Implicit Learning of Regularities in Western Tonal Music by Self-Organization

Barbara Tillmann*, Jamshed J. Bharucha* & Emmanuel Bigand**
*Dartmouth College, Hanover **Université de Bourgogne, Dijon
Barbara.Tillmann@dartmouth.edu
Jamshed.J.Bharucha@dartmouth.edu
Emmanuel.Bigand@u-bourgogne.fr

Abstract

Western tonal music is a highly structured system whose regularities are implicitly learned in everyday life. A hierarchical self-organizing network simulates learning of tonal regularities by mere exposure to musical material. The trained network provides a parsimonious account of empirical findings on perceived tone, chord and key relationships and suggests activation as a unifying mechanism underlying a range of cognitive tasks.

1. Introduction

Implicit learning is the acquisition of knowledge in an incidental manner without complete verbalizable knowledge of what is learned [39]. It has been investigated in the laboratory with artificial material based on statistical regularities. For example, letter strings have been generated by finite state grammars on the basis of restricted sets of letters. After passive exposure to grammatical strings, participants differentiated better than chance new grammatical strings from nongrammatical ones. Most were unable to explain the rules underlying the grammar in verbal free reports [1, 16, 35]. These results have been extended to more complex material attempting to bridge implicit learning of artificial grammars and environmental event sequences. In [37], for example, participants became sensitive to regularities in artificial language-like auditory sequences based on the transition probabilities of syllables.

Western tonal music is an example of a highly structured system in our environment. It constitutes a constrained system of regularities (i.e., regularities of co-occurrence, frequency of occurrence and psychoacoustic regularities) based on a limited number of elements. Experimental studies in music cognition have provided indirect evidence that listeners acquire implicit knowledge about these regularities through mere exposure. As the number of opportunities to listen to musical pieces obeying this system of regularities is so great in everyday life it is plausibly learned by the same implicit processes as investigated in the laboratory. The present paper reviews a neural network that simulates implicit learning of tonal regularities via mere exposure to musical material through self-organization. Once learning has occurred, the network in combination with a spreading activation mechanism accounts for a range of empirical findings on music perception. Before presenting the self-organizing network, basic regularities of the Western tonal system and some empirical findings on music perception are reviewed.

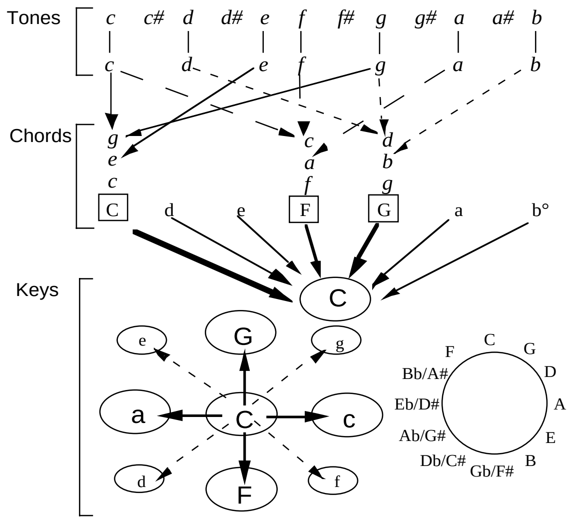

Figure 1: Schematic representation of the three levels in the tonal system using the example of C Major. Top) the 12 chromatic tones, followed by the diatonic scale in C Major. Middle) construction of three major chords, followed by the chord set forming the C Major key. Bottom) relations of C Major with close major and minor keys (left) and with all major keys forming the circle of fifths (right). (Tones are represented in italics, minor and major chords/keys in lower and upper case respectively)

2. Regularities in Western Tonal Music

The Western tonal system may be conceived of as a three-level hierarchical grammar that generates strong regularities in musical pieces (Fig. 1). It is based on a set of 12 pitch classes, which are organized in subsets of seven (defining major and minor diatonic scales). For each note of a scale, chords (e.g., major or minor) are constructed by adding two notes – creating a second order of musical units. In the C major key for example, the chord C major is constructed by combining the notes E and G to the root note C. Based on tones and chords, keys define a third order of musical units. Keys sharing numerous tones and chords are said to be harmonically related. The strength of harmonic relationship depends on the number of shared events. For example, the *C* major key shares six tones and four chords with the *G* major key, but only one tone with the *F#* major key. In music theory, major keys are conceived spatially as a circle (i.e., the circle of fifths), with harmonic distance represented by the number of steps on the circle. Inter-key distances are also defined between major and minor keys. A major key (*C* major) is harmonically related to both its relative minor key (*a* minor) and its parallel minor key (*c* minor). The three levels of musical units occur with strong regularities of co-occurrence. Tones and chords belonging to the same key are more likely to co-occur in a musical piece than tones and chords belonging to different keys. Changes between keys 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).

Within each key, tones and chords have different tonal functions creating tonal and harmonic hierarchies. For example, chords built on the first, fifth, and fourth scale degrees (referred to as tonic, dominant and subdominant respectively) usually have a more central function than chords built on other scale degrees. From a psychological point of view, the hierarchically important events of a key act as stable cognitive reference points [24] to which others are anchored [2]. Interestingly, these within-key hierarchies are strongly correlated with the frequency of occurrence of tones and chords in Western musical pieces. Tones and chords used with higher frequency (and longer duration) correspond to events that are defined by music theory as having more important functions in a given key [12, 19, 25]. Finally, as based on a restricted set of events, the Western tonal system has the characteristic that functions of tones and chords depend on the established key context. For example, the *C* major chord functions as a stable tonic chord in a *C* major context, as a less stable dominant or subdominant chord in an *F* or *G* major contexts respectively and as an out-of-key chord in a D major context.

3. Perception of Tonal Relations by Musicians and Nonmusicians

The multilevel relations between tones and chords, chords and keys, and between keys define a complex set of possible relations between musical events [25, 30]. Despite its complexity, numerous experimental studies in music cognition have provided evidence that even nonmusician listeners without explicit musical training or theoretical background are sensitive to musical structures and functions. While musicians usually exhibit better performance than nonmusicians, the overall responses show the same pattern. The sensitivity to tonal hierarchy was found to be similar in groups with different musical expertise [15, 22] as was the sensitivity to harmonic relations of chords [4, 8] or the ability to detect modulation [13, 14, 42]. Harmonic priming studies showed that independently of listeners' musical expertise, the processing of a target chord was influenced by its harmonic relatedness with the previous context. This finding suggests that factors governing musical expectation are based on cognitive processes that do not require explicit knowledge of musical structure [6, 10, 43].

The overall pattern of results in music cognition research suggests that mere exposure to Western musical pieces suffices to develop implicit, but nevertheless sophisticated, knowledge of the tonal system. Just by listening to music in everyday life, listeners become sensitive to the structures of the tonal system without being necessarily able to verbalize them [19]. This implicit learning of tonal regularities may be viewed as an ecological validation of implicit learning analyzed in the laboratory [9].

Connectionist models allow the simulation of how knowledge may be learned through passive exposure and how this knowledge influences perception. The next section presents a connectionist model that learns tonal regularities via mere exposure and simulates experimental results on music perception via activation spreading through the learned representation.

4. Learning and Representing Tonal Knowledge

In the music domain, neural networks have been developed for selective aspects of music perception, e.g. pitch perception [5, 38, 40], melodic sequence learning [7, 20, 32] or the extraction of tonal centers [28, 29]. A model of tonal knowledge representation (named *MUSACT*, standing for *MUS*ic and *ACT*ivation) is proposed in [3], it is interesting for two features: a) the inclusion of the three organizational levels of the tonal system (tones, chords, keys) and b) the simulation of top-down influences through spreading activation. The general architecture of this network mirrors interactive-activation models of word recognition [31] and speech recognition [18] containing three levels of units (features, letters/phonemes, words). The simulation of top-down influences is an important characteristic of a knowledge model, for both language and 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 been present in the context [19, 25].

In *MUSACT*, tonal relations are not stored explicitly but emerge from activation reverberating between tone, chord and key layers. After reverberation, activation of chord units, for example, reflects the harmonic hierarchy of the key context, with higher activation for stable than unstable chords. *MUSACT* provides a framework for understanding how musical knowledge may be mentally represented and how this knowledge, once activated by a musical context, may influence the processing of tonal structure [6, 10]. However, it was conceived with music theoretic constraints and thus represents an idealized end-state of an implicit learning process.

Recent simulations modelled this implicit learning process via neural self-organization [44]. Through mere exposure to musical stimuli, the connection weights of a hierarchical self-organizing map adapt and internalize tonal regularities. The trained network in combination with a spreading activation mechanism simulates empirical data on perceived relations between and among tones, chords and keys. In the following, we briefly review the learning simulations and the simulations of empirical data (detailed presentation in [44]).

Principles of Self-Organization. As learning of tonal regularities presumably occurs without supervision, unsupervised learning algorithms that extract statistical regularities and encode events that occur often together [21, 23, 36] seem to be well suited to simulate implicit learning in music. The Self-Organizing Map (SOM) proposed in [23] is one unsupervised algorithm that creates topological mappings between input data and units of a map. Before learning, no particular organization exists among map units. With repeated presentation of the input data, the specialization of the map units takes place by competition among units. The unit that is best able to represent an input wins the competition and learns the representation even better by adapting its connection weights. The learning algorithm reinforces links coming from active input units and weakens links coming from inactive ones. The unit's response will be subsequently stronger for this same input and weaker for other patterns. In a similar way, other units specialize to respond to other input patterns. In SOM, learning is not restricted to the winning unit, but extended to its neighbor units. As neighbor units gradually specialize for similar inputs, the representation becomes topographically ordered on the map.

A Hierarchical Self-Organizing Map Learns Tonal Regularities. For the simulation of implicit learning of tonal regularities, a hierarchical self-organizing map was defined in order to allow the extraction of different levels of tonal organization. The input layer consisted of 12 units tuned to the 12 chromatic tones. The input layer was connected to a second layer map which in turn was connected to a third layer map (Fig. 2). Before training, all units between two layers were fully interconnected with weights initialized to random values. Four training simulations were realized with this network architecture. The network was trained with either simple harmonic material (i.e., set of chords belonging to a key) or more realistic short chord sequences. The input was defined by either a sparse coding (i.e., coding the presence of tones) or a psychoacoustically richer coding (i.e., including subharmonics of tones [33]). Training consisted of two phases, during which chords and groups of chords were presented repeatedly to the input layer.

In all four networks, units of the second layer became specialized in the detection of chords, and units of the third layer in the detection of sets of chords (keys). The maps were calibrated by naming each winning unit after the stimulus for which it won. For both maps, a topographic organization was observed (Fig. 2). In the second layer, neighbor units represented chords that share component tones[1]. In the third layer, neighbor units represented keys sharing tones and chords. The organization of key units reproduced the topology of the circle of fifth (cf. Fig. 1). During training, the initially randomized connections changed to reflect the regularities of occurrence between tones and between chords. With simple input coding, for example, 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 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.

The trained network was used as a feedforward system and as a reverberation system.[2] The feedforward activation consisted of bottom-up

---

[1] After training with rich coding, the chord layer showed a more global organization than just the differentiation based on shared pitches. The specialized units on one half of the map represent chords from one side of the circle of fifths, units on the other half the other side of the circle.

[2] As in [3], reverberation was defined by phasic activation spreading between units until equilibrium was reached. Equilibrium was defined to be reached when phasic activation was less than a threshold of .005 for each unit. For event sequences, activation due to each event is accumulated and weighted according to recency. The total activation of a unit is

information only: chord units' activation reflected the number of component tones shared with the input. After reverberation, the activation pattern changed qualitatively due to top-down influences of learned, schematic structures and reflected tonal relations. Across major chord units, for example, activation decreased monotonically with distance around the circle of fifths.

The different training material (sets of chords vs. chord sequences) and input coding (sparse vs. rich) impressed on the learned weights in the connection matrices. For example, for matrices learned with rich coding feedforward profiles had a shape closer to top-down profiles than for matrices learned with sparse coding. However, with reverberation top-down processes imposed a pattern of activation that was analogous for sparse and rich coding and overwrote influences of coding richness.

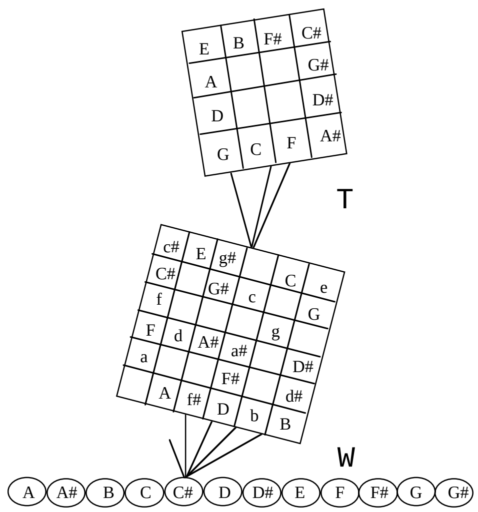

Figure 2: The hierarchical SOM: Input units code the 12 chromatic tones, the first map specialized in the detection of chords, the second map in the detection of keys. The represented maps are the result of a network trained with simple input coding and sets of chords. *W* and *T* refer to matrices with the learned connection weights between units of two layers (cf. text), with only some links represented for convenience.

In sum, the hierarchical self-organizing map extracted underlying regularities of musical material through mere exposure. Specialized representational units were formed for combinations of musical events (tones, chords) that occur with great regularity. Further analyses of activation patterns after reverberation indicated strong correlations between the four trained networks and the *MUSACT* model ($.984 < \underline{r} < .999$; $\underline{p} < .01$; $\underline{df} = 10$). This outcome suggests that the activation patterns of *MUSACT* that had been based on theoretical constraints can emerge automatically through self-organization.

---

thus the sum of three terms: the stimulus activation, the phasic activation accumulated during reverberation and the decayed activation due to previous events.

5. Simulation of Tonal Perception by Activation Spreading in Tonal Knowledge Representation

A crucial test for a neural net model of knowledge representation is to simulate the performance of human participants, in our case of Western tonal listeners in music perception studies. Simulations were run with the experimental material used in a set of empirical studies on the processing of tone, chord, and key relationships. The activation levels of network units 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. The hypothesis was that the level of tonal stability affects performance in the tasks; for example, a higher activated, more stable, event would be more expected or be judged as being more similar to the previous event. The simulations showed that the trained self-organizing network behaved much as human participants do in experiments on tonal perception. This outcome suggests that activation spreading in tonal knowledge representation can be seen as a unifying mechanism underlying a range of cognitive tasks.

<u>Perceived Relations between Chords.</u> The network simulated results of similarity judgments on chord pairs [27], recognition memory [4] and harmonic expectation. The development of harmonic expectation has been studied in harmonic priming [6, 10, 41] and neurophysiological studies [34]. It has been reported that the processing of a target chord is facilitated if harmonically related to the preceding local and/or global context, and that the level of harmonic incongruity between target and context is reflected in the amount of positivity of event-related potentials (ERPs). When the experimental material was presented to the network, activation levels of chord units mirrored facilitation patterns in the priming task (e.g., with higher activation for chord units representing facilitated targets), and activation changes in chord units due to the target mirrored the amount of positivity in the ERP waveforms (e.g. with stronger activation changes for distant key targets).

<u>Perceived Relations between Keys.</u> Further simulations showed that the network accounts for listeners' implicit knowledge of key distances and reproduces listeners' changing sense of key [13, 26]. For example, [13] showed that listeners distinguish sequences that modulate only one step in distance on the circle of fifths from sequences modulating two steps, with a stronger changing sense of key for counterclockwise than for clockwise modulations. After the presentation of modulating sequences to the network, the activation pattern of key units changed more strongly for a two-step than for a one-step modulation (compared to the initial key of the musical sequence). The network also simulated the perceived asymmetry in modulation: for each distance, counterclockwise modulations caused a stronger change than clockwise modulations.

<u>Perceived Relations between Tones.</u> Even if the network was not trained with melodies, it simulated results on the perception of tones. For example, once a key context is instilled, listeners perceive a hierarchy of stability between tones [26]. Simulations realized with a one-chord context showed that the activation received by tone units after reverberation reflected the differences between in-key and out-of-key tones of the context. Due to reverberation, the model made this

distinction also for tones that had not been present in the context. In addition to simulating similarity judgments on tone pairs [24], the network simulated melodic recognition memory [17]. The network did not learn the melody in itself, but it captured elements linked to tonal stability that influenced human participants in differentiating between targets and foils. The state of the network was, for example, more strongly correlated between items for which human participants showed lower performance.

6. Conclusion

The presented connectionist network simulates implicit learning of tonal regularities by self-organization and proposes activation spreading through tonal knowledge representation as mechanism underlying different cognitive tasks. Based on self-organization, the structure of the network adapted to tonal regularities through repeated exposure to tonal material. The trained network provides a low dimensional and parsimonious representation of tonal knowledge. For example, to account for the context dependency of the tonal system the network does not represent tones and chords several times, but contextual stability changes are reflected in activation levels. As a consequence, cognitive reference points (that also change with key contexts) are not stored separately, but emerge from activation. Finally, a further parsimonious feature is related to key induction. The network does not need to fall back on supplementary processes (e.g. template matching); rather, the underlying key emerges from activation in the key layer.

The network offers a framework for generating new testable predictions that relate to important issues in music perception. The experimental testing can be related to key identification (e.g., number of notes necessary to establish a key, disturbing effect of an unrelated event), to key modulation (e.g. how long the trace of a key remains) or to the eventual link between activation decay and musical short-term memory span. The self-organizing algorithm that allowed the network to learn underlying regularities conforms to principles of cortical information processing, such as the formation of spatial ordering in sensory processing areas. In auditory cortex, tonotopic organization has been shown among cells responding best to different frequencies [11, 45]. The outcome of the present simulations, together with aspects of cortical organization, leads to the question, whether there exist higher order maps such as a tonotopic organization of key centers.

Acknowledgments

This research was supported in part by a grant to B.T. from the Deutsche Akademische Austauschdienst DAAD, by grants to J.J.B. from the National Science Foundation (SBR-9601287) and NIH (2P50 NS17778-18), and by a grant to E.B. from the International Foundation for Music Research.

References

1. Altmann, G.T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: LMC*, *21*, 899-912.
2. Bharucha, J. J. (1984). Anchoring effects in music. *Cognitive Psychology*, *16*, 485-518.
3. Bharucha, J. J. (1987). Music cognition and perceptual facilitation. *Music Perception*, *5*, 1-30.
4. 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.
5. 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.
6. Bharucha, J.J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception & Psychophysics*, *41*, 519-524.
7. Bharucha , J.J., & Todd, P. (1989). Modeling the perception of tonal structures with neural nets. *Computer Music Journal*, *13,* 44-53.
8. Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics*, *59*, 1098-1107.
9. Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Current Psychology of Cognition*, *17*, 577-600.
10. Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Exp. Psychology: HPP*, *25*, 184-197.
11. Brugge, J. F., & Reale, R. A. (1985). Auditory cortex. In A. Peters & E. G. Jones, (Eds.), *Cerebral Cortex: Association and Auditory Cortices*. New York: Plenum Press.
12. Budge, H. (1943). *A study of chord frequencies*. Teacher College.
13. 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.
14. Cuddy, L. L., & Thompson, W. F. (1992b). Perceived key movement in four-voice harmony and single voices. *Music Perception*, *9*, 427-438.
15. Cuddy, L. L. & Badertscher, B. (1987). Recovery of tonal hierarchy: Some comparisons across age and levels of musical expertise, *Perception & Psychophysics*, *41*, 609-620.
16. Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: LMC*, *17*, 875-887.
17. Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341-354.
18. 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.
19. Francès, R. (1958). *La perception de la musique*. (2nd ed., 1984) Paris: Vrin.
20. Griffith, N. (1994). Development of tonal centers and abstract pitch as categorizations of pitch use. *Connection Science*, *6*, 155-175.
21. 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.

22. 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 Exp. Psy.*, *49*, 193-209.
23. Kohonen, T. (1995). *Self-Organizing Maps*. Springer: Berlin.
24. Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346-374.
25. Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: University Press.
26. 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.
27. Krumhansl, C. L., Bharucha, J. J., & Castellano, M. (1982). Key distance effects on perceived harmonic structure in music. *Perception & Psychophysics*, *32*, 96-108
28. Leman, M. (1995). *Music and Schema Theory*. Springer: Berlin.
29. Leman, M., & Carreras, F. (1997). Schema and Gestalt. (pp. 144-168), In: Leman, M. (Ed.) *Music, Gestalt, and Computing*. Springer: Berlin.
30. Lerdahl, F. (1988). Tonal Pitch Space. *Music Perception*, *5*, 315-345.
31. 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.
32. Page, M. A. (1994). Modeling the perception of musical sequences with Self-Organizing neural networks. *Connection Science*, *6,* 223-246.
33. Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the roots of a musical chord. *Music Perception*, *6,* 65-94.
34. Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music. *Journal of Cognitive Neuroscience*, *10*, 717-733.
35. Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855-863.
36. Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, *9*, 75-112.
37. Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning. *Psychological Science*, *8*, 101-105.
38. 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: MIT Press.
39. Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*, 163-169.
40. Taylor, I., & Greenhough, M. (1994). Modeling pitch perception with adaptive resonance theory artificial neural networks. *Connection Science*, *6*, 135-154.
41. Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: HPP*, *24*, 252-260.
42. 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.
43. Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of local and global context on harmonic expectancy. *Music Perception*, *16*, 99-118.
44. Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit Learning of Tonality: A Self-Organizing Approach. *Psychological Review. 107(4)*.
45. Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in human auditory cortex examined with fMRI. *Human Brain Mapping*, *5*, 18-25.