Psychological Review
2000, Vol. 107, No. 4, 885–913

Copyright 2000 by the American Psychological Association, Inc.
0033-295X/00/$5.00 DOI: 10.1037//0033-295X.107.4.885

Implicit Learning of Tonality: A Self-Organizing Approach

Barbara Tillmann
Université de Bourgogne and Dartmouth College

Jamshed J. Bharucha
Dartmouth College

Emmanuel Bigand
Université de Bourgogne

Tonal music is a highly structured system that is ubiquitous in our cultural environment. We demonstrate the acquisition of implicit knowledge of tonal structure through neural self-organization resulting from mere exposure to simultaneous and sequential combinations of tones. In the process of learning, a network with fundamental neural constraints comes to internalize the essential correlational structure of tonal music. After learning, the network was run through a range of experiments from the literature. The model provides a parsimonious account of a variety of empirical findings dealing with the processing of tone, chord, and key relationships, including relatedness judgments, memory judgments, and expectancies. It also illustrates the plausibility of activation being a unifying mechanism underlying a range of cognitive tasks.

Natural environments contain highly structured systems to which we are exposed in everyday life. The human brain internalizes these regularities by passive exposure, and the acquired implicit knowledge influences perception and performance. Aspects of language and music provide two examples of highly structured systems that may be learned in an incidental manner. In each case, there is a paradox. On the one hand, a thorough formal description of the structure has proven to be extremely challenging. On the other hand, native speakers or nonmusician listeners internalize the regularities underlying linguistic or musical structures with apparent ease. A substantial corpus of research has been devoted to the learning process of language, but little has been devoted to the learning of music. The central purpose of the present article is to investigate how implicit knowledge of some basic features of Western musical grammar may be acquired and mentally represented.

Barbara Tillmann, Department of Psychology, Université de Bourgogne, LEAD-CNRS, Dijon, France, and Department of Psychological and Brain Sciences, Dartmouth College; Jamshed J. Bharucha, Department of Psychological and Brain Sciences, Dartmouth College; Emmanuel Bigand, Department of Psychology, Université de Bourgogne, LEAD-CNRS.

This research was supported in part by National Science Foundation Grant SBR-9601287 and National Institutes of Health Grant 2P50 NS17778-18 to Jamshed J. Bharucha and by a grant from the International Foundation for Music Research to Emmanuel Bigand. Barbara Tillmann has been supported by the French Ministry of Education and Research and by the Deutsche Akademische Austauschdienst DAAD. We thank Hervé Abdi, Pierre Perruchet, and Philippe Schyns for insightful comments at different stages of the present work and Carol Krumhansl and Fred Lerdahl for comments on the article.

Correspondence concerning this article should be addressed to Barbara Tillmann, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, New Hampshire 03755. Electronic mail may be sent to barbara.tillmann@dartmouth.edu.

We present a connectionist model that (a) simulates the implicit learning of pitch structures occurring in Western harmony and (b) accounts for a range of empirical findings in music perception. The article is organized in four parts. First, we summarize the regularities underlying the Western tonal musical system that may be internalized by implicit learning and review existing models of tonal knowledge presentation. Second, we propose a connectionist model, based on self-organizing maps (SOMs), that simulates the learning of tonal regularities by mere exposure. Third, we present tests of the trained network with experimental tasks on the perception of tonality. Fourth, we discuss the proposed model and some of its future developments.

IMPLICIT LEARNING OF REGULARITIES IN WESTERN TONAL MUSIC

In the acquisition and representation of knowledge, a distinction is made between explicit and implicit. Explicit learning is the acquisition of declarative knowledge. This is thought to occur by hypothesis testing or rule instruction. Implicit learning is the acquisition of knowledge in an incidental manner without complete verbalizable knowledge of what is learned (Seger, 1994). Implicit learning is seen as a fundamental characteristic of the cognitive system, enabling the acquisition of highly complex information that may not be acquired in an explicit way (Reber, 1989). Implicit learning processes have been studied in the laboratory with artificial material based on statistical regularities. One of the most frequently used experimental situations consists of presenting participants with sequences of events generated by an artificially defined grammar. For example, a finite state grammar generates complex letter strings based on a restricted set of letters (Reber, 1967). After passive exposure to grammatical letter strings, participants were better than chance at differentiating new grammatical letter strings from new nongrammatical ones. Most were unable to explain the rules underlying the grammar in verbal

free reports (e.g., Altmann, Dienes, & Goode, 1995; Dienes, Broadbent, & Berry, 1991; Reber, 1967, 1989).

Artificial materials are simpler than environmental sequences of events. However, the same basic principles of learning may serve as a model for understanding the implicit learning processes in natural environments (Winter & Reber, 1994). For example, several studies attempted to bridge implicit learning of artificial grammar and language learning. Saffran, Newport, Aslin, Tunick, and Barrueco (1997) showed with auditory sequences that participants were able to use the statistical regularities such as transition probabilities of syllables. After the presentation of an artificial language-like auditory sequence (e.g., bupadapatubitutibu . . .), children and adults performed above chance in distinguishing artificial words (e.g., bupada, patubi) from nonwords. Even a brief exposure to a complex natural language induces a sensitivity to structural constraints: After a 12-min presentation of a cartoon film narrated by a native Mandarin speaker, Dutch adults discriminated above chance between real Mandarin words and pseudowords (Zwisterlood, 1990, reported by Altmann et al., 1995). Further results obtained for children's sensitivity to orthographic regularities (Pacton, Perruchet, Fayol, & Cleeremans, in press) support the extension of implicit learning conclusions issued from artificial laboratory research to natural material.

In several domains, implicit learning has been studied with both artificial and natural events. In the music domain, only a few studies investigated implicit learning directly with artificial strings of musical events (Bigand, Perruchet, & Boyer, 1998). However, many studies addressed this issue indirectly with music processing (see Dowling & Harwood, 1986, for a review). Western musical grammar is more complex than the finite state grammars used in implicit learning studies. It may be conceived of as a three-level hierarchical grammar that generates strong regularities in musical pieces. Let us consider some of the most basic rules and regularities.[1]

In Western music, a restricted set of 12 pitch classes (referred to as the tones C, C♯/D♭, D, D♯/E♭, E, F, F♯/G♭, G, G♯/A♭, A, A♯/B♭, B) are combined in highly constrained ways. This set of 12 tones is organized in subsets of seven, called diatonic scales. Depending on the pattern of intervals separating the seven tones, diatonic scales may be major or minor. Applying the two diatonic scales to each of the 12 pitch classes leads to the distinction of 12 major and 12 minor musical keys. The tones that form a musical key (e.g., B-C♯-D♯-E-F♯-G♯-A♯) are more likely to co-occur in Western melodies than tones that do not form a musical key (e.g., B-C-D♯-E-F♯-G♯-A). For each scale, seven diatonic chords may be defined on each of the seven degrees of the scale according to specific harmonic rules. A chord is a simultaneity of three tones, usually called the root, third, and fifth. In the major keys, the chord built on the first, fourth, and fifth scale degree (I, IV, and V) are major, those built on the second, third, and sixth scale degree are minor (ii, iii, vi), and the chord built on the seventh degree is diminished (vii°).[2] Chords form a second order of musical units, and their occurrence is strongly constrained by Western musical grammar. Chords belonging to the same musical key are more likely to co-occur in a given musical piece than chords belonging to different keys.

Keys define a third order of musical units. Some keys share numerous chords and tones. For example, the C-major key shares four chords and six tones with the G-major key, two chords and five tones with the D-major key, and only one tone with the F♯-major key. Keys sharing chords or tones are said to be harmonically related. The strength of these harmonic relationships depends on the number of shared chords or tones. In music theory, keys are conceived spatially as a circle, referred to as the cycle of fifths (Figure 1). The number of steps separating two keys on this circle (whatever the direction of the rotation) defines their harmonic distance. These interkey distances form the basis for strong regularities in pieces of Western music. Key changes are more likely to occur between closely related keys (e.g., C and F or G major) than between less related ones (e.g., C and F♯ major). Interkey distances are also defined between major and minor keys. A major key (e.g., C major) is harmonically related to both its relative minor key (a minor) and its parallel minor key (c minor). These multilevel relations among tones and chords, chords and keys, and major and minor keys define a complex set of possible relations between musical events (see Krumhansl, 1990; Lerdahl, 1988), and they strongly constrain the transition probabilities between musical events as a piece unfolds (Francès, 1958; Piston, 1978).

Another important feature of Western musical grammar is that tones and chords have different structural functions within a key. According to Meyer (1956), "In the major mode in Western music the tonic tone[3] is the tone of ultimate rest toward which all other tones tend to move. On the next higher level the third and fifth of the scale, though active melodic tones relative to the tonic, join the tonic as structural tones; and all the other tones, whether diatonic or chromatic, tend toward one of these" (pp. 214–215). These differences in musical functions create within-key hierarchies. Interestingly, within-key hierarchies are strongly correlated with the frequency of occurrence of tones in Western musical pieces. Tones that occur with greater frequency (the tonic, the fifth, and, to a lesser extent, the third) are those described by music theory as being the most important in a given key. From a psychological point of view, the hierarchically important tones of a key act as stable cognitive reference points (Krumhansl, 1979, 1990) to which the other tones are anchored (Bharucha, 1984).

A within-key hierarchy is also found among the seven chords of a key (Bharucha & Krumhansl, 1983; Krumhansl, Bharucha, & Castellano, 1982). Chords built on the first, fifth, and fourth scale degrees (referred to as, respectively, the tonic, dominant, and subdominant chords) usually have a more central syntactic function than those built on the other scale degrees. For example, a dominant chord followed by a tonic chord is an authentic cadence, which marks finality. In contrast, a subdominant chord followed by a dominant chord is a half cadence, which marks a temporary

[1] A complete description of the Western musical system is beyond the scope of this article. The main purpose of the following section is to highlight the organizational features that give rise to the most salient regularities in Western musical pieces.

[2] Major chords consist of intervals of a major third (four semitones) and a perfect fifth (seven semitones with respect to a reference pitch [the root]). In minor chords, the third is minor (three semitones), and in diminished chords the third is minor and the fifth is diminished (six semitones).

[3] The tonic is the tone on the first degree of the scale. It also gives its name to the key. For example, the tonic of the key of C major is the tone C.

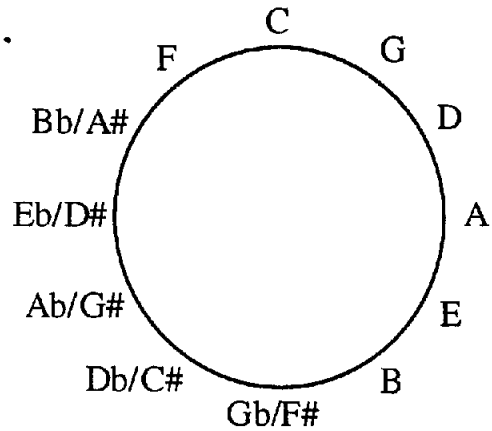

Figure 1. Cycle of fifths representing the distances between major keys.

ending. Both music theorists (Deliège, 1984; Schenker, 1979) and music psychologists (Sloboda, 1985) consider that the authentic cadence acts as a basic syntactic structure in Western music (see Deliège, 1984, and Sloboda, 1985, for a further development).

A critical feature of Western music is that the functions of musical events change with the key context. A C-major chord functions as a stable tonic chord in a C-major context and as a less stable dominant or subdominant chord in the F- or G-major key contexts, respectively. Similarly, a G–C chord sequence forms an authentic cadence in a C-major key context but not in a G-major key context. This context dependency is a fundamental aspect of the Western tonal system: Understanding the function of events with respect to the musical context is crucial to assessing musical grammaticality.

Despite the complexity of the system, sensitivity to musical structure does not require explicit learning. Because musically naive listeners are constantly exposed in everyday life to the regularities underlying the music of their culture, they acquire implicit knowledge of them (Bharucha, 1984; Dowling & Harwood, 1986; Francès, 1958). This implicit knowledge embodies the functions of tones and chords in a key (Tillmann, Bigand, & Madurell, 1998), the relations between different keys (Bartlett & Dowling, 1980; Cuddy & Thompson, 1992a, 1992b; Thompson & Cuddy, 1989), and the change in function of events depending on the key context (Bharucha & Krumhansl, 1983; Bigand, 1993; Bigand, 1997; Bigand & Pineau, 1997; Krumhansl et al., 1982). The internalized representation influences musical memory (Bigand & Pineau, 1996; Cuddy, Cohen, & Mewhort, 1981; Dowling, 1978; Dowling, 1991), musical expectancies (Bharucha & Stoeckig, 1986, 1987; Cuddy & Lunney, 1995), and the restoration of missing musical events (DeWitt & Samuel, 1990). Results generally reveal strong consistency for listeners with different levels of musical expertise. Although musicians usually exhibit better performance than nonmusicians, their overall responses show the same patterns (see also Bigand, Parncutt, & Lerdahl, 1996; A. Cohen, 1994; Croonen & Houtsma, 1994). Questionnaire data even suggest that tonal knowledge is tacit for both groups of participants (Holleran, Jones, & Butler, 1995). Finally, event-related potentials (ERP) studies provide further evidence that musicians and nonmusicians show similar electrophysiological responses to subtle changes in the harmonic function of a target chord (Regnault, Bigand, & Besson, in press).

Our main purpose in this article is to investigate how this implicit knowledge of Western pitch regularities may be represented and learned through passive exposure to musical exemplars. We argue that a model of distributed knowledge offers a possible explanatory framework that accounts for learning in the absence of explicit tutoring and that suggests an underlying mechanism—activation—that unifies a range of psychological tasks. It should be noted that, beyond the complex regularities of pitch structure, the Western tonal system also contains complex temporal regularities defined by metrical and rhythmic structures. In music cognition, the processing of pitch and temporal structures have often been considered independently. Even though this methodological independence remains a matter of debate (Boltz, 1999; Jones & Boltz, 1989; Peretz & Kolinsky, 1993), we focus on pitch regularities only and not on temporal regularities. However, we return to this issue in the General Discussion by considering possible extensions of the connectionist model to integrate some temporal regularities of Western music.

MODELS OF DISTRIBUTED KNOWLEDGE REPRESENTATION

Connectionist models have two principal advantages over traditional rule-based models: (a) The rules governing the domain are not explicit but rather emerge from the simultaneous satisfaction of multiple constraints represented by individual connections, and (b) these constraints themselves can be learned through passive exposure. In the field of artificial grammar learning, a knowledge representation of training stimuli may emerge from associative learning mechanisms of connectionist models. Autoassociator networks memorize stimuli generated by artificial grammars, classify new stimuli, and simulate experimental results even better than exemplar-based models (Dienes, 1992). This approach allows us to interpret the notion of "abstract knowledge" differently than only in the strong sense of rule-like knowledge or of simple sensitivity to stored examples (Cleeremans, 1994). In the language domain, McClelland et al. developed one class of neural net models of knowledge representation. These interactive-activation models of word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) and speech recognition (Elman & McClelland, 1984; McClelland & Elman, 1986) simulate the interaction between knowledge and perception without storing linguistic rules explicitly. Three levels of units represent features, letters (or phonemes), and words. Rule-like behavior emerges from the interactions of a set of word units and letter (or phoneme) units.

In the music domain, a growing number of neural network models have been developed during the last decade. Models of this nature have been offered for pitch perception (Sano & Jenkins, 1991; Taylor & Greenhough, 1994), octave equivalence (Bharucha & Mencl, 1996), chord classification (Laden & Keefe, 1991), and melodic sequence learning (Bharucha & Todd, 1989; Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Page, 1994). Some models simulate more complex aspects of music learning and perception, such as categorization and memory of feature patterns. Gjerdingen (1990) exposed a four-level network based on Grossberg's (1987) adaptive resonance theory to early works of Mozart. The input layer codes music theoretic concepts (i.e.,

harmonic tritone, contrapuntal dissonance) and low-level music features (i.e., melodic contour, pitch of the major diatonic scale plus a unit for the alterations of flat and sharp). The four levels include dynamic short-term memory (Level 1) that leads to the formation of stable categories (Level 2), and a second temporary store (Level 3) that categorizes on a higher level (Level 4). The model develops memories of critical feature patterns and derives categorizations comparable to complex music theoretic concepts (i.e., voice-leading combinations).

Few models attempt to formalize how the multiple relationships among tones, chords, and keys may be represented in a single framework. Griffith's (1994) model simulates how keys are induced from patterns of pitch use and how abstract pitch identities are established from interval use. Supervised and unsupervised paradigms are used in a modular combination, allowing the model to use its own derived information to guide subsequent processes. The model formalizes an inductive mechanism for learning key and scale degrees from melodic sequences. It mostly concentrates on the links between pitch and key and is not designed to address the relationships between tones and chords and chords and keys. The output of the model is compared with music theory but not with empirical data.

Leman (1995; Leman & Carreras, 1997) simulated the perceptual learning of tonal centers by presenting chords and real musical pieces to a framework of two modules. In these simulations, the musical acoustical signal is first processed by an auditory model, and the transformed information defines the input for a self-organizing map. In Leman (1995), three auditory models define three types of input vectors for a variety of chords, and a self-organizing map is trained with these different inputs. Whatever the auditory model used, map units specialize in the detection of chords. After training, the activations on the map in response to a given stimulus reflect the harmonic relatedness of the represented chords to the stimulus. On the basis of activation regions found on the map, the tonal centers are inferred. The trained network is exposed to musical pieces, and the detected changes in tonal centers are compared with music theoretical analyses. In Leman and Carreras (1997), the input signal is derived from neural firing patterns in response to real sound recordings of Bach pieces. An SOM is trained to extract the regularities in these input patterns. After training, tonal centers activated by a given musical stimulus are inferred on the basis of activation regions on the map. The output of the model conforms generally to music theory, and some empirical data are simulated by the model.

Leman's (1995) and Leman and Carreras's (1997) models focus on chords and on tonal centers but do not account for the relationships between tones, chords, and keys. The main interest is to show that higher order units of Western music (i.e., chords or tonal centers) may be learned by passive exposure to a rich acoustic input. The SOMs extract the invariant features of musical sounds that lead to the formation of abstract units. In other words, these models formalize how learning processes may be driven by psychoacoustic features in a bottom-up manner. However, they do not investigate how the acquired knowledge may, in turn, influence the processing of musical events, leading to predictions that can be tested experimentally. The crucial benefit of learning is the use of knowledge to react to environmental stimuli better. Top-down influences facilitate the processing of environmental events, as has been shown in different domains of cognition, including music. For example, once the key of a musical context is recognized, the tones belonging to that key are perceived as more stable than other tones, even if they were not present in the stimulus context (Francès, 1958; Krumhansl, 1990). A model of knowledge representation should be able to account for these top-down effects and for the way they combine with bottom-up influences. In word recognition, connectionist models simulate the influence of knowledge by interactive activation between higher level units (words) and lower level units (letters; McClelland & Rumelhart, 1981). In music, Bharucha's (1987a, 1987b) model of spreading activation (referred to as MUSACT) relies on a comparable architecture.

In this model, a pattern of connections constitutes a knowledge representation of Western harmony (Bharucha, 1987b, 1994). The units of the network are organized in three layers corresponding to tones, chords, and keys (Figure 2, top). Each of the 12 tone units is connected to three major and three minor chords, of which that tone is a component. Analogously, each chord unit is connected to three major key units representing keys of which it is a member. Western musical rules are not stored explicitly but emerge from the activation that reverberates by connected links between tone, chord, and key units.

When a chord consisting of three tones (e.g., C-E-G) is played, the units representing these tones are activated, and phasic activation is sent toward the chord units (see Figure 2, top). The chord unit connected to all three tones receives the strongest activation (the C-major chord in this example). During a second cycle (Figure 2, middle), phasic activation from the active chord units spreads toward the key units (bottom-up activation) and starts to reverberate toward tone units (top-down activation). During the next cycle (see Figure 2, bottom), activated key units send top-down activation to chord units that simultaneously receive activation from the tone units. Phasic activation continues to spread through the network between all layers until an equilibrium is reached. Early activation cycles reflect bottom-up influences: Activated chord units contain at least one of the component tones of the stimulus chord. For example, after a C-major chord, the chord unit E will be more activated than the chord unit D because it shares one tone with the stimulus chord (i.e., the tone E). It is noteworthy that the E-major chord unit is more activated than the D-major chord unit, even though the C-major chord and the E-major chord are harmonically less related in theory (they have no parent key) than are the C-major chord and the D-major chord (which both belong to the G-major key). Similarly, key units are activated if they have an activated chord as a member (Figure 3, left). During reverberatory cycles from initial activation to equilibrium, activation patterns change qualitatively. At equilibrium, the state of the network incorporates the influences of top-down processes and mirrors theoretical Western hierarchies. Activation tends to decrease with increasing harmonic distance between chords around the cycle of fifths. These top-down influences are clearly illustrated by the D- and E-major chord units. At equilibrium, the chord unit D receives stronger activation than the chord unit E. Similarly, the activation of key units decreases monotonically with increasing harmonic distance between major keys around the cycle of fifths (see Figure 3, right).

For chord sequences, activation resulting from each chord is accumulated. After the offset of an event, the activation begins to decay exponentially over time. If another event occurs before activation has decayed appreciably, the phasic activation resulting

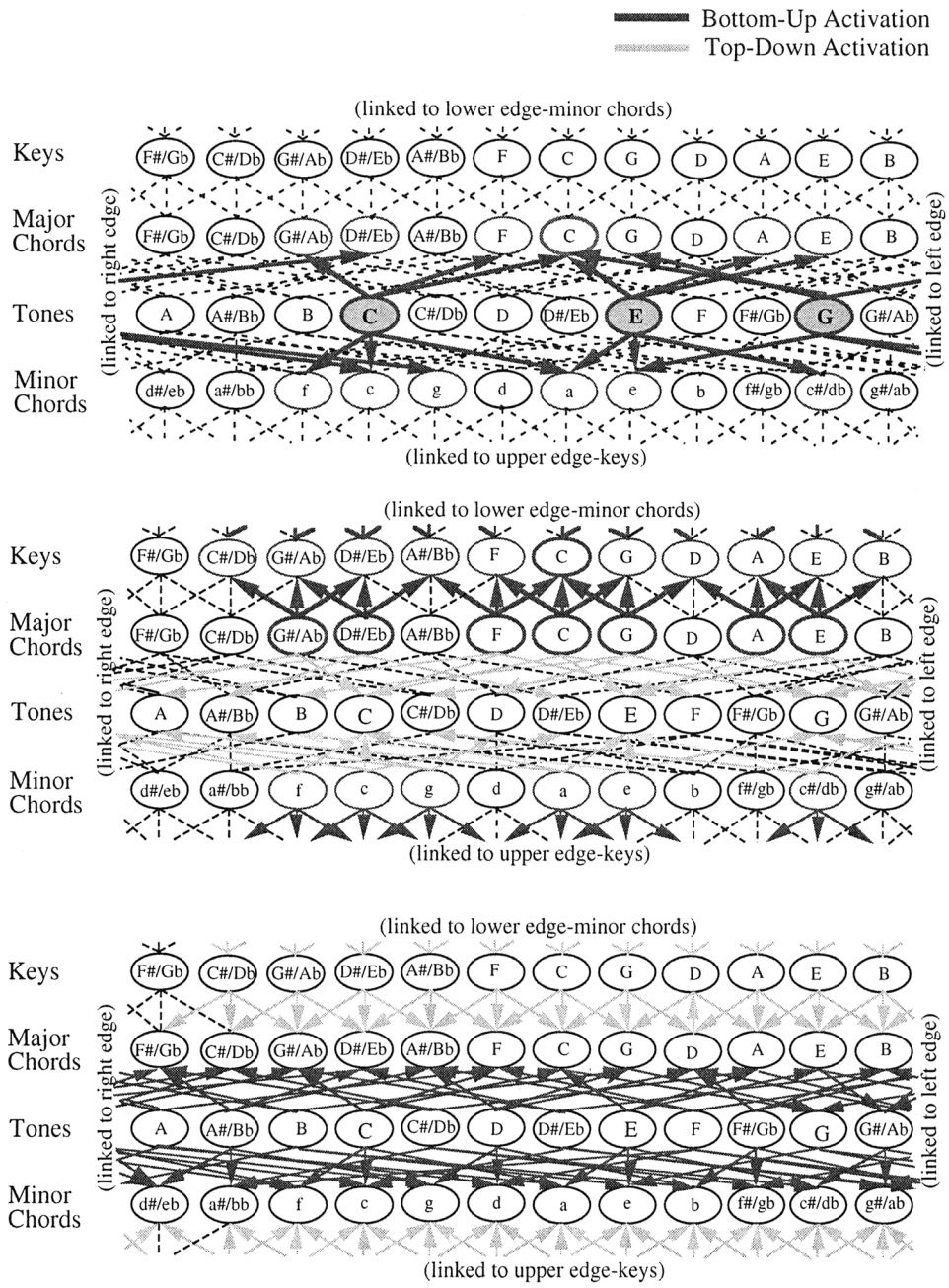

Figure 2. Bottom-up and top-down activation spreading in the MUSACT model after the presentation of a C-major chord. Top: Activated tone units send bottom-up activation to connected chord units (first cycle). Middle: Phasic activation spreads to key units and reverberates to tone units (second cycle). Bottom: Chord units receive activation from both key units and tone units (third cycle). From "MUSACT: A Connectionist Model of Musical Harmony," by J. J. Bharucha, 1987a, in *Program of the Ninth Annual Conference of the Cognitive Science Society,* pp. 508–517, Figure 7, Hillsdale, NJ: Erlbaum. Copyright 1987 by the Cognitive Science Society, Incorporated. Used by permission.

from that next event is added to the residual activation from the previous event, thereby creating a pattern of activation that can be influenced by an entire sequence of events, weighted according to recency. In other words, the activation of a unit i in the network is a function of not just the most recent event e but also of the previous event, $e - 1$, the activation of $e - 1$ being itself a function of event $e - 2$ and so on. The total activation, $a_{i,e}$, of a unit i (a tone, a chord, or a key) after an event e is an additive function of three quantities: (a) the bottom-up activation caused directly by the stimulus itself (i.e., the tones), (b) the indirect

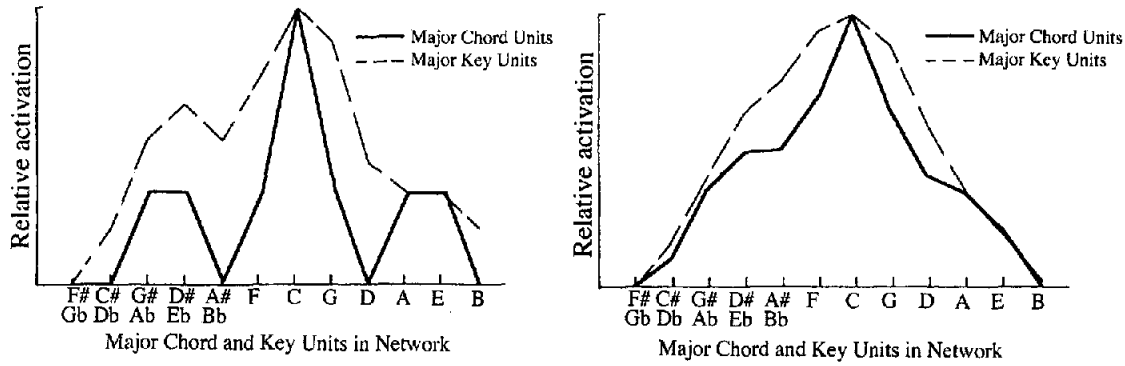

Figure 3. Left: The state of the network just after hearing the tones C, E, and G (a C-major chord) but before activation has had a chance to reverberate back from key units to chord units. Right: The state of the network after activation has reverberated to a state of equilibrium. From "Music Cognition and Perceptual Facilitation: A Connectionist Framework," by J. J. Bharucha, 1987b, *Music Perception, 5,* p. 1, Figures 8 and 9. Copyright 1987 by The Regents of the University of California. Adapted with permission.

activation received from other units in response to event e (i.e., the phasic activation spreading in the system), and (c) the decayed activation caused by previous events $e - 1$ (being itself a function of event $e - 2$ and so on). The total activation, $a_{i,e}$, of a unit i is given by the following equation:

$$a_{i,e} = A + \sum_{c=1}^{q} \Delta a_{i,e,c} + a_{i,e-1}(1 - d)^t \qquad (1)$$

where A represents the stimulus activation; $\Sigma\, \Delta a_{i,e,c}$ is the sum of phasic activation of unit i in response to event e, accumulated over the q reverbatory cycles that are necessary to reach equilibrium; d represents the rate (varying between 0 and 1) at which activation decays after the offset of the last event; and t is the time elapsed since the last offset.

MUSACT is an idealized simplification of aspects of the Western musical grammar because it does not incorporate units for minor keys and several other types of chords that may occur in Western music (i.e., diminished chords). A further simplification is the absolute pitch class coding at the input layer and the representation of chord units independent of the relative position of the component tones.[4] In spite of these simplifications, the model provides a relevant framework for understanding how musical knowledge may be mentally represented and how this knowledge, once activated by a given musical context, may influence the processing of tonal structures (Bharucha, 1987b).

Support for the MUSACT model has come from empirical studies using a harmonic priming paradigm. The rationale of these studies is that a previous chord primes harmonically related chords so that their processing is speeded up. The extent to which a chord is primed by a context is a function of the activation of the unit representing this chord in the model. The more a chord unit is activated, the more the chord is primed. To test this hypothesis, Bharucha and Stoeckig (1986, 1987) asked participants to decide as quickly as possible whether or not a target chord following a prime chord was in tune (see also Tekman & Bharucha, 1992). Participants heard a prime chord followed by a harmonically closely or distantly related target chord. For example, if the prime chord is C major, B♭ major would be a related target and F♯ major an unrelated target. In MUSACT, the B♭-major chord would be more strongly primed by a C-major chord than would be the F♯-major chord (see Figure 3). Empirical data confirmed this prediction: The priming effect was shown by (a) a bias to judge targets to be in tune when related to the prime and out of tune when unrelated, and (b) shorter response times for in-tune targets when related and for out-of-tune targets when unrelated. A previous musical context (a single chord in these experiments) thus generates expectancies for related chords to follow, resulting in greater consonance and faster processing for expected chords.

The priming of a chord depends not only on the previous chord but also on its harmonic function within an extended temporal context (Bigand & Pineau, 1997; Bigand, Madurell, Tillmann, & Pineau, 1999; Tillmann, Bigand, & Pineau, 1998). Bigand and Pineau (1997) presented participants with eight-chord sequences. The expectations for the last chord (the target) were varied by changing the harmonic context created by the first six chords. The target and its flanking chord were always held constant. In one context, the last chord acted as an harmonically stable tonic chord, part of an authentic cadence (V-I). In the other context, the last chord took the form of a less stable fourth harmonic degree following an authentic cadence (I-IV). Participants were faster in their intonation judgment of the target chord when it acted as a tonic chord. Simulations were performed with MUSACT using the first seven chords of the sequences, and the activation of the target chord unit was read off (Bigand et al., 1999). The target chord unit received stronger activation in the former context (i.e., when the

[4] The model does not deny the structural importance of harmonic voice leading (i.e., the melodic motion between the tones of successive chords) and chordal position (open vs. close) in Western music theory. It postulates that the introduction of these factors is not critical to account for how musical knowledge may be mentally represented and how this knowledge, once activated by a given musical context, may influence the processing of tonal structures. Empirical studies that have manipulated voice leading, chordal position, or the harmonic spectrum of the sound (e.g., piano-like sound vs. pure tones) suggest a weak influence of these factors on chord processing that is overridden by harmonic relationships (Bigand, Tillmann, Manderlier, & Poulain, 2000; Rosner & Narmour, 1992; Stoeckig, 1990).

target acted as a stable tonic chord) than in the latter context. The activation pattern thus mimics human performance and manages to account for global context priming effects. Other research indicates that MUSACT also accounts for priming effects at several levels of global musical context (Bigand et al., 1999) or that occur when both global and local context are factorially manipulated (Tillmann, Bigand, & Pineau, 1998).

MUSACT highlights a crucial issue of Western music: whether the relations between chords are driven by similarities based on acoustic properties of tones or by implicit knowledge of cultural conventions and usage (see Bigand et al., 1996, and Parncutt, 1989; for a discussion). MUSACT disentangles these two factors by charting the time course of bottom-up and top-down influences. It predicts that the activation pattern reflects bottom-up influences at early activation cycles, whereas top-down influences are predominant when the model has enough time to reach equilibrium. Tekman and Bharucha (1998) tested this prediction by pitting shared tones against conventional relatedness. Two types of target were selected: One was psychoacoustically more similar to the prime, the other more closely related on the basis of harmonic convention. For example, a C-major prime shares a tone with an E-major target but does not share a tone with a D-major target; yet D major is more closely related to the prime in conventional usage. Priming results revealed facilitation for psychoacoustically similar targets when they followed after a short (50 ms) stimulus onset asynchrony (SOA) and facilitation for conventionally related targets after a longer SOA (500 ms or longer). Although both psychoacoustic similarity and conventional relatedness drive priming, the influence of the former is short lived, precisely as predicted by the temporal course of activation in the MUSACT model.

As originally conceived, the model was based on music theoretic constraints; neither the connections nor their weights resulted from a learning process. In this respect, the model represented the idealized end state of an implicit learning process. To be compelling, a cognitive model of Western harmony should simulate the internalization of Western pitch regularities by mere exposure, allowing the connection weights to adapt to the musical environment. It has been suggested that MUSACT ideally can internalize these regularities by passive self-organization (Bharucha, 1991, 1992). In the next section, we describe how this can happen. Specifically, we show how a hierarchical SOM manages to learn the Western pitch regularities, comparable to those in MUSACT. The learned model will then be tested for its capacity to simulate a variety of empirical data concerning the perceived relationships between and among tones, chords, and keys (cf. SIMULATIONS OF EMPIRICAL DATA section).

A MODEL THAT LEARNS WESTERN HARMONY BY SELF-ORGANIZATION

General Principles of SOMs

In connectionist models, unsupervised learning algorithms extract statistical regularities and encode events that occur often together (Grossberg, 1970, 1976; Kohonen, 1995; Rumelhart & Zipser, 1985; Von der Malsberg, 1973). These algorithms are well suited to music perception because the organization of chords or tonalities presumably occurs without supervision. One unsupervised learning algorithm is the SOM proposed by Kohonen (1995). It creates topographic mappings between the input data and neural net units of a map. For two similar input patterns, the map units that respond maximally are located near each other. This conforms to principles of cortical information processing, such as the formation of spatial ordering in sensory processing areas (i.e., somatosensory, vision, and audition). In the primary visual cortex, the orientation of stimuli to which cells respond best changes in an orderly fashion across the context: Nearby cells respond best to similar orientations (Hubel & Wiesel, 1962). The auditory cortex displays a tonotopic organization in which cells responding best to different frequencies are arranged in an orderly fashion (Brugge & Reale, 1985; Wessinger, Buonocore, Kussmaul, & Mangun, 1997). In the auditory system, tonotopic organization can be found at almost all major stages of processing (i.e., inner ear, auditory nerve, cochlear nucleus, auditory cortex).

SOM is based on competitive learning, an algorithm for data-driven self-organized learning. With this algorithm, the neural net units gradually become sensitive to different input stimuli or categories (Rumelhart & Zipser, 1985). The specialization occurs by competition among the units. When an input arrives, the unit that is best able to represent it wins the competition. The winning unit is then allowed to learn the representation of this input even better, as is described later. The unit's response will be subsequently stronger for this same input pattern and weaker for other patterns. In a similar way, other units learn to specialize to respond to other input patterns.

The competitive learning algorithm can be generalized, if there exists a spatial layout of the units. On an SOM, for example, the units are located on a discrete lattice. The generalization implies that not only does the winning unit learn, but its neighbor units are also allowed to learn. Neighbor units will gradually specialize to represent similar inputs and the representation becomes ordered on the map. After learning, each unit is specialized to detect a particular input pattern, and a topographic organization of the input data can be discovered on the map, such that similar input patterns activate nearby map units.

In such a neural network, the input layer and the two-dimensional map layer are fully interconnected by synapses. Before learning, the connection strengths (the weights, w) are initialized to random values. When a stimulus is presented, the input units, i, that are tuned to the features of the stimulus, are activated. These activations of the input units $a(i)$ spread to the map units, j, by the connected links, $w(i, j)$. Each unit of the map layer, j, accumulates the activation it receives from the input units. The activation of each map unit $a(j)$ is given by:

$$a(j) = \sum_i a(i)*w(i, j). \tag{2}$$

The unit j with the highest activation defines the winning unit. During the learning phase, the associated weight vectors of the winning unit w and those within a neighborhood set N are updated. The weights of units outside the neighborhood set are kept constant. Learning consists of updating the weights feeding into the winning unit and its neighbors with the following algorithm (Kohonen, 1995):

$$w_{(t+1)} = \frac{w_{(t)} + \eta_{(t)}*a_{(t)}(i)}{w_{(t)} + \eta_{(t)}*a_{(t)}(i)\|}, \tag{3}$$

where $w_{(t+1)}$ represents the weight vector at time $t + 1$, $w_{(t)}$ at time t, and $\eta_{(t)}$ represents the learning rate. This learning rule reinforces links coming from active input units and weakens links coming from inactive input units. In other words, it moves the weight vectors closer to the input vector, making the winning unit and its neighbors more likely to win the competition when this input or one similar to it is presented again.

The neighborhood N is set to be wide at the beginning of learning. During learning, it decreases monotonically until it consists of the winning unit alone. As learning begins, a large neighborhood allows a global organization to emerge. With a smaller neighborhood radius, the units become adapted to the individual patterns and its close relatives, and a local organization emerges.

Before learning, there is no particular organization among the map units. When the net is trained by repeated presentation of the input data, it begins to self-organize. A topographic pattern begins to appear, such that units that are topographically close in the array will be activated by similar input stimuli. SOM can be conceived of with one map layer or be adapted to multilayer hierarchical self-organizing maps (HSOMs; Lampinen & Oja, 1992).

Simulating Implicit Learning of Western Music With SOMs

For the simulations of implicit learning of Western harmony, we defined a hierarchical SOM. Its structure is similar to models of language perception (McClelland & Rumelhart, 1981; Elman & McClelland, 1984) and of music perception (Bharucha, 1987b). The input layer is tuned to octave equivalent pitch classes. The second and third layers will learn to specialize in the detection of chords and keys, respectively. The hierarchical map is inspired by feature detectors found in the brain, with elementary feature detectors at the sensory periphery (e.g., frequency), and more abstract feature detectors in the primary auditory cortex, for example, pitch (Pantev, Hoke, Lütkenhöner, & Lehnertz, 1989) or contour (Weinberger & McKenna, 1988). The abstract pitch-class coding was based on the demonstrated ability of neural net models to extract pitch from frequency (M. A. Cohen, Grossberg & Wyse, 1995; Sano & Jenkins, 1991) and to learn octave equivalent pitch classes (Bharucha & Mencl, 1996). M. A. Cohen et al.'s (1995) model transforms a spectral representation of an acoustic source into a spatial distribution of pitch strengths. In Bharucha and Mencl's (1996) model, octave categories with abstract pitch classes are learned by self-organization operating on a spectral representation. Our present model can be conceived of as subsequent to these phases of auditory preprocessing (cf. Bharucha, 1991, for an expansion of Bharucha, 1987b).

The self-organizing algorithm permits the formation of units representing events that are frequently associated. In music, this association consists of either the simultaneous occurrence of tones (which define a chord) or the temporal proximity of events (such as arpeggiated tones of a chord or the chords forming a key). In the proposed model, self-organization leads to a hierarchical encoding in which tones occurring together are represented by chord units and, similarly, chords occurring together are represented by key units. To model the influence of knowledge on perception, the neural net structure resulting from learning simulations is then used with a spreading activation mechanism. After the presentation of a stimulus, activation reverberates between the three layers until an equilibrium is reached. The use of a neural net structure as a reverberation system implies two constraints.

The first constraint deals with the simulation of top-down influences and has its origin in previously proposed spreading activation models (i.e., Bharucha, 1987b; McClelland & Rumelhart, 1981). Three-layer models of word recognition (i.e., features, letters, words) and of musical knowledge representation (i.e., tones, chords, keys) manage to simulate top-down influences by favoring the spread of activation between the higher levels of representation, namely the second and the third layers. In the word-recognition model, the interactive processing concerns only the letter and word levels, without feedback to the feature level. In MUSACT, the weights of the connections between chord and key layers are stronger than between tone and chord layers, yielding an initial influence of the tone input, followed by a strong influence of the two abstract layers and only a weak additional influence of the tone layer. A similar constraint is implemented in the simulations presented later.

The second constraint results from units on the SOM that are not specialized in the detection of a stimulus after learning. Because reverberation should occur only between specialized units, the connections feeding into unspecialized units (e.g., units that do not win for any of the training patterns) are treated by a pruning mechanism. Pruning is a procedure supported by the developmental neural principle that connections weaken from disuse, and it is generally used in supervised learning by weight decay or weight elimination (Haykin, 1994; LeCun, Denker, & Solla, 1990; Setiono, 1997).

Four learning simulations respecting both constraints are presented. All simulations were based on the same network architecture that learned with the self-organizing algorithm. The network was trained with either simple harmonic material (cf. Learning Simulations With Simple Harmonic Material section) or more realistic chord sequences (cf. Learning Simulations With Short Chord Sequences section). The input was defined by either a sparse coding or a psychoacoustically richer coding scheme (sparse input coding in simulations SIC-1 and SIC-2 vs. rich input coding in simulations RIC-1 and RIC-2). We start by considering learning based on simple harmonic material and a sparse input coding.

Learning Simulations With Simple Harmonic Material

SIC-1

Network architecture. A three-layer hierarchical system was defined as follows: The input layer consisted of 12 units, the second layer was a map of 36 units, and the third layer was a map of 16 units. The input units were tuned to the 12 chromatic scale tones, representing octave equivalent pitch-class detectors. The units of the first and second layers were fully interconnected with a connection matrix. The units of the second and third layers were fully interconnected with a second connection matrix. All connec-

tions were bidirectional, and their strengths were initialized to random values between 0 and 1 before learning.[5]

Input coding. During training, chords (consisting of three tones) were presented to the input layer (see later discussion). In the sparse input coding, the presence of tones was coded directly: A tone unit was activated if the tone to which it was tuned occurred in the stimulus, and was set to 0 otherwise. By convention, the 12 tone units were ordered as follows: A-A♯-B-C-C♯-D-D♯-E-F-F♯-G-G♯. The three component tones of a C-major chord were represented by the following vector: {0-0-0-1-0-0-0-1-0-0-1-0}.

Training. The network was allowed to self-organize as indicated previously (see Equations 2 and 3). At the outset of training, the neighborhood radius was set to 5 and 3.6 for chord and key layers, respectively (for both maps the neighborhood was defined by euclidean distances between units). During training, the neighborhood radius decreased until reaching 0, at which point only the winning unit learned. The learning rate η was decreased over the course of learning. Whenever the neighborhood radius decreased, the learning rate was divided by two. In the convergence phase (i.e., when only the winning unit learned), the learning rate decreased over the number of training cycles (one training cycle consisted of the presentation of the whole set of stimuli), c, as follows:

$$\eta' = 1/(c + 1/\eta). \tag{4}$$

Training consisted of two phases. In the first training phase, the second layer was trained with sets of three tones (e.g., C-E-G) corresponding to the 12 major and 12 minor chords of Western music. Each triplet of tones was presented separately to the input layer. In this phase, units of the second layer learned to detect chords. In the second training phase, the third layer was trained with 12 sets of 6 chords presented to the tone layer. One set of chords consisted of 3 minor and 3 major chords of a given key (e.g., the 3 major chords C, F, and G and the 3 minor chords d, e, and a, all of which belong to the C-major key). The six chords were presented individually to the tone layer. For each input chord, the activation of the winning chord unit in the second layer (referred to by the index b) was stored in memory without decay until the end of the presentation of the chord set. The pattern of indices, b, defined the input for the training of the third layer. In this second phase, the units of the third layer learned to detect keys. In both phases the training patterns were presented in random order during each training cycle.[6]

Calibration of the maps. During training, units became specialized in the detection of chords in the second layer and in the detection of sets of chords (keys) in the third layer. For both training phases, the weight changes decreased over the training cycles and with decreasing neighborhood. When weights converged to stationary values, the maps were calibrated by naming each winning unit after the stimulus for which it won. For example, the unit that won for the three tones C-E-G was called the C-major chord unit. After training, the average quantization error (i.e., the mean of the euclidean distances between each input vector and the weight vector of its corresponding winning unit) was less than .01 for each map.

Topographic organization and connection matrices. The calibration phase revealed a topographic organization of both maps. After training, chords that share component tones were represented by neighboring units in the second layer (Figure 4, left). Chords

c#	E	g#		C	e
C#		G#	c		G
f				g	
F	d	A#	a#		D#
a			F#		d#
	A	f#	D	b	B

E	B	F#	C#
A			G#
D			D#
G	C	F	A#

Figure 4. The calibration maps of the second layer (left) and the third layer (right) of the network SIC-1. For the second layer, winning units are labeled by chord names (minor chords in lowercase letters, major chords in uppercase letters). For the third layer, names of winning units indicate major keys.

that do not share tones were not represented by neighboring units. In the third layer (see Figure 4, right), key units were organized in a way that reproduces the topology of the cycle of fifths (also see Figure 1). Keys sharing chords and tones were represented close to each other on the map. The quality of the learned representation was tested for both maps. In a coherent learning solution, the distance between specialized units on the map should reflect the relationships between the input stimuli. To address this issue, the euclidean distances between all specialized units E and the correlations between all input vectors C were calculated separately for each map. The correlation between the two indices C and E was $r(298) = .69$ for the chord layer and $r(76) = .95$ for the key layer. The learned representations on the two-dimensional maps reflected the regularities within the input spaces.

After training, each tone unit was strongly linked to six winning units in the chord layer, and each chord unit was strongly linked to three winning units in the key layer. The connections feeding into unspecialized units were eliminated during the pruning phase. In the pruning procedure, all stimuli (chords and sets of chords) were presented in random order to the input layer. After the presentation of each stimulus, the weights feeding into losing units were decreased, and the weights feeding into the winning unit were reinforced by a parameter varying as a function of the number of winning units in the layer. After pruning, only connections feeding into winning units remained.

Feed-forward and reverberation system. After learning, the model was used both as a feed-forward system and as a reverber-

[5] For chord and key layers, the number of units on the maps were set somewhat relative to the number of stimuli in the training set. In supplementary simulations with SIC-1, the second and third layers were each defined by a map of 36 units. The outcome suggested that the basic spatial structure learned by the key map holds for maps of both sizes (i.e., 16 and 36 units). In the input layer, the invariant pitch-class coding leads us to conceive melodies as sets of scale degrees in reference to the tonic and to consider chords as being invariant under inversions.

[6] The number of necessary training cycles varied from 200 to 1,000 over the four learning simulations. At the outset of training, the learning rate η was set to .5 and .1 for chord and key layers, respectively, in the reported simulations. Other simulations with SIC-1 and SIC-2 showed that a learning rate of .2 for both layers led to similar results.

ation system. Reverberation was defined by phasic activation spreading between the units by the weights until equilibrium was reached. Equilibrium was defined to be reached when phasic activation was less than a threshold of .005 for each unit (cf. Bharucha, 1987b). Because of the normalization process (see Equation 3), the learned weights of the two matrices generated reverberating phasic activation that exceeded the bounds within which equilibration occurs.[7] The phasic activation reverberating within each of the matrices was thus scaled by a parameter designed to favor reverberation in the chord-key matrix (see the first constraint discussed earlier). This parameter was set first for the chord-key matrix by placing it within the boundary conditions for equilibration for this matrix alone. The corresponding parameter for the tone-chord matrix was then titrated so that the system as a whole reached equilibrium within approximately 100 reverberation cycles.

When used as a feed-forward system, the stimulated tone units send activation toward chord units, which in turn send activation toward key units. This feed-forward activation reflects bottom-up information (tones present in the stimuli) and does not incorporate top-down influences. Figure 5 (left) represents activation levels for the 12 major chord units and the key units. When the three tones of a C-major chord are presented, the unit in the second layer that becomes most active is the one specialized in the detection of the C-major chord, followed by the units specialized for major chords sharing one of the three tones. Units of chords that do not share component tones were not activated. A similar pattern was found for the key layer: The most activated unit was the C-major key, followed by the F-major key, the G-major key, and the other keys linked to one of the activated chord units. The key of F♯ was not activated, because it does not share any tones with the C-major chord.

When used as a reverberation system, the pattern of activation changes qualitatively because of the top-down influence of learned, schematic structures (see Figure 5, right). In the chord layer, the difference between bottom-up and top-down activation was manifest in the comparison of activations for the E- and D-major chords. When the three tones of a C-major chord were presented, feed-forward activation caused the E-major chord unit to be more strongly activated than the D-major chord unit. During reverberation, however, the D-major chord unit overtook the E-major chord unit by virtue of its closer harmonic relationship to C major as mediated by the key units. Across all major chord units, activation decreased monotonically with distance around the cycle of fifths from C, with the lowest for F♯ and B. In the third layer, the most activated key unit was C major (for which the presented chord is the tonic), and activation decreased monotonically with distance along the cycle of fifths and, equivalently, with distance along the topographic key map.

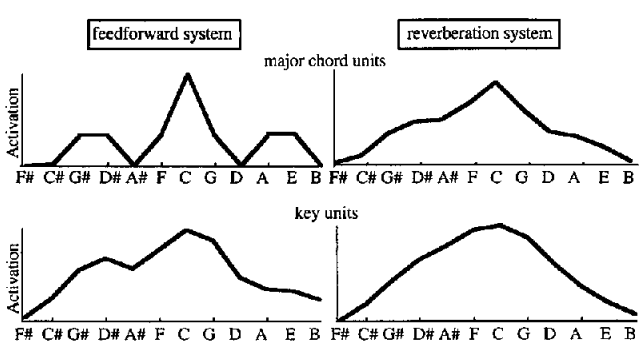

Figure 5. Activations of major chord units and key units in the network SIC-1 after the presentation of a C-major chord for bottom-up activation only (left) and for when the network has reached equilibrium (right).

RIC-1

The sparse input coding considered only the presence of the component tones of the chords. However, each tone contains a complex harmonic spectrum, which may influence the perception of harmony (Helmholtz, 1885/1954; Parncutt, 1989; Terhardt, 1974). Consequently, a richer input coding scheme was used in the following learning simulation (RIC-1). This coding, based on the psychoacoustic theory of Parncutt (1988), was used by Leman (1995) in his simple auditory model. In Parncutt's theory, subharmonic virtual pitches are assigned to each component of a complex tone, and the frequency of the most commonly generated subharmonic determines the perceived pitch. Harmonic relationships between two chords are estimated by two indices based on virtual pitch cues: pitch salience (intensity of a virtual pitch cue) and pitch commonality (number of cues shared by two pitch vectors). Following Parncutt (1988) and Leman (1995), a chord was coded as a pitch-class vector whose values are weighted sums of the subharmonics corresponding to those pitch classes. For example, a C-major chord was represented by the following pattern: {.85-.2-0-1.83-.1-.45-.33-1.1-.7-.25-1-.33} (representing the tone units A-A♯-B-C-C♯-D-D♯-E-F-F♯-G-G♯). A learning simulation, as described for SIC-1, was run again with this richer input coding.

Topographic organization and connection matrices. The rich input coding influenced only the topographic organization of the second layer and the matrix connecting the tone layer to the chord layer. Twenty-four units in the second layer were specialized in the detection of the chords. As in SIC-1, neighboring units had tones in common. However, the richer coding yielded a more global organization: The map was globally divided in two (upper and lower halves), each half corresponding to chords from one side of the cycle of fifths (Figure 6, left). This outcome was in accord with Petroni and Tricarico (1997), who analyzed the influence of input coding on the location of winning units. They found that, in contrast to a simple local representation (that indicates which pitch is activated in the triad), adding the subharmonic sum of the tones produced an organization close to the cycle of fifths. In our simulations, 11 of the 12 tone units were linked to a winning chord unit, even if some of the links remained small. This pattern of connections reflected the new input coding in which 1 of the 12 subharmonics was not present in the coding of one chord: For example, in the C-major chord the tone unit B was not activated

[7] A reverberation system with two matrices reaches equilibrium if two conditions are satisfied. First, each matrix used separately as a reverberation system reaches equilibrium on its own. This constraint is satisfied if the maximum eigenvalue of the matrix is less than 1. Second, the mean of weights over both matrices is small enough to permit an equilibrium for reverberation in the total system. This constraint is satisfied if the phasic activation sent back from a layer in a second cycle is smaller than the phasic activation sent in the first cycle.

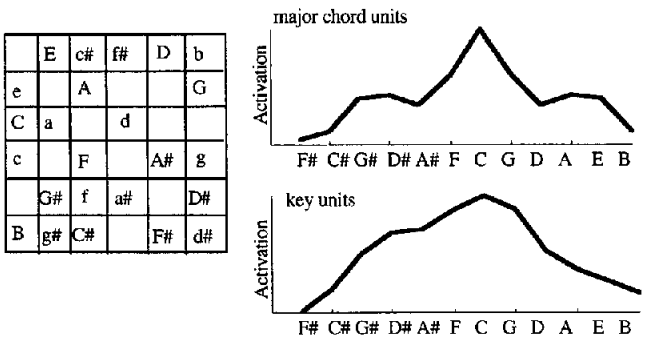

Figure 6. The calibration map of the second layer of the network RIC-1 (left). Winning units are labeled by chord names (minor chords in lowercase letters, major chords in uppercase letters). Activations of major chord units and key units after the presentation of a C-major chord for bottom-up activation only (right).

(see prior discussion). The first connection matrix thus developed quite differently in this simulation from the way it did in SIC-1. Interestingly, these differences had no influence on the formation of the second matrix. The second matrix and the topographic organization of the third layer had the same characteristics as in SIC-1.

Feed-forward and reverberation system. The feed-forward activation for major chord and key units revealed a slightly different pattern from SIC-1 (see Figure 6, right). In the chord layer, these differences were observed especially at the A♯- and D-major units, whose activations were higher than in SIC-1 although still lower than the G♯- or A-major units. The global shape of the curve was thus preserved with a stronger feed-forward activation for the E-major chord than for the D-major chord. The key layer differed in two aspects from the feed-forward simulation in SIC-1. First, activation decreased with distance around the cycle of fifths, because the A♯ key was more strongly activated than the D♯ key. Second, the difference between the dominant key and the subdominant key disappeared.

For both layers, subharmonic information changed bottom-up profiles so that they were closer to the shape of top-down profiles than with sparse coded inputs (see Figure 5, left). With reverberation, however, the activation profiles of chord and key layers were analogous for rich and sparse coding (see Figure 5, right). The comparison between SIC-1 and RIC-1 suggests that the asymptotic behavior of the reverberating model is not particularly sensitive to the richness of the input, even though the bottom-up effects are. The top-down processes driven by the more abstract learned knowledge of Western harmony impose a pattern of activation that is similar regardless of whether the structure within the harmonic and subharmonic series is used by the system.

Learning Simulations With Short Chord Sequences

In simulations SIC-1 and RIC-1, the model was exposed to sets of six chords presented randomly. This material may be viewed as somewhat artificial for two reasons. First, it did not mirror the transition probabilities of chords in Western music. Second, each of the chords was presented with equal salience. The material used in the following simulations, SIC-2 and RIC-2, was more ecologically valid for learning the connection matrix between chord and key units. Short chord sequences were presented to the network, and the activation of chord units decayed as the chords receded into the past, simulating a decaying memory.

Sequences of seven chords were constructed by conforming to standard harmonic root progressions (Piston, 1978) and statistical chord distributions (Budge, 1943). The selection of chords was restricted to the following set of chords in a major key: I, ii, iii, IV, V, and vi. Tonic chords (I) occur more often than dominant chords (V), followed by subdominant chords (IV), then vi, and ii, and least often iii. Ten sequences were constructed by selecting chords at random using a probability distribution based on the prior constraints. The last chord in each sequence was the tonic chord of the key. Consistent with conventions of harmonic progression, each tonal sequence ended with a final V-I or IV-I cadence. These sequences were transposed to all 12 major keys, resulting in 120 training sequences.

The chords of a sequence were presented to the input layer one by one. For each input chord, the activation of the winning chord unit (referred to by the index b) was stored in memory until the end of the sequence and was decreased by a decay parameter. The decreasing activation pattern of these indexes b defined the input for the third layer. Simulations were run with the sparse input coding (SIC-2) and the rich input coding (RIC-2).

SIC-2

Topographic organization and connection matrices. The new training material influenced only the matrix linking chord to key units. In the calibration phase, the winning key units were determined by presenting sets of six chords (together forming a key) to the network. Presenting the training sequences after learning created the highest activation in the key unit representing the tonic key of the sequence. The representation of the key units in the third layer mirrored the cycle of fifths. As in the simulations RIC-1 and SIC-1 with artificial tonal material, keys sharing chords and tones were still represented close to each other on the map. Each winning key unit was linked to six chord units representing chords belonging to that key. The strengths of these connections mirrored the statistical chord distributions of the stimuli: Links from the tonic chord unit were the strongest (.663), followed by those from the dominant chord (.542), the subdominant (.34), then vi (.296), ii (.24), and iii (.08). The strengths of the connections depended on the functions of the chords in the corresponding key: The same chord (e.g., C major) had a stronger link to the key for which it is the tonic chord (the C-major key) than to a key for which it is the subdominant chord (the F-major key).

Feed-forward and reverberation system. When used as a feed-forward system, the presentation of a C-major chord activated the C-major key most strongly, followed by F, G, and others sharing one of the activated chords. The pattern of activation across the key layer generally mirrored that of SIC-1, except that (a) the dominant and subdominant keys were equally activated and (b) the activation of C, F, and G major keys were more pronounced. This difference was due to higher rates of occurrence of the tonic, subdominant, and dominant chords in the sequences, resulting in stronger weights to these chords and correspondingly higher activation (Figure 7).

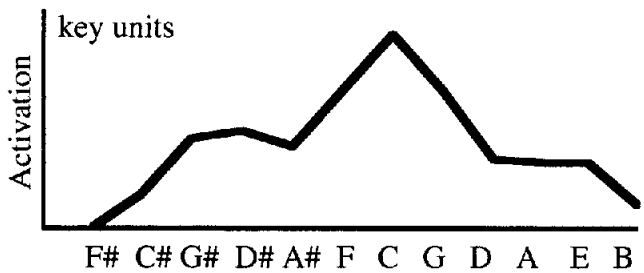

Figure 7. The activations of key units after the presentation of a C-major chord for bottom-up activation only in the network SIC-2.

When used as a reverberation system, the activation patterns of chord and key layers at equilibrium were consistent with the results of SIC-1. The unit of the presented chord and the corresponding tonic key unit were the most activated, and activation decreased with increasing distance around the cycle of fifths. In SIC-2 the basic structure and connections were learned successfully (albeit with varied connection strengths), and the harmonic relationships that emerged after reverberation were consistent with activation patterns obtained previously.

RIC-2

Finally, simulation RIC-2 was run with richer input coding (as described previously for RIC-1), using the chord sequences from SIC-2. The purpose of this simulation was to determine how the complexity of both input coding and statistical regularities of musical chord sequences could be combined during the learning process.

Topographic organization and connection matrices. The second matrix reflected the characteristics of the new material but was not influenced by the rich input coding. The second matrix and the organization of the third map paralleled the results obtained with sparse coding in SIC-2.

Feed-forward and reverberation system. The feed-forward activation pattern of chord and key layers was similar to the results of RIC-1 (see Figure 6, right). After reverberation, the activation levels of chord and key units decreased with increasing distance around the cycle of fifths. In the chord layer, the decrease of activation was smoother than in the previous simulations (Figure 8). Unlike SIC-1, the differences in activations between D♯ and A♯ chord units (and D and A chord units) were no smaller than those between the other chord units' neighbors on the cycle of fifths (see Figure 5, right). The chord activation pattern seems to reflect the combined influence of rich coding and complex learning material. However, the curve reflects the same rank order of activation between chord units as in the other networks. In the key layer, the results mirrored those of the previous simulations, including those based on simple material.

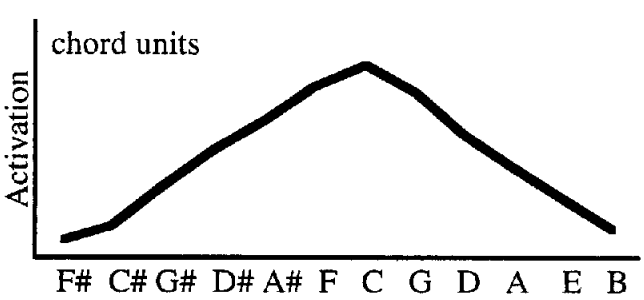

Figure 8. Activation of major chord units after the presentation of a C-major chord after reverberation in the network RIC-2.

Comparison of the Four Trained Networks With MUSACT

The four learning simulations demonstrate that a basic representation of Western harmony can be learned by mere exposure through self-organization. For both sparse and rich input coding, specialized representational units were formed for combinations of musical events (tones, chords) that occur with great regularity. Table 1 displays the correlation between the activation patterns of MUSACT and activation patterns of the four trained networks after the presentation of a single chord: SIC-1, embodying the most constraints (simple learning material, sparse input coding), matches MUSACT almost perfectly, and the other three networks correlate strongly. The activation patterns of the MUSACT model, originally constructed on the basis of theoretical musical constraints and originally proposed as an idealized end state of a learning process, can emerge automatically through self-organization.

A comparison of the four simulations suggests that the model learns the structural differences of the input coding (sparse vs. rich) and of the learning materials (six chords per key vs. chord sequences). The connection matrix between the tone and chord layers reflected differences as a result of the input coding. The matrix that issued from rich input coding (RIC-1 and RIC-2) was more complex than that issued from sparse input coding (SIC-1 and SIC-2). The connection matrix between the chord and key layers reflected the learning materials. After learning with sets of chords (SIC-1 and RIC-1), all links between chords and keys had the same strength, and no distinction was made between major and minor chords. After learning with chord sequences (SIC-2 and RIC-2), links between chords and keys varied in strength, reflecting the statistical distribution of chords within the corpus. However, this higher level learning was not influenced by the input coding.

The differences in the structure of the input coding were seen essentially in the feed-forward activation of chord units. After

Table 1
Correlation of Activation Patterns in the Four Learned Networks and in MUSACT for Major Chord and Key Units After a Single-Chord Presentation

	Feed-forward system		Reverberation system	
Networks	Chords	Keys	Chords	Keys
SIC-1	1	.999	.999	.999
RIC-1	.915	.977	.984	.998
SIC-2	1	.974	.989	.995
RIC-2	.915	.973	.971	.995

Note. $p < .01$ for all correlations. SIC = sparse input coding; RIC = rich input coding.

reverberation, the four trained networks yielded comparable activation patterns. The psychoacoustically rich input coding is thus especially important for bottom-up processes but makes little difference after the abstract knowledge by top-down processes has exerted its influence. We examined this quantitatively using Parncutt's (1989) psychoacoustic measure of harmonic relatedness. In Parncutt's model, the strength of a harmonic relationship is predicted from the degree to which two chords have virtual pitches in common. To analyze the importance of input coding, the chord activations of all networks as feed-forward and reverberation systems were correlated with Parncutt's pitch commonality values. Correlations between feed-forward activation of chord units and pitch commonality values were higher for networks based on rich input coding $r(10) = .90$ for RIC-1 and RIC-2, than for networks based on simple input coding, $r(10) = .73$ for SIC-1 and SIC-2. However, after reverberation, these correlations changed for all networks: They decreased for rich coding networks, $r(10) = .85$ and $r(10) = .88$ for RIC-1 and RIC-2, respectively, and increased for simple coding networks, $r(10) = .87$ and $r(10) = .84$ for SIC-1 and SIC-2, respectively. This confirms that the importance of psychoacoustic information in coding seems to be overwritten by top-down activation in the reverberation process.

SIMULATIONS OF EMPIRICAL DATA

In the previous section, we established that the underlying structure of Western harmony can be learned via mere exposure by a self-organizing neural net. In this section, we report simulations showing that the learned model behaves much as human participants do in a range of experiments on the perception of tonality.

Once the key of a musical context is established, tones and chords are perceived in a hierarchy of stability. In particular, the tones of the tonic triad and the chords built on the first, fifth, and fourth scale degree of the key are perceived as being more stable than the other events (Bharucha & Krumhansl, 1983; Krumhansl, 1979, 1990; Krumhansl et al., 1982; Krumhansl & Kessler, 1982). The psychological distance between two tones (or chords) decreases as the stability of these events in the key context increases. The psychological distance between two keys decreases as the correlation between the tone (or chord) stability profiles of these keys increases. The Western hierarchy of stability influences psychological distances between events according to three contextual principles articulated by Krumhansl (1990; Bharucha & Krumhansl, 1983; Krumhansl et al., 1982). The first principle, contextual identity, states that the psychological distance between two instances of the same event decreases as the stability of the event increases in the key context. For example, the psychological distance between two *C*-major chords is smaller in a *C* than in an F♯-major key context. The second principle, contextual distance, states that the psychological distance between two distinct events decreases as the stability of the events increases in the key context. The psychological distance between the C- and G-major chords is smaller in a C- than in an F♯-major key context. The third principle, contextual asymmetry, reflects the importance of temporal order on psychological distance. The distance between two events decreases when the more stable event occurs after the less stable one. For example, the psychological distance from an F♯- to a C-major chord is smaller in a C-major key context than in an F♯-major key context. These three principles govern the organization of musical events in perception and memory, such that highly stable events are perceived as being more closely related to, and more easily confused with, each other and previously presented less stable events than the reverse. They are also more expected and are preferred as good endings.

The simulations were run with the experimental material used in the chosen set of empirical studies, and the activation levels of the units in the network were interpreted as levels of stability. The more a unit (i.e., a chord unit, a tone unit) is activated, the more stable the musical event is in the corresponding context. If the trained networks are to serve as good models, they should mirror the human data.

Simulations of chord sequences are presented first because these are the types of stimuli with which the network was trained. They are followed by simulations on perceived relations between keys and between tones. Most of the simulations were run with the four trained networks described previously. Given the strong similarity between these networks, the outcomes of these simulations did not vary qualitatively as a function of the network. In the interest of brevity, only simulations run with the SIC-2 network are presented here.[8]

Perceived Relations Between Chords

The perceived relations between chords have been studied with different experimental paradigms. The following section presents simulations of experimental data that were based on similarity judgments of pairs of chords, memory performances, and harmonic expectations.

Similarity Ratings

Human Performance

In similarity rating experiments, listeners were presented with a pair of target chords after a given key context (Bharucha & Krumhansl, 1983; Krumhansl et al., 1982). Listeners rated the similarity of the two target chords on a 7-point subjective scale. The target chord pairs were all pairwise combinations of the chords from the C- and F♯-major keys. The key context was established by a set of three chords ending with a perfect cadence (IV-V-I). Five context keys (C, G, A, B, and F♯ major) were manipulated to vary in terms of their distances from C and F♯ major on the cycle of fifths (see Figure 1): G major is one step away from C major and five from F♯ major; the reverse is true for B major, and A major is as far from C major as from F♯ major (three steps).

As shown in Figure 9 (top left), the average ratings for pairs of chords in C major were the highest in the C-major key context. These ratings decreased linearly as the distance between the context key and the C-major key increased on the cycle of fifths. An inverted pattern was observed for pairs of chords in F♯ major. Results supported the contextual distance principle: The perceived similarity of two chords is proportional to their stability in the key context. Contextual asymmetry was supported by ratings for mixed

[8] The parameters (cf. Equation 1) were adapted from Bharucha (1987b) with $d = .4$ for decay rate and $t = 1$ for the time elapsed since the last offset, unless otherwise noted.

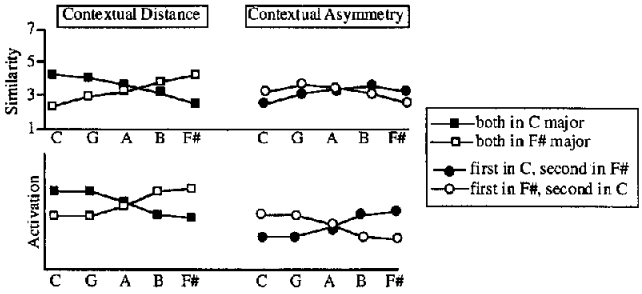

Figure 9. Top: Average ratings for pairs of diatonic triads in the same key (left) and in different keys (right) presented as a function of the key contexts of C major, G major, A major, B major, or F♯ major (data from Bharucha & Krumhansl, 1983, and Krumhansl, Bharucha, & Castellano, 1982). Bottom: Average activations of the second target chord unit after a first target in the same key (left) or in a different key (right) presented as a function of the five key contexts.

pairs of chords (e.g., with one chord in C major and one in F♯ major). In a C-major key context, the similarity ratings were higher when a chord from the C-major key occurred after a chord from an F♯-major key. This asymmetry vanished and reversed when the key context was progressively shifted from the C- to the F♯-major key (Figure 9 top right).

Simulation

In the simulation, as in the experiment, the five key contexts (C, G, A, B, and F♯ major) were established by a IV-V-I cadence. The target chord pairs were defined by all combinations of the major and minor chords from the C- and F♯-major keys. The three-chord cadences plus the first target chord of each pair were presented to the SIC-2 network. The activation level of the unit representing the second target chord, before its occurrence, was taken as output. Figure 9 (bottom) displays the average activation of units representing the second target chord. When both chords of the pair came from the C-major key, the activation of the second target chord was high in the C-major key context and decreased linearly as the distance between the context key and the C-major key increased around the cycle of fifths. An inverted pattern was observed for pairs of chords in F♯ major. The network thus replicated the contextual distance principle observed for human participants. Contextual asymmetry was also reproduced by the network. The activation of the second target chord was stronger when the second chord of the pair was more stable in the context than the first. When the second chord belonged to the C-major key, this asymmetry vanished and then reversed as the key context was progressively shifted from the C- to the F♯-major key. When the second chord belonged to the F♯-major key, this asymmetry increased when the key context was progressively shifted from the C- to the F♯-major key.

Recognition Memory

Human Performance

In recognition memory experiments, participants were required to judge whether or not standard and comparison sequences of seven chords were identical (Bharucha & Krumhansl, 1983). Eight experimental conditions were distinguished. In Conditions 1 to 3, the comparison sequence was identical to the standard ("same" condition). In Condition 1, all chords of the standard sequence belonged to a same key. In Condition 2, one chord was nondiatonic. In Condition 3, chords were randomly chosen from different keys. In Conditions 4 to 8, the comparison sequences differed by one chord from the standard sequence ("different" conditions). A diatonic chord of the standard sequence was replaced in the comparison sequence by either a diatonic chord (Condition 4) or a nondiatonic chord (Condition 5). In Conditions 6 and 7, the standard sequences contained one nondiatonic chord that was replaced by either a diatonic chord (Condition 6) or another nondiatonic chord (Condition 7). In Condition 8, the standard sequences contained a random set of nondiatonic chords, and one of them was replaced by another randomly chosen chord.

Figure 10 (top) displays the percentage of correct responses for each condition in the original experiment (Bharucha & Krumhansl, 1983). When the standard and comparison sequences were identical, the correct "same" judgments were more numerous for Diatonic Condition 1 than for Nondiatonic Condition 2 and Random Condition 3. This finding illustrates the principle of contextual identity: Recognition of two instances of the same event is easier when the event is stable in the key context. When the standard and comparison sequences differed, the correct "different" judgments were inversely related to stability: Fewer correct responses were observed when a chord of a standard diatonic sequence was replaced by a chord belonging to the key context (Condition 4). The number of correct "different" judgments increased when no key context was established in the standard sequence (Condition 8) and when a nondiatonic chord of the standard sequence was replaced by another nondiatonic chord in the comparison sequence (Condition 7). These findings illustrate the principle of Contextual Distance: Stable diatonic chords in a tonal context are more easily confused (low performance for Condition 4) than chords without a tonal context (Condition 8) or unstable nondiatonic chords (Condition 7). The data were also consistent with the contextual asymmetry principle. Fewer "different" judgments were observed when a nondiatonic chord was

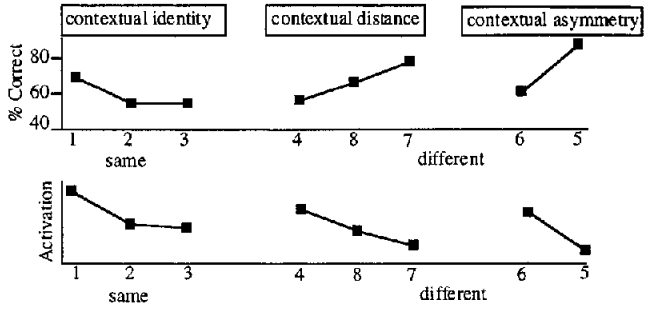

Figure 10. Top: Percentages of correct responses in the eight experimental conditions (see explanations in the text), separated for same and different targets. From "The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context," by J. J. Bharucha and C. L. Krumhansl, 1983, *Cognition, 13*, p. 86, Experiment 2, Table 1. Copyright 1983 by Elsevier Science. Reprinted with permission from Elsevier Science. Bottom: Average activation values of the target chord unit in the comparison sequence for the eight experimental conditions.

replaced by a diatonic chord in the comparison sequence (Condition 6) than when a diatonic chord was replaced by a nondiatonic chord (Condition 5). This indicates that stable events are more easily confused with previously presented unstable ones than the reverse.

Simulation

For each experimental condition, simulations were run with a sample of 18 sequences taken from Bharucha and Krumhansl (1983). In the original experiment, the changed target chord was located in Serial Positions 2 through 6 in the sequence. Given that no interaction of serial position was found, the target chord was always placed in the fourth serial position for the present simulations. For one half of the "different" trials, the fourth chord of the standard sequence was major and was replaced by a minor chord in the comparison sequence. For the other half, the fourth chord was minor and was replaced by a major chord.

Standard sequences were presented to the SIC-2 network followed by the first three chords of the comparison sequences. Activation was read off the chord unit corresponding to the fourth chord of the comparison sequence. According to the principle of contextual identity, stable events are more easily recognized as being identical than are less stable ones. Therefore, activation in the network should be positively correlated with the number of correct responses in the three "same" experimental conditions. According to the principles of contextual distance and contextual asymmetry, however, stable events in a given context are more easily confused with previously presented events than are less stable ones (resulting in lower performance for stable events in the "different" conditions), the strength of this confusion varying as a function of the temporal order of the event (with a stable event being more easily confused with an unstable event than the reverse). Accordingly, activation in the network should be negatively correlated with the number of correct responses in the "different" experimental conditions. The network is not trained to recognize the specific sequences. Instead, the activation levels that reflect tonal stability of the target in the context are predicated to mirror participants' pattern of performance.

Figure 10 (bottom) displays the activation of the target chord units in the comparison sequence. For the "same" conditions, higher activation was observed for diatonic targets (Condition 1) than for nondiatonic targets in tonal contexts (Condition 2) or targets in random contexts (Condition 3). The activation pattern was in accordance with the human data, showing better recognition performance for stable events. For the "different" conditions, activation for diatonic targets (Conditions 4 and 6) was higher than for nondiatonic targets (Conditions 7 and 5) or random targets (Condition 8). Once again, these activation levels mirror behavioral results: In comparison to unstable nondiatonic target chords, stable diatonic targets are more easily confused with other chords (diatonic or nondiatonic).

Harmonic Expectation

Harmonic expectation has been studied in different ways. Harmonic priming studies have explored how a previous context influences the processing of a target chord. These studies addressed the influence of a single chord (Bharucha & Stoeckig, 1987; Tekman & Bharucha, 1998), long chord sequences (Bigand & Pineau, 1997), and the mutual influence of global and local contexts (Tillmann, Bigand, & Pineau, 1998). Harmonic expectation has also been observed at a neurophysiological level. Using ERP, unexpected events were associated with a larger late positivity than expected ones (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). In a neural network, expected events should correspond to the units most activated by the previous context.

Harmonic Priming

Human performance in single-chord priming. In single-chord priming experiments, participants heard a prime chord followed by a target chord and had to decide as quickly as possible whether the target was or was not in tune. In Bharucha and Stoeckig (1987), the prime and target were either closely related (e.g., C- and B♭-major chords) or distantly related harmonically (C- and F♯-major chords). Priming was shown by more accurate and faster processing of in-tune targets when they were related to the prime. In Tekman and Bharucha's (1998) study, two types of target were selected: One was more psychoacoustically similar to the prime (e.g., E- and C-major chords), the other more closely related on the basis of harmonic convention (D- and C-major chords). The results revealed facilitation for psychoacoustically similar targets for a short (50 ms) SOA, and facilitation for harmonically related targets after a longer SOA (500 ms or longer).

Simulation. To simulate the priming results of Bharucha and Stoeckig (1987), the prime chords were presented to the SIC-2 network and the activation of target chord units was read off. For convenience, the results in Figure 11 are presented with reference to a C-major chord prime. After a C-major chord prime, activation was higher for harmonically related targets (B♭ major) than for harmonically unrelated targets (F♯ major). The SIC-2 network thus accounts for the asymptotic facilitation of the processing of related chords. To model the findings reported by Tekman and Bharucha (1998), the network was presented with the prime chords, and activation of the target chords was read off in early activation cycles (representing bottom-up activation) and at equilibrium (with top-down influences). The former simulated the short SOA and the latter, the longer SOA. As shown in Figure 11, the pattern of activation changed qualitatively during the reverberatory cycles from initial activation to equilibrium. At the outset, the pattern of activation reflected the number of tones shared by the prime and the target, whereas at equilibrium the pattern reflected conventional relatedness: the distance around the cycle of

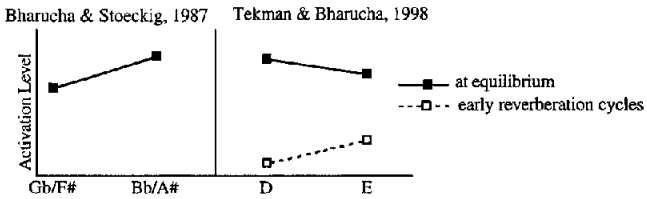

Figure 11. Left: Activation level for target chord units G♭/F♯ and B♭/A♯ after the presentation of a C-major chord prime (see Bharucha & Stoeckig, 1987). Right: Activation level for target chord units D and E after the presentation of a C-major chord prime at early reverberation cycles and at equilibrium (see Tekman & Bharucha, 1998).

fifths. The temporal course of activation in the model mirrored the priming results.

Human performance in chord sequence priming. Harmonic priming has also been examined using longer contexts. In a study by Bigand and Pineau (1997), the local context surrounding the target chord was kept constant, but the harmonic function of the target was manipulated by varying the global context (see prior discussion for more details). Participants were faster and more accurate in their intonation judgment when the target chord acted as a stable tonic chord rather than as a less stable subdominant chord. Tillmann, Bigand, and Pineau (1998) investigated the combined influence of global and local context. The target's relatedness to the prime was manipulated at both a global and a local level. For example, in a C-major key, the target chord was globally and locally related (GRLR) when it was a tonic chord (C) and was preceded by a dominant chord (G). It was globally related but locally unrelated (GRLU) when the preceding dominant chord was played one semitone higher (G♯). In this case, the target and the preceding chord did not belong to the same key. The target was globally unrelated but locally related (GULR) when only the first six chords of the sequences were transposed one semitone above (i.e., in the C♯-major key). Here the key of the first six chords was weakly related to the keys of the target chord and its preceding chord (i.e., C and G major keys). Finally, the target chord was both globally and locally unrelated (GULU) when the first seven chords were transposed one semitone higher (in the C♯-major key). The performance of participants demonstrated a strong effect of both global and local contexts. Target chords were processed more accurately and quickly when they were locally or globally related to the previous context. Furthermore, the effect of global context tended to be more pronounced at a fast than at a slow tempo.

Simulation. The first simulation was of the global context effects reported by Bigand and Pineau (1997). The first seven chords of the 20 sequences used in each experimental context were presented to the SIC-2 network, and the activation of the target chord unit was read off. As shown in Figure 12 (left), the target chord unit received stronger activation in the expected context when it acted as a stable tonic chord (I) than in the unexpected context when it was a less stable subdominant chord (IV). The second simulation was of the combined effects of global and local contexts reported by Tillmann, Bigand, and Pineau (1998). Simulations were run with the first seven chords of each sequence for two tempi. For the fast-tempo condition, the t parameter (cf. Equation 1) was set to 1 and in the slow condition to 1.5. This change in t (from 1 to 1.5) corresponded to the percentage change in tempo from the slow to the fast condition. The activation of the target chord unit depended on whether one or two sources of priming were present (see Figure 12, right). It was the highest for the GRLR condition, because both contexts were related to the target chord. Activation decreased for GRLU and GULR, respectively, with only one context related to the target. It was the lowest for GULU, in which the target chord had no relation to the previous context. At a slow tempo, the global and the local contexts exerted roughly similar effects. At a fast tempo, however, the global context strongly prevailed over the local context. The activation pattern in the neural net globally mirrored the combined influence of global and local context, as well as the impact of tempo on the global context effects observed with human participants (Tillmann, Bigand, & Pineau, 1998).

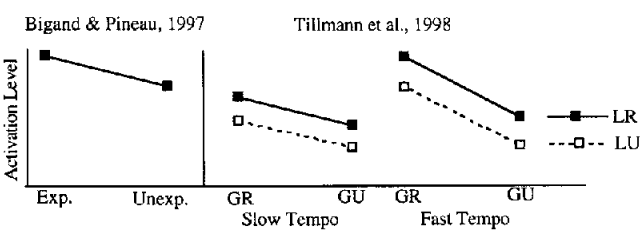

Figure 12. Left: Activation level for target chord units in the expected (Exp.) and unexpected (Unexp.) conditions used in Bigand and Pineau (1997). Right: Activation level for target chord units as a function of global context, local context, and tempo (Tillmann, Bigand, & Pineau, 1998). GR = globally related; GU = globally unrelated; LR = locally related; LU = locally unrelated.

ERP

Patel et al. (1998) measured ERPs for target chords occurring in the middle of musical sequences. The target chord varied in the degree of relatedness to the context: It was the tonic of the actual key, the tonic of a nearby key (three steps around the cycle of fifths), or the tonic of a distant key (five steps around the cycle of fifths). The out-of-key targets elicited late positivities of different amplitudes, depending on their harmonic distance from the context key. The ERP waveforms between 500 ms and 800 ms were clearly distinguished in positivity for the three target chords: The positivity was the largest for distant-key targets, and its amplitude decreased for the nearby-key targets and then for the same-key targets. The authors interpreted the positivity as manifesting the difficulty of fitting a given chord into the context. In addition, participants judged the musical sequences with the same-key, nearby-key, and distant-key target chords as acceptable on 80%, 49%, and 28% of the trials, respectively.

Simulation. The simulation was run by presenting the chords of the musical sequence preceding the target chord (Patel et al.'s 1998, Figure 5). The activations of chord units representing the three targets were then read off. Activation was the highest for the same-key target followed by the nearby-key target and was the lowest for the distant-key target (Figure 13). This activation hierarchy for the targets reflects the pattern of ERP data and subjective judgments about musical coherence. The more the target chord was activated by the context, the smaller was the amplitude of the ERP waveform and the greater was the likelihood of judging the whole sequence as being acceptable. The activations are interpreted as the levels of expectation for incoming events. Strictly speaking, the ERP data reflect the response to the target itself, with stronger positivity for unexpected targets. A second simulation was run to analyze the change of the network's activation as a result of the presentation of the target. For each of the three target chord units, the difference in activation before and after the target was calculated. The activation difference was the strongest for the distant-key target, decreased for the nearby-key target, and then decreased further for the same-key target. This result thus mirrored the amount of positivity in the ERP waveforms that was observed for the three targets.

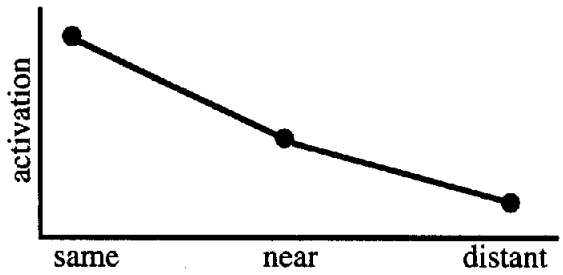

Figure 13. Activations for the three target units (same, near, and distant key) used in the musical sequence in Patel, Gibson, Ratner, Besson, and Holcomb (1998).

Perceived Relations Between Keys

In the simulations presented thus far, the underlying key remained the same in a given context. The network model simulates the perceived changes in chord stability caused by different key contexts. The changes in stability, in turn, affect similarity ratings, recognition accuracy, and harmonic expectation. In most Western musical pieces, however, temporary changes in key (referred to as modulation in music theory) often occur. The extent of a modulation depends on the harmonic distance between the main key and the new key. Modulation is a crucial feature of Western music that is relevant from both stylistic and expressive points of view. In the classical period (Haydn and Mozart), modulations occur primarily between closely related keys. In the last romantic period (Wagner), however, shifts toward harmonically unrelated keys became more sudden and frequent. Empirical research suggests that listeners possess implicit knowledge of the distance between keys (Bharucha & Krumhansl, 1983; Cuddy, Cohen, & Miller, 1979; Holleran et al., 1995; Krumhansl et al., 1982) and manage to detect modulations occurring in musical pieces (Cuddy & Thompson, 1992a, 1992b; Krumhansl & Kessler, 1982, Experiment 2; Platt & Racine, 1994; Thompson & Cuddy, 1989). The following simulations investigated whether the network can account for psychological distances between keys and can reproduce listeners' changing sense of key.

Quantifying Interkey Distances From Harmonic Hierarchy

Human Data

According to Krumhansl (1990), psychological distances between keys may be derived from the perceived tone or chord hierarchy induced by a given key context. In one experiment (Krumhansl, 1990, p. 182), participants rated on a 7-point scale how well the 12 major chords fit with a given major key context established by a short sequence. Harmonic hierarchies found for a given key were then shifted to the 11 other major keys. These 12 hierarchy profiles were then intercorrelated. The pattern of correlation (Figure 14) approximates the music theoretical distances between major keys on the cycle of fifths.

Simulation

A simulation was run with the same C-major context as in the experiment. The activation of the 12 major chord units of the network defined the activation profile specific to the C-major key. To compare the network performance and Krumhansl's findings, the activation profile specific to the C-major key was shifted to the different tonics of the 11 other major keys, and the 12 activation profiles were then intercorrelated. As shown in Figure 14, the pattern of correlations derived from the network reflected the cycle of fifths. Although the peaked curve based on participants' ratings differed slightly from the smooth curve based on activations, the correlation between human data and the model was significant, $r(10) = .793, p < .01$. Data based on both participants' ratings and activation of the model suggest that a key context imposes a hierarchical organization of chords, and these hierarchies are more similar for closely related keys than for distant keys. Similarity ratings and the activations of neighboring keys on the cycle of fifths (e.g., C and either G-major or F-major keys) were positively correlated. Negative correlations were observed for distant keys on the cycle of fifths (e.g., C- and F♯-major keys). The local peaks in the participants' profile observed for A major and D♯ major can be understood as deriving from the parallel and relative major-minor key relationships with C major (e.g., A major is the parallel major of a minor, which is the relative minor of C major). The network does not include minor keys, which may be the origin of the missing peaks in the network's correlation profile.

Perceiving Key and Modulation

Human Performance

Krumhansl and Kessler (1982, Experiment 2) investigated how listeners develop a sense of key and how they assimilate modulations to new keys. The chord sequences used in the experiment remained in the same key (no modulation), modulated toward a closely related key (close modulation), or modulated toward a more distant key (remote modulation). Table 2 displays examples of the three types of sequences used. In the modulating sequences, the first three chords established the first key with a perfect cadence (V-I). The last three chords formed a perfect cadence in the new key either one step away (close modulation) or two steps

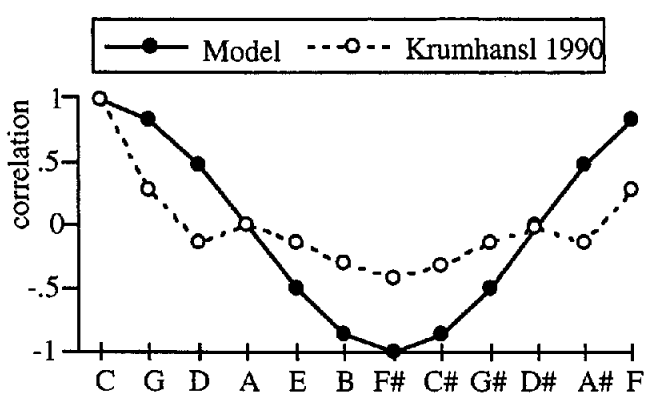

Figure 14. Correlation between the C-major context and the 12 major contexts for participants' ratings and for activation values of major chord units in the learned connectionist model. From *Cognitive Foundations of Musical Pitch* (p. 183, Table 7.10), by C. L. Krumhansl, 1990, Oxford, England: Oxford University Press. Copyright 1990 by Oxford University Press. Used by permission.

Table 2
Three of the Chord Sequences Used by Krumhansl and Kessler (1982, Experiment 2)

Sequences	Chord number 1	2	3	4	5	6	7	8	9
No modulation (C major)									
Sequence	F	G	a	F	C	a	d	G	C
Probe-tone method	F	C	C	F/C	C	C	C	C	C
Model	F	C	C	F	C	C	C	C	C
Direct and close modulation (C major → G major)									
Sequence	F	G	C	a	e	b	e	D	G
Probe-tone method	F	C	C	C	C/G	G	G	G	G
Model	F	C	C	C	G	G	G	D	G
Indirect and remote modulation (C major → B♭ major)									
Sequence	F	G	C	a	F	g	E♭	F	B♭
Probe-tone method	F	C	C	C	C	C	B♭	B♭	B♭
Model	F	C	C	C	F	F	B♭	B♭	B♭

Note. The second line indicates the inferred keys based on probe-tone judgments. The third line indicates the most activated key units in the connectionist model.

away (remote modulation) from the first key on the cycle of fifths. Several intermediate pivot chords belonging to both keys allow a smooth modulation from one key to another. In the remote modulation, only the fifth chord of the sequences was a pivot chord. However, the fifth and the sixth chords may also be analyzed as belonging to the F-major key, a key between the C- and the B♭-major keys on the cycle of fifths.

After each chord of the sequence, participants rated on a subjective scale how well each of the 12 tones of the chromatic scale fits with the preceding context (probe-tone ratings). The development of participants' sense of key was tracked by correlating these probe-tone ratings with the tonal hierarchy profile specific to each key.[9] The strongest correlation between the probe-tone ratings and a given key profile indicates which key is induced at each point in the chord sequence. Tracing the change in probe-tone ratings thus yields the trajectory of the key over the course of the sequence. Krumhansl and Kessler's (1982) probe-tone ratings analysis may be summarized as shown in Table 2. For all three major key sequences, the probe-tone judgments after the first chord correlated strongly with the F-major key. For the nonmodulating sequence, the sense of the C-major key started to become instilled after the second chord (G major) and increased thereafter. Some local effects of the individual chords were evident in the probe-tone ratings. For example, the fourth chord (F major) tended to increase the correlation between the probe-tone ratings and the F-major key profile. For the two modulating sequences, the sense of the first key (C major) was clearly instilled with the third chord of the sequence (C-major chord). For the close modulation sequence, the sense of the second key appeared on the fifth chord and was clear from the sixth chord until the end of the sequence. For the remote modulation sequence, the sense of the new key was firmly established with the seventh chord of the sequence.

Simulation

Simulations were run with the three sequences displayed in Table 2. After each chord, the activation pattern of the key layer was read off. The key unit with the maximum activation at each point in the sequence was taken to be the key most strongly induced in the network at that point in time. Table 2 shows the tracked keys for each chord in the three sequences. For all sequences, the key tracked after the first chord reflected the key of which this chord was the tonic (i.e., F major). Adding a G-major chord shifted the maximum activation to the C-major key unit. It is worth noting that the model detected the underlying C-major key even though the actual C-major chord was not presented. For the subsequent chords, the most activated key unit changed in the course of the sequence. In the nonmodulating sequence, C major remained the most activated key unit until the end of the sequence, with only one exception: on the fourth chord (F-major chord) the key of F major was more active in the network than the key of C major. This finding reflects the local influence of the fourth chord, also observed by Krumhansl and Kessler (1982) with human participants. For the two modulating sequences, the key of C major remained the most active key unit until the fourth chord. For the close modulation sequence, the activation of both G- and C-major key units was high on the fifth chord (the pivot chord) with a slight advantage for the G-major key unit. For the remaining chords, the G-major key was the most active unit, except on the eighth chord. This chord (D major) produced the highest activation for the D-major key, a key closely related to the G-major key.[10] For the remote modulation sequence, the key unit of F major was the most active on the fifth and sixth chords of the sequence. After the seventh chord, the B♭ major key unit received the strongest activation until the end of the sequence. The network thus tracked the modulation around the cycle of fifths: starting in C major and going over F major to B♭ major.

As shown in Table 2, the changing sense of key in human participants shares several features with the network's performance. The critical feature for both is that the inferred key was generally not instantiated by the actually presented chord. The sense of a key prevailed even though the tonic chord was not present. This suggests that, for both humans and the network, a key refers to a more abstract level of structure than just a collection of chords. This finding is even more remarkable when a modulation occurs in the chord sequence. For example, in the close modulation sequence, both the model and the participants inferred the key of G major on the fifth chord of the sequence, even though no G-major chord had been presented. A similar result was obtained in the remote modulation, in which both participants and the model inferred the key of B♭ major on the seventh chord of the sequence, even though no B♭ chord had been presented. The last main correspondence between the performance of the participants and the model concerned the time course of the changing sense of key.

[9] In a previous experiment (Krumhansl & Kessler, 1982, Experiment 1), probe-tone ratings were obtained for contexts that unambiguously instilled a major or minor key. These rating profiles showed highest ratings for the tonic tone, followed by the third- and fifth-scale degrees, with the lowest ratings for nondiatonic tones. They agreed with tonal functions defined by music theory and were considered distinctive markers of the instilled key.

[10] In the present case, this strong local influence of the D-major chord might be due to the fact that the preceding b minor and e minor chords both belong to the G-major and D-major keys.

The sense of a change in key generally occurred for the same chords in the sequences in both human and model performance.

Perceived Asymmetry in Modulation

Other empirical studies investigated how listeners' sensitivity to key change may be influenced by both the distance and the direction of modulation on the cycle of fifths (Cuddy & Thompson, 1992a, 1992b; Thompson & Cuddy, 1989). In these studies, short chord sequences were adapted from Bach chorales and either did or did not modulate. In modulating sequences, the new key established at the end of the sequence was either one or two steps on the cycle of fifths from the first key in either a clockwise or counterclockwise direction. In other words, the chord sequences modulated either to the dominant key (one step clockwise), to the subdominant key (one step counterclockwise), to the supertonic (two steps clockwise), or to the flattened seventh (two steps counterclockwise).

Human Performance

In Cuddy and Thompson's (1992a) study, probe-tone ratings were obtained at the end of each sequence. With nonmodulating sequences, participants' ratings were highly correlated with Krumhansl's profile for the key of the sequence (Table 3). The correlation decreased as the distance between the first and the second keys increased on the cycle of fifths. Listeners differentiated between slight (one step) and strong (two steps) key movement. The critical new finding of this experiment was an asymmetry in key movement. For both distances, the decrease in correlation was more pronounced for counterclockwise than for clockwise modulations. This outcome demonstrated that the changing sense of key was stronger when the sequence moved toward the subdominant than toward the dominant key or when the sequence moved toward the flattened seventh than toward the supertonic key. Thompson and Cuddy (1989) reported converging evidence for perceived asymmetry in modulation, obtained with a different experimental task.

Simulation

Simulations were run with the modulating and nonmodulating sequences that Cuddy and Thompson (1992a) used. The dominant seventh chords that introduced the counterclockwise modulations were presented to the input (tone) layer, even though the model had not learned this chord type. To compare human and model performance, the state of the key layer at the end of the sequence was correlated with the state of the key layer after a context in the first key. These correlations reflect the network's sensitivity to key change: The higher the correlation, the smaller is the shift in tonal organization between the first and second keys. As shown in Table 3, the highest correlation in the network was found for nonmodulating sequences. The correlation then decreased as the distance between the first and second keys increased on the cycle of fifths. The network also reproduced the perceived asymmetry in modulation: For each distance, counterclockwise modulations caused a stronger decrease in correlation than clockwise modulations. In addition, the correlation with the second key was evidence that the network correctly identified the second key of the sequence. Interestingly, the correlation with the second key showed a weak asymmetry, comparable to the one observed with human participants (see Table 3).

Perceived Relations Between Tones

Once the key of a musical context is established, the 12 tones of the chromatic scales are perceived in a hierarchy of stability (Cuddy et al., 1981; Francès, 1958; Krumhansl, 1979, 1990). In major key contexts, the tonic (first scale degree) is perceived as more stable than the dominant (fifth scale degree), which in turn is

Table 3
Correlation Between Probe-Tone Ratings and Tonal Hierarchies of the Principal Key in Nonmodulating Sequences and of the First or Second Key in Modulating Sequences (Cuddy & Thompson, 1992a) and Between the State of the Key Layer in SIC-2 at the End of the Sequence and After a Context in the First or Second Key

Sequences	Cuddy & Thompson	SIC-2
Nonmodulating	.91	.978

	First key		Second key	
	Cuddy & Thompson	SIC-2	Cuddy & Thompson	SIC-2
Clockwise				
One step	.87***	.92**	.83***	.9917***
Two steps	.53**	.86**	.91***	.9998***
Counterclockwise				
One step	.71***	.61*	.78***	.9898***
Two steps	.31	.52	.83***	.9993***

Note. Cuddy and Thompson (1992a) data are from "Asymmetry of Perceived Key Movement in Chorale Sequences: Converging Evidence From a Probe-Tone Analysis," by L. L. Cuddy and W. F. Thompson, 1992, *Psychological Research, 54,* p. 57, Table 3. Copyright 1992 by Springer-Verlag. Reprinted with permission. SIC = sparse input coding.
* $p < .05$. ** $p < .01$. *** $p < .001$.

perceived as more stable than the mediant (third scale degree). Tones of the other scale degrees are perceived as less stable yet more stable than nondiatonic tones. These empirical findings lead to a challenging hypothesis for the present network. Given that tones are linked to chords, which in turn are linked to keys, the network may have internalized some of the Western relationships between tones and keys even though it had never been exposed to Western melodies. Because activation reverberates from key to chord units and from chord to tone units, the network potentially may account for top-down influences of the key context on perception and memory for individual tones.

Probe-Tone Ratings

Human Data

In Krumhansl and Kessler's experiment (1982), participants were presented with a key context followed by each of the 12 tones of the chromatic scale as probe tones. Participants rated on a 7-point scale how well the probe tone fits with the previous key context. The key context was established by either an ascending or descending scale, a single major or minor chord, or a short chord sequence in a major or minor key. Average ratings defined a key profile specific to a given key. The C-major and c-minor key profiles are shown in Figure 15 (bottom). In both cases, diatonic tones provided a better fit with the context than did nondiatonic tones. Among diatonic tones, the highest ratings were observed for the tonic tone, followed by the tones of the third and fifth scale degrees, which together with the tonic form the tonic triad chord.

Simulation

In the following simulations, the tones of a C-major chord (containing the tones C, E, G) or a c-minor chord (containing the tones C, D♯/E♭, G) were presented to the network. During reverberation, chord units sent phasic activation to the tone layer. The units in the tone layer accumulated phasic activation through spatial and temporal summation until equilibrium was reached. The accumulated phasic activation of tone units at equilibrium is displayed in Figure 15 (top). After the presentation of a C-major chord, several tone units other than the three sounded tones (C, E, and G) were activated. The diatonic tone units (C, D, E, F, G, A, B) received more activation than the nondiatonic tone units (C♯, D♯, F♯, G♯, A♯), with a reduced activation for the tone B. This activation profile reflected the top-down influences of the tonic key of the presented chord. After the presentation of a c-minor chord, the tone units corresponding to the descending minor scale (C, D, D♯/E♭, F, G, G♯/A♭, A♯/B♭) received stronger activation than the other tone units, even though several tones were not present in the c-minor chord that was played. The C-major and c-minor activation profiles were clearly different. Interestingly, these differences were reflected in the activation of noncomponent tone units of the two chords. For example, the tone A belongs to neither the C-major chord nor the c-minor chord. Nevertheless, the tone A was more activated by a C-major chord than by a c-minor chord. This difference reflects key membership: the tone A belongs to the C-major key but not to the c-minor key. The opposite was observed for the tone B♭, which belongs to the c-minor key but not to the C-major key.

At equilibrium, the profiles of phasic activation reflected globally the relative stability of the 12 chromatic tones in a given key context and were correlated with Krumhansl's key profiles, $r(10) = .89, p < .01$, for major key contexts, $r(10) = .83, p < .01$, for minor key contexts. Some differences were nevertheless observed between the profiles of humans and of the model. Although the model clearly inferred the remaining diatonic tones from the tonic triad input, the difference in accumulated phasic activation between the tonic triad tones and the other diatonic tones was not as great as in the human data, almost as if the learned cultural knowledge exerted an influence so strong that the stimulus itself was no more strongly represented than what was inferred from it. It has been suggested that listeners' responses in a probe-tone situation reflect not only implicit knowledge of Western hierarchy but also the influence of short-term memory of the tones played in the stimulus context (Butler, 1989; Deutsch, 1972). Given that participants may have been partly influenced by short-term memory of the stimulus tones, multiple regression analyses were performed to predict human probe-tone ratings from phasic activation and activation from the stimulus alone (Figure 15 bottom). A linear

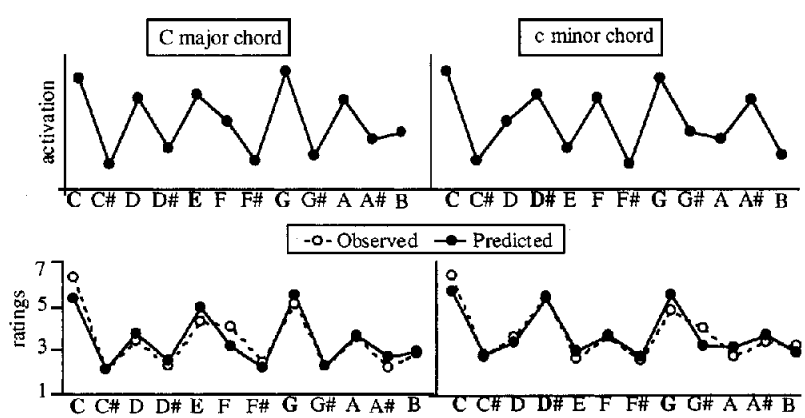

Figure 15. Top: Activations received by the tone layer during reverberation after the presentation of a C-major and a c-minor chord. Bottom: Observed similarity ratings from Krumhansl and Kessler (1982, Experiment 1; adapted by permission of the author) and results of regression analysis for C-major and c-minor chord.

combination of these factors provided a good fit for probe-tone ratings for major key, $R = .93$, $F(2, 9) = 27.39$, $p < .001$, and for minor key, $R = .93$, $F(2, 9) = 27.10$, $p < .001$. This suggests that Krumhansl's key profiles may be reproduced in the network by an additive combination of phasic activation reverberating to equilibrium and activation resulting directly from the stimulus.

Perceived Asymmetry Between Tones

Human Performance

Stable tones in a given key context act as cognitive reference points (Krumhansl, 1979, 1990) to which the other less stable tones are anchored (Bharucha, 1984, 1996). Stable reference points influence similarity ratings and recognition memory of tones and account for the contextual asymmetry principle. In a study by Krumhansl (1979, Experiment 1), participants rated, on a scale from 1 to 7, the similarity of a tone pair presented after a tonal context. The key context was established by a C-major chord, an ascending scale, or a descending scale. Table 4 displays the pairs of tones used in this experiment. Tone pairs contained a tone of the major triad (C, E, or G) plus another diatonic scale tone, a tone of the major triad plus a nondiatonic tone, or a nondiatonic tone plus another diatonic scale tone. The temporal order of the tones was counterbalanced. Similarity ratings reflected the influence of tonal hierarchies: Ratings were higher when the pair ended on a tone of the major triad (C, E, G) than on another diatonic scale tone (D, F, A, B). They were higher when the pair ended on a diatonic scale tone (D, F, A, B) rather than on a nondiatonic tone (C♯, D♯, F♯, G♯, A♯). Moreover, participants' responses revealed the importance of the temporal order of the two tones. Similarity ratings were higher when the first tone in the pair was less stable than the second tone compared with the reverse order. The strongest asymmetry was observed for pairs containing a nondiatonic tone and a major triad tone: The similarity ratings were 11.13% higher when the second tone belonged to the major triad. A smaller asymmetry was found for pairs made of nondiatonic and diatonic tones: Ratings increased 10.79% when the pair ended on a diatonic tone. The smallest asymmetry was found for pairs made of major triad tones and diatonic tones: Similarity ratings were 7.48% higher when the second tone belonged to the major triad.

Simulation

In the simulations, each of the key contexts used in Krumhansl's experiment was presented to the network followed by the first tone of the experimental pairs. The phasic activation received by the tone unit representing the second tone of the pair was read off. Results were averaged over the three contexts (C-major chord, ascending scale, descending scale). As shown in Table 4, the activation at the second tone of the pair decreased as its stability in the key context decreased. Moreover, the activation of the tone layer reproduced asymmetries in similarity judgments. The strongest asymmetry appeared for pairs consisting of a major triad tone and a nondiatonic tone: The increase in activation was 13.9% higher when the second tone of the pair was a major triad tone than when it was nondiatonic. The asymmetry was smaller for nondiatonic-diatonic pairs of tones: The increase in activation was 8.73% higher when the second tone of the pair was diatonic than when it was nondiatonic. The smallest asymmetry was observed for pairs consisting of a major triad tone and a diatonic tone: The activation of a triad tone unit after hearing a diatonic tone was 5.3% higher than was the activation of a diatonic tone unit after hearing a major triad tone.

Memory for Melodies

The aspects of the perception of and memory for tones considered previously greatly underdetermine the perception of melodies, which involves many factors. Beyond the tonal hierarchy, melodic contour (the up and down movement of pitches in a melody) is a crucial feature for melody recognition. Although the model does not at present include pitch height or contour, the purpose of this last simulation was to investigate the extent to which the network may account for results obtained in experiments involving melodies, such as those reported by Dowling (1978).

Human Performance

In Dowling's (1978) experiment, participants made same-different judgments for pairs of melodies (a standard melody followed by a comparison melody). In the "same" comparison condition (referred to as the "exact transposition" condition), the comparison melody was played in a different key from the standard melody but preserved the interval sizes and the melodic contour of the standard. The comparison melody was thus an exact transposition of the standard. Three "different" comparison conditions were distinguished. In the first different condition (the "tonal answer" condition), the comparison melody was shifted to the sixth scale degree of the standard. As a consequence, the comparison melody shared the same melodic contour but differed

Table 4
Asymmetries of Similarity Data and Activation Data

First tone	Second tone	Similarity	Difference	Activation	Difference
Nondiatonic	Major triad	3.675		8.76	
Major triad	Nondiatonic	3.308	.368	7.72	1.0
Nondiatonic	Other scale	3.717		8.47	
Other scale	Nondiatonic	3.355	.362	7.79	.68
Other scale	Major triad	4.871		9.24	
Major triad	Other scale	4.532	.339	8.75	.49

Note. Major triad = tones C, E, G; other scale = tones D, F, A, B; and nondiatonic = tones C♯, D♯, F♯, G♯, A♯. All activation values were multiplied by 10^3. Similarity data are from "The Psychological Representation of Pitch in a Tonal Context," by C. L. Krumhansl, 1979, *Cognitive Psychology, 11*, p. 346, Table 3. Copyright 1979 by Academic Press. Adapted with permission.

from the standard melody by one interval. In the second "different" condition (the "atonal contour foil" condition), the comparison melody shared the same contour as the standard but was made of randomly selected intervals. In the third "different" condition (the "random foil" condition), the comparison melody differed from the standard in both melodic contour and diatonic scale intervals. Listeners' performance in each of the "different" conditions was compared with that of the "exact transposition" condition. An analysis of area under the memory operating characteristic curve (MOC) showed that participants performed at chance in the "tonal answer" condition (the areas under the MOC curve were .48 and .49 for experienced and inexperienced listeners, respectively). Recognition performance increased for atonal contour foils (areas under MOC curve were .79 and .59 for experienced and inexperienced listeners, respectively). The highest performance was found for random foils (areas under MOC curve were .84 and .83 for experienced and inexperienced listeners, respectively). Results suggest that both the musical scale and the melodic contour governed same-different judgments for melodies.

Simulation

The simulations were run with pairs of melodies (Dowling, 1978, Figure 3) consisting of the standard melody followed by one of the four experimental conditions: "exact transposition," "tonal answer," "atonal contour foil," and "random foil." The network was presented with the standard melody immediately followed by the comparison.[11] To compare human and network performance, the phasic activation received by the tone layer in the exact transposition condition (following the presentation of the standard and its exact transposition) was correlated with the phasic activation received by the tone layer in each of the three "different" conditions. High correlation values would indicate that the "different" condition did not substantially affect the state of the tone layer in comparison to the exact transposition condition. The correlation values thus reflect the capacity of the network to differentiate the "exact transposition" condition from each of the "different" conditions. These correlation values should thus be inversely related to the participants' performance. A high correlation, $r(10) = .99$, was found between exact transpositions and tonal answers, suggesting that the network barely differentiated between the comparison melodies in these conditions. The correlation decreased for exact transpositions and atonal contour foils, $r(10) = .27$. The lowest correlation was found for exact transpositions and random contour, $r(10) = .16$. Network performance thus mirrored human performance: The distinction between the standard and the comparison melodies failed when the key and the melodic contour of the standard were preserved. Differentiation increased when the comparison shared the same contour as the standard but was made of randomly selected intervals regardless of the tonal scale. Differentiation was easiest when the comparison melody had a different melodic contour and different diatonic scale intervals than the standard. The neural net simulation accounts for the tonality effects in this memory experiment. Its outcome suggests that the melodic contour effects observed in participants' performance may have had some tonal implications that are being picked up by the network.

GENERAL DISCUSSION

This study investigated the implicit learning of a highly structured system (the Western tonal music) to which we are frequently exposed in our natural environment. In this system, tones cluster in characteristic ways to form chords, and both tones and chords cluster over longer temporal windows to form keys. Tones and chords have no external meaning (as do words), but they have structural functions. Among the factors that determine the structural function of a tone or a chord is its tonal stability. Structural function is context dependent. For instance, the C tone is a stable tonic in a C-major context, a less stable dominant and mediant in F- and A♭-major key contexts, respectively, a moderately unstable leading tone in a D♭-major key context, and a very unstable nondiatonic tone in an F♯ key context. Similarly, the functions of all tones of the chromatic scale and of all 24 major and minor chords change in the different key contexts. Keys are the largest building blocks of the Western tonal system. They are related in specifically defined ways. Each of the 24 keys has close relationships with four other keys. For example, the C-major key is very close to the G- and F-major keys (cycle of fifths relationship), to the a-minor key (relative major-minor relationship), and to the c-minor key (parallel major-minor relationship). In sum, Western tonal music rests on a sophisticated network of functional relationships among tones, chords, and keys.

Given this apparent complexity, it is frequently asserted that perceiving Western musical structures requires extensive explicit learning of music theory (see Levinson, 1997, for a debate). However, the experimental data show strikingly that explicit learning is unnecessary to exhibit a sensitivity to subtle musical structures (see, e.g., Bharucha & Stoeckig, 1986; Bigand, 1997; Cuddy & Thompson, 1992a; Hébert, Peretz, & Gagnon, 1995). This suggests that mere exposure to Western musical pieces suffices to develop implicit but sophisticated knowledge of Western harmony. Implicit learning, which Francès (1958) referred to as a "tonal acculturation process," is thought to be developed like "un système d'habitudes perceptives . . . sans que les termes objectifs dont dépend leur acquisition soient consciemment posés par les sujets" [a system of perceptual habits . . . without a conscious acquisition by the participants] (p. 108).

Tonal acculturation illustrates in the music domain the power of the implicit learning process investigated by Reber (1967, 1989), a process that has received increasing attention during the last decade (e.g., Altmann et al., 1995; Cleeremans, 1993; Saffran et al., 1997). According to Reber (1967), implicit learning is a fundamental characteristic of the cognitive system, enabling the acquisition of highly complex information without awareness. It has been investigated in the laboratory with artificial material that conforms to regularities defined by finite state grammars, for example. Western tonal music contains a highly sophisticated system of regularities (i.e., regularities of co-occurrence, frequency of occurrence, and psychoacoustic regularities). Of course, musical systems are much more complex than artificial finite-state grammars. However, the number of opportunities to listen to musical pieces obeying this system of relationships is so great in

[11] To account for the longer duration of the last note in Dowling's melodies (half note), the input vector of this tone was multiplied by a factor of 2.

everyday life that it is plausibly learned by the same process. In this way, the tonal acculturation process may be viewed as an ecological validation of implicit learning (Bigand et al., 1998).

The general interest of the present connectionist model is to provide a formal account of Western music knowledge representation that is based on a plausible learning process. In the simulations reported here, the Western tonal system is taken as one specific example of the learning of complex systems in general. A musical system contains both general constraints as suggested by commonalties between music of different cultures and specific constraints that vary across cultures. The most obvious commonalities between music of different cultures are the presence of a finite number of sound units and the presence of octave equivalence (see Dowling & Harwood, 1986). In our connectionist framework, these two general constraints are expressed by the presence of a finite number of input units that represent pitch chroma. The specific constraints that vary across cultures pertain to the way in which these discrete units are combined simultaneously or sequentially. These are the constraints that are learned by the model described previously.

The learning constraints given to the model are general. They are grounded in fundamental processes in neurophysiology: frequency-tuned units in auditory cortex, a layered architecture, plasticity in auditory cortex, and hebbian learning (Hebb, 1949). With the added constraint that hebbian learning is restricted to the most active output and its neighbors, self-organization is obligatory and automatic. The musical environment impresses on the brain the correlational structure of music, and this internalized structure then serves to filter and facilitate subsequent music perception. With the help of these constraints, the model adapts to the specific rules of Western harmony through mere exposure to typical musical exemplars. As the learning rules are general, the model would also work with different kinds of musical systems as well as with other systems based on events that have correlational structures.

The implicit learning of the Western tonal system was simulated by an unsupervised learning algorithm. At the outset, the model did not incorporate knowledge of the tonal system (e.g., weights of the connections were set to random values). The structure of the system adapted to the regularities of harmonic relationships through repeated exposure to tonal material. The connectionist algorithm changes connections to allow units to become specific detectors of combinations of musical tones over brief temporal windows. The entire learning process is guided by bottom-up information only, and there is no external teacher. Furthermore, there are no explicit rules or concepts stored in the model. The learned connections encode collectively how the events appear together in music. By mere exposure, these regularities are extracted and represented in the pattern of connections between the three layers. The first connection matrix reflects which pitch (or virtual pitch in the case of the rich input coding simulations RIC-1 and RIC-2) is part of a chord; the second matrix reflects which chord is part of a key. The overall pattern of connections reflects how tones, chords, and keys are interrelated in Western music. Just as for nonmusician listeners, the knowledge is acquired without explicit instruction or control. It may be argued that layer-by-layer training is an oversimplification of the learning process. However, preliminary simulations with SIC-1 suggested that simultaneous training of the two layers leads to the same network structure, although the training takes longer.

An interesting feature of the model is its ability to account for Western music knowledge with a low-dimensional and parsimonious representation. Low dimensionality is important for a cognitive model of knowledge representation. The internal representation should capture similarities and redundancies in the environment with as few dimensions as possible. Projections of high-dimensional input onto lower dimensional representations have been proposed in models of visual perception (S. Edelman & Intrator, 1997) and have been found in brain structures for different types of sensory stimuli (Kaas, 1995). Techniques to model dimensionality reduction include principal-components analysis, multidimensional scaling, and SOMs. Besides the derivation of a low-dimensional representation from an initially high-dimensional perceptual space, the Kohonen map creates topographic representations reflecting similarities between objects with distances on the map. Furthermore, after learning, the map can be used to simulate classifications and representations of new objects or object combinations. In our model, lower dimensions are extracted from the input space (the chords and the specific combinations of chords in time) without supervision, creating topological representations on two hierarchical levels (the chord level and the key level). For example, a 12-dimensional vector codes the pitch combinations for one chord. The 24 vectors coding the pitch combinations for minor and major chords are projected on a topologically ordered two-dimensional map.

A major problem for a model of musical knowledge is to provide a parsimonious account of the contextual dependency of musical functions. Some geometrical models represent musical events (tones or chords) in a multidimensional space separately in each parent key (e.g., Longuet-Higgins, 1978; Shepard, 1982). For instance, a C-major chord would be represented at least four times to account for its different functions in the C-, G-, F-major and a-minor key contexts. Because each chord and tone might occur in different key contexts, representing all potential functions of chords and tones in the same model results in considerable redundancy. In the present model, the learned structure of the tonal representation is quite simple: The tones are linked to the chords that are linked to the keys, reflecting their statistical properties. The different contextual functions of a musical event emerge from the patterns of activation spreading through this simple architecture. In addition to its nonredundant representation, the interactive activation patterns also account for the three basic contextual principles shown by Krumhansl (1990). For example, contextual asymmetry is captured in the following way: In a given musical context (e.g., C-major key context), a stable event (e.g., a C-major chord) sends weaker activation toward an unstable one (e.g., an F♯-major chord) than the reverse. Musical cognitive reference points are thus not stored explicitly in the model. They are emergent properties of activation spreading through the network.

A further emergent property relates to the identification of the key in a given context. Several key finding algorithms have been proposed in the literature to address this issue (e.g., Huron & Parncutt, 1993; Krumhansl & Schmuckler, cited in Krumhansl, 1990; Vos & Van Geenen, 1996). Our model takes an alternative approach. In the key layer, the most activated key unit represents the underlying key induced in listeners at any given point in a musical piece. The underlying key is not computed separately but

emerges from the activation pattern. Activation propagated toward the key units may be caused by tone sequences (melody) or chords (harmony) or both. The simulations provide evidence that the model is not entirely dependent on the presence of the tonic in the stimuli to detect the underlying key. For example, an F-major chord (the subdominant chord in a C-major context) followed by a G-major chord (the dominant chord in a C-major context) leads to the C-major key unit being the most strongly activated, even if the C-major chord has not been played yet. As activation accumulates in the network, the model also tracks the development of the underlying key during the ongoing sequence. The activation pattern of the most activated key units reflects the dynamic process of tonality sensation that constantly evolves over time. Consequently, the model does not need to encounter the problem of delineating appropriate segments for an analysis (see Shmulevich & Yli-Harja, 1999, for a discussion). In addition, the model generates predictions for the most probable key to occur next during a piece.

The model also provides a parsimonious account of musical tension–relaxation and expectancy. The organization of tension–relaxation relationships in Western music (see Lerdahl & Jackendoff, 1983), and its link to expectation (Lerdahl, 1996), has received some attention but most writings of music aesthetics have focused on expectation. According to Meyer (1956, 1973), musical expressivity derives from the way musical expectation may or not be resolved in the course of the piece. The present connectionist framework establishes how musical tension–relaxation and musical expectancy may derive from abstract knowledge of Western harmony. The simulations account for the influence of a short context (one chord) on the development of harmonic expectancy (cf. Bharucha & Stoeckig, 1987) as well as for the growth of expectancy during a longer musical sequence (cf. Bigand et al., 1999). They also provide a possible account of the mutual influence of global and local harmonic relationships on harmonic expectancy (Bigand et al., 1999; Tillmann, Bigand, & Pineau, 1998), and of the influence of the temporal order of chords within a chord sequence (Tillmann & Bigand, 1999). A more critical feature of the model is its ability to account for the time course of bottom-up and top-down processes during expectancy formation. Western musical structure reflects the influences of both acoustic properties of tones and cultural tradition in combining pitches in specific ways. The respective influences of sensory and knowledge driven processes thus have been a matter of debate in music psychology (Krumhansl, 1990; Parncutt, 1988), and both play roles in the processing of musical signals (Bigand et al., 1996). The question remains to determine when each influence is the most active (see Regnault et al., in press).

The present model sheds light on the time course of both processes. The first cycles of reverberation reflect bottom-up influences, and only units sharing actual sounded events are activated. After several cycles, activation spreads through the network, and the top-down influences of higher representation levels become apparent. The key layer influences the chord layer that in turn impresses on the tone layer. The lower levels send activation back to the higher levels, and the process continues until an equilibrium is reached, at which point activation patterns reflect tonal and harmonic relations in the actual tonal context. In some cases, the difference between bottom-up and top-down influences simulated by the model may be sources of potential empirical falsification. Thus far, human data fit nicely with the model (Tekman & Bharucha, 1998).

In sum, by simulating context effects and taking into account a range of experimental data, the model provides a formal and parsimonious way of unifying music theoretic and psychological constructs with the help of a common mechanism, namely spreading activation (cf. in language, e.g., Dell, 1986; McClelland & Elman, 1986; McClelland & Rumelhart, 1981).

The present model also offers a framework for generating new testable predictions that relate to important issues in music perception. A first prediction pertains to the loci of key identification in long harmonic contexts. This issue is a matter of debate for both music and cognitive theories (Brown, 1988; Butler, 1989; A. J. Cohen, 1991; Thompson & Cuddy, 1989; Vos & Van Geenen, 1996). The model enables us to investigate this question in a way that may not be anticipated by those theories. For example, degrading the harmonic context is likely to affect the identification of the key, just as degrading a visual stimulus affects its identification (e.g., Abdi, Valentin, Edelman, & O'Toole, 1995; B. Edelman, Valentin, & Abdi, 1998, for faces; Biederman, 1987; Biederman & Cooper, 1991, for objects). If tonally important notes are eliminated from the sequences, the difference in processing of the target would diminish or vanish. Simulations will permit us to determine which tones of the context are responsible for establishing the tonality and for creating the observed difference in activation of the target.

Two further predictions are linked to the model's simulation of dynamic aspects of harmonic processing such as the influence of a previous key on the processing of events occurring later in the piece. Theories of tonal music cognition make the strong assumption that listeners keep the first main key in memory throughout the entire piece, even if the piece modulates to another key (Cook, 1987, 1990; Lerdahl & Jackendoff, 1983; Marvin & Brinkman, 1999; Meyer, 1956, 1973; Schenker, 1935/1979). Accordingly, the processing of an event should be influenced by the first instilled main key and the local key. It seems likely that the memory trace of the main key depends both on the harmonic relationships between main key and intervening local key and on the time elapsed since the first key. Bigand and Parncutt (1999) reported that listeners tend to lose the feeling of the main key rapidly after a distant modulation has occurred. The authors suggested that the short-term memory span "might increase when musical sequences modulate smoothly to closely related keys" (p. 252). The connectionist model addresses this suggestion by predicting the strength of the main key as a function of both harmonic distance and elapsed time. In the model, activations of both main key and intervening key are accumulated and are weighted by a decay that increases with elapsed time. A preliminary study along these lines was conducted by Bigand et al. (1999, Experiment 3).

In a related vein, the model predicts the effect of an interfering unrelated event (such as an erroneous note) on the processing of later events. Every musician would agree that such an unrelated event would more or less disturb the listening process, but neither music nor cognitive theories formalize the strength of this disruption. The connectionist model predicts that the interference caused by an unrelated event is a function of its harmonic incongruity, the time elapsed since its occurrence, and the tempo of the sequence (see Tillmann, Bigand, & Pineau, 1998, for a preliminary empirical investigation of this issue). In the three examples, the connec-

tionist model proposes a formal account making precise predictions about aspects of music perception that are important from both cognitive and music theoretical perspectives.

The present study has focused on how Western regularities in the pitch dimension may be internalized through passive learning processes. Although pitch is the most obvious form-bearing dimension of Western tonal music, regularities in other musical dimensions might also contribute to listeners' perceptual experience and may be internalized through similar processes. Beyond pitch, time is a crucial form-bearing dimension in music. Temporal regularities include the organization of event-onset intervals through time leading to a sensation of meter, a sensation of a regular succession of strong and weak beats superimposed over an isochronous pulse. Temporal regularities also include the temporal patterns of onset intervals creating rhythms that are perceived against a metrical background. To date, psychological accounts of music perception have typically treated pitch and temporal structures separately. Models of pitch have been proposed without regard to the potential influence of meter or rhythm (Deutsch & Feroe, 1981; Lerdahl, 1988; Longuet-Higgins, 1978; Krumhansl, 1990; Shepard, 1982) and vice versa (Fraisse, 1974; Longuet-Higgins & Lee, 1984; Povel, 1981). A growing number of researchers have nevertheless attempted to investigate the mutual contribution of pitch and time from both an empirical and theoretical perspective (e.g., Jones, 1976; Jones & Boltz, 1989; Lerdahl & Jackendoff, 1983). Temporal regularities have been shown to influence the perception of musical events in many ways, including performance in recognition tasks (Bigand & Pineau, 1996; Boltz, 1993), recall (Boltz, 1991), completion judgments (Boltz, 1989a, 1989b; Laden, 1994; Palmer & Krumhansl, 1987a, 1987b), evaluations of musical tension (Bigand, 1993, 1997), and musical expectations. For instance, a listener familiar with Western tonal-metric music will develop expectations about the "what" and the "when" of the events (Jones & Boltz, 1989), resulting in greater facilitation for expected events that occur at the expected time (Bigand et al., 1999, Experiment 3; Schmuckler & Boltz, 1994).

From this standpoint, the present model gives an incomplete account of music perception because it focuses on pitch regularities. The model is not atemporal but rather quasitemporal. Pitch information is integrated over time, as governed by the temporal persistence of activation. Persistence is controlled by the decay rate of a neural unit, thereby determining the temporal window (Bharucha, 1999). The model is, therefore, sensitive to some sequential information, encoding duration by exponential decay. Furthermore, activation is phasic, making the model sensitive to event onsets, which are the defining moments in meter and rhythm. The model thus goes further in incorporating time than do most models of musical pitch, which are atemporal. Nevertheless, meter and rhythm are missing, requiring further developments of the model. We have proceeded as we have in this article on the grounds that, although pitch and time may interact, we know of no findings pointing to interactions wherein temporal structure changes qualitatively the hierarchies of tonality established by pitch structures.

The question of how pitch and time work in combination arises not only for perception and memorization but also for the implicit learning of regularities underlying environmental events. In the research domain of implicit learning, few studies have analyzed learning processes with stimuli that vary on two simultaneous dimensions. Mayr (1996) studied the learning of experimental stimuli that vary independently on two dimensions: spatial regularities (object location) and nonspatial regularities (object type). The results reveal that two independent and relatively complex sequences can be learned. The author concluded that learning is not a unitary process but that several at least partially independent systems are included in the learning of spatial and nonspatial regularities. In light of this result, the question arises as to what type of processes mediate the learning of material like music that follows regularities displayed on several more or less independent dimensions such as time and pitch: Do listeners learn the regularities developed on the two dimensions independently or in an integrated manner? In conclusion, we would like to suggest some avenues for investigating this issue in the music domain through neural net computing.

Two frameworks have been distinguished to account for the respective contributions of temporal and nontemporal regularities in the perception of musical events (Peretz & Kolinsky, 1993). On the one hand, Jones and Boltz (1989) presented a single-component model in which the temporal accents and the harmonic accents are integrated in "a joint accent structure" that guides the attention of the listener during the ongoing musical piece. The tonal and temporal structures are not processed independently, and the processing of one dimension interferes with the processing of the other dimension. As a consequence, the modification of one dimension influences irrepressibly the processing on the other dimension (Bigand, 1997; Bigand & Pineau, 1996; Boltz, 1991, 1993). On the other hand, Peretz and Kolinsky (1993) described some independence between the processing of temporal and nontemporal information, both types of information being integrated at a later stage. Empirical support for a two-component model comes from neuropsychological data (with the documentation of a double dissociation, that is amelodia without arrhythmia and vice versa), and a large body of experimental findings (Palmer & Krumhansl, 1987a, 1987b; Peretz, 1990; Peretz & Kolinsky, 1993; Peretz & Morais, 1989).

Two types of neural net architecture may be potentially distinguished to account for the learning of Western musical structures that would mimic these two theoretical models. Following a single-component model (i.e., an interactive-integration model), a neural net should learn the two regularities conjointly without a supplementary integrative step. A possible solution consists of adapting the input codings proposed by Berger and Gang (1997, 1998) or by Stevens and Wiles (1994). In the work of Berger and Gang (1997, 1998), the input units of a sequential net are tuned to the 12 tones of the chromatic scale, and seven supplementary units code the metrical position of a chord. For example, three units code a meter with three beats and four units code a meter with four beats. When a chord is presented to the network, one of these "metrical" units is activated (e.g., the first unit is activated when the event arrives on the first beat of a triple meter). Stevens and Wiles (1994) added to the input layer units that code either the duration of an event or its metrical accent (strong, weak). Although these types of coding are relatively abstract, they mimic the conjoint learning of the two dimensions. Following a two-component model (independent processes followed by integration), the learning of temporal and nontemporal regularities would be separated during the training session. The SOM model described previously might learn the Western pitch regularities independently of another

network model designed to extract temporal regularities from rhythmic patterns, as does the model proposed by Large and Kolen (1994). Following Large and Kolen (1994), the output of the temporal model may be interpreted as cycles of attention or pulses of attention. The two models would learn separately the harmonic information and the temporal information. Their respective output would then be integrated to model the development of expectation; for example, the oscillatory cycles might reinforce or weaken the activation of events in the harmonic model. The presently discussed extension of the integration of the temporal dimension might provide additional insight into learning processes and perception of the two dimensions in music.

References

Abdi, H., Valentin, D., Edelman, B. E., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception, 24,* 539–562.

Altmann, G. T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 899–912.

Bartlett, J. C., & Dowling, W. J. (1980). The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 501–515.

Berger, J., & Gang, D. (1997, September). *A neural network model of metric perception and cognition in the audition of functional tonal music.* Paper presented at the 1997 International Computer Music Conference, Thessaloníki, Greece.

Berger, J., & Gang, D. (1998, October). *A computational model of meter cognition during the audition of functional tonal music.* Paper presented at the 1998 International Computer Music Conference, Ann Arbor, Michigan.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16,* 485–518.

Bharucha, J. J. (1987a). MUSACT: A connectionist model of musical harmony. In *Program of the ninth annual conference of the Cognitive Science Society* (pp. 508–517). Hillsdale, NJ: Erlbaum.

Bharucha, J. J. (1987b). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5,* 1–30.

Bharucha, J. J. (1991). Pitch, harmony and neural nets: A psychological approach. In P. Todd & G. Loy (Eds.), *Music and connectionism* (pp. 84–99). Cambridge, MA: MIT Press.

Bharucha, J. J. (1992). Tonality and learnability. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 213–223). Washington, DC: American Psychological Association.

Bharucha, J. J. (1994). Tonality and expectation. In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 213–239). Oxford, England: Oxford University Press.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13,* 383–400.

Bharucha, J. J. (1999). Neural nets, temporal composites and tonality. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 413–441). San Diego, CA: Academic Press.

Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13,* 63–102.

Bharucha, J. J., & Mencl, W. E. (1996). Two issues in auditory cognition: Self-organization of octave categories and pitch-invariant pattern recognition. *Psychological Science, 7,* 142–149.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 403–410.

Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception and Psychophysics, 41,* 519–524.

Bharucha, J. J., & Todd, P. (1989). Modeling the perception of tonal structures with neural nets. *Computer Music Journal, 13,* 44–53.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393–419.

Bigand, E. (1993). The influence of implicit harmony, rhythm and musical training on the abstraction of "tension–relaxation schemes" in a tonal musical phrase. *Contemporary Music Review, 9,* 128–139.

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm and musical expertise. *Journal of Experimental Psychology: I Human Perception and Performance, 21,* 808–822.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 184–197.

Bigand, E., & Parncutt, R. (1999). Perception of musical tension in long chord sequences. *Psychological Research, 62,* 237–254.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics, 58,* 125–141.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Current Psychology of Cognition, 17,* 577–600.

Bigand, E., & Pineau, M. (1996). Context effects on melody recognition: A dynamic interpretation. *Current Psychology of Cognition, 15,* 121–134.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59,* 1098–1107.

Bigand, E., Tillmann, B., Manderlier, D., & Poulain, B. (2000). *Harmonic priming versus chord repetition effect in music.* Manuscript in preparation.

Boltz, M. (1989a). Perceiving the end: Effects of tonal relationships on melodic completion. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 749–761.

Boltz, M. (1989b). Rhythm and "good endings": Effects of temporal structure on tonality judgments. *Perception and Psychophysics, 46,* 9–17.

Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition, 19,* 239–251.

Boltz, M. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception and Psychophysics, 53,* 585–600.

Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimension? *Journal of New Music Research, 28,* 67–79.

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception, 1988, 5,* 219–250.

Brugge, J. F., & Reale, R. A. (1985). Auditory cortex. In A. Peters & E. G. Jones (Eds.), *Cerebral cortex: Association and auditory cortices* (pp. 229–271). New York: Plenum.

Budge, H. (1943). *A study of chord frequencies* (Contributions to Education No. 882). New York: Columbia University, Teacher College.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal of intervalic rivalry. *Music Perception, 6,* 219–242.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing.* Cambridge, MA: MIT Press.

Cleeremans, A. (1994). The representation of structure in sequence pre-

diction tasks. In C. Umiltà & M. Moscovithc (Eds.), *Attention and performance XV: Conscious and non-conscious information processing* (pp. 783–809). Cambridge, MA: MIT Press.

Cohen, A. J. (1991). Tonality and perception: Musical scales primed by excerpts from the well tempered clavier of J. S. Bach. *Psychological Research, 28,* 255–270.

Cohen, A. J. (1994). Musical training and pitch schemata: Tonal inferences in micro-tonal contexts. In *Proceedings of the Third International Conference for Music Perception and Cognition* (pp. 81–82). Liège, Belgium: ESCOM.

Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America, 98,* 862–879.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception, 5,* 197–206.

Cook, N. (1990). *Music, imagination and culture.* Oxford, England: Clarendon Press.

Croonen, W., & Houtsma, A. (1994). Recognition of tone series: Effects of tonal structure, inference and training. In *Proceedings of the Third International Conference for Music Perception and Cognition* (pp. 247–248). Liège, Belgium: ESCOM.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 869–883.

Cuddy, L. L., Cohen, A. J., & Miller, L. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology, 33,* 148–157.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception and Psychophysics, 57,* 451–462.

Cuddy, L. L., & Thompson, W. F. (1992a). Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Psychological Research, 54,* 51–59.

Cuddy, L. L., & Thompson, W. F. (1992b). Perceived key movement in four-voice harmony and single voices. *Music Perception, 9,* 427–438.

Deliège, C. (1984). *Les fondements de la musique tonale.* Paris: J. C. Lattès.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Deutsch, D. (1972). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93,* 156–162.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

DeWitt, L. A., & Samuel, A. G. (1990). The role of knowledge-based expectation in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General, 119,* 123–144.

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science, 16,* 41–79.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 875–887.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341–354.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception and Psychophysics, 50,* 305–313.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* Orlando, FL: Academic Press.

Edelman, B., Valentin, D., & Abdi, H. (1998). Sex classification of face areas: How well can a linear neural network predict human performance? *Journal of Biological Systems, 6,* 241–264.

Edelman, S., & Intrator, N. (1997). Learning as extraction of low-dimensional representations. In R. L. Goldstone, D. L. Medin, & P. G. Schyns (Eds.), *Perceptual learning* (pp. 353–379). San Diego, CA: Academic Press.

Elman, J. L., & McClelland, J. L. (1984). The interactive activation model of speech perception. In N. Lass (Ed.), *Language and speech* (pp. 337–374). New York: Academic Press.

Fraisse, P. (1974). *La psychologie du rythme.* Paris: PUF.

Francès, R. (1958). *La perception de la musique.* Paris: Vrin.

Gjerdingen, R. O. (1990). Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception, 8,* 339–370.

Griffith, N. (1994). Development of tonal centers and abstract pitch as categorizations of pitch use. *Connection Science, 6,* 155–175.

Grossberg, S. (1970). Some networks that can learn, remember and reproduce any number of complicated space-time patterns. *Studies in Applied Mathematics, 49,* 135–166.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121–134.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11,* 23–63.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* Englewood Cliffs, NJ: Prentice Hall.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hébert, S., Peretz, I., & Gagnon, L. (1995). Perceiving the tonal ending of tune excerpts: The roles of pre-existing representation and musical expertise. *Canadian Journal of Experimental Psychology, 49,* 193–209.

Helmholtz, H. L. von. (1954). *On the sensations of tone as a physiological basis for the theory of music.* (A. J. Ellis, Trans.). London: Longmans, Green. (Original work published 1885)

Holleran, S., Jones, M. R., & Butler, D. (1995). Perceiving implied harmony: The influence of melodic and harmonic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 737–753.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology, 160,* 106–156.

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12,* 154–157.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review, 83,* 323–355.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491.

Kaas, J. H. (1995). The reorganization of sensory and motor maps in adult mammals. In M. S. Gazzaniga (Ed.), *The cognitive neuroscience* (pp. 51–71). Cambridge, MA: MIT Press.

Kohonen, T. (1995). *Self-organizing maps.* Berlin: Springer.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11,* 346–374.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* Oxford, England: Oxford University Press.

Krumhansl, C. L., Bharucha, J. J., & Castellano, M. (1982). Key distance effects on perceived harmonic structure in music. *Perception and Psychophysics, 32,* 96–108.

Krumhansl, C. L., & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89,* 334–368.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectancy in Finnish spiritual hymns: Convergence of statistical, behavioral and computational approaches. *Music Perception, 17,* 151–195.

Laden, B. (1994). Melodic anchoring and tone duration. *Music Perception, 12,* 199–212.

Laden, B., & Keefe, D. H. (1991). The representation of pitch in a neural

net model of chord classification. In P. Todd & G. Loy (Eds.), *Music and connectionism* (pp. 64–83). Cambridge, MA: MIT Press.

Lampinen, J., & Oja, E. (1992). Clustering properties of hierarchical self-organizing maps. *Journal of Mathematical Imaging and Vision, 2,* 261–272.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6,* 177–208.

LeCun, Y., Denker, J. S., & Solla, S. A. (1990). Optimal brain damage. In *Advances in neural information processing systems* (Vol. 2, pp. 598–605). San Mateo, CA: Morgan Kaufmann.

Leman, M. (1995). *Music and schema theory.* Berlin: Springer.

Leman, M., & Carreras, F. (1997). Schema and gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), *Music, gestalt, and computing* (pp. 144–168). Berlin: Springer.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5,* 315–345.

Lerdahl, F. (1996). Calculating tonal tension. *Music Perception, 13,* 319–363.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Levinson, J. (1997). *Music in the moment.* Ithaca, NY: Cornell University Press.

Longuet-Higgins, H. (1978). The perception of music. *Interdisciplinary Science Review, 3,* 148–156.

Longuet-Higgins, H., & Lee, C. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1,* 424–441.

Marvin, E. W., & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception, 16,* 389–408.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 350–364.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 86,* 287–330.

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

Meyer, L. B. (1973). *Explaining music: Essay and exploration.* Berkeley: University of California Press.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (in press). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General.*

Page, M. A. (1994). Modeling the perception of musical sequences with self-organizing neural networks. *Connection Science, 6,* 223–246.

Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 116–126.

Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception and Psychophysics, 41,* 505–518.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science, 246,* 486–488.

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the roots of a musical chord. *Music Perception, 6,* 65–94.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach.* Berlin: Springer.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10,* 717–733.

Peretz, I. (1990). Processing of local and global musical information in unilateral brain-damaged patients. *Brain, 113,* 1185–1205.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A,* 301–327.

Peretz, I., & Morais, J. (1989). La musique et la modularité. In S. McAdams & I. Deliège (Eds.), *La musique et les sciences cognitives* (pp. 393–414). Bruxelles, Belgium: P. Mardaga.

Petroni, N. C., & Tricarico, M. (1997). Self-organizing neural nets and the perceptual origin of the circle of fifths. In M. Leman (Ed.), *Music, gestalt, and computing* (pp. 169–180). Berlin: Springer.

Piston, W. (1978). *Harmony* (4th ed.). New York: Norton.

Platt, J., & Racine, R. (1994). Detection of implied harmony changes in triadic melodies. *Music Perception, 3,* 243–264.

Povel, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.

Regnault, P., Bigand, E., & Besson, M. (in press). Event-related brain potentials show top-down and bottom-up modulations of musical expectations. *Journal of Cognitive Neuroscience.*

Rosner, B., & Narmour, E. (1992). Harmonic closure: Music theory and perception. *Music Perception, 9,* 383–412.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. *Psychological Review, 89,* 60–94.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9,* 75–112.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning. *Psychological Science, 8,* 101–105.

Sano, H., & Jenkins, B. K. (1991). A neural network model for pitch perception. In P. Todd & G. Loy (Eds.), *Music and connectionism* (pp. 42–49). Cambridge, England: MIT Press.

Schenker, H. (1979). *Free composition* (E. Oster, Trans.). New York: Longman. (Original work published 1935)

Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception and Psychophysics, 56,* 312–325.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–169.

Setiono, R. (1997). A penalty function approach for pruning feedforward neural networks. *Neural Computation, 9,* 185–204.

Shepard, R. N. (1982). Geometrical approximations to the structure of pitch. *Psychological Review, 89,* 305–333.

Shmulevich, I., & Yli-Harja, O. (1999). Localized key-finding: Algorithms and applications. In *Proceedings of the Expert Workshop on Tonality Induction,* (pp. 109–120). Nijmegen, the Netherlands: University of Nijmegen.

Sloboda, J. (1985). *The musical mind: The cognitive psychology of music.* Oxford, England: Oxford University Press.

Stevens, C., & Wiles, J. (1994). Tonal music as a componential code: Learning temporal relationships between and within pitch and timing components. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in neural information processing systems* (Vol. 6, pp. 1085–1092). San Francisco, CA: Morgan Kaufmann.

Stoeckig, K. (1990). *The effects of octave transformations: Evidence from harmonic priming concerning the functional equivalence of chords and their inversions.* Unpublished doctoral thesis, Dartmouth College, Hanover, New Hampshire.

Taylor, I., & Greenhough, M. (1994). Modeling pitch perception with adaptive resonance theory artificial neural networks. *Connection Science, 6,* 135–154.

Tekman, H. G., & Bharucha, J. J. (1992). Time course of chord priming. *Perception and Psychophysics, 51,* 33–39.

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 252–260.

Terhardt, H. (1974). Pitch, consonance, harmony. *Journal of the Acoustical Society of America, 55,* 1061–1069.

Thompson, W. F., & Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception, 7,* 151–168.

Tillmann, B., & Bigand, E. (1999). *Global relatedness effect in normal and scrambled chord sequences.* Manuscript submitted for publication.

Tillmann, B., Bigand, E., & Madurell, F. (1998). Influence of global and local structures on the solution of musical puzzles. *Psychological Research, 61,* 157–174.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of local and global context on harmonic expectancy. *Music Perception, 16,* 99–118.

Von der Malsberg, C. (1973). Self-organizing of orientation sensitive cells in the striate cortex. *Kybernetic, 14,* 85–100.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception, 14,* 185–224.

Weinberger, N. M., & McKenna, T. M. (1988). Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception. *Music Perception, 5,* 355–389.

Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping, 5,* 18–25.

Winter, B., & Reber, A. S. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 114–145). San Diego, CA: Academic Press.

Zwisterlood, P. (1990, November). *Speech segmentation during the first minutes of second-language acquisition.* Paper presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans.

Received October 11, 1999
Revision received February 9, 2000
Accepted February 22, 2000 ■