Don L. Scarborough, Ben O. Miller, and Jacqueline A. Jones
Department of Psychology
Brooklyn College of the City University of New York
Brooklyn, New York 11210 USA
BITNET:DOSBC@CUNYVM

Connectionist Models for Tonal Analysis

Introduction

When people listen to music from the Western tonal tradition, their sense of the music depends on perceiving the tonal structure. A piece is written in a particular key (though it may change), and the choice of key specifies a subset of 7 of the 12 notes of the chromatic scale as the primary elements of the composition. A key defines a schema for the individual notes of the scale and for chords within the key. Some notes and chords of a key are considered to be consonant and stable, while others are considered to be dissonant and unstable. Besides the tonic, several of the other six notes of the scale play especially prominent roles. For instance, the third and fifth notes of the scale, along with the tonic, define the tonic or root chord, which is considered to be the most stable chord within the key. Thus, a key defines a complex set of interrelationships among a subset of the notes of a chromatic scale.

This tonal structure plays a central role in a listener's perception in that individual pitch events and chords are heard within the framework or context provided by the tonal structure (Krumhansl 1979; Dowling and Harwood 1986). Furthermore, the organization of an entire piece is tied to this tonal structure, and the sense of harmonic development, progression, and resolution within this framework is a central aspect of the expressive quality of music. It is not clear, however, how listeners infer the key of a piece. A performer does not begin a piece by playing the appropriate scale or the tonic chord to define the tonal context. Rather, listeners must induce the key of the piece from the individual notes and chords as they are heard.

Computer Music Journal, Vol. 13, No. 3, Fall 1989,

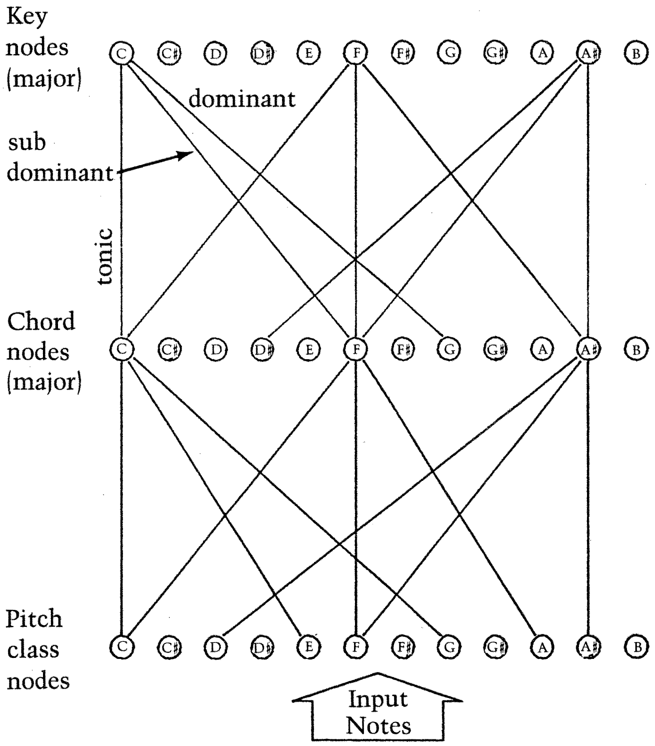
Induction of Tonality

Simon (1968) developed a computer program, called LISTENER, which attempted to identify tonality and phrasing from a single melodic line. Simon noted that the frequency with which various notes occurred provided a clue to the tonality. The LISTENER program therefore simply counted how often each note of the chromatic scale occurred. The program then counted how often the notes that formed the tonic chord (e.g., C-E-G in C major) of each key occurred. The chord with the highest note count established the tonality. In three of four test cases (Beethoven, Mozart, and Schumann) this algorithm correctly identified the key of the piece. In a fourth case (Brahms), the program's choice of key (A minor) disagreed with the key signature (E minor). However, a large portion of the Brahms's piece modulates to A minor, although it begins and ends in E minor.

LISTENER produces a single tonality judgment for a piece, an approach that has problems with modulations because it sums the note count over the entire piece. Another problem is that it is insensitive to the order in which notes occur. Deutsch (1984) and Bharucha (1984a; 1984b) point out that the order of a note sequence can influence the perceived tonality. This seems plausible from considerations of voice leading (Piston 1948), whereby an unstable note resolves by moving to a stable note. Thus, with the ascending sequence, D♯-E-F♯-G-B-C there is a tendency to hear D♯ as resolving to E, F♯ to G and B to C, yielding the notes E-G-C of the first inversion of the C-major tonic triad. In descending order, people hear the triad D♯-F♯-B and the tonality as either B major or E minor (Deutsch 1984).

A later effort to develop a computer program for musical analysis (Jones, Miller, and Scarborough 1988; Scarborough, Jones, and Miller 1988) based on Lerdahl and Jackendoff's generative theory of tonal

Fig. 1. Linear network for tonal induction.



music (1983) builds on Simon's approach. Jones, Miller, and Scarborough (1988) implemented a version of Simon's algorithm in terms of a simple connectionist network (Rumelhart and McClelland 1986). A network of interconnections was used as a simple method for counting notes and weighing the evidence for one key versus another. In this network, the input layer consists of pitch-class nodes that represent the notes of the chromatic scale independent of octave. These pitch-class nodes are connected to a layer of chord nodes, each of which receives input from just three pitch nodes. Finally, sets of three chord nodes—representing the tonic, subdominant, and dominant chords of a particular key— are connected to key nodes.

In the model, the occurrence of one or more notes in the music activates the corresponding pitch nodes. Activation then flows from the pitch nodes to each chord node that includes the active pitches. Finally, activation flows from the chord nodes to any key nodes for which those nodes are the tonic, subdominant, or dominant chords. At any point in the piece, the most active key node defines the perceived key.

The degree to which the activity at one node influences other nodes is determined by the weight between the nodes. This weight reflects the strength of the connection between nodes, which in turn reflects how much one node influences or depends on another node. A schematic diagram of the network for major keys is shown in Fig. 1. The diagram shows, for example, that at the key node level, the key node for C major is activated by a C chord (tonic), an F chord (subdominant) and a G chord (dominant). Other key nodes would be activated by the chords representing the tonic, subdominant, and dominant for those keys. A complete network contains additional nodes for minor chords and keys.

This network algorithm differs from Simon's in several respects. First, the amount of activation provided by an input note is proportional to the note's duration: i.e., a half-note has more influence than a quarter-note. Second, once a note stops, the activation of the corresponding pitch node does not stop immediately but rather decays with time. Similarly, the activations of chord and key nodes also decay with time. This decay of activation means that the activity of the network is influenced most by recent notes, but the overall pattern of activity represents a weighted sum over all notes that have occurred. The tonality at any point in a piece is taken to be specified by the most active of the key nodes at that point. The ratio of this key node's activity to the sum of the activities of all the key nodes provides a measure of the certainty associated with the choice of that key node as the tonality.

In general, this network does a better job than Simon's counting algorithm because more of the input is considered. The duration of a note, for instance, contributes to the activation of a node, and this reflects the fact that tonally important notes are often sustained. In addition, it can respond to modulations and is sensitive to note order because of the decay parameters. The performance of the network, however, depends critically on the weight and decay parameters. Thus, an important issue is how the best set of parameters can be selected. This is fairly easy in the case of this simple network because it is linear: the output at time t is simply a weighted sum of the inputs up to that time, where the weights include a decay parameter. For example, for a two-note sequence, if $o_i(2)$ is the output of key node unit i at time 2, then:

$$\begin{aligned} o_i(2) &= \sum_j w_{ij} \cdot o_j(2) + \beta \sum_j w_{ij} \cdot o_j(1) \\ &= \sum_j w_{ij} \cdot (o_j(2) + \beta o_j(1)), \end{aligned}$$

where w_{ij} is the strength of the connection or weight from chord node unit j to key node unit i, and β is the decay parameter. The output, o_j, of a chord node unit depends, in turn, on the sum of the inputs it receives from the pitch node layer, but this is also just a linear weighted sum of the same form. Because this network is linear, the middle chord layer adds no computational power in the determination of the key but exists only to represent chord perception. The output produced by the composition of two matrices of the same shape can be duplicated by a single matrix. Jordan (1986) provides a good discussion of basic linear algebra.

In principle, we can estimate the best weights by training the network using the Widrow-Hoff delta rule (Rumelhart, Hinton, and Williams 1986). However, this requires knowing what the network output should be at any given point in a piece, data that does not currently exist except at an intuitive level. What is needed is systematic data on perceived tonality to be able to provide good parameter estimates. Instead, parameters were picked largely based on intuitive guesses. Pitch node activation decays quickly, chord nodes more slowly, and key nodes slowest of all. In terms of the input evidence for a particular chord, the root is weighted most heavily, and the third and fifth somewhat less. For a key node, a tonic chord provides the strongest evidence, with the dominant and subdominant counting less. Other chords of a key could easily be included but were assigned zero weights in testing the algorithm. With these parameters, the network can identify the tonality of many pieces.

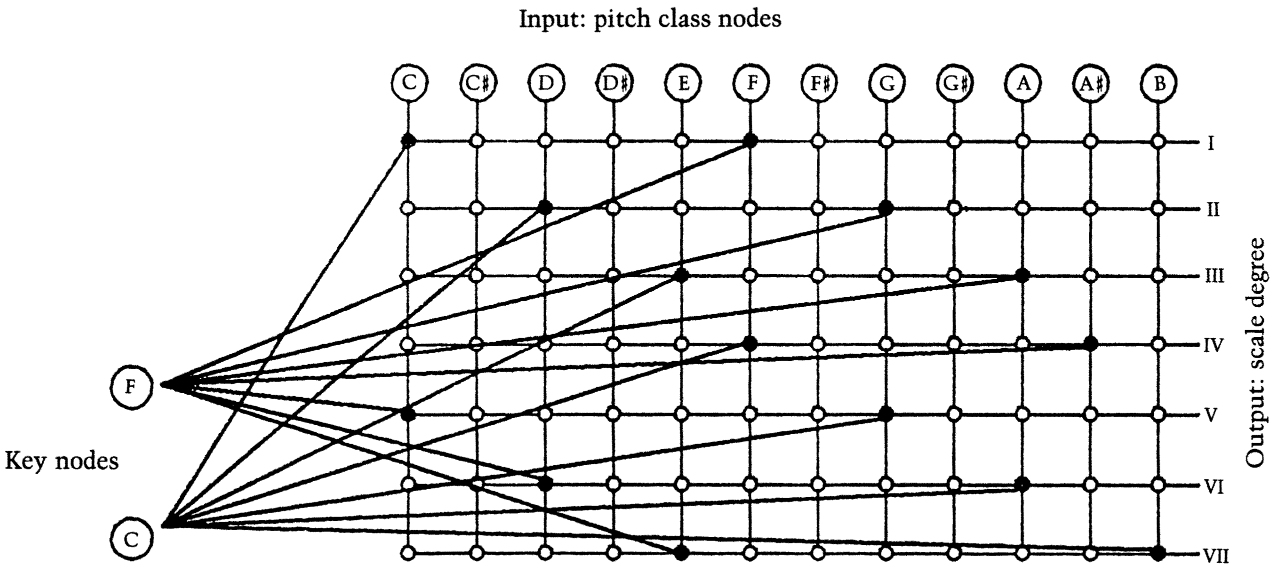
As an extreme example, it correctly guesses the key of *Auld Lang Syne* on the first note. The piece is in the key of F, and the first note is C, the fifth degree of the scale, an important note that occurs in both the tonic and dominant chords of the key of F. The particular choice of weights made F the most plausible key, a hypothesis that later gained confirmation. In most other cases, it takes only a few notes before the network can correctly identify the key.

The network can handle monophonic and polyphonic music and modulations. Because of the decay parameters, this network is also sensitive to note order and makes the tonal judgments described by Deutsch (1984) for ascending and descending sequences. In terms of many of the connectionist models appearing today (e.g., Grossberg 1980; 1988), this model is crude and simple. Nonetheless, the network does a more than creditable job in identifying the tonality of simple pieces of music. It is not clear how well this network simulates human performance because we know very little about how people identify tonality.

Bharucha (1987a; 1987b; 1988) has also proposed a connectionist network to solve the tonality induction problem. His network is similar in topology to the linear network described above, though it differs at the computational level. Bharucha uses *phasic* interactions so that a node responds only to changes in other nodes to which it is connected. Also, his network has both bottom-up and top-down connections. That is, while pitch nodes can activate chord nodes, which in turn can activate key nodes in a bottom-up sequence, a key node can prime chord nodes, which in turn can prime pitch nodes in a top-down fashion. A consequence of this architecture is that there are interactions between bottom-up and top-down effects, and it takes the network several simulation cycles to settle into a stable activation state in response to an input. Because of the top-down activation in his model, Bharucha can readily account for some chord priming effects on chord perception that he has found (Bharucha and Stoeckig 1987). Bharucha reports that his network exhibits many reasonable properties with respect to simulating tonal induction, although the simulation results have not been reported in detail. Bharucha has also proposed ways in which such a network might be tied into a larger model of music processing (Bharucha 1987b; 1988).

An even earlier network model was proposed by Deutsch (1969), which at first glance has several

Fig. 2. *Network to map pitch nodes into scale degrees based on key identification.*



similarities to the linear network model described above. For example, her chord analyzers are similar to the chord node array in the linear model. There are two main differences between her network and the linear network, however. First, her chord analyzers respond only to simultaneously sounded notes rather than note sequences, which the linear network does by integrating information over time. (The structure of the Deutsch network was designed to account for several other aspects of chord and interval perception and transposition.) Second, the Deutsch network has no provision for extracting the tonality of music.

A Network for Tonal Schemata

A problem with the network approaches described so far is that they fail to deal with some aspects of human music perception. For example, the networks do not explain how we recognize a familiar piece despite transposition. Another problem is that the networks do not explain how the definition of a key then provides a schema for the interpretation of the notes with respect to the key. The tonal induction networks described above provide a good basis for an extended network to address these problems, however. Bharucha (1988) has suggested a scheme similar to the one described below.

To place any note within a tonal schema, we need to identify that note with respect to its position within the scale defined by a particular key. If we use the network in Fig. 1 to identify the key of a piece, we can then use this information to add an additional layer to the network that will map individual notes onto scale degrees. The idea is that the key node that is most active will control the mapping of the notes. That is, if the key node for C major is most active, then a pitch of C should be mapped to the first degree of the scale, while if the F key node is most active, then C should be mapped to the fifth degree. The simple cross-bar switching matrix shown in Fig. 2 can do this.

In Fig. 2, the pitch-class nodes are shown along the top. Each of these pitch nodes feeds into this network (in parallel with their input into the tonality network in Fig. 1) and has a connection with each of seven output units representing the seven degrees of a scale. Which one of the seven connections is active is determined by key nodes that selectively gate these connections. For example, the C-major key node would gate the input from C to the first scale degree, from D to the second degree, and so on, while an F-major key node would gate D to the sixth degree. Only two of a complete set of key nodes are shown in Fig. 2.

This gating can be accomplished in several ways. One way is to use sigma-pi connections (Rumelhart and McClelland 1986) between the key nodes and the pitch nodes shown in Fig. 2. A second approach is to have an output threshold for each degree node so that an output is produced only if the sum of the input to it from the key and pitch nodes is large enough. For either of these two approaches to work well, it is necessary to have only one key node active at a time. For the simple linear tonality network described earlier, however, the output of a key node is proportional to the strength of the evidence for that key. Singling out one key node and disabling the others can be accomplished by letting the output of key nodes be a non-linear sigmoidal function of the input, and by adding inhibitory connections between key nodes. With these changes and with appropriate choice of parameters, we can get a "winner-take-all" (Grossberg 1988) output from the strongest key node.

Another approach to the gating problem is to have a processing unit at each of the intersections shown in Fig. 2 that receives input from the key node and the pitch node. Each such unit, in addition to sending an excitatory output to the appropriate degree node, has inhibitory weights to all other units in the same column (satisfying the constraint that a given pitch node can only be mapped to a single degree). Each unit also has inhibitory weights to all other units in the same row (satisfying the constraint that a single degree node can receive input from only one pitch node). Again, with a nonlinear, sigmoidal output function, we can achieve a clear mapping from the input pitch nodes to the degree nodes even if several key nodes show some activation.

Figure 2 illustrates only two major key nodes, although this scheme permits mapping the pitch nodes into any mode with seven degrees. For example, by adding a set of minor key nodes, the pitch nodes can be mapped to the degree nodes in the appropriate way for minor scales.

Tonal Pitch Space

Tonality involves more than a set of pitches that constitute a scale. A theory of tonal perception should account for the perceptual structure of tonal pitch space such as the perceived proximity of pitches, chords, and keys (Lerdahl 1988). At first glance, the networks described above seem to require additional mechanisms to address such issues, although this may not be necessary. If we think of the nodes in a network as components of a vector, then the state of all the nodes at any moment defines a state vector for the network. A state vector can be thought of as defining a point in a multidimensional space, where each node of the network represents a dimension of the space. From this perspective, if we choose an appropriate network model, pitches, chords, or keys that are perceptually proximal should be represented by nearby points in the vector space.

We can compare state vectors representing two different states of the system (corresponding to, say, different chords) by calculating the inner product of the vectors (Jordan 1986). If the two states are similar, the inner product (which is related to the correlation coefficient) will be large, while vectors representing dissimilar states will tend to have smaller inner products. For example, suppose we compare the state of the system in response to the pitch C in the key of C major, with the response of the system to the pitch G in C major. Many of the nodes of the network will be in a similar state of activation for the two notes. That is, the same key node will be active, the same C-E-G chord node will be activated by the two notes, and the same diatonic set of pitch-class nodes that occur in C major will have been activated by notes prior to the C and the G to a similar degree. The state of the network will differ in terms of the activation of the specific pitch nodes for C and G and of the diatonic degree nodes (I versus V) represented in Fig. 2. The two notes will also tend to activate overlapping but not identical sets of chord nodes: e.g., F-A-C for C versus G-B-D for G.

In contrast, the state vector for the note C in the key of C major will show a larger difference from the state vector for the note G if the key is F major, for example, because the set of active pitch nodes (i.e., B versus B♭), chord nodes and key nodes will differ, and the mapping of pitch-class nodes to diatonic degree nodes will differ. The closeness of C to G within the network as defined by the state vectors thus depends on the tonal context, as is also true for human listeners (Krumhansl 1979). It seems, then, that the state vector for the network taken as a whole may provide an appropriate representation of tonal relationships.

Learnability

The tonal structure of music is culturally determined (Dowling and Harwood 1986). Thus another test of the network approach is to ask how such a network could be learned. In fact, much of the network structure may be acquirable through a competitive learning mechanism (Grossberg 1987). Such a learning mechanism could develop chord nodes that corresponded to the regular simultaneous occurrence of note patterns in the input. We can also bypass the chord nodes with a linear network, as noted earlier, and let the key nodes be directly activated from pitch nodes. This means that learning the pitch collections that correspond to various keys can parallel learning chord configurations.

On the other hand, the network described in Fig. 2 above presents a more serious challenge for learning because it is a multilevel, nonlinear architecture that depends upon having a key node level of representation. In this case, a simple competitive learning algorithm is inadequate, and more complicated learning algorithms are required (Grossberg 1987).

Conclusion

Connectionist approaches offer a promising and seemingly natural way to deal with the problem of tonality identification, though currently we do not know how well the networks described here simulate human performance. Another question is whether these networks can account for the rich perceptual structure of tonal pitch space, such as the proximity relations of individual notes, chords and keys (Lerdahl 1988). A further issue hinges on the fact that tonal relations are culturally determined (Dowling and Harwood 1986). This raises the question of how such networks could be learned (Grossberg 1987).

Acknowledgments

This work was sponsored in part by a PSC-CUNY Research Award to Jacqueline A. Jones.

References

Bharucha, J. J. 1984a. "Anchoring Effects in Music: The Resolution of Dissonance." *Cognitive Psychology* 16(4):485–518.

Bharucha, J. J. 1984b. "Event Hierarchies, Tonal Hierarchies, and Assimilation: a Reply to Deutsch and Dowling." *Journal of Experimental Psychology* 113(3):421–425.

Bharucha, J. J. 1987a. "MUSACT: A Connectionist Model of Musical Harmony." *Proceedings of the Ninth Annual Conference of the Cognitive Science Society.* Hillsdale, New Jersey: Lawrence Erlbaum Associations, pp. 508–517.

Bharucha, J. J.1987b. "Music Cognition and Perceptual Facilitation: A Connectionist Framework." *Music Perception* 5(1):1–30.

Bharucha, J. J. 1988. "Neural Net Modeling of Music." *Proceedings of First Workshop on Artificial Intelligence and Music.* Menlo Park, California: American Association for Artificial Intelligence, pp. 173–182.

Bharucha, J. J., and K. Stoeckig. 1987. "Priming of Chords: Spreading Activation or Overlapping Frequency Spectra?" *Perception and Psychophysics* 41(6):519–524.

Deutsch, D. 1969. "Music Recognition." *Psychological Review* 76(3):300–307.

Deutsch, D. 1984. "Two Issues Concerning Tonal Hierarchies: Comment on Castellano, Bharucha, and Krumhansl." *Journal of Experimental Psychology* 113(3):413–416.

Dowling, W. J., and D. L. Harwood. 1986. *Music Cognition.* Orlando, Florida: Academic Press.

Grossberg, S. 1980. "How Does the Brain Build a Cognitive Code?" *Psychological Review* 87(1):1–51.

Grossberg, S. 1987. "Competitive Learning: From Interactive Activation to Adaptive Resonance." *Cognitive Science* 11(1):23–63.

Grossberg, S. 1988. "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures." *Neural Networks* 1(1):17–61.

Jones, J. A., B. O. Miller, and D. L. Scarborough. 1988. "A Rule-based Expert System for Music Perception." *Behavior Research Methods, Instruments and Computers* 20(2):255–262.

Jordan, M. I. 1986. "An Introduction to Linear Algebra in Parallel Distributed Processing." In D. Rumelhart and J. McClelland, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vol. 1. Cambridge, Massachusetts: MIT Press, pp. 365–422.

Krumhansl, C. L. 1979. "The Psychological Representation of Musical Pitch in a Tonal Context." *Cognitive Psychology* 11(3):346–374.

Lerdahl, F. 1988. "Tonal Pitch Space." *Music Perception* 5(3):315–350.

Lerdahl, F., and R. Jackendoff. 1983. *A Generative Theory of Tonal Music.* Cambridge, Massachusetts: MIT Press.

Piston, W. 1948. *Harmony.* Revised ed. New York: Norton.

Rumelhart, D., and J. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vol. 1. Cambridge, Massachusetts: MIT Press.

Rumelhart, D. E., G. Hinton, and R. Williams. 1986. "Learning Internal Representations by Error Propagation." In D. Rumelhart and J. McClelland, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* vol. 1. Cambridge, Massachusetts: MIT Press, pp. 151–193.

Scarborough, D., J. Jones, and B. Miller. 1988. "An Expert System for Music Perception." *Proceedings of the First Workshop on Artificial Intelligence and Music.* American Association for Artificial Intelligence. Menlo Park, California: pp. 9–19.

Simon, H. A. 1968. "Perception du Pattern Musical par AUDITEUR." *Sciences de l'art,* Tome V-2:28–34.