Musical Tonality, Neural Resonance and Hebbian Learning

Edward W. Large

Center for Complex Systems and Brain Sciences, Florida Atlantic University
large@ccs.fau.edu

Abstract. A new theory of musical tonality is explored, which treats the central auditory pathway as a complex nonlinear dynamical system. The theory predicts that as networks of neural oscillators phase-lock to musical stimuli, stability and attraction relationships will develop among frequencies, and these dynamic forces correspond to perceptions of stability and attraction among musical tones. This paper reports on an experiment with learning in a model auditory network. Results suggest that Hebbian synaptic modification can change the dynamic responses of the network in some ways but not in others.

Keywords: Tonality, Stability, Attraction, Neurodynamics, Oscillation, Network, Learning.

1 Introduction

Tonality is a set of stability and attraction relationships perceived among musical tones. Stability and attraction relationships are thought to function analogously to the meaning of words in language. As Zuckerkandl [1] asserts, "... musical tones point to one another, attract and are attracted" and these dynamic qualities "make melodies out of successions of tones and music of acoustical phenomena." Dynamical theories [2,3,4] aim to explain which aspects of tonality may be universal, and which are likely to be learned, based on neurodynamic principles. The current theory claims that 1) the auditory system is a complex, nonlinear dynamical system, 2) the way that the auditory system resonates to sound determines the highly structured perceptual responses that we collectively call "tonal cognition," and 3) Hebbian synaptic modification can change these responses in some ways, but not others.

2 Tonal Theory

Music combines individual sounds into well-formed structures, such as melodies. One feature that the melodies of most musical systems share is that they elicit tonal percepts: Listeners experience feelings of *stability* and *attraction* among tones in a melody. Stability, in this sense, means that one or more tones are perceived as points of repose. One specific tone, called the tonic, provides a

C. Agon et al. (Eds.): MCM 2011, LNAI 6726, pp. 115–125, 2011.

focus around which the other tones are organized in a hierarchy of relative stability (Fig. 1), such that some tones are perceived as more stable than others. Less stable tones provide points of dissonance or tension, more stable tones provide points of consonance or relaxation. Less stable tones are heard relative to more stable ones, such that more stable tones are said to attract the less stable tones [5]. Some theorists describe tonal attraction by analogy to physical forces (e.g., [6]), others as resolution from dissonance to consonance [7].

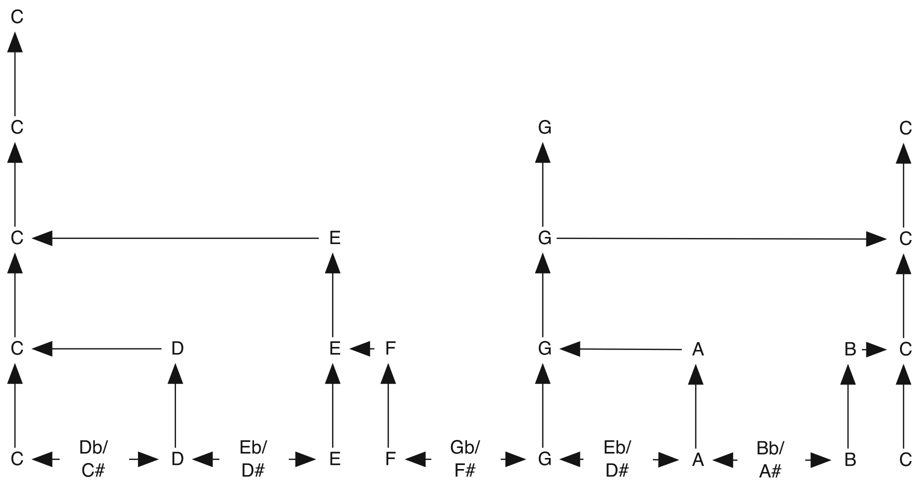

Fig. 1. Music theoretic depiction of tonal stability and attraction for the key of C Major. More stable tones occupy higher positions (upward pointing arrows). Attractions occur at each hierarchical level (horizontal arrows; adapted from [5]).

2.1 Theories of Musical Tonality

Small Integer Ratio (SIR) Theory. The oldest theory of musical tonality states that perceptions of consonance and dissonance are governed by ratios of whole numbers. Pythagoras used the principle of small integer ratios (SIRs) [8] to explain the musical scale that was in use in the West at the time, and Pythagoras and his successors proposed SIR systems for tuning musical instruments, such as just intonation (JI). The study of frequency ratios and tuning systems formed the basis for theoretical approaches to music cognition until the middle of the nineteenth century. Around that time, modern Western equal temperament (ET) came into use. ET intervals closely approximate JI ratios. However, apart from octaves, the intervals are not small integer ratios, they are irrational numbers.

Helmholtz Theory. Ohm [9] proposed that pitch was a consequence of the auditory system's ability to perform Fourier analysis—a linear transformation—on acoustical signals. Helmholtz [10] noted that from a linear processing point of view, there is nothing special about small integer ratios. He hypothesized that the dissonance of a pair of simultaneously sounding complex tones was due to the interference of its pure tone components [10]. So-called *sensory dissonance* occurs when simultaneous tones interact within an auditory critical band [11], and the interaction of pure tone components correctly predicts ratings of consonance for pairs of complex tones [12]. Helmholtz went on to observe that ET intervals (irrational ratios) are approximately as consonant as nearby SIR intervals. He used such arguments to build a strong case against SIR theory.

Modern Tonal Theory. Relating Helmholtz's theory of consonance to musical perceptions of stability and attraction, however, has proved problematic. *Musical* consonance and dissonance differ across cultures and styles, and have changed over the history of Western music [13]. *Musical* (as opposed to sensory) consonance and dissonance now tend to be defined in sequential terms: "...a dissonance is that which requires resolution to a consonance" [13]. Such definitions rule out Helmholtz theory as a potential explanation because the sensory dissonance phenomenon is heard only for simultaneously sounded tones, not for sequences (e.g., melodies). Mainly for these reasons, it is argued that—beyond transient sensations of roughness—particular frequency ratios do not matter *at all* in higher-level music cognition. Rather, the auditory system transforms musical notes into *abstract symbols* unrelated to frequency, in much the same way that spoken words are transformed into abstract symbols whose meaning is unrelated to their sound. Importantly, correlational evidence is interpreted to mean that stability and attraction relationships are internalized solely through long term exposure to the music of one's culture through a process of statistical learning, e.g. [14,15,16,17,18,19].

2.2 Some Issues with Modern Tonal Theory

According to modern tonal theory, the sounds themselves don't matter, much as the sounds associated with spoken words do not determine what they mean. However, this view does not make intuitive sense; in music, it feels like the sounds *do* matter. Modern theories displaced SIR theories based largely on early evidence that the auditory system analyses sound by linear frequency decomposition. However, this argument fails to account for at least two lines of important recent evidence.

The auditory system is highly nonlinear. The cochlea produces sharp mechanical frequency tuning, exquisite sensitivity and nonlinear distortion products that cannot be explained by linear systems, but can be explained as self-tuned critical oscillations of hair cells, e.g. [20]. In the central auditory system, action potentials phase-lock to acoustic stimuli at many different levels, including cochlear nucleus, superior olive, inferior colliculus (IC), thalamus and A1 [21,22]. Nonlinear response curves have been identified in the central auditory systems of animals [23,24,25]. In humans, nonlinear responses to musical intervals have been measured in the auditory brainstem response [26]. Such results provide evidence of nonlinear resonance in the auditory system all the way from the cochlea to the primary auditory cortex. In nonlinear systems there is something special about integer ratios.

There is a cross-cultural tendency toward SIR intervals. A recent review observed "a strong propensity for perfect consonances (2:1, 3:2, 4:3) in the scales of most cultures, the main exceptions being the ofttime highly variable scales of cultures that are either preinstrumental, or whose main instruments are of the xylophone type" [8] (p. 249). Moreover, the three largest non-Western musical traditions—Chinese, Indian, and Arab-Persian—all employ tuning systems based on small integer ratios [8]. Such tuning systems may be more than 9,000 years old [27]. While there is no one "natural scale", the strong cross-cultural tendency toward SIR frequency relationships may point toward an underlying universal.

3 Dynamical Model of Auditory Processing

In the current approach, nonlinear cochlear responses are modeled as nonlinear oscillations of hair cells (cf. [20]), and phase-locked responses of auditory neural populations are modeled as phase-locked neural oscillations [28]. Here, we briefly present the basic conceptual components of the model.

Recently, nonlinear models of the cochlea have been proposed to simulate the nonlinear responses of outer hair cells. It is important to recognize that outer hair cells are thought to be responsible for the cochlea's extreme sensitivity to soft sounds, excellent frequency selectivity and amplitude compression [20]. Models of nonlinear resonance that explain these properties have been based on the Hopf normal form for nonlinear oscillation, and are generic [29]. Normal form (truncated) models have the form

$$\frac{dz}{dt} = z(\alpha + i\omega + \beta|z|^2) + x(t) + h.o.t. \tag{1}$$

where z is a complex-valued state variable, ω is the intrinsic oscillator frequency ($\omega = 2\pi f$, f in Hz), α is a bifurcation parameter, and the value $\alpha = 0$ is the critical value, or bifurcation point. When $\alpha > 0$, the system can spontaneously oscillate. $\beta < 0$ is a nonlinear damping parameter, which prevents oscillation amplitude from blowing up when $\alpha > 0$. $x(t)$ denotes linear forcing by an external signal. The term $h.o.t.$ denotes *higher-order terms* of the nonlinear expansion that are truncated (i.e., ignored) in normal form models. Nonlinear oscillator address behaviors that linear filters do not, such as extreme sensitivity to weak signals, amplitude compression and high frequency selectivity.

A canonical model was recently derived from a model of neural oscillation in excitatory and inhibitory neural populations [28,30]. The canonical model (Eq. 2) is related to the normal form [29] (Eq. 1), but it has properties beyond those of Hopf normal form models because the underlying, more realistic oscillator model is fully expanded, rather than truncated. The complete expansion of higher-order terms produces a model of the form

$$\frac{dz}{dt} = z(\alpha + i\omega + (\beta_1 + i\delta_1)|z|^2 + \epsilon\frac{(\beta_2 + i\delta_2)|z|^4}{1 - \epsilon|z|^2}) + cP(\epsilon, x(t))A(\epsilon, \bar{z}) \tag{2}$$

There are, of course, similarities with the normal form model. The parameters, ω, α and β_1 correspond to the parameters of the truncated model. β_2 is an additional amplitude compression parameter, and c represents strength of coupling to the external stimulus. Two frequency detuning parameters δ_1 and δ_2 are new in this formulation, and make oscillator frequency dependent upon amplitude. The parameter ϵ controls the amount of nonlinearity in the system. Here, $x(t)$ represents a generic input, that could represent either external input (i.e., a sound), or network input from afferent, efferent, or internal network connectivity. In the latter case, $x = \sum a_j z_j$ where a_j ranges over a row of a connectivity matrix A (i.e., a_j is a row vector) and z_j is the j^{th} oscillator in a column vector representing a network state.

The canonical model (Eq. 2) is more general than the Hopf normal form (Eq. 1) and encompasses a wide variety of behaviors that are not observed in linear resonators, including compressive nonlinearities and frequency detuning. Most importantly, coupling to the stimulus is nonlinear and has a passive part, $P(\epsilon, x(t))$, and an active part, $A(\epsilon, z)$, producing higher order resonances. In such nonlinear systems, higher order resonances—also known as "distortion products"— produce neural responses at frequencies that are not physically present in the stimulus, which may correspond to harmonics ($k \times f_1$), subharmonics (f_1/m), summation frequencies (e.g., $f_1 + f_2$), difference frequencies (e.g., $f_2 - f_1$), and integer ratios (e.g., $k \times f_1/m$), where f_1 and f_2 are stimulus frequencies, and k and m are integers.

Higher order resonance is a key feature of this model, leading to predictions of stability and attraction. Fig. 2 illustrates entrainment regions and stability/attraction relationships in a tonotopically arranged network of oscillators, each stimulated with a signal external frequency. Gray regions are regions in which oscillators resonate, or entrain, to the stimulus. Higher-order resonances are found at frequencies that form integer ratios with the external stimulus, and the entrainment regions increase in extent with the strength of coupling. More *stable* resonances have larger resonance regions, and these are colored in darker shades. Where regions overlap, the stronger resonance overpowers the weaker resonance, and attraction effects are observed [4]. Higher-order resonance produces the types of frequency responses that have been reported in the central auditory nervous system. Recently, a variant of this model predicted precisely the nonlinearities observed in brainstem responses to musical intervals [26], and did so with high accuracy for both consonant and dissonant intervals [31]. Such models predict ratings of consonance/dissonance [3] as well as empirical ratings of stability (e.g., [32]) for major and minor modes, with high precision [4].

We have recently developed a canonical version of Hebbian learning that can dynamically evolve connections between oscillators of different frequencies through detection of multifrequency phase coherence [4].

$$\dot{c}_{ij} = -\delta_{ij} c_{ij} + k_{ij} \frac{z_i}{1 - \sqrt{\epsilon z_i}} \frac{\bar{z}_i}{1 - \sqrt{\epsilon \bar{z}_i}} \tag{3}$$

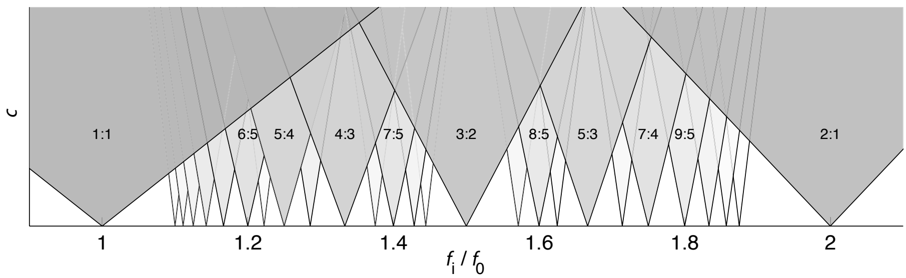

Fig. 2. Bifurcation diagram showing natural resonances in a tonotopic nonlinear oscillator array as a function of connection strength and frequency ratio (adapted from [4])

where c_{ij} is complex, representing the magnitude and phase of the connection between any two nonlinear oscillators at a point in time [33], and δ_{ij} and k_{ij} are real parameters representing the speed of change of the connection. The variables, z_i and z_j are the complex valued state variables of the two oscillators connected by c_{ij}. Previous analysis of the learning rule suggested that as a listener experiences Western musical sequences, connections within and between networks would essentially be pruned, learning the frequencies of Western equal temperament (ET). However, no experiments with learning were reported [4].

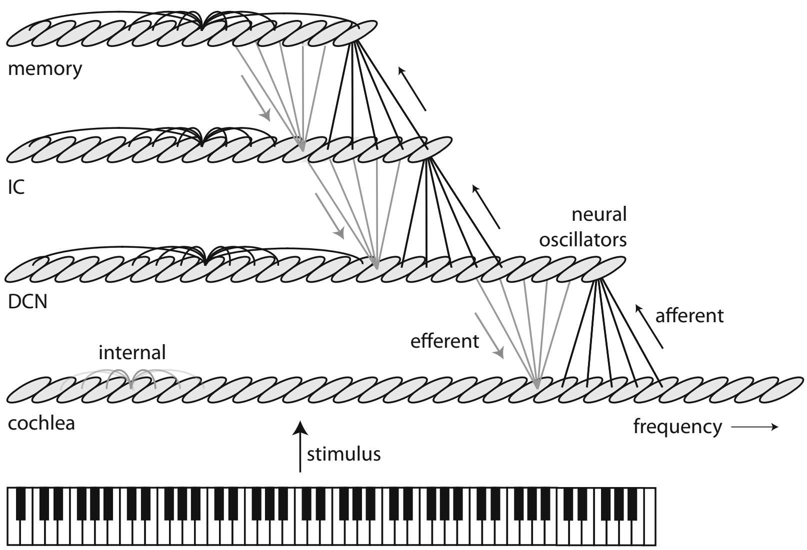

Fig. 3. Model architecture of phase-locking neural dynamics in the auditory system. Fixed connections are shown in gray, learned connections in black. Higher levels operate only at lower frequencies, modeling deterioration of phase-locking at higher auditory levels.

With these building blocks in place, is possible to construct complex models of phase-locking neural networks in the central auditory pathway, and make predictions about listeners with differing levels of auditory experience. Our simple model of the early auditory system includes a model cochlea, dorsal cochlear nucleus (DCN) and inferior colliculus (IC). Cochlear analysis is modeled with a network of locally coupled outer hair cell oscillators, each tuned to a distinct intrinsic frequency [34,35]. Phase-locked responses in the DCN and IC are modeled as the phase-locking of neural oscillators. To this network we add a memory network, which could be interpreted as MGB / A1, but is justified here mainly based on psychological grounds, as memory is a key component in tonal cognition and perception. High-frequency phase-locking deteriorates as the auditory pathway is ascended. To capture this, the upper frequency limit of GFNNs decreases at each processing level, as illustrated in Fig. 3. Note that the multi-layered architecture is not redundant: the transformations are nonlinear, so information is added at each level.

4 A Learning Experiment

To study the behavior of the learning algorithm, a stimulus was generated consisting of two complex, steady state tones, shown in Fig. 4(A). Tone 1 was a harmonic complex consisting of frequencies 500, 1000, 1500, 2000, and 2500 Hz. Tone 2 was a harmonic complex consisting of frequencies 600, 1200, 1800, 2400, and 3000 Hz. The interval was a minor third, tuned according to JI (6:5).

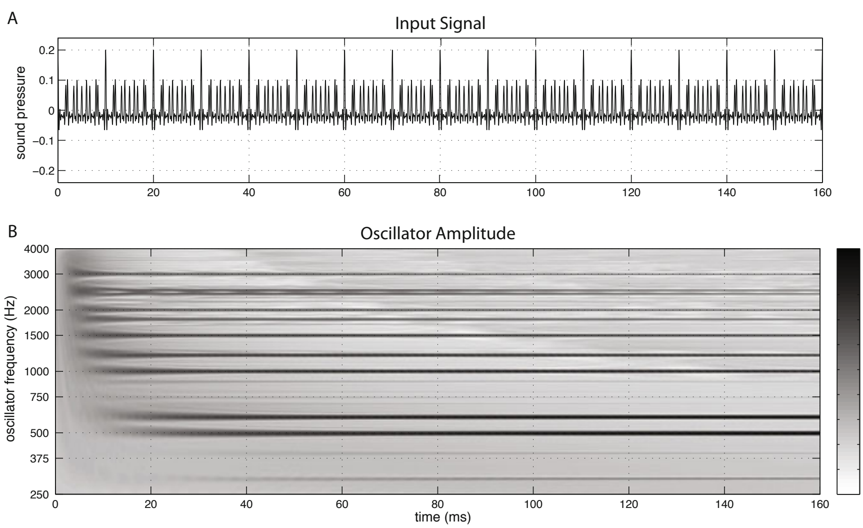

Fig. 4. (A) A harmonic musical interval (minor third) made up of two complex tones. (B) A spectrogram-like representation—oscillator amplitude as a function of time (horizontal) and natural frequency (vertical axis)—showing the response of the Layer 3 (IC) network to this stimulus.

The network of Fig. 3 processed the sound mixture. Layers 1 and 2 of the network of oscillators operated in the critical parameter regime (i.e., $\alpha = 0$), with the oscillators poised between damped and spontaneous oscillation, Layer 3 operated in the active parameter regime (i.e., $\alpha > 0$) resulting in low amplitude spontaneous oscillations. Layer 4 operated in a generalized Hopf (a.k.a. Bautin) parameter regime, (i.e., $\alpha < 0$, $\beta_1 > 0$, $\beta_2 < 0$), enabling a persistent memory of the stimulus frequencies (cf. [4]). Internal connectivity in the cochlea (gray) was local and fixed, simulating local basilar membrane coupling among outer hair cells. All efferent connections (gray) were set to zero. The connections in black were initialized to all-to-all connectivity at low amplitude and zero phase; these connections were learned. The response of the Layer 3 (IC) network to this stimulus (oscillator amplitude, $|z|$, as a function of time) is shown in Fig. 4(B).

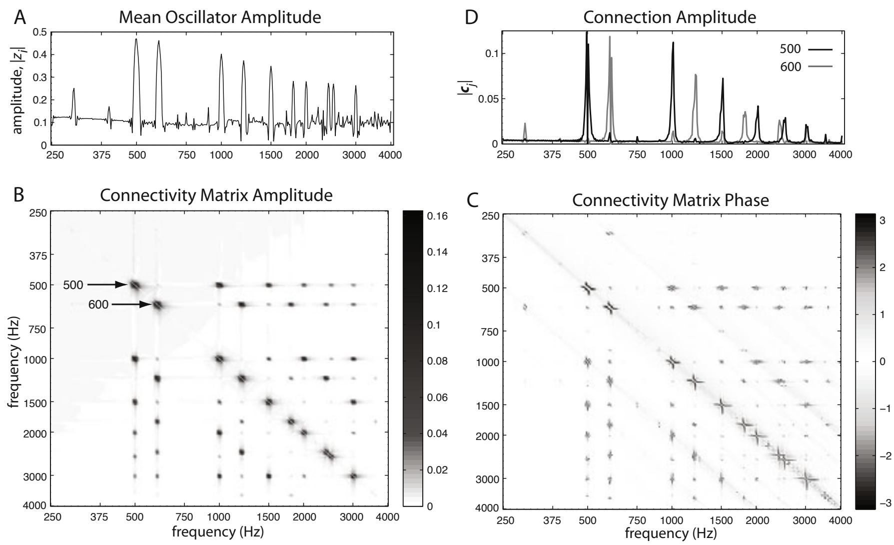

Fig. 5. Internal connection matrices learned in the third layer of the network after the processing the minor third. Amplitude response of the oscillator network, averaged over the last 20 ms. Panels B and C show the magnitude and phase of the connection matrix respectively. Panel D focuses on two rows of the amplitude matrix (Panel B), showing the amplitudes as a function of frequency.

The learning algorithm discussed above was run synchronously, evolving connections as the network processed the stimulus. The result of learning is shown in Fig. 5. Panel A shows the amplitude response of the oscillator network, averaged over the last 20 ms. Reading counterclockwise, Panels B and C show the amplitude and phase of the connection matrix. Note that in the amplitude matrix (Panel B) the peaks in the rows corresponding to the 500 Hz and 600 Hz oscillators are different. Connections are learned from those oscillators whose activity is phase coherent with the oscillators of interest (500 and 600 Hz) over the relevant time scale. Panel D focuses on two rows of the amplitude matrix (Panel B), showing the amplitudes as a function of frequency. Note that the strongest connections are to harmonics, revealing the components of the two different sources, Tone 1 and Tone 2. However, lower amplitude connections also exist connecting the fundamental frequencies (and some harmonics) of the minor third interval. Thus the algorithm learns appropriate connections associating the harmonics of individual tones, as well as connections associating the fundamentals of simultaneously presented intervals. The latter connections are weaker, because the 6:5 ratio is a weaker high-order resonance. This initial study used simple stimuli and short training protocols to analyze the behavior of the algorithm. Ongoing studies are using longer pieces of music and learning sequential contingencies.

5 Conclusions

Tonality is a universal feature of music, found in virtually every culture, but tonal "languages" vary across cultures with learning. Here, a model auditory system, based on knowledge of auditory organization and general neurodynamic principles, was described and studied. The model provides a direct link to neurophysiology and, while simplified compared to the organization and dynamics of the real auditory system, it makes realistic predictions [31]. Analysis of the model suggests that certain musical universals may arise from intrinsic neurodynamic properties (i.e., nonlinear resonance). Moreover, preliminary results of learning studies suggest that different tonal languages may be learned in such a network through passive exposure to music. In other words, this neurodynamic theory predicts the existence of a dynamical, universal grammar (cf. [36]) of music.

Acknowledgements

This work was supported by NSF grant BCS-1027761. The author thanks Nicole Flaig for her comments on this manuscript.

References

1. Zuckerkandl, V.: Sound and Symbol: Music and the External World. Princeton University Press, Princeton (1956)
2. Large, E.W., Tretakis, A.E.: Tonality and Nonlinear Resonance. In: Avanzini, G., Lopez, L., Koelsch, S., Majno, M. (eds.) The Neurosciences and Music II: From Perception to Performance. Annals of the New York Academy of Sciences, vol. 1060, pp. 53–56 (2005)
3. Shapira Lots, I., Stone, L.: Perception of musical consonance and dissonance: an outcome of neural synchronization. Journal of The Royal Society Interface 5(29), 1429–1434 (2008)
4. Large, E.W.: A dynamical systems approach to musical tonality. In: Huys, R., Jirsa, V.K. (eds.) Nonlinear Dynamics in Human Behavior. Studies in Computational Intelligence, vol. 328, pp. 193–211. Springer, Heidelberg (2010)

5. Lerdahl, F.: Tonal Pitch Space. Oxford University Press, New York (2001)
6. Larson, S.: Musical forces and melodic expectations: Comparing computer models and experimental results. Music Percept 21(4), 457–498 (2004)
7. Bharucha, J.J.: Anchoring effects in music: The resolution of dissonance. Cogn. Psychol. 16, 485–518 (1984)
8. Burns, E.M.: Intervals, scales, and tuning. In: Deustch, D. (ed.) The Psychology of Music, pp. 215–264. Academic Press, San Diego (1999)
9. Ohm, G.S.: Uber die Denition des Tones, nebst daran geknufter Theorie der Sirene und ahnlicher tonbildender Vorichtungen. Ann. Phys. Chem. 135(8), 513–565 (1843)
10. Helmholtz, H.L.F.: On the sensations of tone as a physiological basis for the theory of music. Dover Publications, New York (1863)
11. Plomp, R., Levelt, W.J.M.: Tonal consonance and critical bandwidth. J. Acoust. Soc. Am. 38, 548–560 (1965)
12. Kameoka, A., Kuriyagawa, M.: Consonance theory part II: Consonance of complex tones and its calculation method. J. Acoust. Soc. Am. 45, 1460–1471 (1969)
13. Dowling, W.J., Harwood, D.L.: Music Cognition. Academic Press, San Diego (1986)
14. Krumhansl, C.L.: Cognitive foundations of musical pitch. Oxford University Press, NY (1990)
15. Cuddy, L.L., Lunney, C.A.: Expectancies generated by melodic intervals: Perceptual judgements of melodic continuity. P&P 57, 451–462 (1995)
16. Schellenberg, E.G.: Expectancy in melody: Tests of the implication realization model. Cognition 58, 75–125 (1996)
17. Knopoff, L., Hutchinson, W.: An index of melodic activity. Interface 7, 205–229 (1978)
18. Temperley, D.: The cognition of basic musical structures. MIT Press, Cambridge (2001)
19. Pearce, M.T., Wiggins, G.A.: Expectation in Melody: The Influence of Context and Learning. Music Perception 23(5), 377–405 (2006)
20. Eguìluz, V.M., Ospeck, M., Choe, Y., Hudspeth, A.J., Magnasco, M.O.: Essential nonlinearities in hearing. PhRvL 84(22), 5232 (2000)
21. Langner, G.: Periodicity coding in the auditory system. Hear. Res. 60, 115–142 (1992)
22. Joris, P.X., Schreiner, C.E., Rees, A.: Neural Processing of Amplitude-Modulated Sounds. Physiol. Rev. 84(2), 541–577 (2004)
23. Escabi, M.A., Schreiner, C.E.: Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. J. Neurosci. 22(10), 4114–4131 (2002)
24. Langner, G.: Temporal processing of periodic signals in the auditory system: Neuronal representation of pitch, timbre, and harmonicity. Z. Audiol. 46(1), 80–21 (2007)
25. Sutter, M.L., Schreiner, C.: Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. J. Neurophysiol. 65(5), 1207–1226 (1991)
26. Lee, K.M., Skoe, E., Kraus, N., Ashley, R.: Selective Subcortical Enhancement of Musical Intervals in Musicians. J. Neurosci. 29(18), 5832–5840 (2009)
27. Zhang, J., Harbottle, G., Wang, C., Kong, Z.: Oldest playable musical instruments found at Jiahu early Neolithic site in China. Nature 401, 366–368 (1999)
28. Large, E.W., Almonte, F., Velasco, M.: A canonical model for gradient frequency neural networks. Physica D: Nonlinear Phenomena 239(12), 905–911 (2010)
29. Hoppensteadt, F.C., Izhikevich, E.M.: Weakly Connected Neural Networks. Springer, New York (1997)

30. Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13, 55–80 (1973)
31. Large, E.W., Almonte, F.: Phase-locked neural oscillation predicts human auditory brainstem responses to musical intervals. PNAS Under Review (2011)
32. Krumhansl, C.L., Kessler, E.J.: Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol. Rev. 89(4), 334–368 (1982)
33. Hoppensteadt, F.C., Izhikevich, E.M.: Synaptic organizations and dynamical properties of weakly connected neural oscillators II: Learning phase information. Biol. Cybern. 75, 126–135 (1996)
34. Duke, T., Julicher, F.: Active traveling wave in the cochlea. PhRvL 90(15), 158101 (2003)
35. Kern, A., Stoop, R.: Essential role of couplings between hearing nonlinearities. Phys. Rev. Lett. 91(12), 128101–128104 (2003)
36. Prince, A., Smolensky, P.: Optimality: From Neural Networks to Universal Grammar. Science 275, 1604–1610 (1997)