Chapter 7
Neurodynamics of Music
Edward W. Large

7.1 Introduction

Music is a high-level cognitive capacity, similar in many respects to language (Patel 2007). Like language, music is universal among humans, and musical systems vary among cultures and depend upon learning. But unlike language, music rarely makes reference to the external world. It consists of independent, that is, self-contained, patterns of sound, certain aspects of which are found universally among musical cultures. These two aspects – independence and universality – suggest that general principles of neural dynamics might underlie music perception and musical behavior. Such principles could provide a set of innate constraints that shape human musical behavior and enable children to acquire musical knowledge. This chapter outlines just such a set of principles, explaining key aspects of musical experience directly in terms of nervous system dynamics. At the outset, it may not be obvious that this is possible, but by the end of the chapter it should become clear that a great deal of evidence already supports this view. This chapter examines the evidence that links music perception and behavior to nervous system dynamics and attempts to tie together existing strands of research within a unified theoretical framework.

The basic idea has three parts. The first is that certain kinds of musical structures tap into fundamental modes of brain dynamics at precisely the right time scales to cause the nervous system to resonate to the musical patterns. Exposure to musical structures causes the formation of spatiotemporal patterns of activity on multiple temporal and spatial scales within the nervous system. The brain does not “solve” problems of missing fundamentals, it does not “compute” keys of melodic sequences, and it does not “infer” meters of rhythmic input. Rather, it resonates to music. The second part is that certain aspects of this process can be described with concepts that are already well-developed in neurodynamics, including oscillation of neural populations, rhythmic bursting, and neural synchrony. Dynamical analysis
enables the description of the time-varying behavior of neural populations at the level of macroscopic variables. This approach also provides a means for moving between physiological and psychological levels of description, allowing a rather direct link between universal principles of neurodynamics and universal elements of music. The third and final part is that dynamic pattern formation corresponds directly to our experience of music. In other words, perceptions of pitch and timbre, feelings of stability and attraction, and experiences of pulse and meter arise as spatiotemporal patterns of nervous system activity. Section 7.2 introduces some of the relevant concepts from neurodynamics. The subsequent three sections consider, respectively, three essential and universal elements of music.

7.2 An Introduction to the Neurodynamics of Music

7.2.1 Dynamical Systems in Neuroscience and Psychology

Over the past several years, enormous progress has been made toward detailed understanding of nervous system dynamics, and mathematical models are now available that capture this behavior with considerable precision. Models of single neurons at the level of ion channels have now been available for more than 50 years (Hodgkin and Huxley 1952), and more recently the dynamical analysis of single-neuron models has explained and categorized the various kinds of behaviors observed in single neurons (Hoppensteadt and Izhikevich 1997; Izhikevich 2007). Starting in the 1960s and 1970s, analyses of small networks of neurons began to clarify the behavior of local neural populations (Wilson and Cowan 1973; Kuramoto 1975; e.g., Amari 1977). For example, Fig. 7.1a shows the connections between individual members of local excitatory and inhibitory subpopulations that are sufficient to sustain oscillation. Dynamical systems analyses have shown how such connectivity leads to the emergence of various types of dynamic behaviors from such a simple system; these include spiking, oscillation, bursting, and even more complex patterns as shown respectively in Fig. 7.1b (e.g., Crawford 1994; Strogatz 2000; Stefanescu and Jirsa 2008). Recently, with the aid of massive computing power, large-scale simulations have begun to investigate global interactions among local neural populations. In one large-scale simulation of thalamocortical dynamics based on models of various individual neuron types, realistic connectivity among local populations (derived from diffusion tensor imaging, see Fig. 7.1c) led to spontaneous emergence of global spatiotemporal patterns, including waves and rhythms, and functional connectivity on different scales (Izhikevich and Edelman 2008).

Because most cognitive functions are subserved by interactions among brain networks distributed over various subcortical and cortical areas, the studies described above have the potential to elucidate the neurodynamic underpinnings of cognition. It has even been argued that certain features of the complex dynamics observed in neural systems correlate well with key aspects of conscious experience.
Nevertheless, what we know about brain dynamics often seems to be disconnected from the observations we make and theories that we test at the level of human behavior (e.g., Barrett 2009). For example, linguistic theories of prosody, syntax, and semantics are not easily conceived in terms of the neurodynamics of synapses, neurons, and circuits (Poeppel and Embick 2005). Empirically, individual neurophysiological events are often observed to correlate with certain predictions about cognitive function (e.g., Kutas and Hillyard 1980; Tallon-Baudry and Bertrand 1999). However, behavior-level theories are generally not described at the level of neurodynamics (with some exceptions, e.g., Kelso 1995; Large and Jones 1999); rather, attempts are made to explain how brain dynamics might implement abstract computational mechanisms required by cognitive theories (see, e.g., Prince and Smolensky 1997; Jackendoff 2003).

Recent empirical and theoretical results suggest that, unlike linguistic communication, musical behavior may not require postulation of abstract computational mechanisms, but may be explainable directly in terms of neurodynamics. To facilitate understanding of this approach, this section introduces a few of the basic concepts of neurodynamics. The first is the notion of a local population of excitatory and inhibitory neurons, as illustrated in Fig. 7.1a. Such populations can give rise to several behaviors, illustrated in Fig. 7.1b, the simplest and most well-understood of which are oscillation, bursting and resonance. Because each of these behaviors has psychological significance, the remainder of this section describes oscillation and resonance in some detail, while bursting is visited toward the end of the chapter.
Because there are many different mathematical models that can be used to describe neural behavior, the principal concern is to choose a level of mathematical abstraction that is appropriate for the type of data that are available and the type of predictions that are desired. At the physiological level, individual neuron dynamics can be modeled in detail by Hodgkin–Huxley equations (Hodgkin and Huxley 1952), and more specialized models of neural oscillation are also available (FitzHugh 1961; Nagumo et al. 1962; Wilson and Cowan 1973; Hindmarsh and Rose 1984). It is important to keep in mind that actual neurons and neural networks are real dynamical systems. At any given time, a dynamical system has a state that can be modeled as a point in an appropriate state space. A dynamical model is a mathematical formalization – often a differential equation – that describes the time evolution of the point’s position in state space. Stability is a fundamental property of a dynamical system, which means that the qualitative behavior of its time evolution is not affected by small perturbations of the trajectory. Figure 7.1b shows four types of stable trajectories in excitatory–inhibitory neural networks. Section 7.2.3 discusses two important stable states: resting states (equilibria) and periodic trajectories (limit cycles). An attractor is a stable state to which a dynamical system evolves after a sufficiently long time. Thus, points that are close enough to an attractor return to the attractor even if temporarily disturbed, for example, by an external stimulus.

Returning to neurons and neural networks, resting states correspond to stable equilibria, and tonic spiking states correspond to limit cycle attractors (Izhikevich 2007). Analysis of the transition between states in dynamical models is called bifurcation analysis (Wiggins 1990). Bifurcation analysis is facilitated by the transformation of a complex dynamic model to a generic form, called a normal form. Interestingly, this analysis transforms virtually any model of neural oscillation into the same normal form, under certain assumptions that are generally reasonable for neural systems. This analysis reveals that neural oscillations share a set of universal properties, independent of many details (Wiggins 1990; Hoppensteadt and Izhikevich 1997). A canonical model is the simplest (in analytical terms) of a class of equivalent dynamical models, and can be derived using normal form theory. The canonical model we introduce in Eq. (7.3) was derived, using normal form theory, from a model of the interaction between excitatory and inhibitory neural populations (Wilson and Cowan 1973; Large et al. 2010). However, it is generic, so it could also be derived from other models of nonlinear oscillation (including outer hair cell models; see Julicher 2001). The canonical model uncovers universal properties, making predictions that hold under a rather general set of assumptions (Hoppensteadt and Izhikevich 1997). This makes the canonical model especially attractive from the point of view of modeling human perception and behavior. Some relevant generic properties of neural oscillation are described in Sect. 7.2.4.

Section 7.2.3 describes how the nervous system can resonate to sound, at various frequencies and on multiple timescales. The conceptual model is a network of
Neurodynamics of Music

7.2.3 Networks of Neural Oscillators Resonate to Sound

One way to understand nonlinear resonance is to first consider linear resonance. A common signal processing operation is frequency decomposition of a complex input signal, for example, by a Fourier transform. Often this operation is accomplished...
via a bank of linear bandpass filters processing an input signal, \( x(t) \). For example, a widely used model of the cochlea is a gammatone filter bank (Patterson et al. 1992), which – for comparison with our model – can be written as a differential equation:

\[
\dot{z} = z(\alpha + i\omega) + x(t)
\]  

(7.1)

where the overdot denotes differentiation with respect to time (e.g., \( dz/dt \)), \( z \) is a complex-valued state variable, \( \omega \) is radian frequency (\( \omega = 2\pi f \) in Hz), and \( \alpha < 0 \) is a linear damping parameter. The term \( x(t) \) denotes linear forcing by a time-varying external signal. Because \( z \) is a complex number at every time, \( t \), it can be rewritten in polar coordinates revealing system behavior in terms of amplitude, \( r \), and phase, \( \phi \). This transformation is not reproduced here, but amplitude and phase of oscillations are discussed in Sect. 7.2.4. Resonance in a linear system means that the system oscillates at the frequency of stimulation, with amplitude and phase determined by system parameters. As stimulus frequency, \( \omega_0 \), approaches the oscillator frequency, \( \omega \), oscillator amplitude, \( r \), increases, providing band-pass filtering behavior.

Recently, nonlinear models of the cochlea have been proposed to simulate the nonlinear responses of outer hair cells. It is important to note that outer hair cells are thought to be responsible for the cochlea’s extreme sensitivity to soft sounds, excellent frequency selectivity, and amplitude compression (e.g., Eguiluz et al. 2000). Models of nonlinear resonance that explain these properties have been based on the \textit{Hopf normal form} for nonlinear oscillation, and are generic. Normal form (truncated) models have the form

\[
\dot{z} = z(\alpha + i\omega + \beta |z|^2) + x(t) + h.o.t.
\]  

(7.2)

Note the surface similarities between this form and the linear resonator of Eq. (7.1). Again \( \omega \) is radian frequency, and \( \alpha \) is still a linear damping parameter. However, in this nonlinear formulation, \( \alpha \) becomes a bifurcation parameter that can assume both positive and negative values, as well as \( \alpha = 0 \). The value \( \alpha = 0 \) is termed a bifurcation point and is discussed further in Sect. 7.2.4.1. \( \beta < 0 \) is a nonlinear damping parameter, which prevents amplitude from blowing up when \( \alpha > 0 \). Again, \( x(t) \) denotes linear forcing by an external signal. The term \textit{h.o.t.} denotes higher-order terms of the nonlinear expansion that are truncated (i.e., ignored) in normal form models. Like linear resonators, nonlinear oscillators come to resonate with the frequency of an auditory stimulus; consequently, they offer a sort of filtering behavior in that they respond maximally to stimuli near their own frequency. However, there are important differences in that nonlinear models address behaviors that linear ones do not, such as extreme sensitivity to weak signals, amplitude compression and high frequency selectivity; these are discussed in detail in the Sect. 7.2.4. The compressive gammachirp filterbank exhibits similar nonlinear behaviors, described within a signal processing framework (Irino and Patterson 2006; see also see Patterson et al., Chap. 2).
A canonical model was recently derived from a model of neural oscillation in excitatory and inhibitory neural populations (Wilson and Cowan 1973; Large et al. 2010). The canonical model (Eq. [7.3]) is related to the normal form (Eq. [7.2]; see e.g., Hoppensteadt and Izhikevich 1997), 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

\[ \dot{z} = z(\alpha + i\omega + (\beta_1 + i\delta_1)|z|^2 + \frac{(\beta_2 + i\delta_2)\epsilon|z|^4}{1 - \epsilon|z|^2}) + cP(\epsilon, x(t))A(\epsilon, \overline{z}) \]  

(7.3)

There are again surface similarities with the previous models. The parameters, \(\omega\), \(\alpha\), and \(\beta_1\) correspond to the parameters of the truncated model. \(\beta_2\) is an additional amplitude compression parameter, and \(c\) represents strength of coupling to the external stimulus. Two frequency detuning parameters, \(\delta_1\) and \(\delta_2\), are new in this formulation, and make oscillator frequency dependent on amplitude (see Fig.7.4c). The parameter \(\epsilon\) controls the amount of nonlinearity in the system. Most importantly, coupling to a stimulus is nonlinear (not discussed in further detail here, but see Large et al. 2010) and has a passive part, \(P(\epsilon, x(t))\) and an active part, \(A(\epsilon, \overline{z})\), producing nonlinear resonances that are discussed in Sect. 7.2.4.4. Helmholtz’s (1863; see Sect. 7.3) difference tone, proposed to explain the pitch of the missing fundamental, was a passive nonlinearity. The three-frequency resonance of Cartwright et al. (1999a; see Sect. 7.3), proposed to explain residue pitch shift (Schouten et al. 1962), arises through the interaction between passive and active nonlinearities in this system. The canonical model given by Eq. (7.3) is more general than the Hopf normal form and encompasses a wide variety of behaviors that are not observed in linear resonators, some of which are discussed next.

### 7.2.4 Some Universal Properties of Nonlinear Oscillation

#### 7.2.4.1 Andronov–Hopf Bifurcation

In the absence of stimulation, a nonlinear oscillator can display two qualitatively different stable states, both of which depend upon the specific value of the bifurcation parameter, \(\alpha\). Figure 7.3a illustrates the transition between a stable equilibrium and a stable limit cycle, called the Andronov–Hopf bifurcation. When \(\alpha < 0\) the system behaves as a damped oscillator, but when \(\alpha > 0\) (negative damping) the system generates a spontaneous oscillation. \(\alpha = 0\) is the bifurcation point – also referred to as the critical value of the parameter – the value at which behavior changes from damped to spontaneous oscillation or vice versa. Other kinds of bifurcations that also lead to spontaneous oscillation can be found in this canonical model (see Guckenheimer and Kuznetsov 2007). Models of neural oscillation often assume spontaneous activity, i.e., \(\alpha > 0\). Models of cochlear outer hair cells assume critical oscillation, i.e., \(\alpha = 0\).
When the system oscillates spontaneously ($\alpha > 0$) and a stimulus is present, the oscillation will phase-lock, or entrain, to the stimulus. Figure 7.3b is a bifurcation diagram showing some phase-locked regions for an active oscillator network. Phase-locked states (Fig. 7.3b), are found at higher-order resonances (e.g., integer

\begin{align*}
1:2 & \\
1:1 & \\
3:2 & \\
2:1 & \\
3:1 & 
\end{align*}

For active oscillation, there are well-defined boundaries between phase-locked (shaded areas) and non-phase-locked states.

### 7.2.4.2 Entrainment

When the system oscillates spontaneously ($\alpha > 0$) and a stimulus is present, the oscillation will phase-lock, or entrain, to the stimulus. Figure 7.3b is a bifurcation diagram showing some phase-locked regions for an active oscillator network. Phase-locked states (Fig. 7.3b), are found at higher-order resonances (e.g., integer
ratios, discussed in Sect. 7.2.4.4) and are stable. On the horizontal axis, \( f/f_0 \) denotes the ratio of oscillator frequency to stimulus frequency, and \( A \) denotes the stimulus amplitude. The diagram shows regions of attraction, where an oscillator will adopt an instantaneous frequency that is different from its natural frequency. These oscillations are attracted to integer ratios of the stimulus frequency, such as 1:2, 1:1, 3:2, and 3:1 (see also Fig. 7.4b and c). For active oscillation, well-defined boundaries are found between phase-locked (shaded areas) and non-phase-locked states.

7.2.4.3 Nonlinear Amplitude Responses

Figure 7.4a illustrates the response of three different resonator models to sinusoidal stimulation, presented at their own natural frequencies. The curves show the amplitude responses for a linear filter (Eq. [7.1]), and two versions of a critical (i.e., \( \alpha = 0 \)) nonlinear resonator, namely the Hopf normal form (Eq. [7.2]) and the fully expanded canonical model (Eq. 7.3). Linear filters have linear amplitude response. By contrast, both the Hopf normal form (truncated) model and the fully expanded canonical model exhibit extreme sensitivity to weak signals, one of the characteristic properties thought to explain nonlinear cochlear responses (e.g., Eguíluz et al. 2000), discussed in Sect. 7.3. While both also exhibit amplitude compression, amplitude is fully compressive in the canonical model, but not in the Hopf normal form.

7.2.4.4 Higher-Order Resonance

Figure 7.4b shows the response of the three different resonator networks to a complex tone comprising two frequency components \((f_1, f_2)\). Resonances are shown for a linear filter bank (Eq. [7.1]), and two versions of a critical oscillator array (i.e., \( \alpha = 0 \) for all oscillators), namely the Hopf normal form (Eq. [7.2]) and canonical model (Eq. [7.3]). Higher-order resonances are found only in the canonical network, due to the nonlinear coupling. Higher-order resonance means that a nonlinear oscillator network responds to a pure tone at the frequency \( f \) with activity not only at \( f \) but also at harmonics \((2f, 3f, ...)\), subharmonics \((f/2, f/3, ...)\) and integer ratios \((2f/3, 3f/4, ...)\) of \( f \). Further, if a complex tone is presented that contains multiple frequencies, a nonlinear network will respond at combination frequencies \((f_2 - f_1, 2f_1 - f_2, ...)\) as well. These responses follow orderly relationships and can be predicted given stimulus amplitudes, frequencies, and phases.

This feature of nonlinear resonance has important implications for understanding the behavior of such systems. A nonlinear oscillator network does not merely transduce signals; it actually adds frequency information, which may account for pattern recognition and pattern completion, among other things. The cochlea is known to produce audible higher-order resonances, including difference tones and harmonics (e.g., Robles et al. 1997), as produced by the canonical model. Neural pattern
**Fig. 7.4** Amplitude responses predicted by different resonance models. (a) A linear filter bank model (dashed line; Eq. [7.1]) vs. a critical Hopf normal form model (gray solid line; Eq. [7.2]) vs. a critical canonical model (solid black line; Eq. [7.3]) responding to stimuli at their own natural
completion based on nonlinear resonance may explain the perception of pitch in missing fundamental stimuli (Cartwright et al. 1999a), the perception of tonal relationships (e.g., Large and Tretakis 2005; Shapira Lots and Stone 2008), and the perception of pulse and meter in rhythmic patterns (Sect. 7.5; for a review, see McAuley, Chap. 6).

### 7.2.4.5 Frequency Selectivity and Detuning

Figure 7.4c presents the results of three simulations of an array of critical ($\alpha = 0$) nonlinear oscillators, based on Eq. (7.3). The frequencies of the oscillators in the array vary from 250 to 4,000 Hz, along a logarithmic frequency gradient, and the stimulus is a sinusoid with a frequency of 1,000 Hz. Each simulation shows the result for a different stimulus amplitude. These simulations illustrate two important properties of nonlinear resonance. First, the response at low stimulus amplitude levels reveals that high-frequency selectivity is achieved. As stimulus amplitude increases, frequency selectivity deteriorates due to nonlinear compression ($\beta_1, \beta_2 < 0$). Second, due to frequency detuning ($\delta_1, \delta_2 \neq 0$) the peaks in the resonance curve begin to bend as oscillator amplitude ($r$) increases. Both types of response agree with measurements in living intact cochleae (e.g., Ruggero 1992; see Fig. 7.5a). Also, as the stimulus amplitude increases, higher-order resonances appear at harmonics, subharmonics, and integer ratios.

### 7.2.4.6 Connectivity and Learning

Connections between oscillators can be modified, for example, via Hebbian learning (Hoppensteadt and Izhikevich 1996b), providing a mechanism for synaptic plasticity wherein the repeated and persistent coactivation of a presynaptic cell and a postsynaptic cell lead to an increase in synaptic efficacy between them. The number of possible synapses between excitatory and inhibitory subpopulations implies that a connection between two oscillators has both a strength and a natural phase (Hoppensteadt and Izhikevich 1996a). Both connection strength and phase can be learned by the Hebbian mechanism if a near-resonant relationship exists between their frequencies.

---

Fig. 7.4 (continued) frequency. Amplitude response is linear for the linear filter, partially saturates for the normal form, and fully saturates for the canonical model. (b) Three resonator networks (linear, Hopf, and canonical) responding to a two-frequency stimulus ($f_1$ and $f_2$). Oscillator amplitude is shown in logarithmic units as a function of resonator frequency. The canonical network produces harmonics and combination tones of the stimulus frequencies unlike the linear filter or the normal form model. (c) A canonical network (Eq. [7.3]) stimulated with a sinusoid at 1,000 Hz, for three different stimulus amplitudes (different curves). As stimulus amplitude increases, frequency selectivity deteriorates, frequency detuning is observed, and higher-order resonances appear.
The Hebbian learning mechanism can learn connections between oscillators of different frequencies (Large in press).

7.2.5 Summary

Neural resonance can arise from the interaction between excitatory and inhibitory subpopulations. Canonical models of neural oscillation capture universal properties that are independent of physiological details. The same generic properties are also

Fig. 7.5  (a) Laser velocimetric data from a living chinchilla’s cochlea displaying the root-mean-square velocity of one point on the basilar membrane as a function of stimulus frequency. Each curve represents a different level of stimulation (dB SPL). Note the dramatic increase in bandwidth and the detuning as intensity increases. (From Ruggero 1992, with permission) (b) Hopf resonance. The amplitude response, \( r \), to different levels of forcing is obtained from Eq. (7.2); the amplitude of forcing increases in increments of 10 dB for successive curves from bottom to top. At resonance the response increases as the one-third power of the forcing, whereas away from the resonance the response is linear in the forcing (From Eguíluz et al. 2000, with permission)
found in other kinds of nonlinear oscillations, such as mechanical oscillations at the cellular scale (Choe et al. 1998; Eguíluz et al. 2000; Julicher 2001). Canonical models are also available for burst oscillation, and these share some of the basic properties of limit cycle oscillation described in the preceding text (Izhikevich 2007). Although a detailed discussion of mathematical models of burst oscillation is beyond the scope of this chapter, the potential role of burst oscillation in rhythm perception is considered toward the end of the chapter. Gradient frequency networks of nonlinear oscillators can resonate to sound. Nonlinear resonators share some filtering properties with linear resonators, but also exhibit many properties that are not found in linear resonators. These include spontaneous oscillation, nonlinear amplitude responses, and higher-order resonance. Higher-order resonance is of critical importance; it implies a sort of pattern-formation behavior that is appropriate for describing the perception of structured patterns in musical sounds. For neural oscillation, there is also a canonical version of the Hebbian learning rule, enabling the development of connectivity among neural oscillators. The following sections will consider cochlear resonance, central auditory nonlinearities and entrainment of cortical rhythms from a dynamical systems point of view. The dynamic approach will lead to an understanding of the relationship between such phenomena and experiences of pitch, tonality, and rhythm in music perception.

### 7.3 Cochlear Resonance, Neural Resonance, and Pitch Perception

The first attempts to explain the physical basis of music perception concerned pitch. Shortly after Fourier methods were developed, Ohm (1843) proposed that pitch was a consequence of the auditory system’s ability to perform Fourier analysis on acoustical signals. In Ohm’s view, the pitch of a complex tone was a Fourier component of the sound. Helmholtz (1863) agreed that the ear acts as a rough Fourier analyzer and proposed the hypothesis that the analysis was performed by the basilar membrane. He described the cochlea as a time-frequency analysis mechanism that decomposes sounds into sinusoidal components for subsequent analysis by the central auditory nervous system. In the 1960s von Békésy (1960) demonstrated experimentally that the hypothesis of Helmholtz was essentially correct, that is, the basilar membrane carries out a frequency analysis of acoustic stimuli. Von Békésy’s observations – using measurements on human cadavers – suggested that cochlear responses are linear over the range of physiologically relevant sound intensities.

Since then, however, a number of problems have arisen with the notion that the cochlea performs a passive, linear analysis. The weakest audible sounds impart energy per cycle no greater than that of thermal noise (Bailek 1987), and the system operates over a range of intensities that span at least 14 orders of magnitude. Gold (1948) recognized that these properties were incompatible with a passive, linear cochlea; rather, additional energy must be added into the system by active feedback. He also noted that if an active resonator underwent a Hopf bifurcation (see Fig. 7.4a),
it would oscillate spontaneously, and the ear would emit sound. Recently, the discov-
ery of spontaneous otoacoustic emissions (Kemp 1979; Murphy et al. 1996) con-
firmed Gold’s prediction. Moreover, laser-interferometric velocimetry performed on
living, intact cochleae has revealed exquisitely sharp mechanical frequency tuning,
which deteriorates with increasing stimulus amplitude (Ruggiero 1992; Ruggiero
et al. 1997), as illustrated by different curves in Fig. 7.5a. These and related discov-
eries have led to the proposal that active amplification, in the form of Andronov–
Hopf type nonlinearities, is the basic mechanism of the mammalian cochlear
response (Choe et al. 1998; Camalet et al. 1999). The sharp mechanical frequency
tuning, exquisite sensitivity, and operating range of the cochlea are now explained
as self-tuned critical oscillations of hair cells (Eguíluz et al. 2000; Fig. 7.5b). It
appears that the cochlea performs a type of active, nonlinear time-frequency trans-
formation, using a network of locally coupled outer hair cell oscillators, each tuned
to a distinct intrinsic frequency (eigenfrequency), and driven by an external stimulus
(Duke and Julicher 2003; Kern and Stoop 2003; see also Irino and Patterson 2006).

Regarding perception, Seebeck (1841) demonstrated that if most of the energy
at the fundamental frequency is removed from the complex spectrum of a periodic
sound, the perceived pitch remains unchanged, matching the pitch of a sinusoid
with the frequency of the missing fundamental. Seebeck (1843) proposed a
periodicity detection theory for pitch perception in complex sounds. However,
Helmholtz (1863) embraced Ohm’s approach, proposing that a physical compo-
nent at the missing fundamental frequency, a “difference combination tone,” could
be generated by passive nonlinearities of the ear (similar to that in Eq. [7.3]). But
Schouten et al.‘s (1962) famous pitch-shift experiments demonstrated that the
missing fundamental is not a difference tone. Schouten’s theory of pitch was based
on the periodicity properties of the nonresolved “residue” components of the
stimulus. Eventually, because peripheral theories failed to explain psychophysical
experiments and because dichotically presented stimuli also elicit pitch perception
(e.g., Houtsma and Goldstein 1972), central processor theories for pitch percep-
tion arose (e.g., Goldstein 1973; Terhardt 1974). Complex pitch perception is still
debated by theorists. It is determined neither solely by the spectral content of
sound nor solely by its temporal structure (Plack and Oxenham 2005).

Recently, key theoretical advances have been made in understanding multifre-
quency resonance behaviors of nonlinear oscillators (Cartwright et al. 1999b), and
this may have relevance for auditory perception. In experiments and numerical simu-
lations, Cartwright and colleagues worked out the organization of higher-order reso-
nances in representative nonlinear oscillators, and argued convincingly that such
organization is universal across a large class of systems. They further showed that
nonlinear resonance explains the “pitch shift of the residue,” one of the important
unexplained cases of pitch perception (Schouten et al. 1962; Cartwright et al.
1999a). If pitch depends on a difference tone (a passive nonlinearity), then when the

---

1 Schouten (1938) showed that removing the fundamental component completely from the acoustic
stimulus did not alter the pitch, and Licklider (1956) showed that the same pitch was heard even when
the frequency region that would normally be occupied by the fundamental was masked by noise.
components of a missing fundamental harmonic complex are all shifted by the same amount the pitch should not change, because their difference remains the same. But Schouten et al. showed that perceived pitch does indeed shift, as illustrated in Fig. 7.6 for harmonics of 200 Hz. Figure 7.6 also shows that physical frequencies produced by generic nonlinear oscillators, acted upon by two independent periodic excitations, can reproduce the experimental data from Schouten’s famous pitch-shift experiments with impressive precision. This provides strong evidence that nonlinear resonance is a viable neural mechanism for pitch perception. In Eq. (7.3), the pitch-shift resonance of Cartwright and colleagues arises through the interaction between passive and active nonlinearities and is nontrivial. Thus, higher-order resonance of neural oscillation could explain important aspects of pitch perception.

Nonlinear oscillations can arise through the interaction of excitatory and inhibitory neural populations, as illustrated in Fig. 7.1a and b, and there is a growing body of evidence consistent with nonlinear oscillation in the central auditory system. In mammals, action potentials phase-lock to both fine time structure and temporal envelope modulations at many different levels in the central auditory system, including cochlear nucleus, superior olive, inferior colliculus (IC), thalamus, and A1 (Langner 1992; Joris et al. 2004), and recent evidence points to a key role for synaptic inhibition in maintaining central temporal representations. Hyperpolarizing inhibition is phase-locked to the auditory stimulus and has been shown to adjust the temporal sensitivity of coincidence detector neurons (Grothe 2003), while stable pitch representation in the
IC may be the result of a synchronized inhibition originating from the ventral nucleus of the lateral lemniscus (Langner 2007). Such evidence suggests that nonlinear oscillation may be a good model for phase-locked central auditory responses.

Recent evidence also supports higher order resonance in neural activity. Multipeaked spectrotemporal receptive field (STRF) curves have been identified in cat primary auditory cortex, some with responses to second and third harmonics of the fundamental frequency (Sutter and Schreiner 1991). Modulation-rate selective cells in the auditory midbrain of Pollimyrus, which receive both excitatory and inhibitory input, have been successfully modeled as nonlinear oscillators (Large and Crawford 2002). Nonlinear STRFs have been identified in cat IC (Escabi and Schreiner 2002), and neurons in the IC of the gerbil have been observed to respond at harmonic ratios (e.g., 3:2, 2:1, 5:2; cf. Fig. 7.4c) with the temporal envelope of the stimulating waveform (Langner 2007). Nonlinear $2f_1 - f_2$ difference tones (see Fig. 7.4b) have been identified in brain stem auditory evoked potentials of guinea pigs (Chertoff and Hecox 1990), in human frequency-following responses using electroencephalography (EEG; Pandya and Krishnan 2004), and in auditory cortex, using steady-state methods in magnetoencephalography (MEG) (Purcell et al. 2007). These results provide evidence of higher-order resonance in the auditory system all the way from the cochlea to the primary auditory cortex.

7.3.1 Summary

The auditory nervous system is highly nonlinear, and observed responses are consistent with the generic predictions of nonlinear resonance, possibly arising in excitatory–inhibitory networks of the auditory nervous system. One potentially important functional consequence would be the perception of pitch, which may arise through an active nonlinear mechanism that is generic to nonlinear oscillators (i.e., Eq. [7.3]). Fourier-based approaches rely on linear systems theory almost exclusively, thus they describe human perceptual capabilities only approximately. However, generic models of neural oscillation (e.g., Eqs. [7.2] and [7.3]) are available, which are able to capture functionally important nonlinearities. As a result, such models may be able to capture many human perceptual and cognitive capabilities in a physiologically realistic way, but without strong dependence on physiological details. This observation has important implications not only for pitch perception, but also for other aspects of musical experience.

7.4 Neurodynamics of Tonality

The preceding section focused on responses to individual tones. But music is more than the perception of isolated tones; it involves the combination of tones into larger structures, such as melodies. Musical melodies typically involve discrete
217

neurodynamics of music

tones, organized in archetypal patterns that are characteristic of musical genres, styles, and cultures. These patterns may be related to a *scale*, an ordered collection of all the tones used in a given melody, which summarizes the frequency ratios that govern the intervals between tones in a melody. One feature the melodies of most musical systems share is that they give rise to tonal percepts. Listeners experience feelings of *stability* and *attraction* among tones in a tonal melody. Stability means that one or more tones are perceived as points of repose. One specific tone, called the tonic, provides a focus around which the other tones are dynamically organized, and there is a hierarchy of relative stability, 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 (e.g., Lerdahl 2001). Some theorists have described tonal attraction by analogy to physical forces, such as gravity (Larson 2004); others link it to the resolution of musical dissonance (Bharucha 1984). Zuckermandl (1956) argued that these dynamic tonal qualities make “melodies out of successions of tones and music out of acoustical phenomena (p. 21).” But what processes in the nervous system could give rise to such perceptions in music?

The oldest theory of musical consonance is that perceptions of consonance and dissonance are governed by ratios of whole numbers. Pythagoras is thought to have first articulated the principle that intervals of small integer ratios (cf. Figs. 7.3b and 7.4c) are pleasing because they are mathematically pure (Burns 1999). He used this principle to explain the musical scale that was in use in the West at the time, and Pythagoras and his successors proposed small-integer-ratio systems for tuning musical instruments, such as Just Intonation (JI). Modern Western equal temperament (ET), divides the octave into 12 intervals that are precisely equal on a log scale. ET approximates JI, and transposition in ET is perfect, because the frequency ratio of each interval is invariant. Apart from octaves, however, the intervals are not small integer ratios, they are irrational. The fact that intervals based on irrational ratios are approximately as consonant as nearby small integer ratios is generally considered *prima facie* evidence against the theory that musical consonance derives from the mathematical purity of small integer ratios.

Helmholtz (1863) hypothesized that the dissonance of a pair of simultaneously sounding complex tones was due to the interference of its pure tone components, explaining dissonance as a sensation of roughness produced by the beating of sinusoids. This phenomenon, called sensory dissonance, is heard when simultaneous tones interact within an auditory critical band (Plomp and Levelt 1965), and the interaction of pure tone components correctly predicts ratings of consonance for pairs of complex tones (Kameoka and Kuriyagawa 1969). However, there are a number of problems that arise with Helmholtz theory as a theory of musical consonance (Dowling and Harwood 1986). For one thing, the sensory dissonance phenomenon is heard for isolated clusters of simultaneously sounded tones, but not for sequentially presented tones (i.e., melodies). Moreover, musical consonance and dissonance are intrinsically dynamic: “… a dissonance is that which requires resolution to a consonance” (Dowling and Harwood 1986). Recently, Shapira Lots and
Stone (2008) used the theory of coupled neural oscillators to explain why simple frequency ratios are important for music perception. They used the width of the resonance regions (cf. Fig. 7.3b) for higher order resonances to predict the consonance of the intervals of chromatic scales. Their analysis revealed that this method of ordering higher order resonances corresponds to the standard ordering of consonance often listed in Western music theory (Helmholtz 1863), suggesting that neural synchrony may be important in music perception. One piece of relevant evidence comes from a recent study in which nonlinear responses to harmonic musical intervals were measured in the auditory brain stem response. Two musical intervals, the major sixth (E3 and G2) and the minor seventh (E3 and F#2), were found to give rise to highly nonlinear responses including difference and summation tones (cf. Fig. 7.4b), revealing nonlinear processing of simultaneously sounded musical intervals in the auditory system (Lee et al. 2009).

Tonal perceptions such as stability, attraction, perceptual categorization, and learning of tonal relationship may depend on neural resonance as well (Large and Tretakis 2005; Large in press). Perceptual categorization and discrimination experiments reveal that musicians show categorical perception of melodic intervals (Burns and Campbell 1994), and nonmusicians also perceive pitch categories (Smith et al. 1994). Resonance regions (Fig. 7.3b) predict perceptual categorization of musical intervals, because resonances not only affect oscillators with precise integer ratios; they also establish patterns of resonant neighborhoods (Fig. 7.7c). Thus, even if resonance center frequencies do not precisely match stimulus frequencies, as connection strength increases, larger regions of the network resonate, emanating from integer ratios, and encompassing nearby ratios (Large and Tretakis 2005).

Simplicity of frequency ratios has been shown to account not only for judgments of consonance and dissonance, but also for judgments of similarity and discrimination of tone patterns across a wide range of tasks and listeners (Schellenberg and Trehub 1994). In one study, 6-month-old infants detected changes to sequentially presented pairs of pure tones (intervals) only when the tones were related by simple frequency ratios (Schellenberg and Trehub 1996). In adults as well, changes from patterns with simple frequency ratios to those with more complex ratios were more readily detected than were changes from complex ratios to simpler ratios. This implies that memories for tone sequences with small integer ratio relationships are more stable than memories for complex integer relationships. Large (in press) found a similar result in an oscillator network simulation. Tones with small integer ratio relationships (1:1, 5:4 and 3:2 – a tonic triad) produced a stable memory in the neural oscillator network (cf. Fig. 7.2). Although a leading tone (8:15 ratio with the tonic frequency) could be stabilized through external stimulation, when the external stimulus was removed, the leading tone frequency lost stability as those oscillators that had responded at the leading tone frequency began to resonate at the tonic frequency. In other words, the tonic frequency functioned as an attractor of nearby oscillations. Thus, nonlinear resonance predicts both memory stability of small integer ratios and tonal attraction among sequentially presented frequencies (Large in press).

Krumhansl and Kessler (1982) measured the stability of each tone within a musical key directly, by asking listeners to rate how well individual pitches fit within a
tonal context (see Fig. 7.7a, b). Higher goodness-of-fit ratings imply higher stability, so, for example, C and G are the most stable in both the C-major and C-minor tonal contexts. When applied to Western music, the measured hierarchies are consistent with music-theoretic accounts and agree with frequency-of-occurrence statistics for tonal songs (Krumhansl 1990). It is possible to apply a dynamic analysis to predict tonality rating data. Nonlinear resonance predicts that the relative stability of higher order resonances is given by $\varepsilon^{(k+m-2)/2}$, where $k$ and $m$ are the numerator and denominator, respectively, of the frequency ratio (Hoppensteadt and Izhikevich 1997). Here $\varepsilon$ is a parameter that controls coupling nonlinearity (see Eq. [7.3]). One could use this fact and assume that tones heard in a tonal context would be stabilized in memory (based on the simulation results described above), to create a single-parameter ($\varepsilon$) fit to the stability judgments. Theoretical predictions of stability based on this analysis matched perceptual judgments well (Large in press), as shown in Fig. 7.7a and b.

The tuning systems of the world’s largest musical cultures, Western, Chinese, Indian, and Arab-Persian, are based on small integer ratio relationships (Burns 1999). However, each tuning system is different, and this has led to the notion that

---

**Fig. 7.7** Comparison of theoretical stability predictions and human judgments of perceived stability for two Western modes: (a) C-major and (b) C-minor. Open circles denote mean goodness-of-fit ratings from Krumhansl and Kessler (1982), and solid lines represent nonlinear resonance predictions. (c) An “Arnold tongues” bifurcation diagram showing natural resonances in a gradient frequency nonlinear oscillator array as a function of connection strength and frequency ratio (similar to Fig. 7.3b). An infinite number of resonances are possible on this interval; shown here are the unison (1:1), the octave (2:1), and the 25 most stable resonances in between. (From Large in press, with permission) Shading of each resonance region reflects the intrinsic stability of the ratio, used for the predictions shown in (a) and (b). Where regions overlap, less stable frequencies are attracted to more stable frequencies.

---

2 ET in the West is designed to approximate small integer ratio tuning and has been in widespread use for less than 150 years.
frequency relationships do not matter in high-level music cognition; rather, auditory transduction of musical notes results in abstract symbols, as in language (see, e.g., Patel 2007). If this were true, stability and attraction relationships would also have to be learned presumably based solely on the frequency-of-occurrence statistics of tonal music (for a current overview, see Krumhansl and Cuddy, Chap. 3). However, Hebbian learning of multifrequency relationships can provide a theoretical basis for the acquisition of frequency relationships. As the music of one’s culture is heard, auditory networks would learn the most stable attractors whose center frequencies closely approximate the experienced relationships. Natural resonances predict significant constraints on which frequency relationships can be learned, as illustrated in Fig. 7.7c. Hebbian synaptic modification would effectively prune some resonances, while retaining or enhancing others (Large in press). This reasoning suggests that frequency relationships are learned depending on the frequency relationships employed in the music of a particular style or culture. However, stability and attraction relationships are not learned per se, but are intrinsic to neural dynamics given a particular set of frequency relationships.

7.4.1 Summary

Nonlinear resonance predicts the perceived dynamics of tonal organization and important aspects of neurophysiological responses, qualitatively and quantitatively. Thus, nonlinear resonance may provide the neural substrate for a substantive musical universal, similar to the concept of universal grammar in linguistics (Prince and Smolensky 1997). However, in the case of music, perceptual universals are predicted by universal properties of nonlinear resonance, properties that provide direct links to neurophysiology. Learning would alter connectivity to establish different resonances and different tonal relationships. According to this approach, stability and attraction relationships would not be learned on the basis of statistical properties of tone sequences; instead, because nonlinear resonance predicts stability and attraction, and because stability and attraction are correlated with sequence statistics, nonlinear resonance predicts tone frequency statistics. Thus, higher-order resonances may create resonant tonal fields in the central nervous system, and musical melodies may be perceived in relation to such fields, creating a dynamical context within which perception of tone sequences takes place.

7.5 Resonating to Rhythm

Musical structure is found not only in the pitch dimension, but also in the time dimension. Jones (1976) originally proposed that neural rhythms entrain to the temporal structure of environment stimuli. Entrainment of intrinsic neural dynamics would enable
dynamic attending, providing a basis for temporal expectancy and facilitating perception of events that occur at expected points in time. Musical rhythms are highly temporally structured sequences of acoustic events, and in most musical rhythms people perceive periodicity, called pulse or beat, and structured patterns of accentuation among pulses, called meter (London 2004). Pulse can be thought of as a frequency, and meter as a pattern of frequencies, which can be transcribed as arrangements of dots (reflecting beats) aligned with a musical score, as shown in Fig. 7.8a. The fundamental pulse periodicity (the rate at which one taps with a rhythm) is notated as a single row of beats, and the pattern of strong and weak pulses as additional rows of beats at related frequencies (for a more thorough discussion of pulse and meter, see McAuley, Chap. 6). Sometimes metrical frequencies are physically present in stimulus rhythms; sometimes they are not. For example, in the clave rhythm of Fig. 7.8a, the frequencies of the pulse and meter are almost completely absent.

The temporal relationships observed in human musical interactions are among the most elaborate observed in nature (for a review, see Large 2008). When humans

![Fig. 7.8](attachment:image.png)

(a) Pulse and meter of the 3–2 son clave rhythm à la Lerdahl and Jackendoff (1983). At 500 ms/quarter note, the pulse frequency would be 2 Hz or 120 bpm. Results of a linear (b; Eq. [7.1]) and nonlinear (c; Eq. [7.3]) analysis of the 3–2 son clave rhythm. The linear analysis reveals very little energy in this rhythm at the pulse frequency or at other metrical frequencies. The nonlinear analysis responds at all metrical frequencies (as well as many others) via higher-order resonance.
temporally coordinate in musical interactions, we synchronize – or more generally, entrain – pulse frequencies. Entrainment is the process whereby two spontaneously oscillating systems, which have different frequencies when they function independently, assume the same frequency, or integer-ratio related frequencies, when they interact. In general, entrainment of neural oscillations predicts multifrequency coordination at simple frequency ratios such as 1:1, 1:2, 1:3, 2:3, due to higher-order resonance (Figs. 7.3b and 7.4c). Such entrainment is found in everyday musical interactions and has been observed in behavioral studies involving perception (e.g., Vos 1973), attention (e.g., Barnes and Jones 2000), and motor coordination (e.g., Parncutt 1994). Moreover, newborns can perceive pulse (Winkler et al. 2009); by 7 months infants discriminate rhythms and categorize melodies on the basis of meter (Hannon and Johnson 2005); and 9-month-old infants detect changes in the context of metric rhythms but not in sequences that induce a metric percept only weakly or not at all (Bergeson and Trehub 2006). Toddlers as young as 2.5 years are capable of entraining motor rhythms with periodic sequences (Provasi and Bobin-Begue 2003; Kirschner and Tomasello 2009), and even some animals can entrain motor rhythms to music (Patel et al. 2009; Schachner et al. 2009).

The two complexities of rhythm that are the most troublesome for theoretical accounts of pulse and meter are syncopation and temporal fluctuation. Syncopation refers to rhythms in which accented events occur on weaker positions in the metrical structure while leaving nearby stronger positions empty (Fitch and Rosenfeld 2007). This is illustrated by the clave rhythm of Fig. 7.8a, in which note events occur on only half the beats of the basic pulse and occur often on relatively weak beats. Temporal fluctuation refers both to localized temporal nuances and to larger scale tempo changes (i.e., rubato) that arise in music performance due to motoric, perceptual, and expressive constraints (Palmer 1997; Penel and Drake 1998). Temporal fluctuation is correlated with important aspects of musical structure (Sloboda 1983, 1985; Todd 1985; Palmer 1989), exhibits \( 1/f \) (fractal) structure (Rankin et al. 2009), and conveys affect and emotion to listeners (Sloboda and Juslin 2001).

Several studies have compared people’s ability to entrain to simply structured versus syncopated rhythms (Snyder and Krumhansl 2001; Toiviainen and Snyder 2003; Patel et al. 2005). Level of syncopation is a good predictor of pulse-finding difficulty, and syncopation causes some off-beat taps and some switches between on beat and off beat tapping (Snyder and Krumhansl 2001; Patel et al. 2005). Overall, however, humans are quite good at entraining to the pulse of even highly syncopated rhythms. How is this possible? Figure 7.8b and c illustrates two resonance predictions for a highly syncopated rhythm (Fig. 7.8a), one generated by a linear filter bank (Eq. [7.1]) and the other by a critical nonlinear resonator array (Eq. [7.3]). The linear filter bank responds at frequencies that are physically present in the time series, finding very little energy at 2 Hz (the pulse frequency for this rhythm). There are several strong peaks, however, with the strongest at 1.33 Hz, corresponding to the time interval between the first and second notes (i.e., \( 1/0.750 \) s). By contrast, a nonlinear oscillator array finds its strongest peak at 2 Hz, the pulse frequency, due to higher-order resonance. Such observations predict that perceived pulse in highly syncopated rhythms arises through higher order resonance.
Most listeners are also good at synchronizing with music that contains expressive timing fluctuations (Drake et al. 2000). A number of studies have investigated the response of nonlinear oscillators to temporal fluctuation in music (McAuley 1995; Toiviainen 1998; Large and Palmer 2002; Cont 2008), generally finding support for oscillator predictions. However, one surprising recent finding in this area is that people are able to predict, rather than simply react to, expressive temporal fluctuations (Repp 2002; Rankin et al. 2009). It has been hypothesized that listeners exploit musical and fractal structure to predict tempo changes in music (Repp 2002; Rankin et al. 2009). One important aspect of this ability may be the covert monitoring of multiple metrical frequencies during entrainment (Large et al. 2002; Repp 2008) One nonlinear resonance model captures this phenomenon as coupling between nonlinear oscillators as they respond at different metrical frequencies (Large and Jones 1999; Large and Palmer 2002; Jones 2008).

Recent functional imaging studies have shown that the perception of rhythmic sequences involves multiple, spatially distinct brain regions. Rhythmic information is represented across broad cortical and subcortical networks in a manner that is dependent upon task and rhythmic complexity (Sakai et al. 1999; Grahn and Brett 2007; Jantzen et al. 2007; Chen et al. 2008). Metric rhythms are easier to reproduce, and elicit higher activity in the basal ganglia and supplementary motor area (Grahn and Brett 2007), suggesting that these motor areas play a role in mediating pulse and meter perception. Both performance and neural activity are modulated as musicians and nonmusicians tap in synchrony with progressively more syncopated auditory rhythms (Chen et al. 2008). In perception, secondary motor regions were recruited in musicians and non-musicians, and the dorsal premotor cortex appeared to mediate auditory–motor interactions (Chen et al. 2008). The dorsal auditory pathway is also implicated in rhythm performance, regardless of the modality in which the rhythms are trained and paced (Karabanov et al. 2009). Thus, both auditory and motor areas play key roles in both rhythm perception and rhythm production. A set of brain areas including dorsal auditory pathway areas, dorsal premotor cortex, the supplementary and presupplementary premotor areas, the cerebellum, and the basal ganglia are implicated. A key question is: What is happening in this distributed network?

Using EEG, Snyder and Large (2005) observed that peaks in the power of induced beta- and gamma-band activity anticipated tone onset (average ~0 ms latency), were sensitive to intensity accents, and persisted when expected tones were omitted, as if an event had appeared. By contrast, evoked activity occurred in response to tone onsets (~50 ms latency) and was strongly diminished during tone omissions. Recent MEG studies have found subharmonic rhythmic responses in the beta-band when subjects were instructed to impose a subjective meter on a periodic stimulus (Iversen et al. 2009), and anticipatory responses for periodic and metrical sequences, but not for randomly timed sequences in primary auditory cortex (Fujiocka et al. 2009). Thus, the features of high-frequency brain activity match the main predictions for pulse and meter. Such observations could indicate cortical bursting, which can also arise from excitatory–inhibitory neural circuits (Izhikevich 2007; see Fig. 7.1a). Bursting is a dynamic state where neurons repeatedly fire
groups, or bursts, of action potentials, and each burst is followed by a period of quiescence before the next occurs (Izhikevich 2007). Interburst periods, the time interval between one burst and the next, are generally consistent with timescales of musical pulse and meter. Burst oscillation is currently receiving a great deal of attention in the computational neuroscience literature, and mathematical analyses have shown that rhythmic bursting displays key properties (Coombes and Bressloff 2005; Izhikevich 2007) that are necessary to predict pulse and meter. Figure 7.9 shows a computational simulation of burst oscillation (Izhikevich 2000) responding to a simple rhythm, displaying both entrainment to the stimulus sequence (Fig. 7.9a) and oscillatory persistence in the absence of an element in this sequence (Fig. 7.9b).

Moreover, bursts of high-frequency activity could explain communication between different cortical areas (Brovelli et al. 2004). For example, oscillatory activity in the beta range is widely observed in sensorimotor cortex in connection with motor behavior in humans (Pfurtscheller and Lopes da Silva 1999; Salenius and Hari 2003) and nonhuman primates (Rougeul et al. 1979; Sanes and Donoghue 1993; MacKay and Mendonca 1995). Synchrony of beta oscillations is often observed between different areas of sensorimotor cortex (Murthy and Fetz 1992; Sanes and Donoghue 1993). Moreover, synchronized beta oscillations may bind multiple sensorimotor areas into a large-scale network during motor behavior and carry causal influences from primary somatosensory and inferior–posterior parietal cortices to motor cortex (Brovelli et al. 2004). Anticipatory rhythmic bursts of beta activity may enable communication between auditory and motor cortices in rhythm perception and motor coordination as well. Rhythmic bursts of higher frequency gamma activity may also enable functional communication between different cortical regions. The theoretical picture that
emerges is one of communication, through bursts of high-frequency activity, between different neural areas as they resonate to rhythmic patterns.

Entrainment of rhythmic neural bursting could explain how the perception of pulse and meter arise from listening to complex sequences, as well as the development of expectancy for events occurring in a rhythmic context. Dynamic attending theory (DAT) hypothesizes that endogenous attentional rhythms entrain to temporally structured external events (Jones 1976; Large and Jones 1999). DAT has traditionally been discussed in terms of facilitation of perception to certain external events, and this has found support in a number of recent studies (McAuley and Kidd 1995; Jones and Yee 1997; Large and Jones 1999; Barnes and Jones 2000; Jones et al. 2002; Jones and McAuley 2005; Quené and Port 2005). However, conceptualizing attentional rhythms as rhythmic bursting provides a new hypothesis regarding the role of attention in coordinating the interaction between auditory and motor areas (Large and Snyder 2009). Bursts of beta and gamma band activity that entrain to external rhythms could provide a mechanism for rhythmic communication between distinct brain areas, and attention may facilitate such integration among auditory and motor areas.

### 7.5.1 Summary

Entrainment of endogenous neural rhythms and higher order resonance could explain why metrical percepts favor small integer ratios. It can also explain how people perceive a regular pulse in highly syncopated rhythms and how listeners adapt to frequency fluctuations in expressive performances. Rhythmic bursting in higher frequency bands is a plausible neural correlate of pulse and meter. This could explain not only perceptual facilitation of expected events, but also functional integration of auditory and motor areas.

### 7.6 Summary and Conclusions

As noted at the outset, it is informative to compare theories of music with theories of language. Poeppel and Embick (2005) discuss a “conceptual granularity mismatch” between cognitive and neurobiological mechanisms in language. That is, theories that are typically invoked to account for linguistic computation – in terms of syntax, meter, and semantics – are not related in any obvious way to the neurodynamics of synapses, neurons, and circuits. The theoretical picture they paint is potentially bleak and would seem to require a paradigm shift to reconcile the two approaches to language. However, it may be unnecessary to invite such theoretical difficulties into the musical domain. In music, our experiences of the fundamental universals, including pitch, tonality, and rhythm, can be readily conceived in relation to neurodynamic universals, including limit cycle oscillation, resonance, and rhythmic bursting.
Helmholtz (1863) originally envisioned that a proper understanding of auditory physiology should one day form the basis for a theory of music. However, the auditory system is highly nonlinear, and Poincaré, the father of modern dynamical systems theory, was only a boy when Helmholtz penned the preface to the first edition of *On the Sensations of Tone*. Modern theories of auditory and music perception were built on the foundation of linear resonance. Where linear resonance has proven insufficient to explain cognitive and perceptual phenomena, complex mechanisms and general purpose computation have been recruited to fill the explanatory gaps. Known auditory nonlinearity can be described with well-developed concepts of modern neurodynamics. These phenomena are summarized in high-level dynamical models, called canonical models, which are appropriate for describing the macroscopic dynamics of neural populations and for describing key aspects of perception, cognition, and behavior. Neurodynamic models seem to capture many features of music perception and behavior in their own terms, without the need to resort to more abstract computational descriptions. These observations suggest that our qualitative experiences of music arise as a direct consequence of the interaction of sound with the intrinsic dynamics of the nervous system.

References


