Title

MUSACT: A Connectionist Model of Musical Harmony

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 9(0)

Author

Bharucha, Jamshed J.

Publication Date

1987

MUSACT: A Connectionist Model of Musical Harmony

JAMSHED J. BHARUCHA
Dartmouth College

A connectionist model of musical harmony is proposed to account for intuitions of schematic expectancy and sequential consonance. The model distinguishes between schematic and episodic structures, and the current version focusses on the former. Units representing tones, chords and keys are linked in a network. Activation received by tone units from a musical event spreads through the network phasically, reverberating until the network settles into a state of equilibrium. The settled state represents the expectancies for events to follow and the perceived consonance of these events should they occur. The model accounts for psychological data on rating judgments and recognition memory. Evidence for the spread of activation comes from experiments on priming of chords.

Several features of connectionist models--in which simple processing units, connected by weighted links in a network, activate each other in parallel--recommend them as models of music cognition. First, they are highly interactive, so that low level processes can influence higher level processes and vice versa. Thus, a sequence of tones can imply a chord or a key, and the implied chord or key can in turn influence the perception of tones that follow. Second, the architecture of a connectionist system hooks up naturally with a sensory front end that codes frequency as a (spatial or temporal) pattern of activations of neurons in the inner ear. Third, pattern matching of a currently heard sequence can be achieved in parallel by content-addressing all similar memory traces, enabling recognition of sequences and variations thereof. Finally, connectionist networks can internalize persistent regularities, such as occur in our musical environment, without explicit instruction.

A network's ability to internalize regularities--by altering the strengths of connections between units (McClelland & Rumelhart, 1986)--enables the parsimonious view that music cognition is a consequence of general principles of cognition operating on structural regularities in the environment. The average listener has little, if any, explicit knowledge of musical structure, yet shows evidence of considerable tacit knowledge. For example, even listeners with no formal training in music show systematic differences in processing time for chords as a function of the prior harmonic context (Bharucha & Stoeckig, 1986, 1987).

The model proposed here, MUSACT, is designed to capture musical intuitions and psychological data concerning expectancy, sequential consonance and short term memory for musical harmony. Schenker (1906/1954) observed that one of the qualities of the dominant chord is "to indicate that the tonic is yet to come" (p.219). Expectancies of this sort are driven by cognitive structures and processes that have internalized regularities in the musical environment in order to facilitate subsequent perception. Hand in hand with expectancies are intuitions about sequential consonance. The greater an event's expectancy, the greater its context-dependent consonance. A composer may choose to satisfy or violate these expectancies to varying degrees, thereby evoking varying degrees of sequential consonance or dissonance. The aesthetic value of subtle departures from the expected has figured in numerous theoretical writings about music and about emotion (e.g. Mandler, 1984; Meyer, 1956). A new piece is heard as culturally anomalous to the extent that expectancies are violated. Indeed, the connection strengths between units are assumed to be trained by minimizing, over the history of one's exposure, the discrepancy between expectancies generated by the network and transitional probabilities in the music of one's culture.

People exposed to Western music are assumed to acquire a network representation of chord functions (hereafter referred to as chords) and their organization in the form of keys, which serves to schematize subsequent perception. Constraints on the combining of tones into chords and constraints on the sequencing of these chords are among the more obstinate regularities in Western music. Every amateur musician knows that a mastery of only six chords enables you to accompany a vast majority of popular songs. These basic harmonic regularities have even begun to permeate much of the popular music of countries in the East. Given the pervasiveness of these regularities, and given the evidence of tacit knowledge of them (Bharucha & Stoeckig, 1986, 1987), it is reasonable to conclude that they have been internalized as cognitive structures that facilitate and bias subsequent perception.

Schematic & Episodic Structures: Schematic & Veridical Expectancy

Two broad classes of cognitive structures for music are envisioned: schematic structures, which represent abstract structural regularities (sometimes formalized as grammars) of the music of one's culture, and episodic structures, which represent particular musical sequences (Bharucha, 1984b). The former embody typical relationships between types or classes of events, and the latter embody relationships between particular event tokens. The expectation generated by a dominant chord for the tonic to follow arises from schematic structures that encode the typicality of relationships, whereas the expectation for a VI chord to follow a particular dominant chord in a particular familiar piece arises from episodic structures that encode the particular events in that piece. The former generate schematic expectancies and the latter veridical expectancies. The two are usually in agreement but often in conflict, giving rise to the peculiar effect known as the deceptive cadence (a dominant chord followed by a VI chord). The unavoidable effect of schematic expectancies, even when listening to a piece that violates them, provides a resolution of Wittgenstein's puzzle (see Dowling & Harwood, 1985) concerning the possibility of violating expectancies when listening to a familiar piece.

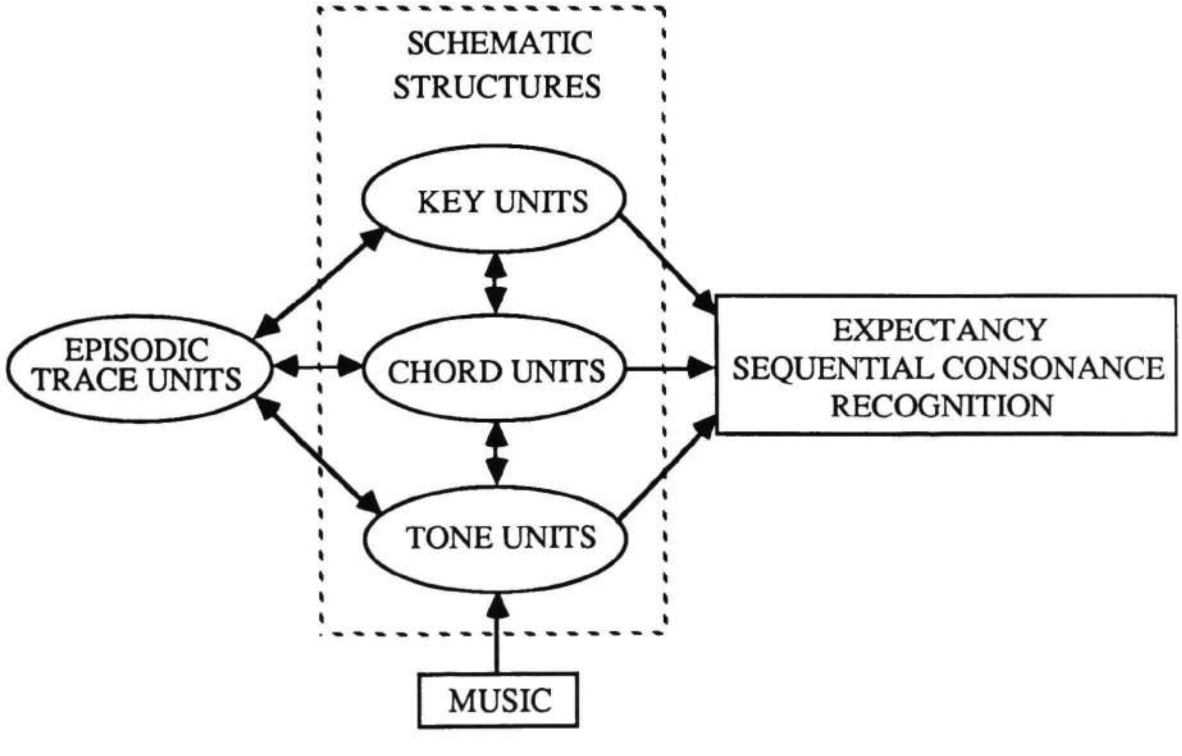

FIGURE 1. A sketch of the model

Although schematic expectancies can in principle be generated from episodic structures by adding veridical expectancies over episodic traces activated in parallel, a model of this sort would require documenting all the episodic memory traces in a typical person's brain. MUSACT focusses on schematic structures, employing episodic structures only to the extent of implementing experiments on short term memory for sequences of chords.

The Model

Architecture

We have the sense that certain typical tone clusters, such as major and minor chords, are heard as unitary. We also have the sense of an even more abstract, meaning-like, state induced by music, called the key. Sequences or clusters of tones may unavoidably suggest a chord or several alternative chords, and combinations of chords establish a sense of key. One of the advantages of the model is that it enables chord and key instantiations to be graded and ambiguous. These ambiguities are an important feature of music, exploited during modulations or other transitions, and used to create graded degrees of expectancy violation.

The schematic network consists of three layers of units, representing tones, chords and keys (see Figures 1 and 2). There are symmetric links between units of adjacent layers (i.e. between tone units and chord units, and between chord units and key units) but no links between units within a layer. The links between tone and chord units reflect the relationships between tones and the chords of which they are components. The links between chord and key units reflect the relationships between chords and their parent keys. In the current version of the model, only major and minor chords are implemented, and only major keys.

The input to the network is a sequence of events, each event being a simultaneous cluster of tones. Input is received via the tone units, which represent the twelve octave-equivalent pitch categories. This layer constitutes a discrete pitch schema to which the pitch continuum is assimilated (see Shepard & Jordan, 1984). The frequency responses of these units are equally spaced along a logarithmic scale of frequency, and are fixed only relative to each other, underlying the relational nature of pitch memory. The sensory front end that provides the input to these units is beyond the present scope, but neural net models that extract octave-equivalent pitch categories (see Deutsch, 1969) can be adapted quite naturally to a connectionist model of more abstract phenomena as proposed here.

The output of the model is the pattern of activation of the chord and key units. A chord unit is activated either by the explicit sounding of some or all of its component tones, or by indirect influences, via its parent keys, from related chords. When only some of the chord's component tones are sounded, the context may help disambiguate the chord by top-down activation from parent key units. A key unit is activated by some or all of its daughter chords, or by indirect influences, via its daughter chords, from related keys. Indirect activation of chord units permits smooth excursions (such as secondary dominants and modulations) from the focus of activation.

Phasic Activation

After an event is heard, activation spreads through the network, via the weighted links, reverberating back to units that were previously activated. In this model, activation is phasic, meaning that units respond only to changes in activation of neighboring units. Phasic activation was selected because of the salience of event onsets in music. On each cycle, units are synchronously updated on the basis of activation levels, from the previous cycle, of neighboring units. Phasic activation eventually dissipates until the network settles into a state of equilibrium. The network will settle if no unit transmits more phasic activation than it received on the previous cycle. This requirement is easily satisfied if the weights are small relative to the fan-in or fan-out.

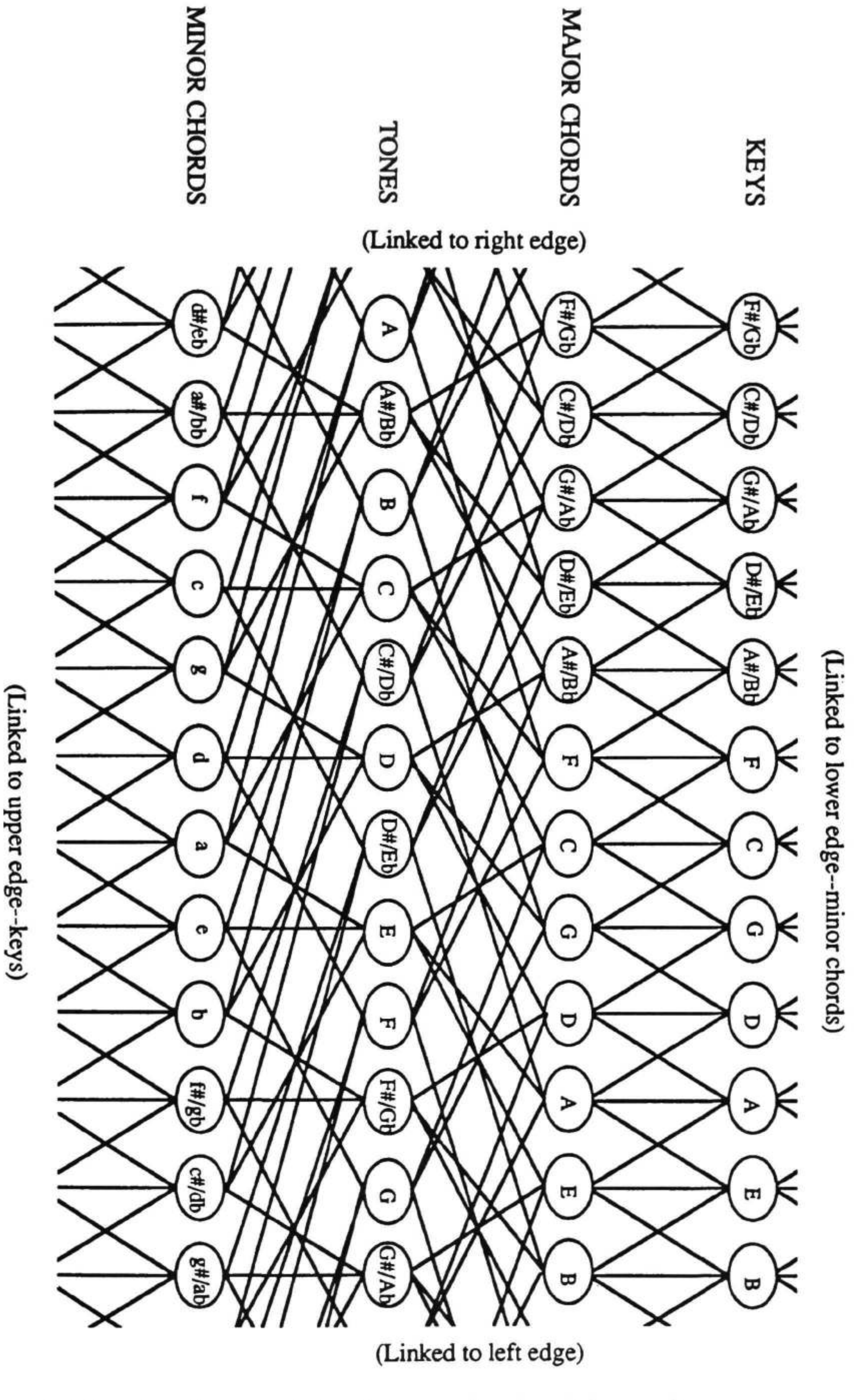

FIGURE 3. The network of tone, chord and key units.

Activation, Tonality, and Expectation

The pattern of activation of key units represents the degree to which keys are established. Tonal music will tend to build up activation in one region of the network, such that one key unit is most highly activated, with activation tapering off with increasing distance from the focal key. Atonal music will typically induce a less focussed pattern, and polytonal music would result in multiple, though not very strong, foci. The model thus allows for gradations of key, and for multiple keys, consistent with the findings of Krumhansl (Krumhansl & Kessler, 1982; Krumhansl & Schmuckler, 1986).

The pattern of activation of chord units represents the pattern of expectancies for chords to follow. A chord whose unit is highly activated is strongly expected, and a strongly expected chord is heard as consonant. The pattern of activation of chord units thus underlies our harmonic expectations as well as our intuitions of context-dependent sequential consonance.

In future versions of the model that incorporate additional pitch structure, such as pitch proximity (see Bharucha 1984a; Deutsch, 1978) for voice leading, tone units will also serve as output units responding to top-down and lateral influences. At present, although the tone units do propogate top-down activation, the model is intended to be tested only for its harmonic, not melodic, adequacy.

Weights

All links between tone units and chord units are assumed to have equal weights. There are six classes of weights between chord units and key units, corresponding to the six chords in each key. All the links of the same class must have the same weight, resulting in a repeating pattern of weights over the network. (For example, the G major chord has the same relationship--the dominant--to the key of C as the C# major chord has to the key of F#.) In music-theoretic terms, a chord's relationships to its parent keys are its functions in those keys. The six classes of weights thus correspond to the six chord functions. Since some chord functions are stronger instantiators of key than others (e.g. the key of C is more strongly instantiated by the C major chord--the tonic--than by the F major chord--the subdominant--even though both are daughter chords), it would be reasonable to assume that links for stronger functions have higher weights. However, it turns out that the model can exhibit all the essential qualitative patterns of behavior even when the weights are not differentiated according to function. Thus, the typical hierarchy of strengths of the three major chord functions (tonic, dominant, subdominant) emerges even if their weights are equal. The pattern of connectivity alone is sufficient to bring about functional differentiation. The model thus generates the desired functional hierarchy of chords simply by knowing which chords are members of which keys and which tones are components of which chords. This is a remarkable and unanticipated property of the model, and points to its power.

The weights are assumed to be higher for major chord units than for minor chord units, since major chords are stronger instantiators of key. As discussed below, with only these elementary constraints on links and their weights, based on fundamental tenets of music theory, a set of weights can be found that enable the model to account for a complex array of psychological data on the perception of harmony as well as some subtle aspects of music theory. Indeed, one of the advantages of a connectionist model is that complexities and interactions may emerge naturally from simple constraints on architecture.

The Spread of Activation

The activation, $\mathbf{a}_{i,e}$, of the $\mathbf{i}^{\mathbf{th}}$ unit after the network has reverberated to equilibrium following the $\mathbf{e}^{\mathbf{th}}$ event is:

$$\mathbf{a}_{i,e} = \mathbf{a}_{i,e-1}(\mathbf{1}-\mathbf{d})^{\mathbf{t}} + \mathbf{sA} + \sum_{c=1}^{equil} \Delta\mathbf{a}_{i,e,c},$$

where $\mathbf{d}$ is the rate of decay ($\mathbf{0} < \mathbf{d}_i < \mathbf{1}$) for one time unit, $\mathbf{t}$ is the time elapsed since the offset of the last event, $\mathbf{A}$ is the source activation due to the stimulus, and s is 1 if the unit is receiving environmental input and 0 otherwise. (For simplicity, the present version of the model assumes that all events are of equal duration; duration is varied simply by repeating events.) $\Delta\mathbf{a}_{i,e,c}$ is the change in activation after reverberation cycle $\mathbf{c}$, and is the output of the unit on cycle $\mathbf{c+1}$. $\Delta\mathbf{a}_{i,e,c}$ is the sum of the outputs of its $\mathbf{n}$ neighboring units, weighted by their links, $\mathbf{w}_{ij}$, ($\mathbf{0} < \mathbf{w}_{ij} < \mathbf{1}$). Thus,

$$\Delta\mathbf{a}_{i,e,c} = \sum_{j=1}^{n} \mathbf{w}_{ij}\, \Delta\mathbf{a}_{j,e,c-1}.$$

Simulations

The simulation results reported below were obtained with a weight assignment in which all chord units of the same mode (major or minor) have equally weighted links with their parent keys. This enables us to observe the differentiation of chord functions without forcing their differentiation via the weights. Major chord units were assigned more strongly weighted links (0.244) with their parent key units than were minor chord units (0.22, 90% of 0.244), since major chords more strongly establish their parent keys. Links between tone units and chord units were also uniform (0.0122, 5% of 0.244), with no preference given to the root of the chord. This weight assignment yields an initial strong influence of the input tones on their local regions of the network, followed by an extended percolation during which the two more abstract layers exert their influence. Thus, if the input tone cluster is {C,E,G}, i.e., a C major chord, the chord units linked to these tone units show an initial prominence, so that, for example, the A major chord unit is more highly activated than the D major chord unit. However, after the activation has had a chance to reverberate for a number of cycles, the D major chord unit overtakes the A major chord unit, by virtue of its greater proximity to the eventual activation peak, which is at C major. Within 40 cycles, all constraints inherent in the network are satisfied, and the pattern of activation does not change qualitatively. Within 50 cycles, the phasic activation has dissipated to the point at which the ratio $\Delta a_{i,e,c}/A$ does not exceed 0.005 for any unit. For the results discussed below, the equilibrium state is stipulated to be the state of the network when this 0.005 criterion is reached.

The tone cluster {C,E,G}, i.e., a C major chord, played without any prior context, generates the following pattern of activation (see Figure 3). The most highly activated key is C, of which the source chord is the tonic. Activations of other key units decrease monotically with distance from C, the lowest being F#. The pattern of activation of the chord units mirrors that of the key units with the same alphabetic name.

Even though the weights between the source chord, C major, and its parent keys, F, C, and G, are equal, the parent keys are not activated to the same degree. In decreasing order, the activations of these three keys are C, F, and G, which are the keys in which the source chord is the tonic, dominant, and subdominant, respectively. This is exactly the hierarchy of harmonic functions to be expected from music theory.

There seems at first to be a paradox here. On the one hand, a chord more strongly instantiates the key of which it is the dominant than the key of which it is the subdominant. On the other hand,

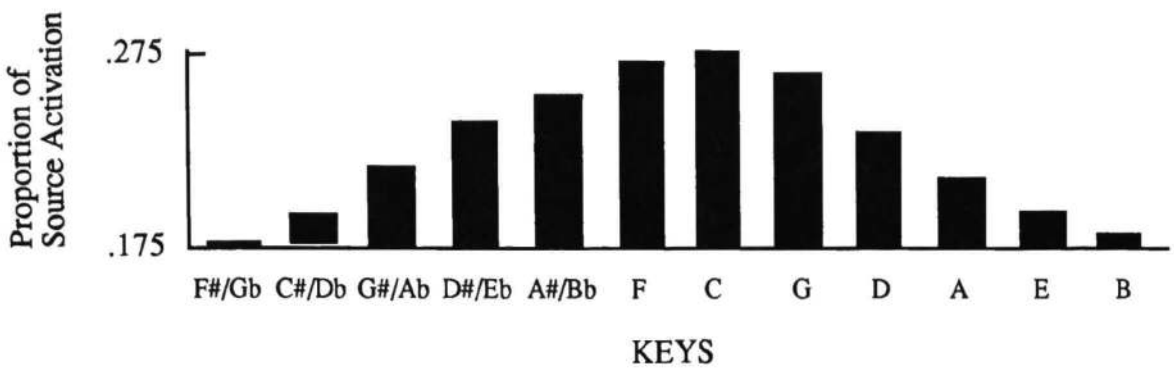

FIGURE 3. The pattern of activation of key units after activating the tone units C, E, and G (heard as a C major chord).

the key of the dominant is generally thought to be closer to the current key than is the key of the subdominant, and should therefore have a higher activation. This paradox is not peculiar to the present model; the model simply forces us to confront it. (Interestingly, the model conforms to the latter alternative if the link between a tone unit and a chord unit is weighted more heavily for the root of the chord than for the other tones, illustrating the unanticipated interdependence of apparently disparate factors.) The resolution of the paradox lies in the fact that dominant chords typically occur much more often than subdominant chords in a piece of music, presumably because dominant chords contribute to a more stabe key. Given the high frequency of dominant chords, the dominant builds up more activation than the subdominant.

For an input consisting of a sequence of events, the rate at which activation decays between events was set at 0.3. If the input sequence is the tone cluster {F,A,C} followed by {G,B,D}, i.e., an F major chord followed by a G major chord, the model shows the most highly activated key unit to be C, even though the tonic chord of that key, the C major chord, has not occurred. This, again, is consistent with what would be expected from music theory.

Psychological Data

The model's performance in experiments eliciting rating judgments and memory confusions is qualitatively equivalent to human performance on these tasks. In a series of experiments on the perception of harmony (Bharucha & Krumhansl, 1983; Krumhansl, Bharucha, & Castellano, 1982), the perceived relationship between chords as a function of context was studied using these tasks.

Rating Judgments

In the rating task, subjects were presented with two chords in succession and rated, on a scale from 1 to 7, how well the second followed the first. For the simulation, the rating judgment consisted of reading off the level of activation of the last chord. On this assumption, the observed rating judgments are equivalent to judgments of expectancy or sequential consonance.

When both chords of the pair shared a parent key, subjects gave higher judgments when the last chord was major than when it was minor. This result was particularly interesting for a pair consisting of one major chord and one minor chord, because the same two chords elicited different judgments depending upon their temporal order. In the model, a major chord activates its unit slightly more than does a minor chord, even though the source activation, **A**, is the same, because of more reverberatory activation from its parent keys. This asymmetry doesn't require asymmetric links, since it follows directly from the fact that major chords establish their parent keys more strongly than do minor chords. In music, this translates into a tendency to end with a major chord, often even if the key is minor.

When the two chords were preceded by a context that established a key, subjects gave higher judgments the more closely related the last chord was to the key of the context. In the model, this occurs because the context activates closely related chords more than distantly related chords. An interesting asymmetry was contained in this result as well: two chords of the same mode (major or minor) elicited a higher judgment if the one closer to the key of the context was played last. Thus, in the key of C, an F major chord followed by a C major chord is a more stable ending than the reverse, even though the reverse may be true in the absence of a prior context.

Recognition Memory

Memory for a sequence does not take the form of an explicit representation, but rather is encoded by weight increases of links between the units in the network and episodic units that are temporally organized. Each event in a sequence increases the weights of links between a proprietary episodic unit and the schematic units, in proportion to the activations of these latter units. The episodic units are then activated every time the tone, chord or key units to which they have strong links are highly activated. Once activated, the episodic units in turn activate the schematic units. This architecture can accomplish pattern completion (recalling a piece given only a few notes or a sketchy rendition), recognition of variations, and retrieval of past musical memories in a parallel, content-addressable fashion rather than through serial search.

In the present implementation, episodic units serve only to simulate short term recognition of a sequence. Consider a to-be-remembered sequence of chords. Each chord token in the sequence has its proprietary episodic unit. Links between this unit and the schematic units have their weights increased in proportion to the phasic activations of the schematic units after that chord is heard. In this way, the pattern of activation of the network after hearing a particular chord token can be recovered simply by activating its episodic unit.

Data from experiments on recognition memory for chord sequences show trends similar to those observed for rating judgments (Bharucha & Krumhansl, 1983; Krumhansl, Bharucha, & Castellano, 1982). In these experiments, subjects judged whether two presentations of a sequence of chords were identical or had different chords (the target chords) in one serial position. If the sequence as a whole established a key, a change was less likely to be detected the more closely related the second target was to the key. This manifest itself in two ways. First, a change was less likely to be detected if both targets were closely related to the key than if both were distantly related to the key. Second, if one target was more closely related to the key than the other, a change was less likely to be detected when the more closely related one occurred in the second presentation rather than the first. In general, if a change is made to a sequence, it is less likely to be detected if it renders the sequence more coherent than more anomalous. Coherence is a consequence of the strong activation of a subsuming unit, such as a key unit in music, or a subsuming semantic unit in language (Bharucha, Olney, & Schnurr, 1985), relative to the other subsuming units in the network.

In the model, the probability with which the second target is judged to be the same as the first is monotonically related to the activation of the unit representing the second target, relative to the activations of the other chord units, when the first target was heard. Since the more closely related the second target is to the key of the context, false alarm rates increase with closeness of the second target to the key of the context.

This short term memory architecture also predicts that if the second presentation is transposed to a different key, false alarm rates should decrease with the distance (along the network) between the two keys. This would be consistent with recognition memory results for sequences of tones (Cuddy, Cohen, & Miller, 1979).

Evidence of Spreading Activation: Priming

Evidence that chords indirectly activate representations of related chords comes from experiments on priming. Bharucha and Stoeckig (1986) presented subjects with two major chords in succession, the first called the prime and the second the target. On half the trials, the chord in the target position was a mistuned foil. Subjects were instructed to judge, as fast as possible, whether the chord in the target position was in-tune or out-of-tune. Subjects first practiced without the prime until a criterion level of accuracy was reached. In the main task, response times were significantly faster when the prime and target were close together along the network than when they were distant. This demonstrates that the prime activates units corresponding to closely related targets, as would be predicted by the model.

Error rates mirrored response times, so that the target was more likely to be judged in-tune the closer it was (along the network) to the prime. The response time and error rate data measure the prime's influence on the target's expectancy and sequential consonance, respectively.

An alternative explanation of the above results is that the priming is due to overlapping harmonic spectra between closely related prime and target chords. In a subsequent study (Bharucha & Stoeckig, 1987), harmonics that overlapped were removed from the stimuli and priming was still observed. This demonstrates that there must be a spread of activation at a fairly abstract cognitive level.

References

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

Bharucha, J.J. (1984b). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. Journal of Experimental Psychology: General, 113, 421-425.

Bharucha, 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., Olney, K.L, & Schnurr, P.P. (1985). Coherence-disrupting and coherence-conferring disruptions in text. Memory and Cognition, 13, 573-578.

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 harmonic spectra? Manuscript under review.

Deutsch, D. (1969). Music recognition. Psychological Review, 76, 300-307.

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. Perception & Psychophysics, 23, 227-230.

Dowling, W.J., & Harwood, D.L. (1985). Music cognition. New York: Academic Press.

Cuddy, L.L., Cohen, A.J., & Miller, J.(1979). Melody recognition: The experimental application of musical rules. Canadian Journal of Psychology, 28, 148-157.

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

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

Krumhansl, C.L., & Schmuckler, M.A. (1986). The Petroushka chord: A perceptual investigation. Music Perception, 4, 153-184.

Mandler, G. (1984). Mind and body: Psychology of emotion and stress. New York: Norton.

McClelland, J.L., & Rumelhart, D. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. M.I.T. Press.

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

Schenker, H. (1954). Harmony (O. Jones, Ed., E.M. Borgese, Trans.). Cambridge: MIT Press. (Original work published 1906).

Shepard, R.N, & Jordan, D.S. (1984). Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. Science, 226, 1333-1334.

Acknowledgements

This paper was written while the author was at Carnegie-Mellon University. The author thanks James McClelland and Herbert Simon for helpful discussions. Katherine Olney contributed to early implementations of the model. The author also thanks Paul Cohen, Diana Deutsch, William Estes, Leonard Meyer, Caroline Palmer, Mark Schmuckler and, especially, Carol Krumhansl, for suggestions and discussions.