Modeling the Perception of Tonal Structure with Neural Nets
Author(s): Jamshed J. Bharucha and Peter M. Todd
Source: *Computer Music Journal,* Vol. 13, No. 4 (Winter, 1989), pp. 44-53
Published by: The MIT Press
Stable URL: http://www.jstor.org/stable/3679552
Accessed: 11/08/2011 09:38

Jamshed J. Bharucha

Department of Music
Dartmouth College
Hanover, New Hampshire 03755 USA
bharucha@eleazar.dartmouth.edu

Peter M. Todd

Department of Psychology
Stanford University
Stanford, California 94305 USA
todd@psych.stanford.edu

Modeling the Perception of Tonal Structure with Neural Nets

What can we say about the perception of music by the silent majority of listeners, those for whom music is written but who neither create music nor can articulate their musical experience? How do they acquire their demonstrably sophisticated intuitions about music patterns typical of their culture? Experiments in the cognitive psychology of music have cast some light on the first question. Recent developments in neural net learning now enable us to explore answers to the second.

In this article, we discuss one aspect of the experience of the nonmusician listener—contextual influences on the perception of pitch. We limit our discussion to tonal implications and expectations and to memory for pitch sequences. We do not presume that this description captures the listener's experience in all its intricacy. We first summarize some psychological research and then explore how neural nets can be employed to model the acquisition of these phenomena through passive exposure.

Two forms of tonal expectancy will be discussed—*schematic* and *veridical.* Schematic expectancies are culture-based expectancies for events that typically follow familiar contexts. Veridical expectancies are instance-based expectations for events that follow in a particular familiar sequence. Schematic and veridical expectancies may conflict, since a specific piece of music may contain atypical events that do not match the more common cultural expectations. This conflict, which was attributed to Wittgenstein by Dowling and Harwood

(1985), underlies the tension between what one expects and what one hears, and this tension plays a salient role in the aesthetics of music (Meyer 1956). Schematic expectancies are driven by structures that have abstracted regularities from a large number of specific sequences. Veridical expectancies are driven by encodings of specific sequences. We briefly discuss models of both forms of expectancy and conclude with a model that subsumes both.

Two of the classes of nets that have promise for this research—auto-associative nets and hierarchical self-organizing nets—are only summarized here since their application to music has been described in detail in earlier papers (see Bharucha 1987a; 1987b; Bharucha and Olney 1989). We focus our modeling account on a third class of nets—sequential nets—that learn specific tone sequences (i.e., veridical expectancies) and in doing so exhibit schematic expectancies as an emergent property. The three classes of nets we discuss—auto-associative nets, hierarchical self-organizing nets, and sequential nets—are neither mutually exclusive nor entirely redundant. We present them as fruitful explorations in musical modeling and consider that one of our goals for future research is to discriminate among them computationally and empirically and, if necessary, to search for models that surpass them.

Psychological Aspects of Tonal Expectation

Most people have strong perceptual intuitions about the structure of the music of their culture. These perceptual intuitions are not typically revealed

overtly in performance, composition, or even verbalization, since most people without formal musical training lack these skills. Through carefully designed psychological experiments, however, listeners who are unaware of or unable to articulate their musical intuitions can nevertheless be shown to be sensitive to rather subtle deviations from typical musical patterns.

The writings of music theorists have given us a powerful set of hypotheses about the perceptual intuitions of the average listener. Given the extensive training of the music theorists, however, and the theoretical constraints implicit in the language with which their theories are constructed, these hypotheses must be subject to rigorous empirical tests before their applicability to the average, untutored listener is established. Although the results of such experiments may typically confirm the hypotheses of music theorists, they are essential for building our corpus of knowledge about the perceptual intuitions of untutored listeners.

Western listeners show from their responses in psychological experiments that they recognize departures from typical tonal patterns. Furthermore, they have tacit knowledge of the distance relationships between tones, chords, and keys in tonal contexts as would be predicted by the work of music theorists. For example, subjects judge chords to be related to each other in accord with the circle of fifths. These intuitions show up in experimental tasks as disparate as:

> Direct subjective measures of relatedness and expectation (Krumhansl and Kessler 1982; Bharucha and Krumhansl 1983; Schmuckler 1988)
> Memory confusions (Cuddy, Cohen, and Miller 1979; Krumhansl, Bharucha, and Castellano 1982; Bharucha and Krumhansl 1983)
> Response time (Bharucha and Stoeckig 1986; 1987)

Although the pattern of data is often more consistent for musically trained subjects on some of these tasks, there seems to be little difference between musically trained and untrained subjects

on reaction time tasks that measure the extent to which a musical context facilitates the perceptual processing of schematically expected events. These tasks reveal systematic patterns of tacit knowledge about the relationship between chords, even in the minds of musically untutored subjects who begin the experiment with profuse apologies about being tone deaf. In these experiments (Bharucha and Stoeckig 1986; 1987; Bharucha 1987b), subjects are instructed to decide whether a target chord is in tune or mistuned. Mistuned chords are constructed by flattening one of the triadic components. When the target chord is preceded by a context (also consisting of a chord), the response time to judge correctly whether the target chord is in tune is monotonically related to the distance of the target from the context along the circle of fifths.

The above result could, in part, be explained by the overlap in harmonic spectra between the context and the target chords. Closely related chords, when played with tones rich with harmonics, have spectral components in common. To test this hypothesis, harmonics shared by context and target were removed, giving closely related targets no acoustic advantage. The target chord was again recognized more quickly following a context to which it was closely related (Bharucha and Stoeckig 1987).

This result establishes definitively that the expectations generated by a tonal context, as measured by the perceptual facilitation of closely related chords, cannot be explained by the harmonic series alone. We are compelled to conclude that the perceived relationships between chords are learned, rather than being somehow inherent in the actual sonic stimuli. Since the perceptual facilitation was found for nonmusicians as well as musicians, the learning must have been passive perceptual learning rather than formal musical training.

Some analogous experiments have been conducted with native listeners of other cultures, though this literature is less conclusive. These experiments have shown, at least for Indian ragas, that native, untutored listeners tend to expect tones that are typical of familiar musical contexts (Castellano, Bharucha, and Krumhansl 1984; Kessler, Hansen, and Shepard 1984; Bharucha 1987b).

Motivation for Neural Net Modeling of Harmonic Expectancies

Neural net models enable us to explore the extent to which these musical intuitions are a consequence of extended passive exposure to the musical regularities of a culture. General purpose learning architectures with units that respond to the presence of musical features can internalize musical regularities by changing the weights of links that connect units. The units themselves can plausibly be shown to develop their specialization to abstract musical features as a result of general principles of self-organization. Examples of these adaptive systems applied to specific musical phenomena are given in the following sections.

One may ask why complicated, often difficult-to-interpret neural net models should be used in the psychology of music when there are many perhaps simpler, symbolic, rule-based models available. Although rule-based models of music have been successful at describing the formal structure of some musical compositions, and have thus provided valuable hypotheses and analytic constraints, they fall short as psychological theories. They fail to account for the acquisition of the rules they postulate, and this ad hoc postulation of rules is not typically limited to a small set of assumptions of which the others are a natural consequence.

Paramount among the psychological constraints on modeling is the constraint that the postulation of cognitive structures must be accompanied by plausible accounts of the innateness or learnability of the structures in question. Few psychological models meet this strict scrutiny, including our own. We suggest, however, that given alternative models or classes of models, the most parsimonious is to be preferred. Neural net models have the capacity to supersede the more traditional rule-based models on parsimony grounds because of their ability to account for the acquisition of intuitions through passive perceptual learning.

Neural net models have the potential to account for perceptual learning of musical structure with only two classes of constraints. First, the net may be constrained by general principles of neural architecture and by constraints specific to the learning

algorithms. Second, there must be pitch-tuned input units see (Linsker 1986) however, for an account of how, in vision, even elementary feature detectors can develop from general constraints on the net). The auditory system reveals a tonotopic mapping of pitch, supporting this constraint. It is important to note that no constraints on the specifics of musical structure are required; they emerge as a result of the net's exposure to music. Rule-based systems, in contrast, typically have as many constraints as specific rules of musical structure, with little justification about the origins of those rules.

Schematic Expectancies

Neural net models can be used to demonstrate the passive learning of schematic expectancies in three different musical domains.

Learning Culture-Specific Modes With Auto-associators

The extent to which patterns of schematic expectancy for the tones in musical scales can be captured by an auto-associative net has been explored in earlier work (Bharucha and Olney 1989). Using the delta rule (Rumelhart and McClelland 1986), this net is taught to map from a complete set of scale tones as input to the same scale set as output. It essentially acts as a pattern completion device that suggests, implies, or "fills in" missing tones at its output when presented with a subset of a scale as input.

Such a net exposed to major and harmonic minor scale sets correctly generates patterns of expectancy consistent with the establishment of keys, exhibits the desired ambiguities of key, and can be shown to tacitly embody the structural constraints abstractly summarized by the circle of fifths. Analogously, a net exposed to Indian ragas fills in expected tones when presented with subsets of the raga tones.

An auto-associative net trained on the scales of one culture can be tested with the scales of another, making predictions about tonal implications

generated in the minds of listeners hearing an unfamiliar form of music. A net trained on the Western major and minor scales seems to assimilate some Indian ragas to the Western scales, sometimes shifting the tonic (Bharucha and Olney 1989).

Learning Hierarchical Representations Through Self-Organization

Hierarchical relationships, such as between tones, chords, and keys, can be learned passively by algorithms for self-organization (Kohonen 1984; Linsker 1986; Rumelhart and McClelland 1986; Carpenter and Grossberg 1987). Most self-organization mechanisms assume the prior existence of abstract units into which the input units feed. These abstract units initially have no specialization, since the links from the input units are initially random. However, repeated exposure to commonly occurring patterns causes some of these abstract units to tune their responses to these patterns.

One of the more straightforward, self-organization algorithms, called *competitive learning* (Rumelhart and McClelland 1986) accomplishes this as follows. For any given pattern some arbitrary abstract unit will respond more strongly than any other, simply because the weights are initially random. Of the links that feed into this unit, those that contributed to its activation are strengthened and the others are weakened. This unit's response will subsequently be even stronger in the presence of this pattern and weaker in the presence of other, dissimilar patterns. In similar fashion, other abstract units learn to specialize to other patterns. This process can be continued to even more abstract layers, at which units become tuned to patterns that commonly occur in the lower layer.

The overwhelming preponderance of major and minor chords in the popular Western musical environment would drive such a net to form units that respond accordingly. Furthermore, the typical combinations in which these chords are used would drive units at a more abstract layer to register larger organizational units such as keys. The notion that individual neurons specialize to respond to complex auditory patterns has some preliminary empirical support from single-cell recording studies on animals (Weinberger and McKenna 1988).

Once these chord and key units have organized themselves, the net models the implication of tones, chords, and keys given a set of tones. A hierarchical constraint satisfaction net built on this organization has been reported in earlier work (Bharucha 1987a; 1987b). In this net, called MUSACT, activation spreads from tone units to chord and key units and reverberates phasically through the net until a state of equilibrium is achieved. At equilibrium, all constraints inherent in the net have been satisfied.
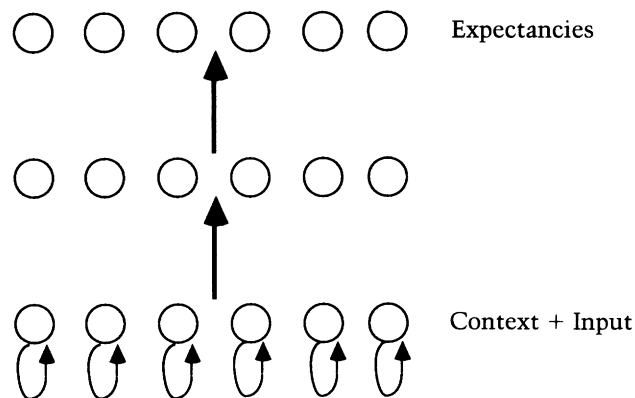
Given a key-instantiating context, the unit representing the tonic becomes the most highly activated. The other chord units are activated to lesser degrees the further they are from the tonic along the circle of fifths.

Two behaviors of the net illustrate its emergent properties. First, the above activation pattern does not require the tonic chord to be played at all. An F major chord followed by a G major chord will cause the C major chord unit to be the most highly activated. Second, the circle of fifths implicit in the activation pattern cannot be accounted for on the basis of shared tones alone. If a C major context chord is played, the D major chord unit is more highly activated than the A major chord unit, even though the latter shares one tone with the sounded chord (C major) and the former shares none at all. A careful tracking of the net's behavior as activation reverberates and before it converges to an equilibrium state reveals a lower initial activiation of D major over A major, reflecting an initial bottom-up influence of shared tones. As activation has a chance to reverberate back from the key units (a top-down influence), this advantage is lost, and D major overtakes A major. So the circle of fifths is truly an emergent property of the simultaneous satisfaction of elementary associations between tones and clusters of tones. See Bharucha (1987a; 1987b) for details.

Learning With Sequential Nets

Some of the schematic expectancies that are essentially sequential, as in chord progressions, can be modeled with sequential nets. The architecture

shown in Fig. 1 supports the learning of schematic expectancies from exposure to sequences. It has three layers of units (input, hidden, and output) and links that feed forward only. The input units register the sequence as it is heard, and the activations generated at the output units represent the learned schematic expectancy for the next event.

The input units, labeled "context + input," represent chords, one unit for each of the six most common chord functions—the three major and three minor chords in a given key, i.e., the triads built upon the first six degrees of the major scale. We use a pitch-invariant representation in which all sequences are normalized to a common tonic. Each unit has a self-recurrent connection with an identical fixed weight between 0.0 and 1.0. These recurrent links implement a decaying memory of the sequence presented up to any given point in time. The first chord in a sequence causes the corresponding input unit to be activated, while the other input chord units remain off. When the second chord is presented, its corresponding input unit is activated, while the unit corresponding to the first chord has its original activation multiplied by the weight on the recurrent link. This process of activating new units and decreasing the activation level of previous units in an exponential fashion continues for the entire chord sequence. If a chord is repeated in the sequence, a new surge of activation is added into the decaying activation already present at the corresponding unit. In this way, the "context + input" vector represents a decaying memory of the sequence (the context) plus the current event (the current input from the environment).

We envision the input units being activated by chord units in MUSACT after they have been normalized to a common tonic. This normalization is necessary because the chord sequences cannot be encoded in a pitch-specific format of the kind used in MUSACT, since most people have no absolute identification of pitch in the long term. The sequences must therefore be encoded in a format that is invariant under transposition. In the short term, however, transpositional invariance is biased by absolute pitch information held by MUSACT. Cuddy, Cohen, and Miller (1979) found that after presenting a standard melody, comparison melodies were more likely to be judged the same when they were transposed to a related key than when they were transposed to an unrelated key. The resulting constraints on modeling are thus: absolute pitch information is held in short-term memory without sequential constraints (as in MUSACT), and sequences are held in long-term memory in a pitch-invariant format with sequential constraints. Only the latter aspect of the model will be discussed here.

Each input unit in the net is linked to each hidden unit, and each hidden unit is linked to each output unit. The activation of unit i is a logistic function of the weighted sum of activations received by the unit plus the unit's bias.

Prior to learning, all weights and biases in the net are initialized to small, non-zero real numbers selected at random. For any given sequence, the "context + input" units register the sequence as it is presented to the net. The presentation of each successive event in the sequence causes activation to propagate through the net beginning at the input units, generating expectancies for the next event as output. Initially, these expectancies will be randomly generated by the untrained net.

Learning is accomplished by changing the weights and biases incrementally after each event so as to reduce the disparity between what the net expects— its output—and what actually occurs—the next event in the sequence. Each event in the sequence is thus the target value used to train the expectancies generated by the previous sequence events. The algorithm employed to change the weights and

Fig. 2. The pattern of activation induced by a context converges on the probability distribution of chords given that context.

Numbers 1–6 represent the six major and minor chords built upon major scale degrees 1–6.

Transition probability Activation

■ ▨

Context = 1

Context = 2

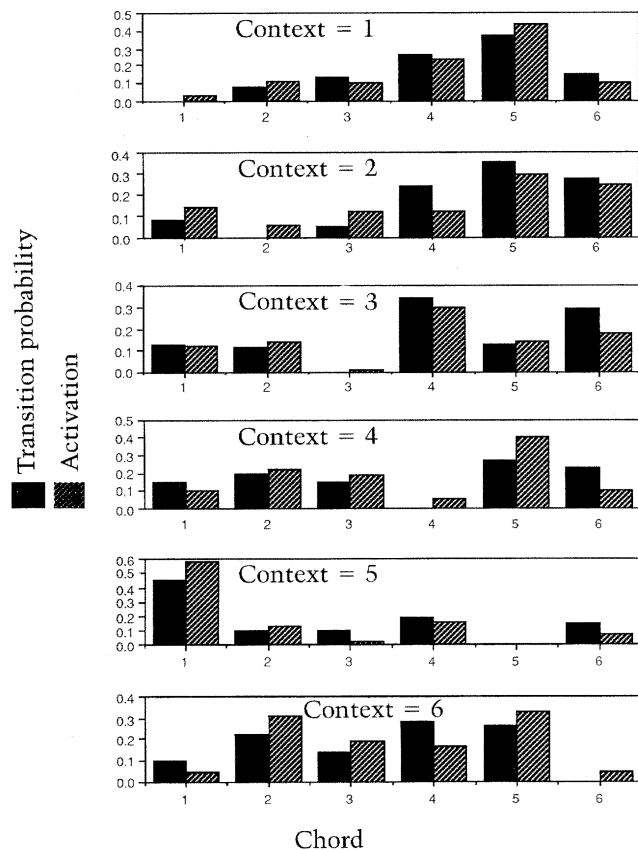Context = 3

Context = 4

Context = 5

Context = 6

Chord

Figure 2 shows the actual chord probability distribution and the net's output activation following each of six single-chord contexts. The net clearly matches the probability distributions in each case. The numbers 1–6 refer to the major and minor triads built upon the first six tones of the major scale. Note that the net has learned some of the sequential regularities of Western harmony. A tonic context chord generates strong expectations for the dominant and subdominant (top panel), a supertonic context chord generates expectations for the dominant and submediant (second from top), and so on. No rules needed to be encoded in order for these patterns to emerge; they simply reflect the internalization of probability distributions through extended exposure to individual sequences. We would argue that this model is considerably more plausible and parsimonious as an account of perception than are rule-based models.

Probability matching, as observed in the above net, has been shown in a number of psychological experiments. The basic result is that when trying to predict the next event after having witnessed events with a certain probability distribution, prediction patterns tend to align themselves with the probability distribution. This result is notable because it is not the optimal prediction strategy from the point of view of maximizing the expected return, which is to always predict the most likely event and never predict any other.

Probability matching accounts for the aura of schematic expectation that is generated at any given point in a musical sequence, in which no one event is expected to the exclusion of others; some events are highly expected, some are highly unexpected, and others have intermediate expectancies. Graded levels of schematic expectancy provide composers with alternatives that are only subtly different in their typicality and induce in the listener a range of expectancy confirmations and violations.

A prediction that derives from this result is that if subjects are asked to rate how appropriate a chord sounds following a single-chord context, the pattern of ratings would resemble the expectancy vector. Figure 3 shows a strong relationship between the rating judgments on this task, obtained by Bharucha and Krumhansl (1983), and the expectancies gener-

biases is the generalized delta rule (also known as back-propagation) developed by Rumelhart, Hinton, and Williams (1986).

The net was exposed to sequences that embody the transition probabilities of chord functions that are representative of Western music of the common practice era, estimated from Piston (1978). Any other set of sequences could have been used, and we plan to explore other actual and possible styles. After repeated exposure to the sequences, the net learns to expect (i.e., produce as output) the schematic distribution of chords for each successive event in a sequence. This net will not learn individual sequences, but will learn to match the conditional probability distributions of the sequence set to which it is exposed. In other words, each output vector approaches a probability vector representing the schematically expected distribution of chords following the sequence context up to that point.

Output activation of network
and relatedness judgments of human subjects
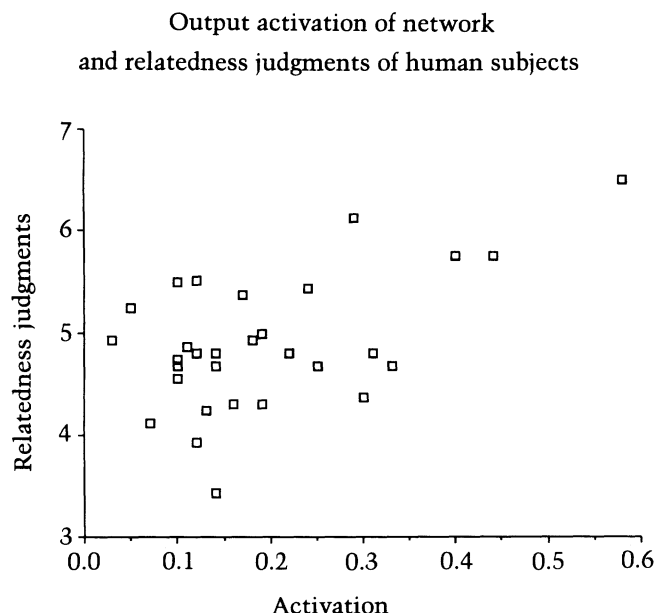
ated by our simulation. The sequential neural net model thus picked up the same sort of schematic expectancies about chord sequences that we find in human listeners.

Veridical Expectancies

The modeling described above focused on the generic cultural expectations and implications embodied in the schematic expectancies of music listeners. But listeners know more than just what musical structures are likely in various contexts in their culture; they know exactly what event is to occur next at particular points in particular pieces of music with which they are familiar.

The sequential schematic expectancy model described above can be modified to learn specific sequences—and thus veridical expectancies—by the addition of input units that serve to distinguish individual sequences in some way. The resulting net structure, with the new group of input units labeled "additional context," is shown in Fig. 4. This architecture is based on the sequential net design proposed by Jordan (1986) and used by Todd (1988) to model melody learning.
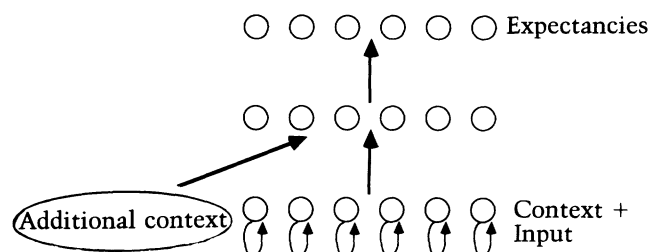
The "additional context" units in Fig. 4 individuate sequences by name or other discriminating context. In our simulations, there is one such unit (which we shall call a *name unit*) for each sequence to be learned by the net. Jordan called these units *plan units*, since he was simulating the production rather than the perception of sequences. Two sequences that are identical up to a point and then diverge can be learned by this net because the inputs for the two sequences would have a different name unit turned on for each sequence. This set of units could also be used to encode richer contexts that might include rhythmic, timbral, and other factors that contribute to the recognition of familiar musical sequences.

We exposed a net of the above sort, with six "context + input" units and six output units representing the six diatonic major and minor chords, to 50 sequences of seven successive chords each. Fifty "additional context" name units were needed to distinguish these 50 learned sequences.

After the net had learned these sequences, we studied its ability to learn two new sequences, one with schematically expected transitions and one with schematically unexpected transitions. The two sequences were matched in terms of the number of distinct chords in each. The sequence with schematic transitions started out with a lower summed squared error (on the output units) than the atypical sequence and was learned more quickly. The net thus learned a novel sequence more quickly if it conformed to familiar regularities. This result is in accord with the prevalent intuition that it is difficult to learn sequences of music from other cultures or from unfamiliar historical periods—that is, sequences that violate schematic expectancies.

50

Fig. 5. Number of cas-
caded activation steps to
reach asymptote for three
types of transitions. The
more common the transi-
tion, the more quickly the
expectancies reach
asymptote.

Combining Schematic and Veridical Expectancies in the Same Sequential Net

Even though the sequential net of Fig. 4 was employed to account for veridical expectancies, the above result suggests that this net acquires properties that are often attributed to a cultural schema. The net seems to inadvertently acquire these schematic properties even when the "additional context" units are operative, that is, even when it serves as a memory for individual sequences. More supportive evidence for this passive acquisition of sequential schematic expectancies from veridical sequence learning can be found by exploring the net's behavior when using cascaded activation.

Cascading was first described by McClelland (1979) and McClelland and Rumelhart (1988) and involves restricting the amount of activation that can pass through the net at a given time, thereby enabling one to observe the development of unit activation levels over time. The time-scale involved in cascading is different from the one involved in the generation of sequences by the net; the multistep cascading process occurs *within* each step of the outer sequence. Cascading is typically performed after the net has been trained to produce the proper sequential outputs, and it is then used to watch the activations of units develop from their initial values to the final asymptotic values they end up with as a result of training. For any given input, the input units exert only a fraction of their influence on the hidden units during the initial cascade time steps. The hidden units in turn exert only a fraction of their influence on the output units. Over succeeding time steps, each layer releases a greater fraction of its activation to the next layer, until the units have reached the asymptotic activations on which they were trained.
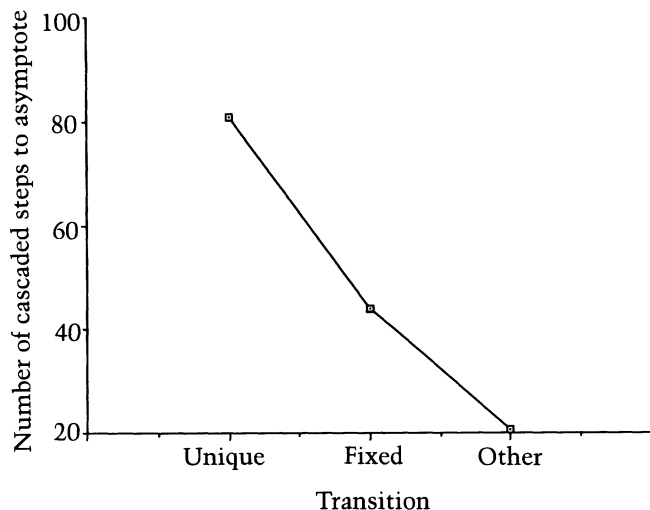
The cascading algorithm is as follows. The net input, $net_{i,t}$ to a hidden or output unit at time-step t is determined as follows:

$$net_{i,t} = k\sum[w_{ij}a_{jt}] + (1 - k)net_{i,t-1}$$

where the standard net input value, computed by summing up the products of the weights w_{ij} and the current activations a_{jt} of the units connected to

unit i, is multiplied by the constant k and added to a fraction of the previous net input.

The constant k is the cascade rate, which determines how fast activations in the net build up to their asymptotic levels.

When cascaded activation is used in a veridical sequence net trained with both schematically expected and unexpected events, the highly expected events reach asymptotic activation much more quickly than unexpected ones. In this way, we can see the effect of learned cultural schemas on the net's performance with particular sequences. In one simulation, we trained the net with sequences that embodied only schematically expected transitions with the exception of one unexpected transition in one sequence. After all the sequences were learned to criterion, including the unexpected transition, the net was tested with the cascading algorithm. Figure 5 shows the number of cascade time-steps it took the activation of the single "on" output unit to reach the trained veridical asymptote (here 0.9), starting from its asymptotic activation at the previous sequence step (usually close to 0.0) for each of three transition-type groups. A unique chord transition X-Y is one in which chord X is followed by chord Y in only one sequence and is therefore highly unexpected. A fixed chord transition X-Y is one in which X is always followed by Y and is therefore highly expected. The third category includes all the

other transitions of intermediate expectancy. As can be seen in the figure, unique transitions took longer to reach asymptote than fixed transitions, and the others fell in between.

These results indicate that the net used was embodying cultural schema information in its weights, which yielded fast cascade response times for expected transitions. In contrast, these schematic biases had to be overcome when unexpected transitions were being produced, leading to longer cascade times in the unique and other transition type cases. We can thus conclude that this net learned to embody cultural schematic expectancies, even though it was trained to produce merely specific veridical expectancies (the sequential outputs).

Conclusion

The studies reported here have demonstrated that neural net models embodying simple assumptions can learn musical schemas by passive exposure. These assumptions include the existence of general purpose learning architectures that implement competitive learning and supervised learning, and the existence of tonotopic pitch mapping in the auditory system. With these assumptions, the psychological regularities that have previously been attributed to rules develop in the net's behavior as an automatic consequence of exposure to a structured musical environment. Even though this environment typically only includes examples of veridical expectancies in the form of specific pieces of music, schematic expectancies of the likelihood of various musical events in the culture are also abstracted from this exposure.

Acknowledgments

References

Bharucha, J. J. 1987a. "MUSACT: A Connectionist Model of Musical Harmony." *Proceedings of the Ninth Annual Meeting of the Cognitive Science Society.* Hillsdale, N.J.: Erlbaum Press.

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

Bharucha, J. J., and C. L. Krumhansl. 1983. "The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context." *Cognition* 13: 63–102.

Bharucha, J. J., and K. L. Olney. 1989. "Tonal Cognition, Artificial Intelligence and Neural Nets." *Contemporary Music Review.* Forthcoming.

Bharucha, J. J., and K. Stoeckig. 1986. "Reaction Time and Musical Expectancy: Priming of Chords." *Journal of Experimental Psychology: Human Perception and Performance* 12:1–8.

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

Carpenter, G. A., and S. Grossberg. 1987. "A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine." *Computer Vision, Graphics, and Image Processing* 37:54–115.

Castellano, M. A., J. J. Bharucha, and C. L. Krumhansl. 1984. "Tonal Hierarchies in the Music of North India." *Journal of Experimental Psychology: General* 113: 394–412.

Cuddy, L. L., A. J. Cohen, and J. Miller. 1979. "Melody Recognition: The Experimental Application of Musical Rules." *Canadian Journal of Psychology* 33:148–157.

Dowling, W. J., and D. L. Harwood. 1985. *Music Cognition.* New York: Academic Press.

Jordan, M. I. 1986. "Attractor Dynamics and Parallelism in a Connectionist Sequential Machine." *Proceedings Eight Annual Conference of the Cognitive Science Society.* Hillsdale, N.J.: Erlbaum Press.

Kessler, E. J., C. Hansen, and R. N. Shepard. 1984. "Tonal Schemata in the Perception of Music in Bali and the West." *Music Perception* 2:131–165.

Kohonen, T. 1984. *Self-Organization and Associative Memory.* Berlin: Springer-Verlag.

Krumhansl, C. L., J. J. Bharucha, and M. A. Castellano. 1982. "Key Distance Effects on Perceived Harmonic

Structure in Music." *Perception and Psychophysics* 32:96–108.

Krumhansl, C. L., and E. J. Kessler. 1982. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." *Psychological Review* 89:334–368.

Linsker, R. 1986. "From Basic Net Principles to Neural Architecture." *Proceedings of the National Academy of Sciences* 83:7509–7512, 8390–8394, 8779–8783.

McClelland, J. L. 1979. "On the Time-Relations of Mental Processes: An Examination of Systems of Processes in Cascade." *Psychological Review* 86:287–330.

McClelland, J. L., and D. E. Rumelhart. 1988. *Explorations in Parallel Distributed Processing.* Cambridge, Massachusetts: MIT Press.

Meyer, L. 1956. *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

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

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

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

Schmuckler, M. A. 1988. "Expectation in Music: Additivity of Melodic and Harmonic Processes." Ph.D. diss., Cornell University.

Shepard, R. N. 1989. "Internal Representation of Universal Regularities: A Challenge for Connectionism." In L. Nadel et al., eds. *Neural Connections and Mental Computation.* Cambridge, Massachusetts: MIT Press.

Todd, P. M. 1988. "A Sequential Network Design for Musical Applications." In D. Touretzky, G. Hinton, and T. Sejnowski, eds. *Proceedings of the 1988 Connectionist Models Summer School.* Menlo Park: Morgan Kaufmann.

Weinberger, N. M., and T. M. McKenna. 1988. "Sensitivity of Auditory Cortex to Contour: Toward a Neurophysiology of Music Perception." *Music Perception* 5:355–390.