Jamshed J. Bharucha
Psychology Department
Dartmouth College
Hanover, New Hampshire 03755 U.S.A.
bharucha@dartmouth.edu

Pitch, Harmony, and Neural Nets: A Psychological Perspective

Two divergent but overlapping goals characterize work on neural net models of music. One is the attempt to understand and develop intelligent musical systems. Networks proposed in pursuit of this goal may be called artificial networks (or models thereof). This is an important goal of research in computer music. The other goal is the attempt to understand a given intelligent musical system: the human brain. Networks proposed in pursuit of this goal may be called human networks (or models thereof). This is the primary goal of research on the psychology and neuroscience of music. Since the brain is only one of presumably many possible intelligent systems, there are necessarily more constraints on models of human networks.

The pursuit of artificial networks follows the machine design methodology of engineering. One begins with a task (such as extracting pitch from an acoustic instrument signal), and then asks the question, How can a machine be designed to carry out this task? The test of the ensuing model's adequacy is whether it carries out the stated task. Alternative models based on very different approaches could in principle carry out the same task. A comparison of alternative models would typically be based on such criteria as how efficient they are and how well they scale up.

The pursuit of human networks follows the hypothesis-testing methodology of basic empirical science. The question here is, How (either physically or functionally) does the brain carry out this task? The test of the ensuing model's accuracy is not only whether it carries out the stated task but whether it also carries out other tasks that are known to be performed by the same system and whether it satisfies all known constraints on the structure and function of the system. Although the additional constraints would result in fewer possible models than for artificial networks, there will nevertheless be alternative models. A comparison of alternative models would be based on the testable predictions that they make and the outcomes of experiments designed to test these predictions.

For example, a neural net that extracts pitch from a frequency fails as a model of human networks if it cannot account for the corpus of laboratory data on shifts of residue pitch (Patterson 1973; Schouten, Ritsma, and Cardoza 1962). A neural net that learns melodies fails as a model of human networks if it cannot account for key-distance effects associated with transposition in the short term (Cuddy, Cohen, and Miller 1979). A neural net that extracts chords fails as a model of human networks if it cannot account for the corpus of data on the perceived relationships between chords (Bharucha and Krumhansl 1983; Bharucha and Stoeckig 1986, 1987; Krumhansl, Bharucha, and Castellano 1982).

This chapter has two goals. The first is to focus attention on some issues pertinent to the development of models of human networks for music perception. The second is to sketch aspects of a particular model. The chapter will begin with a discussion of some issues concerning the choice of algorithm. It will then launch into a brief review of psychological accounts of pitch, followed by a critique of the psychological adequacy of pitch representations adopted in extant models. The remainder of the chapter will summarize continuing work on a model that originated as a spreading activation network for harmony (Bharucha and Stoeckig 1986) and evolved into a constraint satisfaction network called MUSACT (Bharucha 1987a). MUSACT extracts chords from tones and keys from chords. It accounts for the effect of tonal contexts on the perception of tones and chords, and the establishment of keys. The focus in this chapter will be on how MUSACT learns its connectivity, and on how networks that process pitch prior to MUSACT might learn their connectivity. A network for achieving invariance under transposition is also summarized

(Bharucha 1988). This invariance mechanism feeds a sequential memory that is described in another chapter (Bharucha and Todd, this volume).

Supervised Learning Algorithms and Expectancy Violation

The choice of a neural net algorithm (supervised or unsupervised) depends at least as much on the goal (artificial versus human networks) as on the task to be modeled. Supervised learning (e.g., the back-propagation algorithm of Rumelhart, Hinton, and Williams 1986) requires that the output of the network in response to a given input be compared to a target or desired output. The difference is the error signal, which the algorithm seeks to minimize. The pursuit of human networks limits the use of supervised learning algorithms to tasks in which target vectors are clearly available to the system under natural conditions.

Consider a central problem in speech, that of learning to extract discrete phonemes from a speech signal. If the goal is to build an artificial network that can perform this task, one might consider a back-propagation network that learns to identify the phoneme given the speech signal. This option is available because one can supply the network with the target phonemes for each input signal (e.g., Landauer and Kamm 1987). If the goal is to model how humans learn, however, this option is not available. Humans don't have any way of knowing what the target phoneme is except by extracting it from the speech signal, which is of course what we want to learn in the first place.

A similar example in music concerns the ability to extract chords from complex spectra. Laden and Keefe (this volume) invoke back propagation to learn this. The network receives a spectrum as input and the correct chord as the target and thereby learns to identify the chord given the spectrum. A music student who receives feedback from a teacher while learning to classify chords as major or minor would indeed have the necessary target. One can demonstrate in a psychological laboratory, however, that Western subjects with no musical training, who are unable to name chords, are nevertheless able to make judgments about chords and their relationships; these judgments are extraordinarily sophisticated and consistent with the explicit descriptions of music theorists (Bharucha 1987b; Bharucha and Stoeckig 1986). This implicit or tacit knowledge of chords must have been obtained through passive perceptual exposure without feedback; supervised learning is thus inappropriate here.

When the use of supervised learning is unsupported by the learning conditions that would exist for a human, it is necessary to search for unsupervised, self-organizing networks (such as those proposed by Fukushima 1975; Grossberg 1976; Kohonen 1984; Rumelhart and Zipser 1985; von der Malsburg 1973). These networks are compelling psychologically because they do not require target vectors in order to learn. They are compelling neuroscientifically because they involve a learning mechanism about which there is a consensus, namely, Hebbian learning (Hebb 1949). These algorithms have been employed to model the learning of melodic patterns (Gjerdingen, this volume) and to model the acquisition of the MUSACT structure for harmony, as described below.

Are supervised learning algorithms therefore of limited value for modeling human networks? Gjerdingen (this volume) argues in the affirmative, stating that "for higher-level tasks, the notion of an explicit teacher becomes problematical." Gjerdingen's blanket rejection of supervised learning is flawed, however, because it ignores tasks for which target vectors are available and necessary—tasks that depend upon registering an error signal.

Learning musical sequences is a task for which error signals are both available and necessary. They are available because each event is the target to be compared with the expectation generated prior to that event. They are necessary because the error signal plays an important role in the aesthetic or emotional response to music. It is widely held that expectancy violation is an important aspect of the aesthetics of music (Meyer 1956). In the sequential model discussed by Bharucha and Todd (this volume), the error signal that drives learning is precisely the information that would lead to expec-

tancy violation. Thus both learning and the experience of expectancy violation are driven by the same process.

This would suggest that Gjerdingen's model, billed as unsupervised, either fails to account for expectancy violation or doesn't utilize this information for learning. But Gjerdingen is clearly concerned with expectancy violation, and closer examination reveals that, contrary to its billing, the model does indeed rely on this information for learning. The model registers the disparity between what it expects and what actually occurs, the latter being a target in supervised learning. If there is a disparity, "the network's orienting subsystem will automatically react to the mismatch." The model is thus inescapably supervised, and the same error signal that drives the learning also explains expectancy violation.

Psychological Background for Modeling Pitch and Harmony

Psychophysics

The modern scientific study of pitch perception has been conducted primarily by psychologists and neuroscientists. In psychology, most of the work prior to the 1970s was in psychoacoustics, a branch of psychophysics. Psychophysics is the study of how dimensions that characterize the physical world are transformed into dimensions that characterize perception. For example, frequency is a physical dimension whose psychological correlate is pitch. The mapping from frequency to pitch, as employed in music, is logarithmic.

Perhaps the most valuable contribution of psychophysics to our understanding of pitch perception in music is research on the extraction of pitch from complex spectra. The pitch of a complex wave with a harmonic spectrum is generally perceived to be identical to the pitch of a sine wave of the same frequency as the fundamental frequency of the complex wave, whether or not the complex wave contains energy at the fundamental. Helmholtz (1885/1954) explained this "missing fundamental" phenomenon as a distortion product introduced by the auditory system. This view is incorrect because a pitch is perceived at the missing fundamental even when the region of the fundamental is masked with noise (Licklider 1954). Furthermore, the pitch is not necessarily perceived at the frequency of a distortion product (deBoer 1956; Schouten, Ritsma, and Cardoza 1962). It was subsequently thought that perceived pitch (often called the *residue* or *virtual* pitch) is computed from the shape of the waveform resulting from adding the components (deBoer 1956). This theory fails because residue pitch is not sensitive to the relative phases of the components, whereas the shape of the waveform is (Patterson 1973). The perceived pitch thus seems to be the result of a pattern recognition system that performs a best fit of the component frequencies to harmonic spectra that have been learned (Goldstein 1973; Terhardt, Stoll, and Seewann 1982). Because this is a pattern recognition process and because learning seems to be involved, neural nets are ideally suited to modeling pitch extraction.

Aside from pitch extraction, much of the psychophysical study of pitch has dealt with identification, discrimination, and magnitude estimation of pure tones (Green 1976). Although a wealth of knowledge has been gleaned from studies of the pitch of pure tones, the implications for understanding music are limited and can be misleading. For example, scales based on measures of JND (just noticeable differences) between pitches, or on the subjective estimation of magnitudes (the *mel* scale of Stevens and Volkmann 1940), are unable to account for one of the most salient constraints on the perceptual scaling of pitch, namely, the similarity of pitches that are separated by octaves.

Cognitive Psychology

The central limitation of psychoacoustic studies of pitch was that pitch was construed as one dimensional. Cognitive psychologists refer to this dimension as *pitch height*. Pitch height is the dimension of pitch that is described by up and down. As one pitch is raised relative to another, the striking similarity between the two that is noticed at octave in-

tervals requires another dimension of pitch that is described by up and down. As one pitch is raised relative to another, the striking similarity between the two that is noticed at octave intervals requires another dimension in the psychological scaling of pitch (Deutsch 1973; Shepard 1964). When the pitch-height continuum is collapsed across octaves, the relative position of a tone within any given octave is referred to as its *chroma*. It is the chroma dimension that is captured by the octave equivalent letter names, or *pitch classes* (e.g. C, C#, etc.), used in Western music. Other dimensions of psychological pitch space are imposed by the consonance and dissonance of other intervals, such as perfect fifths (Shepard 1982). In the context of a piece of music, pitch classes reveal affinities that define additional dimensions. For example, the three tones of the tonic triad are perceived to be closely related in the context of a piece of Western tonal music (Krumhansl 1979), and diatonic tones are more easily confused with each other than with nondiatonic tones (Dowing 1978; Krumhansl 1979). Analogous effects occur in experiments with the music of other cultures. For example, tones that belong to a *raga* (one of the traditional melodic patterns or modes in Indian music) are perceived to be closely related in the context of a piece of North Indian music (Castellano, Bharucha, and Krumhansl 1984). Cognitive psychologists thus consider pitch to be a multidimensional attribute (Krumhansl 1990).

Cognitive psychologists have also been concerned with the mental processes that compute these pitch relationships. Many of these processes operate automatically (i.e., without conscious awareness) and are not limited to listeners who have had formal musical training. The processes that are of primary interest to this author are those that occur in the minds of people with little or no formal musical training. For example, Western listeners have an elaborate representation of tonal relationships in harmony, even if they have had no formal musical training. This can be shown in a priming task, in which the speed and accuracy with which a musical event is processed can be measured as a function of the preceding musical context. For example, when a major chord is musically related to the prior context, it is processed more quickly and more accurately, and is heard as more consonant, than when it is musically unrelated to the prior context (Bharucha and Stoeckig 1986, 1987). The speed and accuracy of judgments varies monotonically with the distance, around the circle of fifths, of the chord from the prior context (Bharucha 1987b).

Subjects ranging from those with no formal training to professional composers show similar effects. The only apparent prerequisite to exhibiting these data is having grown up in Western society, in which the music that is most pervasive, and impossible to avoid, is overwhelmingly based on triads in tonic, dominant, and subdominant relationships.

Learning

Which aspects of pitch are innately specified and which are learned? Are they universal or culture relative? Although these questions are relevant if one is modeling human networks, they should be asked with care. A mechanism can be innately specified but can fail to develop if the necessary environmental conditions are not present. Conversely, a perceptual phenomenon can be universal simply because the environmental constraints are universal. In this latter case, there may or may not be learning (see Bharucha and Olney 1989).

Harmonic spectra are found universally, because of the physical properties of the human voice and other naturally occurring periodic signals. If complex tones whose pitches are an octave apart are universally judged to be similar, this could be accounted for on the basis of the presence of octave harmonics in natural periodic signals. Octave equivalence would thus be neither innate nor learned but would simply presuppose an auditory mechanism, such as the *place mechanism* in the ear, that registers spectral similarity.

Whereas some relationships can be found in a comparison of the spectra themselves, others can be found in the array of pitches that are extracted from spectra. Parncutt (1989) argues that the perceived relationships in Western harmony can be accounted for in this fashion, eliminating the need for higher level processes, innate or learned, that mediate our perception of harmony. According to this

view, the complex spectrum of a musical chord induces a number of pitches, called *virtual* pitches, as described by Terhardt's pitch extraction model (Terhardt, Stoll, and Seewann 1982). The more pitches there are in common between two chords, the more closely related will the chords be perceived. Thus, although Terhardt's pitch extraction mechanism itself involves learning based on exposure to speech, the relationships that define harmony are driven (in a bottom-up or date-driven fashion) by the structure of harmonic spectra via the extracted virtual pitches, and no additional circuitry specific to harmony is necessary.

Although Parncutt's theory may be sufficient to account for important aspects of harmony, including the origins of Western harmony, it is not sufficient to account for all of the available data. His theory predicts that chords that share tones will be more closely related than chords that don't. Yet the circle-of-fifths relationship between chords violates pitch commonality. The circle of fifths is a spatial arrangement of chords that depicts their proximity in Western tonal music. (The circle of fifths for major chords, expressed in enharmonic equivalents, is as follows: C,G,D,A,E,B,F♯,C♯,G♯,D♯,A♯,F,C). C and D major are closer along the circle of fifths than are C and A major, yet C and D have no tones in common, but C and A have one tone in common. Thus, the commonality of pitches cannot explain the circle of fifths, and the circle of fifths is not just a theoretical construct but emerges from the responses of subjects in psychological experiments.

It is clear, then, that the perceived structure of pitch as it occurs in Western harmony is not driven by characteristics of harmonic spectra alone or by characteristics of the pitch extraction process. Additional evidence comes from an experiment in which spectral components that are shared by related chords are removed. Even without spectral overlap, chords are processed more quickly and more accurately and are perceived as more consonant when preceded by chords that are closely related along the circle of fifths than when preceded by chords that are distant along the circle of fifths (Bharucha and Stoeckig 1987). Relationships such as those described by the circle of fifths thus must have been learned. Since most people do not receive rule-based instruction in the theory of harmony, this learning must occur from passive exposure, presumably over a long period of time.

The view advanced in this chapter is that most of the dimensions of pitch are learned through extended passive exposure. Some aspects of learning are driven by patterns that are universally pervasive (such as octaves in the harmonic spectra of speech) and are thus likely to be manifested universally. Other aspects of learning are driven by patterns that are pervasive only within a culture (such as major and minor chords) and are thus likely to be culture relative. In either case, since the human brain is an extraordinary learning machine, patterns that are pervasive in the environment are likely to be learned inadvertently. This learning can yield emergent phenomena that themselves cannot be accounted for by environmental patterns alone. Neural nets have particular promise as models of this form of learning. The circle of fifths can be shown to emerge from a neural net that learns the typical clustering of tones to form chords and the typical clustering of chords to form keys.

Neuroscience

At least two classes of neuroscientific data are relevant to the development of neural net models of music perception. The first concerns the tuning characteristics of neurons, and the second concerns the anatomical dissociation of musical functions.

In a neural net model, learning consists of changing the weights of links between neural units. This is tantamount to changing the response characteristics or tuning of neural units. In support of this, Weinberger and Diamond (1988) report a response plasticity of neurons in the auditory cortex following associative learning.

Self-organizing algorithms (e.g., Fukushima 1975; Grossberg 1976; Kohonen 1984; Rumelhart and Zipser 1985; von der Malsburg 1973) predict that, as a result of weight changes, certain neurons will become tuned to complex patterns that represent the clustering of simple features. Beyond some preliminary suggestions of the existence of neurons

that respond to pitch contour (Weinberger and McKenna 1988), little is known about whether or not neurons with complex tuning characteristics exist in the auditory system and, if so, what these tuning characteristics are.

Neurons in the early stages of auditory processing have innately specified tuning characteristics. Because of the place coding of the cochlea, a neuron that receives excitation from the basilar membrane is tuned to a particular frequency (its *characteristic* frequency). Adjacent neurons have slightly different characteristic frequencies, with overlapping receptive fields. Collectively, these neurons constitute a tonotopic representation of the audible frequency range, scaled logarithmically (see Sano and Jenkins, this volume).

A tonotopic mapping of log frequency is also found in the auditory cortex (Lauter et al. 1985), albeit with less clarity. The presence of a tonotopic mapping throughout the auditory system and the absence of robust evidence of the existence of more abstract tuning characteristics should not preclude the postulation of abstract representational units in neural net models. Indeed, neural net models can play a valuable role in making predictions for the neuroscientific study of response selectivity, provided the models are of human and not artificial networks. Thus, the MUSACT model described in this chapter predicts the formation of units tuned to tone clusters that are typically encountered, such as chords.

One reason why little other than the tonotopic mapping is known is that neuroscientific studies have typically employed pure tones, thus limiting the discovery of neurons with more complex response characteristics. This predicament was noted by Deutsch (1969) and is not much different today. Commenting on Goldberg and Lavine's (1968) bewilderment at the "surprisingly large number of unresponsive units" (p. 331) in the auditory system, Deutsch (1969) states that this "would hardly be surprising, since animals in their natural environment are much more concerned with auditory pattern recognition than with pure tones" (p.304).

Neuroscientific studies have also provided information about the dissociation of musical functions. Perhaps the greatest consensus concerns the hemispheric dissociation between temporal and atemporal processes, for which the left and right hemispheres, respectively, are dominant. Consequently, in the model sketched in this chapter, an atemporal structure, MUSACT, is dissociated from a temporal memory for sequences.

Another example of a dissociation comes from the study of a stroke patient, named M. S., who lost all of his primary auditory cortex and much of his nonprimary areas as well (Tramo, Bharucha, and Musiek 1990). M. S. was unable to detect spectral changes (in musical chords) that would be simple for normal subjects. Yet he showed evidence of chord priming as described above. These results suggest that tacit knowledge of the relationships between chords is at least partially dissociable from the mechanism that permits fine-grained comparisons of spectra.

The Representation and Organization of Musical Pitch in Neural Nets

How should pitch be represented? What constraints can be brought to bear on this decision? In this section we shall consider the representation of pitch as it funcitons in the perception and encoding of pitch sequences and pitch clusters in music.

The first modern proposal of a neural net model for musical pitch that was motivated by a discussion of a wide range of constraints was due to Deutsch (1969). Although this model is modest by today's standards, it anticipates some central architectural characteristics of contemporary models, including the use of representations for spectra, pitch class, intervals and chords. Deutsch offers a discussion of the merits of different representations. A discussion of selected representations can be found in other chapters (Laden and Keefe, this volume; Sano and Jenkins, this volume; and Todd, this volume). Since several important points have not been touched on in these chapters, the present chapter will attempt to do so, with the perspective of including as many known constraints as possible on the selection of representations.

Spectral Representation

A spectral representation is one that most resembles the acoustic signal in the frequency domain. The existence of tonotopic representations throughout the auditory system, including the cortex, might suggest that the perception and encoding of pitch in music is accomplished directly with spectral representations. Laden and Keefe (this volume), for example, suggest that a spectral representation obviates more abstract representational units such as pitch class. They note that spectral representations "are motivated by the spectral structure of musical sound, the physiological structure of neural activation patterns in the auditory system, and the pattern recognition mechanisms of complex pitch perception theories."

Although there can be no doubt that spectral representations are necessary and do in fact exist, what's at issue is whether they are sufficient for the representation of pitch, particularly as it functions in music. Laden and Keefe argue that a "musical tone needs to be specified as a number of pitch classes that approximate the harmonic partials of the tone rather than as a single pitch." This view obscures the distinction between frequency and pitch and is at odds with what we know from psychophysics. It is difficult even to articulate the known phenomena on pitch perception without assuming two kinds of representations, one for frequency spectra and one for pitch, such that the former begets the latter.

Since a spectral representation encodes a tone as its spectrum, no distinction is made between the many frequency components and the singular pitch that is typically heard. A complex periodic waveform is heard as a singular pitch (the synthetic mode of pitch perception, which is the norm) and not as many pitches, with one for each harmonic (the analytic mode of pitch perception, which requires training). A melody played on an instrument or sung is heard as a sequence of unitary pitches, not as a sequence of pitch clusters.

Furthermore, from one spectral event to the next, a pitch may be heard as rising, falling, or staying the same. Two tones, *x* and *y*, can be constructed such that *x* has a higher pitch than *y* even though the center frequency of *x* is lower, or even though all frequencies in *x* are lower. Laden and Keefe would argue that although *x* has lower acoustic components, the spectrum is filled out via a pattern completion process, hence the psychological representation of the spectrum is richer than the physical one. Although it is plausible that spectral pattern completion occurs and accounts for important psychoacoustic phenomena, it is most unlikely that filled-out spectra alone can account for the perception of pitch. What aspect of the filled-out spectra of *x* and *y* account for *x* being heard as higher than *y*? Perhaps one could compare the (actual or filled-in) fundamental frequencies to determine which is higher. But the spectral representation by itself gives no special status to the fundamental frequency, hence it is not equal to this task. There must be a subsequent stage of processing in which the pitch (usually at the actual or filled-in fundamental) is extracted and represented as unitary. Sano and Jenkins (this volume) postulate just such a mapping from a spectral representation to a unitary pitch representation, although they do not provide a mechanism by which it is learned. In any case, the spectral representation alone is not a representation of pitch at all but rather an elaborated representation of the signal, from which pitch is extracted.

Pitch-Height Representation

If pitch height is extracted from a spectral representation, perhaps this is a level of representation sufficient for representing pitch as it functions in music. In network models, this representation can consist of an array of units, each of which responds selectively to a particular pitch, such that collectively they span the audible pitch range.

There can be little doubt that we have a pitch-height representation, since the up/down dimension of pitch is perhaps the most salient. Although most people have poor long-term memory for pitch height (i.e., most people do not have *absolute pitch*), we do indeed have the ability to remember pitch height over the short term. If two tones are played in succession, a judgment of whether the sec-

ond tone is higher or lower in pitch requires that the first tone be represented in memory long enough to make the comparison.

Although a pitch-height representation may be necessary in models of human musical cognition, it is not sufficient, for at least three reasons. First, it ignores octave equivalence. Often some of the units in a pitch-height representation are labeled with a pitch-class name followed by the octave number (e.g., C3, C♯3, . . . , B3, C4 . . .). This scheme captures octave equivalence only in the labeling, not in the representation. Second, a pitch-height representation entails the ability to remember absolute pitch levels over the long term (this may be called long-term absolute pitch). In other words, if the only pitch representation we have is pitch height, the original pitch-height levels of a tune will be remembered if the tune is remembered at all. But most people show little evidence of this. The third reason why a pitch-height representation is not sufficient (related but not identical to the second reason) is that it ignores *invariance under transposition*, that is, the ability to recognize a sequence transposed to any reference pitch.

The infinite number of pitches in the pitch-height continuum is usually broken up for representational purposes into a finite number of pitch-ranges or bins. For instance, a pitch-height representation could have a unit for every JND of pitch in the audible range. Pitch height representations that are much sparser than this have often been postulated as well. An example of this is a representation with only the twelve chromatic categories in each octave (e.g., C3, C♯3, . . . , B3, C4 . .). Such a representation may be called a *categorical pitch-height* representation. Todd (this volume) has used such a representation as the input to a network that learns musical sequences and then composes new ones. Sano and Jenkins (this volume) assume that there are units only at the semitone foci and that actual pitch deviations from the category foci are assimilated to the categories.

Categorical pitch-height representations have little psychological basis, since pitch-height categories are not anchored at absolute points along the audible frequency continuum. Without a physical reference such as a tuning fork, there is little agreement (across individuals or across occasions for an individual) about the location of the pitch A4, the traditional Western tuning reference of 440 Hz. This contrasts with the situation in vision. Although color categories have fuzzy boundaries, the three main color categories (red, green, blue) are anchored at absolute focal points along the visual wavelength continuum, and there is universal agreement about these foci, even in cultures (such as the Dani) that do not have names for these categories (Heider 1972).

Interval Representation

In an interval representation (discussed by Todd, this volume, as a pitch-interval representation), a pitch sequence is represented in terms of the successive intervals between tones. Each interval defines the pitch distance from one tone to the next, in semitone or log frequency units. A sequence of n tones would thus be represented by $n-1$ intervals (and possibly the starting pitch).

This representation is compelling until one considers further constraints. Although some might contend that we hear intervals, few would deny that we hear individual tones. If two tones are played in succession, separated by a brief period of silence, one's perceptual impression is of hearing two events separated in time, the events beginning at the onsets of the tones. Strong contrary evidence would be required in order to maintain that the memory representation of the sequence is fundamentally different from the perceptual representation, such that only the relationships between the perceptually salient events, and not the perceptually salient events themselves, are encoded in memory.

An interval representation requires that at least the first two tones of a sequence be heard before recognition is possible. This is surely false. A piece can be recognized by its very first event if the combination of tonal, temporal, and timbral cues is sufficiently unique. Since timbre is the property of the acoustic event itself, and not a property of the interval between two acoustic events, the acoustic events must necessarily be encloded. In other words,

it would be bizarre if timbre were indexed by the acoustic event but pitch by the relationship between acoustic events.

An interval representation doesn't account for tonal confusability. If a brief diatonic melody is followed by a (transposed) melody that is either the same melody or has one tone changed (while preserving the contour), listeners are more likely to judge the second melody to be the same as the first if the changed tone is also diatonic than if it is nondiatonic (Dowling 1978). In other words, changes that preserve diatonicity are more difficult to detect than changes that disrupt it. An interval representation alone would not make this prediction, because transitions from diatonic tones to nondiatonic tones traverse intervals that are also found between diatonic tones. That is, the intervals between all possible pairs of tones from the major diatonic scale encompass *all* possible intervals (e.g., me-fa is one semitone, do-re is two semitones, and so on). There are no new intervals introduced by including nondiatonic tones. Thus the interval traversed by a nondiatonic change will, taken alone, provide no information about a violation of diatonicity, unless the scale degree is also encoded as a reference. But if scale degree is to be represented, there's no need for the interval. (If one defines intervals not in terms of number of semitones but in terms of the theory-laden nomenclature, in which, for example, an augmented second is different from a minor third, even though both traverse three semitones, then indeed new intervals will be introduced by nondiatonic tones; however, a representation defined over these theory-laden categories presupposes considerable processing prior to the representation in question.)

An interval representation predicts that a single mistake will cause the key to transpose suddenly, making recovery difficult (Todd, this volume). After playing a wrong note, a performer is likely to recover by playing the next note correctly. An interval representation predicts that in this case the listener hears two errors, the first incorrect interval followed by the performer's compensatory interval. If the performer fails to compensate and skillfully transposes from then on, the listener ought to hear this as a less erroneous performance than the compensatory one, because only one interval error is made. This seems unlikely.

Pitch-Class Representation

In a pitch-class representation, pitch height is collapsed across octaves, but absolute pitch levels within an octave are preserved. This satisfies octave equivalence but is still not sufficient, because it entails long term absolute pitch and fails to account for invariance under transposition, for the same reasons given earlier for pitch height. Furthermore, a pitch-class representation that is restricted to the twelve chromatic categories (a *categorical pitch-class representation*) suffers from the same difficulties as the categorical pitch-height representation.

Although not sufficient, a pitch-class representation is necessary, since transpositional invariance is not uniform in the short term. If a tonal sequence is quickly followed by a transposition to another key, recognition is better for transpositions to related keys than to unrelated keys (Cuddy, Cohen, and Miller 1979). Here relatedness is defined in terms of distance along the circle of fifths, which can be computed in terms of the pitch classes in the keys of the two sequences. Since the perception of the transposed sequence is thus influenced by the relatedness between the pitch-class levels of the two sequences, information about the pitch-class levels of the first sequence must still be available when the second sequence is heard. In the model presented in this chapter, a pitch-class representation resonates over a short period of time and computes a pitch-class representation of keys or tonal centers, thereby producing such key-distance effects. Another representation, an invariant pitch-class representation, supports the encoding of sequences into a sequential memory.

Invariant Pitch-Class Representation

An invariant pitch-class representation is necessary for the long-term encoding of pitch sequences,

in order to account for invariance under transposition. In such a representation, all sequences are normalized into a common set of invariant pitch categories by coding them with reference to a sequence-specific origin. In the case of music with a tonal center, the tonal center would serve as the origin. The invariant pitch categories for the twelve chromatic pitch classes of Western tonal music utilizing equal-tempered tuning would be encoded in a network model by units tuned to twelve equally spaced points, {0,1,2,3,4,5,6,7,8,9,10,11}, along a continuum spanning an octave. (A gating mechanism that will transform a sequence into an invariant pitch-class representation of this type is shown in Fig. 10.) The sequential memory reported by Bharucha and Todd (this volume) for long encoding of sequences presupposes such an invariant pitch-class representation.

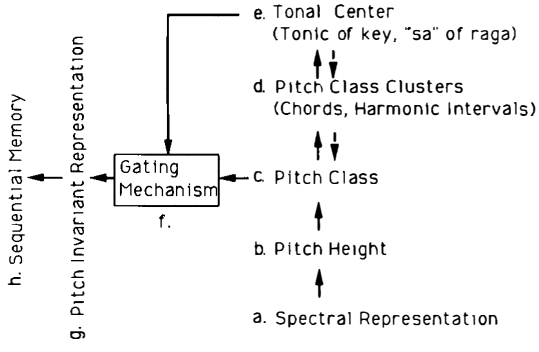
A representation that is denser than the twelve chromatic categories would be able to accomodate other tunings as well. The pitch class representation in Fig. 9, the gating mechanism in Fig. 10, and the sequential memory of Bharucha and Todd (this volume) should be assumed to utilize arrays that are denser than those shown in the figures for convenience.

Some invariant pitch-class representations (e.g., Gjerdingen, this volume) consist of only the diatonic categories (do, re, me, fa, sol, la, ti), with additional units to specify sharp or flat. Aside from the psychological implausibility of units for sharp or flat, such a representation presupposes a mechanism for assigning diatonic categories, is specific to only one kind of scale (e.g., major), and is too sparse to accommodate tuning differences.

A Model

An overview of a model of pitch organization and memory is sketched in Fig. 1. Items *a* through *e* represent hypothetical layers of neuronal units that accomplish the extraction and organization of pitch. Pitch height, *b*, is extracted from a spectral representation, *a*. Layers *d* and *e* represent the organization of pitch in the form of a learned musical schema

Fig. 1. An overview of the model (an expansion of an overview from Bharucha 1987a). Layers a–e *represent the extraction of pitch from signals and the organization of pitch in the form of a learned musical schema of chords and keys. Layers* c–e *comprise the MUSACT network shown in Fig. 9. The gating mechanism,* f, *is shown in Fig. 10; it uses the tonal center to gate absolute pitch class into a pitch invariant format,* g, *for encoding into a sequential memory,* h.



of chords and keys, defined over pitch classes in layer *c*. A gating mechanism, *f*, converts pitch class into an invariant pitch-class format, *g*, that feeds into a sequential memory, *h*. Separating the sequential memory, *h*, from the network that encodes tonal relations, *c* through *e*, is mandated by the known neurological dissociation between these two functions, as described earlier.

The model presupposes spectral, pitch-height, pitch-class, and invariant pitch-class representations. The known psychological constraints on pitch perception in music can be accounted for by attending to different representations to varying degrees or by allocating one's attention in different ways within a representation. For example, the synthetic mode of pitch perception, which is the norm, involves attending only to the most highly activated units in the pitch-height representation. The analytic mode of pitch perception, which enables one to hear multiple pitches within a complex tone but takes practice, involves attending to pitch-height units with low levels of activation.

Portions of this overview are reported elsewhere and will only be summarized here (for layers *c* through *e*, referred to as MUSACT, see Bharucha 1987a, 1987b; for the gating mechanism, *f*, see Bharucha 1988; and for the sequential memory, *h*, see Bharucha and Todd, this volume). The remainder of this

Fig. 2. Each unit in layer c *is connected to each unit in layer* d. *Layer* c *is an array of pitch class units that is dense (i.e., not restricted to chromatic steps) and tonotopically repeating (i.e., each pitch class is represented by many units, each within a one-octave tonotopic strip). Layer* d *is an array of units that are initially agnostic because of random weights from layer* c. *Layer* d *is organized into clusters. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)*

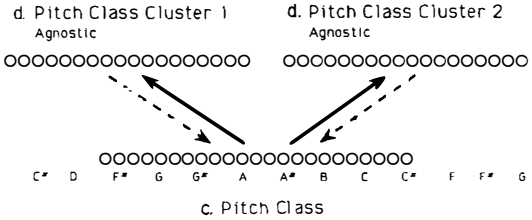
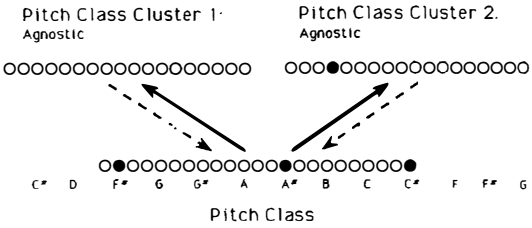


Fig. 3. An arbitrarily selected major chord, F♯ major, is presented, and turns on the pitch class units for F♯, A♯, and C♯. The units in the next layer are activated to varying degrees, and the winner happens to emerge in cluster 2. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)



chapter will focus on how the mapping from *c* to *d* is learned, giving rise to the connectivity that is assumed in the MUSACT model.

The Formation of Chord and Key Units through Self-Organization

In Fig.2, units in the bottom layer, *c*, are tuned to overlapping bands along the pitch-class continuum. This layer is dense (i.e., not restricted to chromatic categories) and tonotopically repeating (i.e., each pitch class is represented by many units, each within a one-octave tonotopic strip). Units in the top layer, *d*, are in clusters. Each cluster is tonotopically mapped but initially without an absolute pitch reference. Units in this layer are initially agnostic (not tuned to recognize any particular group of pitches) because of random weights from the pitch class units.

Now, let us imagine that an arbitrarily selected major chord, F♯ major, is presented and turns on the pitch-class units for F♯, A♯, and C♯ in layer *c* as shown in Fig. 3. The units in the next layer, *d*, are activated to varying degrees, and the winner happens to emerge in cluster 2. Note that any unit in any cluster could have been the winner, because the starting weights were set at random. But one unit must necessarily win, because the starting weights are real numbers that represent a continuum of possible synaptic strengths; hence the probability that two units will have exactly the same activation is infinitesimally small.

Fig. 4 depicts how the weights are changed. Weights to the winner are changed in a manner that is essentially Hebbian learning. Links that fed the winner (i.e., that provided positive input) are strengthened (solid lines); links to the winner that did not feed it are weakened (dashed lines). The winning unit in cluster 2 is on its way to being transformed from an agnostic unit to a unit that responds selectively to an F♯ major triad.

This learning procedure is based on the competitive learning algorithm of Rumelhart and Zipser (1985), which incorporates features from earlier algorithms by von der Malsburg (1973) and Grossberg (1976). Each cluster in Fig. 4 represents a unit in Rumelhart and Zipser's algorithm. What's new about the application of this algorithm to pitch is the tonotopically mapped clusters and the manner in which learning is yoked tonotopically as indicated below.

Parallel links are yoked within a cluster for purposes of weight change (Fig. 5). Thus, the unit in cluster 2 to the right of F♯ unit becomes tuned to a major chord slightly sharper than F♯. The yoking of weight changes on parallel links in cluster 2 results in a dense and tonotopically repeating cluster of units in that cluster that respond selectively to major chords (Fig. 6). A chord type (in this case, major) heard at one absolute pitch level will thus be recognized at any absolute pitch level by the same cluster.

If an arbitrarily selected chord of another type, F♯- minor (a flattened F♯ minor), is presented, the random winner is likely to be in an undeveloped cluster, say, cluster 1 (Fig. 7). Cluster 1 is then transformed by the same yoked weight-change mechanism from agnostic units to a dense and tonotopically repeating cluster of units that respond selectively to minor chords (Fig. 8). An analysis of the circumstances uner which new chord types become

Fig. 4. Weights to the winner are changed according to a modified version of competitive learning (Rumelhart and Zipser 1985). Links that fed the winner are strengthened (solid lines); links to the winner that did not feed it are weakened (dashed lines). The winning unit in cluster 2 is on its way to being transformed from an agnostic unit to a unit that responds selectively to an F♯-major triad. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)

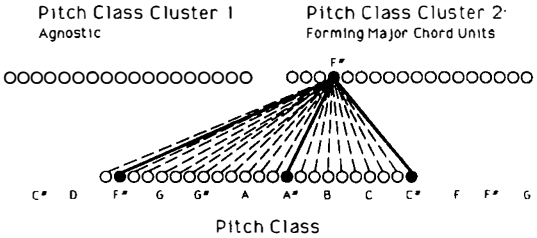
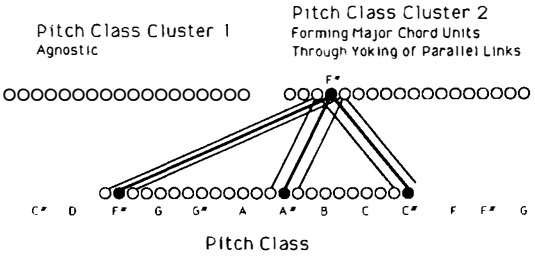


Fig. 5. Parallel links are yoked within a cluster for purposes of weight change. Thus, the unit in cluster 2 to the right of the F♯ becomes tuned to a major chord slightly sharper than F♯. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)



represented by new clusters rather than being assimilated by existing clusters is beyond the scope of this summary chapter. Interested readers are directed to Grossberg (1976) and Rumelhart and Zipser (1985) for formal discussions of this issue.

A similar mechanism will yield units at level *e* (see Fig. 1) that are specialized for typical chord groupings, provided activation at the chord units decays slowly. Thus, chords that are typically heard within the same piece of music become organized around a single unit at level *e*. These are therefore key units. Once again, they are tonotopically repeating.

MUSACT

The final network resulting from the learning in layers *c* through *e* consists of units representing pitch classes, chords, and keys (see Fig. 9, taken from Bharucha 1987a). In Fig. 9, only the chromatic units are shown, for convenience. All layers should be assumed to be denser than these categories. Furthermore, the arrangement of chords according to the circle of fifths is for convenience only. It should be clear from the learning procedure described above that the physical layout is actually tonotopic but can be depicted in any layout provided the links are preserved.

Input to the network is a sequence of events, each event being a simultaneous cluster of tones. Input is received by the spectral units and sent up through the layers. A unit may be activated from the bottom up or from the top down. A chord unit, for example, is activated either by the explicit sounding of some or all of its component pitch classes 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. Indirect activation of chord units permits smooth excursions (such as secondary dominants and modulations) from the focus of activation. A key unit is similarly activated by some or all of its daughter chords or by indirect influences, via its daughter chords, from related keys.

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 the activation of neighboring units. Phasic activation is used because of the salience of event onsets in music. Phasic units are commonly found in the nervous system. 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. Settling will occur provided the weights are small relative to the fan-in or fan-out of the unit connections.

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 along the circle of fifths. Note that the circle of fifths is exhibited as a truly emergent prop-

Fig. 6. The yoking of weight changes on parallel links in cluster 2 results in a dense and tonotopically repeating cluster of units that respond selectively to major chords. A chord type learned at one absolute pitch level will be recognized by the same cluster at any absolute pitch level. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)

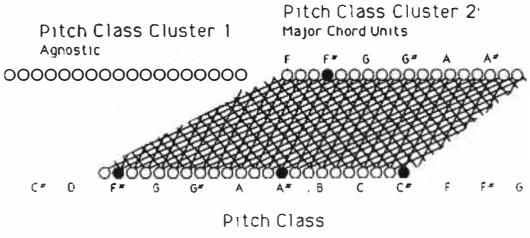


Fig. 7. If an arbitrarily selected chord of another type, F# – minor (a flattened F# minor), is presented, the random winner is likely to be in an undeveloped cluster, say, cluster 1. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)

Fig. 8. Cluster 1 is transformed from agnostic units to a dense and tonotopically repeating cluster of units that respond selectively to minor chords. (Reprinted from Bharucha, in press. Copyright by The Macmillan Press.)

Fig. 7

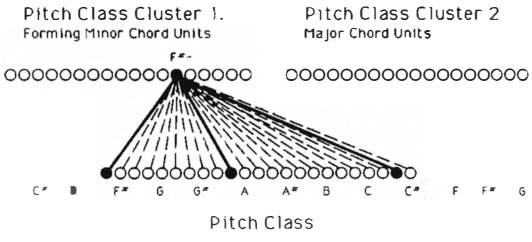
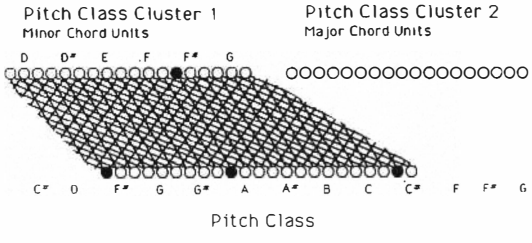


Fig. 8



erty of the network, since the network has learned only the local (temporal) clustering of pitch classes to form chords and the clustering of chords to form keys. The circle of fifths emerges as a consequence of the joint satisfaction of all constraints in the network.

Atonal music will typically induce a less focused pattern, and polytonal music might result in multiple, though not very strong, foci. The model thus allows for gradations of key, and for multiple keys, consistent with evidence from experiments (Krumhansl 1990; Krumhansl and Schmuckler 1986).

A chord is implied or schematically expected to the extent that its unit is activated. The activation of a chord unit also biases judgments about the chord's consonance. This mechanism would explain the finding that the internal consonance of a chord increases when it is schematically expected (Bharucha and Stoeckig 1986, 1987). The model also predicts the pattern of rating and memory judgments obtained in experiments measuring perceived chord relationships (Bharucha and Krumhansl 1983; Krumhansl, Bharucha, and Castellano 1982; see Bharucha 1987a, 1987b, for details of experiment simulations).

In response to a major chord, the network will activate the diatonic tones (of the major scale whose tonic is the root of the chord) more than the nondiatonic tones. It is important to note that these effects are emergent properties of the network—effects that arise out of the joint action of local connectivity based on the clustering of tones and the clustering of chords as they are typically heard. Nothing about the diatonic set was explicitly wired into the network.

Gating Mechanism and Sequential Memory

Fig. 10 shows a gating mechanism that takes pitch-class information from MUSACT and transforms it into an invariant pitch-class representation (Bharucha 1988). Units labeled "pi" multiply activation received from pitch-class units and tonal center (key) units. This gates the pitch classes into an invariant pitch-class format in which the tonic is always "0," so that all tonal sequences have the same tonic or origin.

Musical sequences are then stored in terms of this invariant pitch-class format. Versions of the sequential memory, which is a Jordan (1986) net that learns sequences by back propagation, are presented in the chapters by Todd (this volume) and by Bharucha and Todd (this volume). An invariant pitch-class representation as input to the sequential memory will permit invariance under transposition in long-term memory.

When MUSACT and the sequential memory op-

Fig. 9. MUSACT: The network resulting from a learned clustering of pitch classes into chords and a clustering of chords into keys. Music activates the pitch-class units, and activation reverberates through the network until it settles into a state of equilibrium. The pattern of activation at equilibrium represents the array of chord and key implications, and influences the perception and recognition of events that follow. (From Bharucha 1987a. Copyright by the Cognitive Science Society.)

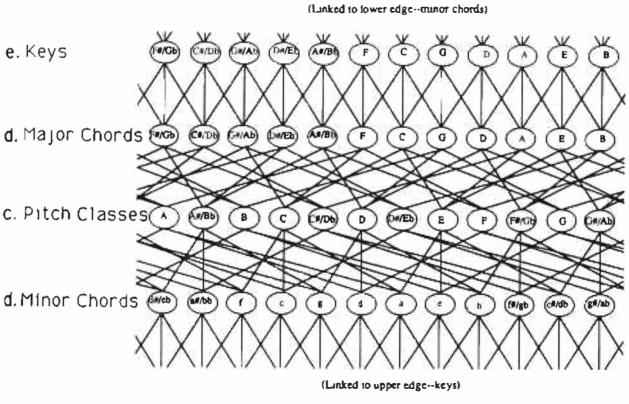
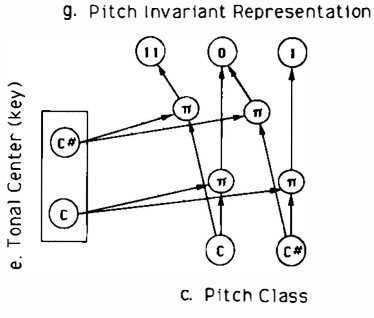


Fig. 10. A gating mechanism (from Bharucha 1988). Units labeled π multiply activation received from pitch-class units and tonal center (key) units. This gates the pitch classes into a pitch invariant format in which the tonic is always 0. Sequential memories are then stored in terms of this pitch invariant format, permitting invariance under transposition. (Versions of the sequential memory are presented in the chapters by Todd and by Bharucha and Todd, this volume.)



erate in tandem, they predict key-distance effects in the transposition of sequences in the short term. If a sequence of chords is transposed immediately, before activation has had a chance to decay completely, the perceived similarity of the two sequences increases with the proximity (around the circle of fifths) of their keys. This is because as each chord in the second sequence is heard, both the sequential memory and MUSACT are checked to see if this chord was expected. The closer the key of the two sequences, the greater the activation of the units in MUSACT that are checked, therefore the greater the degree of expectation will be. The interaction of these two networks is currently being explored.

Conclusion

The development of neural net models of human networks requires attention to a range of empirical and theoretical constraints. In order to account for known constraints on the perception of pitch as it functions in harmony, two networks have been postulated. One (described in the chapter by Bharucha and Todd, this volume) learns musical sequences in an invariant pitch-class format, employing the back-propagation algorithm. The other, summarized in this chapter, learns relationships between pitch classes, chord, and keys, employing a self-organizing algorithm. These networks are linked by a gating mechanism. Together, they account for invariance under transpostion, modulated by key-distance effects in the short-term, while accomplishing their primary tasks of serving as a memory for sequences and a schema for tonal relationships.

Acknowledgments

This work was supported by a grant from the National Science Foundation (BNS-8910778). Portions of this paper have been excerpted from papers published in the proceedings of three conferences as follows: The Ninth Annual Meeting of the Cognitive Science Society (1987, Erlbaum Press); Music, Language, Speech, and Brain (Stockholm 1990, proceedings forthcoming, Macmillan Press); and Music

and the Cognitive Sciences II (Cambridge 1990, proceedings forthcoming in *Contemporary Music Review*). The author thanks Peter Todd for valuable comments.

References

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

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

Bharucha J. J. 1988, "Neural Net Modeling of Music." *Proceedings of the First Workshop on Artificial Intelligence and Music.* St. Paul, Minnesota: American Association of Artificial Intelligence.

Bharucha, J. J. In press. "Cognitive and Brain Mechanisms in Perceptual Learning." *Proceedings of the International Wenner-Gren Symposium on Music, Language, Speech and Brain.* New York: Macmillan Press.

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* 4:341–356.

Bharucha, J. J. and K. Stoeckig. 1986. "Reaction Time and Musical Expectancy: Priming of Chords." *Journal of Experimental Psychology: Human Perception and Performance* 12:403–410.

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

Bharucha, J. J. and P. M. Todd. This volume. "Modeling the Perception of Tonal Structure with Neural Nets."

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.

de Boer, E. 1956. "On The Residue in Hearing." Ph.D. diss., University of Amsterdam.

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

Deutsch, D. 1973. "Octave Generalization of Specific Interference Effects in Memory for Tonal Pitch." *Perception and Psychophysics* 13:271–275.

Dowling, W. J. 1978. "Scale and Contour: Two Components of a Theory of Memory for Melodies." *Psychological Review* 85:341–354.

Fukushima, K. 1975. "Cognitron: A Self-organizing Multilayered Neural Network." *Biological Cybernetics* 20:121–136.

Gjerdingen, R. O. This volume. "Using Connectionist Models to Explore Complex Musical Patterns."

Goldberg, J. M., and R. A. Lavine. 1968. "Nervous System: Afferent Mechanisms." *Annual Review of Physiology* 30:319–358.

Goldstein, J. L. 1973. "An Optimum Processor Theory for the Central Formation of The Pitch of Complex Tones." *Journal of the Acoustical Society of America* 54:1496-1516.

Green, D. M. 1976. *Hearing.* Hillsdale, N.J.: Erlbaum Associates.

Grossberg, S. 1976. "Adaptive Pattern Classification and Universal Recoding: Part I. Parallel Development and Coding of Neural Feature Detectors." *Biological Cybernetics* 23:121–134.

Hebb, D. O. 1949. *The Organization of Behavior.* New York: John Wiley & Sons.

Heider, E. 1972. "Universals of Color Naming and Memory." *Journal of Experimental Psychology* 93:10–20.

Helmholtz, H. 1954. *On the Sensation of Tone.* New York: Dover. Originally published 1885.

Jordan, M. 1986. "Attractor Dynamics and Parallelism in a Connectionist Sequential Machine." *Proceedings of the Eighth Annual Conference of The Cognitive Science Society.* Hillsdale, N.J.: Erlbaum Associates, pp. 531–546.

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

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

Krumhansl, C. L. 1990. *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.

Krümhansl, 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 M. A. Schmuckler. 1986. "The Petroushka Chord: A Perceptual Investigation." *Music Perception* 4:153–184.

Laden, B. and D. H. Keefe. This volume. "The Repre-

sentation of Pitch in a Neural Net Model of Chord Classification."
Landauer, T. K. and C. A. Kamm. 1987. "Teaching a Minimally Structured Back-propagation Network to Recognize Speech Sounds. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society.* Hillsdale, N.J.: Erlbaum Assoicates, pp. 531–536.
Lauter, J. L., P. Hersovitch, C. Formby, and M. R. Raichle. 1985. "Tonotopic Organization in the Human Auditory Cortex Revealed by Positron Emission Tomography." *Hearing Research* 20:199–205.
Licklider, J. C. R. 1954. "Periodicity Pitch and Place Pitch." *Journal of the Acoustical Society of America* 26:945(A).
Meyer, L. 1956. *Emotion and Meaning in Music.* Chicago: University of Chicago Press.
Parncutt, R. 1989. *Harmony: A Psychoacoustic Approach.* Berlin: Springer-Verlag.
Patterson, R. D. 1973. "The Effects of Relative Phase and the Number of Components on Residue Pitch." *Journal of the Acoustical Society of America* 53:1565–1572.
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: Exploration in the Microstructure of Cognition.* Vol. 1. Cambridge, Massachusetts: MIT Press.
Rumelhart, D. E., and D. Zipser. 1985. "Feature Discovery by Competitive Learning." *Cognitive Science* 9:75–112.
Sano, H., and B. K. Jenkins. This volume. "A Neural Network Model for Pitch Perception."
Schmuckler, M. A. 1988. *Expectation in Music: Additivity of Melodic and Harmonic Processes.* Ph.D. diss., Cornell University.
Schouten, J. F., R. J. Ritsma, and B. L. Cardoza. 1962. "Pitch of the Residue." *Journal of the Acoustical Society of America* 34:1418–1424.
Shepard, R. N. 1964. "Circularity in Judgments of Relative Pitch." *Journal of the Acoustical Society of America* 36:2346-2353.
Shepard, R. N. 1982. "Geometric Approximations to the Structure of Musical Pitch." *Psychological Review* 89:305–333.
Stevens, S. S., and J. Volkmann. 1940. "The Relation of Pitch to Frequency." *American Journal of Psychology* 53:329–353.
Terhardt, E., G. Stoll, and M. Seewann. 1982. "Pitch of Complex Tonal Signals According to Virtual Pitch Theory: Tests, Examples, and Predictions." *Journal of Acoustical Society of America* 71:671–678.
Tramo, M. J., J. J. Bharucha, and F. E. Musiek. 1990. "Music Perception and Cognition Following Bilateral Lesions of Auditory Cortex." *Journal of Cognitive Neuroscience* 2:195–212.
Todd, P. M. This volume. "A Connectionist Approach to Algorithmic Composition."
von der Malsberg, C. 1973, "Self-organizing of Orientation Sensitive Cells in the Striate Cortex." *Kybernetik* 14:85–100.
Weinberger, N. M., and D. M. Diamond. 1988. "Dynamic Modulation of the Auditory System by Associative Learning." In G. M. Edelman, W. E. Gall, and W. M. Cowan. *Auditory Function: Neurobiological Bases of Hearing.* New York: John Wiley & Sons.
Weinberger, N. M., and T. M. McKenna. 1988. "Sensitivity of Single Neurons in Auditory Cortex to Contour: Toward a Neurophysiology of Music Perception." *Music Perception* 5:355–390.