Connectionist simulation of tonal knowledge representation

Simulations connexionnistes des représentations des connaissances tonales

Barbara Tillmann

Université de Bourgogne LEAD-CNRS

Boulevard Gabriel

F - 21000 Dijon

tillmann@u-bourgogne.fr

The cognitive system of a listener is able to extract underlying regularities of a complex acoustic environment. Western tonal music is one example of a highly structured system that may be learned in an incidental manner. An adult listener, even without musical formation, has an implicit knowledge about the tonal system which is activated during the listening of music. Internalized representations of structural regularities generates musical expectations and facilitates the processing of harmonically related events. Connectionist models with unsupervised learning algorithms simulate the cognitive capacity to extract statistical regularities and encode events that often occur together. In the present paper, both a connectionist framework for the representation of tonal knowledge and for the simulation of perceptual learning in music are presented.

The cognitive system of a listener is able to extract underlying regularities of a complex acoustic environment. The acquired knowledge forms an implicit knowledge of the environmental world that influences perception and performance. Western tonal music is one example of a highly structured system that may be learned in an incidental manner. An adult listener, even without musical formation, has an implicit knowledge about the tonal system which is activated during the listening of music (Frances, 1958; Bharucha, 1984; Dowling & Harwood, 1986). The content and structure of this knowledge have been exhaustively investigated with different experimental tasks: probe tone techniques (Krumhansl, 1990; Hébert, Peretz, & Gagnon, 1995; Cuddy & Badertscher, 1987), recognition memory tasks (Bharucha & Krumhansl, 1983; Dowling, 1978; Deutsch, 1981), subjective scale judgments (Bigand, 1997; Schmuckler & Boltz, 1994), or harmonic priming (Bharucha & Stoeckig, 1987; Tekman & Bharucha, 1998). Different models of knowledge representation have been developed on the basis of experimental data (Krumhansl, 1990), music theory (Lerdahl, 1988, 1991) and by connectionist modeling (Bharucha, 1987).

The power of connectionist models lies in their capacity to learn statistical regularities of a structured environment by mere exposure and to propose distributed knowledge representations. In contrast to traditional rule-based accounts of knowledge, these models do not store explicit rules, but knowledge is stored in the connections linking the different units representing items which embody these rules. Learning involves the modification of the strength of these interconnections. The first part of the paper presents a connectionist model of tonal knowledge (Bharucha, 1987), this model was built on music theoretic constraints. The second part presents connectionist learning simulations of tonal knowledge by simple exposure to musical material with the help of unsupervised learning algorithms (i.e., Self-Organizing Maps of Kohonen, 1995).

Background: Connectionist Models in Music Perception

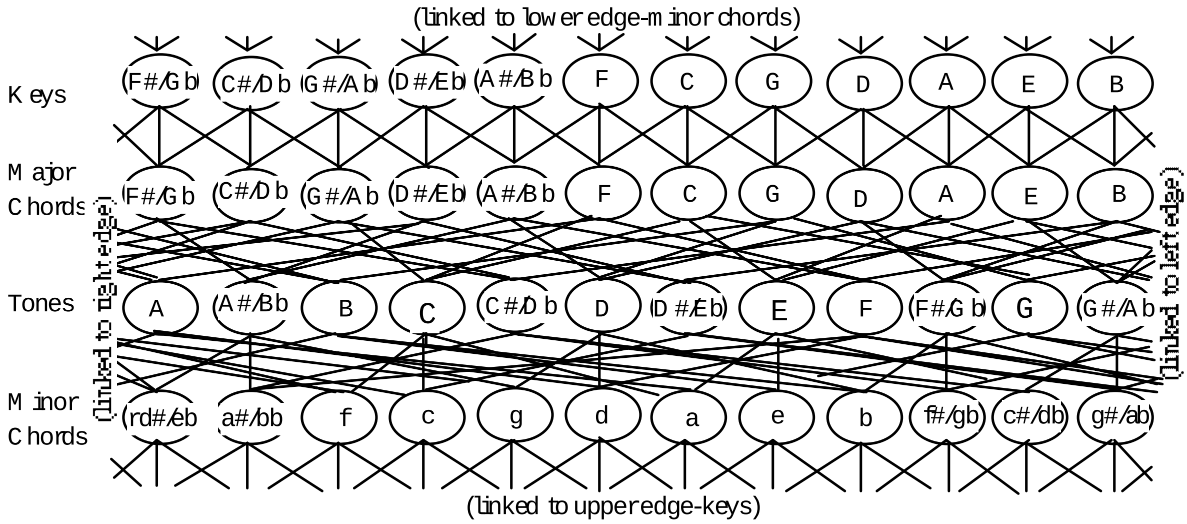
A connectionist model of Western tonal knowledge representation. Bharucha (1987) proposed a connectionist model of musical harmony, called MUSACT (musical activation). This model provides a framework for understanding how musical knowledge may be mentally represented and how this knowledge, once activated by a given musical context, may influence the processing of tonal structures. In this model, the neural net units are organized in three layers corresponding to tones, chords, and keys. Each tone unit is connected to the chords of which that tone is a component. Analogously, each chord unit is connected to key units representing keys of which it is a member. Western musical rules are not stored explicitly but emerge from activation that reverberates via connected links between tones, chords and keys. When three triadic tones are played, the units representing these tones are activated and phasic activation is sent toward the chord units. Phasic activation from the active chord units spreads towards the key units and starts to reverberate in the network. At equilibrium, the state of the network mirrors theoretical Western hierarchies. Activation tends to decrease with increasing harmonic distance between chords around the cycle of fifths. The level of activation in chord units is interpreted as the strength of expectations for further incoming chords - given the previously presented context.

The model also addresses the building up of harmonic expectancies over time. For chord sequences, activation due to each chord is accumulated. Once the model has reached equilibrium after an event, the pattern of activation begins to decay exponentially over time. If another event occurs before activation has decayed appreciably, the phasic activation due to 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: (1) the bottom-up activation caused directly by the stimulus itself (i.e., the tones) (2) the indirect activation received from other units in response to event *e* (i.e. the spreading activation), and (3) 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 an 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, $\Sigma \Delta a_{i,e,c}$ the total 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 following the offset of the last event, and *t* the time elapsed since the last offset.

The activations due to several chords are thus accumulated as the sequence unfolds, yielding an aggregate expectation for further incoming events. In this way, the model takes into account the development of expectations in long harmonic contexts.



<u>Figure 1.</u> Presentation of the MUSACT model (after Bharucha, 1987).

<u>Empirical support from harmonic priming data.</u> Support for this model was provided by empirical studies using harmonic priming paradigm for one chord primes (Bharucha & Stoeckig, 1986, 1987; Tekman & Bharucha, 1992) or longer contexts (Bigand & Pineau, 1997; Bigand, Madurell, Tillmann & Pineau, 1999; Tillmann, Bigand & Pineau, 1998). The rational of these studies is that a harmonic context generates expectancies and primes chords that are harmonically related to the context. The processing of harmonically related chords is facilitated and speeded up. The MUSACT model explains the development of these expectations via activation spreading through a network representing tonal knowledge. 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 by the context. After the presentation of a single prime chord to the model (say C major), activations of harmonically related target chord units (i.e., Bb major) were stronger than of unrelated ones (i.e., F# major). Empirical data with harmonic priming paradigm confirmed this prediction (Bharucha & Stoeckig, 1986, 1987; Tekman & Bharucha, 1992). In these studies, participants heard a prime chord followed by a target chord. The prime and target were either closely related (belong to the same key) or distantly related harmonically. On half of the trials, the target chord was slightly mistuned, and participants were asked to make a speeded intonation judgment, i.e., to decide as quickly as possible whether the target chord was in tune. The priming effect was shown by shorter response times for related targets. The activation pattern of chord units simulates harmonic expectations of human subjects and accounts for the facilitation of the processing of related chords. The harmonic priming effects had been extended to longer

contexts (Bigand & Pineau, 1997; Bigand et al, 1999; Tillmann et al, 1998). The target chord was the last chord of eight- or of fourteen-chord sequences. Expectations for the target were varied by changing the global harmonic context created by the chord sequence. The priming results are in accordance with predictions of the connectionist model, they reflect facilitation effects for the target chord if it is harmonically related to the global context.

The critical point. The model's connectionist representation of tonal knowledge is a powerful framework for understanding the influence of context on harmonic expectations. The fact that this model has received a good deal of support from empirical research suggests that a three layer spreading activation model may account for the way implicit knowledge of Western harmony is mentally represented. The main problem of this model, however, is to represent an idealized end-state of a perceptual learning process (Bharucha & Olney, 1989). The model was based on ad-hoc music theoretical constraints, and neither the connections nor their weights resulted from a learning process. It remains a crucial point to analyze how a representation of tonal knowledge can be learned by mere exposure to musical material. The power of connectionist models lies in their capacity to learn statistical regularities of a structured environment by mere exposure. Unsupervised learning mechanisms may extract underlying regularities of the tonal system, i. e., co-occurrence of notes in chords or of sets of chords in keys. The Self-Organizing Maps proposed by Kohonen (1995), represents one type of unsupervised learning algorithm that was used in the following simulation of perceptual learning of tonal music.

Connectionist Simulation of Perceptual Learning in Tonal Music.

Self-Organizing Maps. Unsupervised learning algorithms extract statistical regularities and encode events that often occur together (Grossberg, 1970, 1976; Kohonen, 1995; Rumelhart & Zipser, 1985; von der Malsberg, 1973). These algorithms seem to be very close to real music perception since no external teacher gives feed-back on the organization of chords or tonalities while listening to music in everyday live. One unsupervised learning algorithm is the Self-Organizing Map SOM proposed by Kohonen (1995). It creates topological mappings between the input data and neural net units of a map. For two similar input patterns, the responding map units are located near to each other. This learning algorithm is based on principles of cortical information processing, such as the formation of spatial ordering in sensory processing areas (i.e. somatosensory, visual and auditory).

The 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 takes place by competition among the net 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. The unit's response will be subsequently stronger for this same input pattern and weaker for other stimuli. In a similar way, other units learn to specialize in other input patterns.

The competitive learning algorithm can be generalized, if there exists an ordering between the units. On a Self-Organizing Map for example, the units are located on a discrete lattice. The generalization implies that not only the winning unit learns, but also that its neighbor units are 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 topological organization of the input data can be discovered on the map, such that similar input patterns activate nearby map units. SOM can be conceived of with one map layer or be adapted to multilayer hierarchical self-organizing maps (HSOM) (Lampinen & Oja, 1992)

Simulation of perceptual learning of Western harmony. A hierarchical self-organizing map will be defined for the learning simulation of tonal music. The hierarchical map was inspired by the hierarchy of feature detectors found in the brain. For auditory processing, hierarchies of feature detectors were suggested with elementary feature detectors at the sensory periphery (i.e. frequency), and more abstract feature detectors in the primary auditory cortex [i.e., pitch (Pantev, Hoke, Lütkenhöner, & Lehnertz, 1989) or contour (Weinberger & Mc Kenna, 1988)].

In the three layer hierarchical system, the input layer is tuned to the 12 chromatic scale tone units that represent octave-equivalent pitch categories. The second and third layer are self-organizing maps that learn to specialize in the detection of chords and keys, respectively. In the input layer, a more abstract coding than just frequency is chosen as it has been shown that neural net models can learn octave equivalent pitch classes (Bharucha & Mencl, 1996). The input unit is activated if the corresponding tone to which it is tuned occurs in the chord, and 0 otherwise. The units of the first and second layers are fully interconnected via a connection matrix; and the units of the second and third layers with a second connection matrix. Before learning, the strengths of all connections are initialized to random values.

In the simulations, the learning set is restricted to 24 chords (12 major and 12 minor chords) and 12 major keys. A major key is defined by a group of six chords (three minor and three major chords) presented to the input layer one by one without decay. The training patterns are presented in random order during each training cycle. Learning consists of two phases. In the first phase, the second layer is trained by the presentation of 24 chords (12 major chords, 12 minor chords) presented individually. In the second learning phase, the third layer is trained with sets of six chords representing major keys. For example the C major key is represented by the major chords C, F, and G, and the minor chords d, e, and a. These six chords are presented individually to the input layer. For each input chord, the best matching unit is chosen from the second-layer map and its index *b* is stored in memory until the end of the presentation of the chord set (Lampinen & Oja, 1992). The pattern of indexes *b* (without decay) defines the input for the training of the third layer. At the beginning of learning, the neighborhood radius is set to its maximum and decreases during training until it reaches 0, i. e. only the winner learns. The learning rate decreases over learning in parallel to the shrinking neighborhood radius (for details see Tillmann, Bharucha & Bigand, in preparation). All simulations were programmed in MATLAB[1].

Results. During training, units specialized for the detection of chords in the second layer and for the detection of sets of chords (referred to henceforth as 'keys') in the third layer. For both training phases, the weight changes decreased over the training cycles and with decreasing neighborhood. When weights had converged to practically stationary values, the maps were calibrated by naming each winning unit after the

[1] MATLAb 5.0 [computer software]. (1997, March). Natick, MA: MathWorks, Inc.

stimulus for which it had won. For example, the unit which win 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.

After training, the weights of the two connection matrices had no random values any longer and both matrices did not have the total interconnectivity defined before training. Each tone unit of the input layer had six connections to the winning units of the chord layer, i. e. to the six chords of which it is a part. Each chord unit was linked to three key units in the third layer, i.e. to the keys to which it belongs. The self-organizing algorithm changed the connections between neural net units in such a way that the units specialized in the detection of a chord or a set of chords. The outcome showed the formation of hierarchical encoding in which tones that often occur together are represented by common chord units and, similarly, chords that often occur together, are represented by key units.

The calibration phase reveals a topographic organization of stimuli represented on both maps (Figure 2). The relatedness between the input stimuli correlated with the distance between the representing units on the map. In the second layer, chord units are organized in a way that neighboring units share component tones. Chords that do not share tones are segregated and presented in different parts on the map. In the third layer, keys sharing chords and tones are represented by units close to each other on the map. Keys sharing only a few elements (chords, tones) were distant on the map. The organization of the specialized key units represents the cycle of fifths, a music theoretic concept. Musical distances between keys are represented on a circle with keys sharing all but one notes as neighbors.

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

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

Figure 2. The calibration maps of the second layer (left) and the third layer (right). For the second layer, winning units are labeled by chord names (minor chords in lower case letters, major chords in upper case letters). For the third layer, names of winning units indicate major keys.

In order to model the influences of knowledge and of context on perception, the neural net structure resulting from learning simulation was used with a spreading activation mechanism (e.g., Bharucha, 1987; Dell, 1986; McClelland & Rumelhart, 1981). After the presentation of a stimulus, activation reverberates between the three layers until an equilibrium is reached. The activation levels of the units reflect the underlying tonality of the context and the corresponding tonal hierarchies of the events (tones, chords).

The outcome of the presented simulation showed that the basic structure of the constraint-satisfaction model MUSACT that originally was proposed as an idealized end-state of a learning process, can be learned by mere exposure via self-organization. The learned matrices globally reflect the links predefined on the basis of music theory. Finally, the learned representation of tonal knowledge generates the same predictions as does MUSACT when the model is used as a feedforward and as a reverberation system (the activation profiles correlated strongly for the key layer and for the chord layer, r (10) = .999, p < . 01, respectively).

Discussion. The present simulation provided evidence that a representation of tonal knowledge can be learned by self-organization. The arising representation associated with a spreading activation process reflects human data on the development of harmonic expectations. The learned knowledge is entirely based on extracted underlying regularities, and no explicit rules were encoded. Without external feedback or supervision, the structure of the material to which the system is exposed to is learned in the connection matrices. As a consequence of these changed connections, specialized representational units are formed for combinations of musical events (tones, chords) that occur with regularity. Interestingly, the units in both maps (for chords and for keys) reveal a topographic organization. Units responding to similar stimuli (i.e. chords or groups of chords) are located in neighborhood on the map.

Self-organizing Maps had been used to model perceptual learning of timbre (Toiviainen, 1996) and of tonal centers (i.e, a concept related to tonality) (Leman, 1995; Leman & Carreras, 1998). In Leman and Carreras (1998) for example, the input signal derived from real sound recordings leads to the formation of tonal centers that are topologically organized and that are compared to music theory. The presented hierarchical SOM simulations lead to a representation of tonal knowledge that allows predictions on three structural levels, namely for keys, chords and tones. The use of the arising structure with a spreading activation model takes in consideration top-down influences of learned knowledge on expectancy formation and perception. Further simulations are in preparation that both extend the learning process to more complex ecological valid material and test the arising structures with empirical data reported in music cognition domain.

References.

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

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

Bharucha, J. J., & Krumhansl, C. (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. & Olney, K. L. (1989). Tonal cognition, artificial intelligence and neural nets, Contemporary Music Review, 4, 341-356.

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

Bharucha, J. J. & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? Perception and Psychophysics, 41, 519-24.

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm and musical expertise. Journal of Experimental Psychology, Human Perception and Performance, 23, 808-812.

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

Bigand, E., & Pineau, M. (1997). Context effects on musical expectancy. Perception and Psychophysics, 59, 1098-1107.

Cuddy, L. L. & Badertscher, B. (1987). Recovery of tonal hierarchy: Some comparisons across age and levels of musical expertise, Perception and Psychophysics, 41, 609-620.

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

Deutsch, D. (1981). The processing of pitch combinations. In D. Deutsch (Ed.), The Psychology of music. (pp. 271-316), New York: Academic Press.

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

Dowling, W. J., & Harwood, D. (1986). Music Cognition. New York: Academic Press.

Francès, R. (1958). La perception de la musique. Paris: Vrin, Transl. W. J. Dowling, (1988), The Perception of Music, Hillsdale, NJ: Erlbaum.

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 recording. I: Parallel development and coding of neural feature detectors. Biological Cybernetics, 23, 121-134.

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.

Kohonen, T. (1995). Self-Organizing Maps. Springer: Berlin.

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

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

Leman, M. (1995). Music and Schema Theory. Springer: Berlin.

Leman, M. & Carreras, F. (1998).Schema and Gestalt : Testing the Hypothesis of Psychoneural Isomorphism by Computer Simulation. In: Leman, M. (Ed.) Music, Gestalt, and Computing. Springer: Berlin, pp. 144-168.

Lerdahl, F. (1988a). Tonal Pitch Space, Music Perception, 5, 315-345.

Lerdahl, F. Pitch-space journeys in two Chopin Preludes. In: M. R. Jones & S. Holleran (Eds.), Cognitive bases of musical communication. (pp. 171-191), A.P.A., 1991.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. Psychological Review, 88, 375-407.

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

Rumelhart, D. F. & Zipser, D. (1985). Feature discovery by competitive learning. Cognitive Science, 9, 75-112.

Schmuckler, M. A. & Boltz, M. A. (1994). Harmonic and rhythmic influences on musical expectancy, Perception & Psychophysics, 56, 313-325.

Tekman, H. G., & Bharucha, J. J. (1992). Time course of chord priming. Perception & 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.

Tillmann, B., Bharucha, J.J. & Bigand, E. (in preparation). Perceptual learning in music: A connectionist framework.

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

Toiviainen, P. (1996). Musical Timbre: Optimizing auditory images and distance metrics for Self-Organizing timbre maps. Journal of New Music Research, 25, 1-30.

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

Weinberger, N. M. & Mc Kenna, T. M. (1988). Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception, Music Perception, 5, 355-590.