk}+q_{ijk}+r_{ijk}-2)/2} z_j^{p_{ijk}} z_k^{q_{ijk}} \bar{z}_i^{r_{ijk}-1} + x_i(t). \quad \text{(A1)}$$

If the oscillator frequencies are near the resonant relationships $m_{ij}f_i = k_{ij}f_j$ (defining two-frequency monomials) and $r_{ijk}f_i = p_{ijk}f_j + q_{ijk}f_k$ (three-frequency monomials) where f_i is the natural frequency of the *i*th oscillator, and m_{ij}, k_{ij}, r_{ijk}, p_{ijk} and q_{ijk} are integers, then the corresponding oscillators interact by mode-locking to each other. The value of $\alpha > 0$ was chosen so that the memory trace for each frequency, once established, persists indefinitely. The fact that τ_i appears on the left-hand side of the equation means that relaxation time (time to reach steady-state amplitude) is measured in oscillator cycles.

All of the connections are determined using a rule for Hebbian plasticity

$$\tau_i \frac{dC_{ij}}{dt} = C_{ij}\left(\lambda + \mu_1|C_{ij}|^2 + \frac{\epsilon_C \mu_2 |C_{ij}|^4}{1-\epsilon_C|C_{ij}|^2}\right) + \epsilon_C^{(k_{ij}+m_{ij}-2)/2} \kappa z_i^{m_{ij}} \bar{z}_j^{k_{ij}},$$

$$\tau_i \frac{dD_{ijk}}{dt} = D_{ijk}\left(\gamma + \nu_1|D_{ijk}|^2 + \frac{\epsilon_D \nu_2 |D_{ijk}|^4}{1-\epsilon_D|D_{ijk}|^2}\right) + \epsilon_D^{(p_{ijk}+q_{ijk}+r_{ijk}-2)/2} \eta z_i^{r_{ijk}} \bar{z}_j^{p_{ijk}} \bar{z}_k^{q_{ijk}}. \tag{A2}$$

Two separate matrices are learned, one for the two-frequency monomials (C_{ij}), one for the three-frequency monomials (D_{ijk}). The learning rules operate in the supercritical Hopf regime (λ, $\gamma > 0$; μ_1, μ_2, ν_1, $\nu_2 < 0$; κ, $\eta > 0$). The system was trained on cadences in the key of C (IV-V$_7$-I), repeated until connection strengths appeared to reach stable values (35 times; Video S2; see footnote 2).

Two tests were run (Video S3; see footnote 2). The first test was the first phrase of *Twinkle Twinkle Little Star*. The movie shows the piano roll notation for the stimulus, the time-varying amplitude of the 25 oscillators in the network, and the correlations of the oscillator amplitudes with the C major tone profile (upper octave only, C through B). A period of silence is simulated at the end, and the amplitude pattern in the network relaxes toward steady state, which resembles the tone profile of C major.

This simulation validates and extends the purely theoretical, closed-form model of Equation 2 (main text). First, it confirms that those tones sounded in the context sequence achieve non-zero values. Second, it shows that the expression used to predict tonal stability, $\epsilon^{(k+m-2)/2}$, gives a quite reasonable approximation to the steady-state amplitudes reached in the simulation, both providing good fits to the C major tone profile.

Moreover, the simulation extends the predictions by including synaptic plasticity and multi-frequency interactions. These two features allow the tone E to achieve greater amplitude than F in this example. This feature of the simulation addresses a systematic flaw in the predictions of the purely theoretical model, in which the F (4:3) is predicted to be more stable than the E (5:4), because the closed-form model assumes only two-frequency interactions. The additional amplitude of E arises as a combination (three-frequency interaction) of C and G. Additionally, the tone B achieves nonzero amplitude in the simulation despite the fact that it does not occur in the context sequence. This also arises as a combination frequency, a combination of G and D. Combination frequencies are learned via synaptic plasticity (Equation A2).

The second test was the theme from Rāga Bilāval, used in the Castellano et al. (1984) experiment. Bilāval was chosen for the simulation because the underlying thāt (scale) uses the same pitches as the Western C major scale. The movie shows the piano roll notation for the stimulus, the time-varying amplitude of the 25 oscillators in the network, and the correlations of the oscillator amplitudes with the Rāga Bilāval tone profile for the Western listeners (upper octave only, C through B). A period of silence is simulated at the end, and the amplitude pattern in the network relaxes toward steady state, which resembles the tone profile for the Western listeners to Rāga Bilāval.