Connection Science
Vol. 21, Nos. 2–3, June–September 2009, 177–192

Taylor & Francis
Taylor & Francis Group

From frequency to pitch, and from pitch class to musical key: shared principles of learning and perception

Jamshed J. Bharucha*

Office of the Provost, Ballou Hall, Tufts University, Medford, MA 02155, USA

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

This article addresses the modelling of harmony and pitch, highlighting the parallels across these two domains. The harmony model involves mapping from tone classes to chords and keys using self-organisation, and accounts for a range of psychological data. The pitch model involves mapping from frequency spectra to pitch (spectral and virtual) using self-organisation and also accounts for key results on pitch perception. The result is a scheme that is built hierarchically from frequency to keys, going through intermediate levels of pitch, pitch class, and chords. Both models account for fusion and for the implication of missing features when sufficient context is present. An architecture is suggested for mapping absolute pitch into a pitch-invariant format, using keys as mapping units. A similar architecture is suggested for mapping frequency spectra into a format in which the relative structure of spectra are preserved across changes in pitch. Key units and pitch units are thus thought to play roles as mapping units.

Keywords: music; harmony; self-organisation; neural networks; pitch

Psychological research in auditory perception has traditionally been polarised into psychoacoustics and cognitive psychology. Yet segregating 'low level' and 'high level' processing prevents us from exploring the existence of common principles. Connectionist models (for example, Griffith 1994; Leman 1995) suggest ways in which similar principles of learning and processing may underlie auditory phenomena at many levels. This approach would compel us to question the easy invocation of mental representations and formal rules when psychoacoustic phenomena may suffice. It would also compel a renewed examination of psychoacoustic phenomena that are context sensitive and thought to be processed centrally rather than peripherally.

A connectionist approach to modelling pitch and harmony enables us to understand a diverse set of psychological phenomena with a small set of principles. These phenomena include: the fusion of features into unitary and hierarchically organised perceptual objects; musical expectations, implications and illusions of pitch; passive perceptual learning; and context-specific tonal relationships. Context-specific tonal relationships have been derived by music theoretic analysis as well as by a range of psychological tasks including memory confusions, probe tone judgments, and priming. Connectionist models have the potential to unify the results of these apparently divergent tasks by postulating underlying mechanisms.

Given fundamental constraints on neural architecture and general purpose neural learning mechanisms, and given passive extended exposure to an auditory environment that is structurally

*Email: Jamshed.bharucha@tufts.edu

ISSN 0954-0091 print/ISSN 1360-0494 online
© 2009 Taylor & Francis
DOI: 10.1080/09540090902733822
http://www.informaworld.com

constrained, physically as well as culturally, what aspects of this structure can plausibly be internalised as representations or codes, and what phenomena emerge as an automatic consequence of how these internalised codes influence subsequent perception? If phenomena automatically emerge from the passive internalisation of environmental structure, then on parsimony grounds learning – and the computational processes entailed by a particular model of learning – are to be preferred to the postulation of innate, symbolic, rule-based or other ad hoc mechanisms.

In this article I bring together a range of phenomena involving harmony and pitch, showing how common principles of neural self-organisation and mapping can function across multiple phenomena to provide a more integrative approach to music perception than has been available under the traditional symbol-manipulation and rule-based approaches. Self-organisation also focuses attention on distinguishing between primitive and derivative concepts. In the harmony network below, for example, there is no special status given to the tonic in either the architecture or the input, and yet tonal centres emerge at the output. Likewise, in the pitch model, there is no special status given to the fundamental frequency over other frequencies: pitch salience simply emerges. The phenomena and approaches included below necessarily constitute a selective set: it is beyond the scope of this article to encompass all that we know about pitch and harmony. I focus on aspects of pitch and harmony that my colleagues and I have studied, with the purpose of providing a glimpse into an integrative approach on which further research can build.

The article is divided into two sections: the first on harmony and the second on pitch. The objective is to show in the course of the narrative how the problem of getting from tones to chords and keys (harmony) is in some ways analogous to the problem of getting from frequency spectra to tones (pitch), and that significant aspects of both levels of abstraction of pitch structure can be approached as neural self-organisation.

1. Harmony

Most cultures divide the octave into 12 roughly equally spaced tones on a logarithmic scale. These 12 tones are used in highly constrained ways. In most Western music that is called 'tonal', a short run of tones (a phrase or shorter) utilises three of the 12 tones in a prominent way. These three tones, played either simultaneously or in succession, constitute a chord, either explicitly or implicitly. The most frequently used chords are the major and minor chords. As the chords change, they do so in a constrained manner. Integrating over the corpus of Western music, a clear hierarchy emerges between the transition probabilities of chords in context. For example, a C major chord, which consists of the tones C, E, G, is most likely to be followed by a G major chord (G, B, D), next most likely to be followed by an F major chord (F, A, C), next most likely to be followed by A minor (A, C, E) or D minor (D, F, A), and so on. As a piece of music unfolds, the clustering of tones over time yields a set of abstract percepts, including keys in Western music, and modes in the music of most cultures. Keys and modes, as well as chords resulting from their component tones played sequentially, are abstract perceptual objects or patterns that generate cultural expectations consistent with past experience. The fulfilment of these expectations enables recognition of cultural schemata, and their violation evokes affective responses that are at the essence of musical aesthetics.

In past research, my colleagues and I have argued that the brain learns these clustering patterns through self-organisation, mapping simpler auditory features onto abstract musical clusterings. Early attempts at modelling neural architectures that map auditory features onto musical categories can be found in the work of Pitts and McCulloch (1947) and Deutsch (1969). The spreading activation models in cognitive science (Anderson 1976; McClelland and Rumelhart 1981) were the inspiration for MUSACT (Bharucha 1987a,b). In MUSACT, units tuned to pitch classes are connected to chord units, which in turn are connected to key units. The connectivity is based

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

on membership: tones connect with all major and minor chords of which they are components, and chords connect with all major keys of which they are components. The connections were hand-wired in the original model, but illustrate how simple hierarchical connectivity can account for a range of phenomena in harmony without postulating rules.

The use of chords as intermediary representations between tones and keys does not preclude direct influences of tones on keys. Some data on the perception of keys can be accounted for on the basis of tones alone, without chords as intermediaries (for example, Parncutt 1989; Krumhansl 1990). And keys can self-organise directly from tone activations without a layer of chords (see Tillmann, Bharucha, and Bigand 2000). However, a significant body of empirical research to date involves the perception of chords. And chords, like keys, appear to have a unitary quality in perception. For example, we hear a familiar chord as a whole, with a distinctive perceptual quality over and above the individual tones that comprise it. An unfamiliar chord is initially heard as a collection of tones, but with familiarity takes on a unitary quality. The objective of this model is to account for as much of the literature on harmony (including the perception of chords) in the most parsimonious way. Adding direct connections from tones to keys will build in redundancies that might actually exist; but from a modelling point of view, the fewer the assumptions, the better. As explained below, a fundamental result (Tekman and Bharucha 1998) cannot be explained without chord units.

When the three component tones of a major triad are played, (for example, tones C, E, and G), they activate the corresponding tone units, which in turn spread activation to the chord units. From the chord units, activation spreads up to the key units but also back down to the tone units. Similarly, activation spreads back down from the key units. In this way, bottom-up (stimulus-drive) and top-down (knowledge-driven) processes interact iteratively until the network settles.

After settling, the activations manifest implicitly some important relationships in harmony. Even though only the tones C, E, and G were presented to the network in the above example, the other diatonic tones of the key of C major (D, F, A, B) also get slightly activated because of top-down processing. Furthermore, if the input tones are C, E, and G, then the C major chord unit receives the most activation, followed in order by the chord units around the circle of fifths. Note that this cannot be predicted directly by the assumptions (which tones are members of which chords, and which chords are members of which keys) that determined the connectivity; in other words, the circle of fifths for major chords does not fall out directly from the overlap of their component tones. In the absence of the model, the assumptions alone predict that activation of chords would be monotonically related to the number of sounded tones they contain. Thus, the D major chord (which consists of the tones D, F#, and A) should not be active at all, because it contains none of the sounded tones (C, E, G). On the other hand, the E major chord (which consists of the tones E, G#, and B), should be more active than the D major chord, because it contains one of the sounded tones (E). Yet, once the model settles, the D major chord unit is more active than the E major chord unit, in accordance with the circle of fifths.

How can this be? After the first bottom-up processing cycle, the D major chord unit is not active at all, while the E major chord is activated by the tone E. This is the purely stimulus-driven activation cycle. But after activation has a chance to percolate iteratively to the key units and back down, the relative activation of these two chord units reverses, in line with the fact that C major is closer to D major than it is to E major around the circle of fifths. In other words, implicit knowledge of the organisation of chords into keys affects the relationships between chords as driven by sounded three simple tones, because of the effect of top-down processing.

We tested this time-based reversal with humans using a priming task (Tekman and Bharucha 1998). We predicted that immediately after hearing the tones C, E, and G together, reaction time would be slower to process a D major chord than an E major chord, because the latter would have been primed by the presence of the tone E. But we also predicted that after the network has a chance to settle, the reaction times would be reversed, reflecting their relationship around the

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

circle of fifths. This is what we found. Fifty milliseconds after hearing the tones C, E, and G, the E major chord is processed more quickly than the D major chords, reflecting stimulus-driven structure. Within 200 ms, the reverse is true, reflecting prior learning. This model thus makes a testable predication that falls out of an emergent property rather than directly from the assumptions built into its connectivity.

Of course, MUSACT does not need to be hard-wired. Functionally equivalent patterns of connectivity can be learned automatically through self-organisation (Bharucha 1991, 1992, 1998; Tillmann, Bharucha, and Bigand 2001). Thus the knowledge built into the system is bootstrapped automatically by repeated exposure to pervasive regularities in our musical environment. We are bombarded on a daily basis with major and minor chords (thereby acquiring, through self-organisation, the mapping of tones onto more abstract units that we call chords). Because most Western (and now even non-Western) popular songs employ constrained sets of chords that evoke keys (for example, a song may utilise the chords C, F, G, A minor, and D minor – the component chords of the key of C major), we are bombarded with the regularities that map chord units onto more abstract units that come to be experienced as keys.

The experimental result of Tekman and Bharucha (1998) is not explained by any other model to date. It does not follow from an analysis of the overlapping patterns of harmonics between tones, chords, and keys, because it requires top-down processing. And it is not explained by neural net models that map directly from frequency to keys without the intervening role of chord units. Thus it appears that a general purpose system of neural self-organisation, with the capacity to hierarchically develop abstract units that integrate pitch patterns over a variety of temporal windows, not only internalises some of the membership regularities of tones, chords and keys but also generates emergent phenomena that can be tested empirically.

The model captures a range of psychological phenomena, some evident introspectively and some demonstrated empirically. Introspectively, we have a strong sense that chords and keys are unitary percepts. A major or minor chord sounds like more than a cluster of three tones; it has a distinct perceptual quality. Similarly, a key (or in other cultures, a mode) has a distinctive perceptual quality over and beyond its component tones or chords. Chords and keys function as Gestalts – holistic percepts that fuse features into integrated, unitary objects. In rich contexts, not all the component features need to be present in order to give rise to the perception of the whole; missing features may be inferred, generating expectations for their occurrence, or filled in, giving rise to illusory features consistent with the whole. When we encounter unfamiliar chords (such as jazz chords), we initially hear them as clusters of individual tones. Only after extended experience of hearing them do we begin to hear them as fused, unitary sound objects. Later in the article we shall consider fusion at a lower level of representation: the integration of frequencies into a unitary percept we call pitch.

Activation of tone and chord units in this model accounts for tonal relationships and expectations that are derived from the temporal integration of tonal information that is not dependent on the specific temporal order of notes. We have addressed the learning of specific musical sequences in other work (see Bharucha and Todd 1991). Indeed, we distinguish between two kinds of expectations. Schematic expectations in harmony lead one to expect tones or chords that typically occur following a context. Veridical expectations lead one to expect tones or chords that actually occur in particular pieces of music. A composer plays with the listener's expectations by producing sequences that diverge occasionally from the schematic expectations. And a performer has to be able to hit a particular note even if it is unexpected. Further discussion of this issue is beyond the scope of this article, and the reader is referred to other sources (Bharucha 1984; Justus and Bharucha 2001).

The model's predictions have been tested by a variety of priming experiments (Bharucha and Stoeckig 1986, 1987; Tillmann, Janata, Birk, and Bharucha 2003). Like all models, it has its share of over-simplifications and challenges (see Tillmann et al., 2008). Modelling by its very nature is intended to provide a new rung on the ladder of understanding, which hopefully enables us to grasp

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

a higher rung. The power of the approach suggested in this article is that it integrates apparently disparate phenomena. There is an unfortunate tendency in some psychological approaches to try to keep coming up with better models for specific experimental tasks, as each successive model fails to account for ever more specific circumstances. In contrast, the present approach has been to try to show how a broad array of tasks can be approximated.

The model learns robustly under a variety of conditions and, like some other recent models (for example, Griffith 1994), predicts the results of a variety of experiments (Tillmann, Bharucha, and Bigand 2001). In a classic study, Dowling (1978) presented subjects with a standard melody followed by a comparison melody that was either identical or had one note changed. Subjects had to say whether the melodies were the same or different. When a note was changed, the change was less easily detected if it preserved the key than if it violated the key. MUSACT makes this same memory confusion: tones that did not occur are more strongly activated by top-down processing if they are consistent with the activated keys than if they are not (Tillmann, Bharucha, and Bigand 2001). The model accounts for data from a similar experiment on chord memory in the same way (Bharucha and Krumhansl 1983). Probe tone judgments following musical contexts (see Krumhansl 1990) are also accounted for by activation spreading through the network. Following the context, the simulation of the model reads out the activation of the probe tone and gives a rating of fitness accordingly (Tillmann, Bharucha, and Bigand 2001). There is also evidence that the implicit knowledge of harmony revealed in the literature is encoded in the brain in a way that is isomorphic with harmonic structure. For example, subjects were presented with modulating sequences, and the changing patterns of activation were recorded with fMRI; the harmonic relationships implied by the circle of fifths and the torus of key relationships are encoded by circuits in the pre-frontal cortex (Janata, Birk, Van Horn, Leman, Tillmann, and Bharucha 2003).

An obvious and challenging concern about the model is that its units are tuned to absolute pitch, not relative pitch. The answer to this is that we must necessarily have, simultaneously, neural records of both the absolute pitches of tones and chords as well as their harmonic functions (the latter being the pitch level relative to the tonic). Indeed, harmonic functions (tonic, subdominant, dominant, etc.) have distinct perceptual qualities, suggesting that they are indeed extracted from the music. But absolute pitch information is clearly present in the short term. It is precisely the relationship between harmonic units based on absolute pitch that determines their harmonic function, just as it is the relationship between successive absolute pitches that determines the interval between them. Indeed, the bulk of the evidence to date about the tuning characteristics of auditory neurons points to absolute frequency or pitch coding. Thus we clearly have absolute pitch coding. The real question is how relative coding is derived from this.

Consider the architecture in Figure 1. The units along the left (labelled 'Tonal Centre Representation' are fed from the key units of MUSACT. The units along the bottom (labelled 'Absolute Pitch Representation') are fed from the tone units of MUSACT. The tonal centre gates activation from the absolute pitch representation into a pitch-invariant representation (shown at the top). The gating is accomplished by sigma–pi units that multiply activation from their two sources. Thus, for example, when C is the tonal centre, tone C activates the tonic (labelled '0') and tone D activates the supertonic (labelled '2'). When D is the tonal centre, tone D activates the tonic, and so on. As the distribution of activation across the key units changes, the harmonic functions become ambiguous, until they are disambiguated by the new key.

This system need not be specific to keys (which are categories of organisation found in Western music, and relatively recent in history). Tonal centres are also found in modal music, which is found in many parts of the world and has existed for thousands of years. In modal music, the tonal centre maps absolute pitches into their slots within the mode.

Elsewhere we have simulated the acquisition of modes of Indian music (see Bharucha and Olney 1989). While the model in that article was an auto-associative network, we have replicated

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

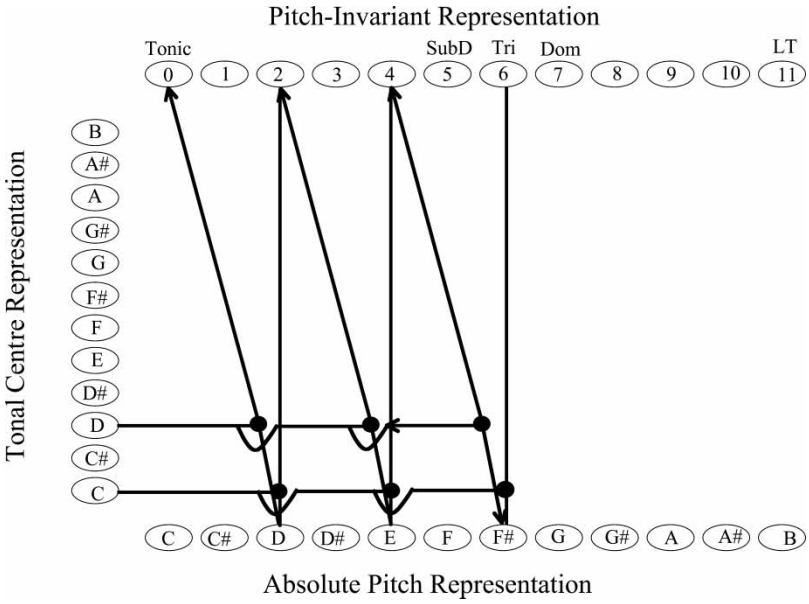

Figure 1. A network that maps tones into a pitch-invariant format that corresponds to a tonic-based harmonic function. The key units (on the left) are from the MUSACT network. They gate the tones (or chords) from the same network (bottom) into the pitch-invariant format (top) via sigma–pi units. An absolute pitch gates into a pitch-invariant pitch slot depending upon how active the corresponding tonic is.

the results using self-organisation. The self-organisation of modes begins with pitch-invariant units as inputs, and learns to map these inputs onto modes. Thus for Indian music, the inputs correspond to the 12 degrees of the scale, in a pitch-invariant format: Sa, Re−, Re, Ga−, Ga, Ma, Ma+, Pa, Dha−, Dha, Ni−, Ni. In a simulation of learning in the Indian musical environment, 10 Indian modes (*thaats)* were learned, producing an output unit for each one. In a control simulation of learning in the Western musical environment, only the major and minor modes were learned. Following learning, both the Indian and Western networks were tested with a unique subset of one of the Indian modes, Bhairav. Bhairav consists of: Sa, Re−, Ga, Ma, Pa, Dha−, Ni. The unique subset omitted Re−. The Indian network recognised it as Bhairav, and when top-down activation was introduced into the network, filled in the Re−, thereby simulating the Indian listener's expectation of Re− even though it was missing. In contrast, top down processing in the Western network did not pick out Re− at all.

The generality of the pitch-invariant mapping system is illustrated toward the end of the article, where a similar system is sketched for preserving the relative frequency relationships in the spectra of complex tones across changes in absolute pitch. Indeed, we have ventured to suggest that abstract units (such as key units) that integrate pitch information over windows of time exist in part to enable such mappings. Key units provide the reference anchors that enable tones and chords to be represented as harmonic functions, thus enabling recognition of tone sequences or chord sequences across keys.

2. Pitch

The modelling efforts discussed above are predicated on the existence of pitch class units, which in turn are predicated on computing octave equivalence and on extracting pitch from frequency

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

spectra. Getting octave equivalence or pitch classes from pitch is not a difficult problem, because complex tones whose fundamental frequencies are separated by octaves share plenty of harmonics, and a self-organising network recognises the similarity (see Bharucha and Mencl 1996). Once self-organisation has achieved a mapping of pitch onto pitch class, it is straightforward to explain how we perceive octave equivalence in music even when octave harmonics are missing or attenuated (as in the case of the clarinet): the network recognises the octave relationship because of the rich context of other harmonics.

Extracting pitch from frequency spectra, by contrast, is an enormous task, as illustrated by a long, distinguished and still vexing history of research dating back at least to Helmholtz (1863/1954). In order to point to the hierarchical parallels between the extraction of harmonic units and the extraction of pitch, a self-organising approach is elucidated for pitch extraction. Self-organising models of pitch extraction already exist in the literature (Taylor and Greenhough 1994; Cohen, Grossberg, and Wyse 1995; Grossberg 1997). My purpose in elucidating such a model is to illustrate the parallel architecture and its psychological consequences. In this way, pitch and harmony can be seen as a hierarchical process of self-organisation beginning with harmonic frequency spectra and culminating in keys or modes. As indicated earlier, going straight from frequency spectra to keys does not explain the Tekman and Bharucha (1998) finding, which seems to rely on the extraction of intermediate units.

The psychological consequences of this parallelism are considerable. At the level of harmony, activation in a network can account for our sense of expectation or the psychological filling-in of tones not present but implied. At the level of pitch, activation in a self-organising network can account for our sense of illusory pitch – as in the perception of a missing fundamental (as studied by psycho-acousticians since Helmholtz), the fundamental bass (as observed by musicians, most notably Rameau), and equivalently, the illusory perception of sub-harmonics.

Furthermore, both harmony and pitch involve fusion. In harmony perception, tones heard simultaneously in familiar combinations fuse to form integrated, unitary percepts that we describe as chords, with perceptual qualities that go beyond those of the tones themselves. The same is true of keys and modes. In pitch perception, frequencies heard simultaneously in familiar (i.e., harmonic) combinations fuse to form integrated, unitary percepts that we describe as pitch.

In the perception of harmony and pitch, mapping to pitch-invariant formats is necessary in order to preserve relative structure under transformation. I have suggested above that abstract organising units such as tonal centres (the tonics of keys in Western music, tonal centres of modes in modal music) can make this transformation possible by gating absolute pitch-based representations into pitch-invariant ones. Toward the end of the article I will suggest a parallel mechanism (described in Bharucha and Mencl 1996) in which an extracted pitch serves to gate a frequency spectrum into a format in which the timbral characteristics of the spectrum are maintained while the pitch deviates.

The psychoacoustic phenomena addressed in this section pertain to the extraction of pitch from complex spectra, in particular, the missing fundamental and the pitch of tones with partials that are not harmonics. The present model is influenced by the seminal work of Terhardt – in particular, by Terhardt's claim that the mapping from frequency spectra to pitch is learned (primarily from exposure to speech). The model below demonstrates how this learning might be done in such a way that some of the standard puzzles of pitch perception fall out automatically.

Periodically vibrating objects to which we attribute pitch have frequency partials that are roughly harmonics (integer multiples) of the fundamental frequency. A harmonic spectrum is typically heard as having a singular or fused pitch that can be matched to that of a pure tone (Stumpf 1898; Thurlow and Rawling 1959; DeWitt and Crowder 1987; Parncutt 1989). This is sometimes called synthetic pitch perception. Yet one's attention *can* be drawn to individual harmonics, leading to the perception of several distinct pitches. This is called analytic pitch perception.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

In most cases the synthetic pitch is matched to that of a pure tone at the fundamental frequency. This is not surprising because, in most natural harmonic sources, the fundamental is the most intense harmonic. However, one can remove the fundamental frequency and still hear the pitch there. This is the central puzzle that has motivated research in pitch perception since Helmholtz (1863/1954) and is the most important empirical constraint on any model of pitch.

Since the early 1970s the prevailing view has been that the missing fundamental is a result of a pattern processing system that fills out the incomplete harmonic spectrum (Terhardt 1972; Goldstein 1973; Wightman 1973b; Terhardt 1974, 1979, 1982, 1986; Terhardt et al. 1982). Early theories that explained the phenomenon in terms of peripheral mechanisms such as distortion (e.g., Helmholtz 1863/1954) have been rejected by at least three classes of evidence: (1) the missing fundamental cannot be masked by noise in the fundamental's critical band (Licklider 1954); (2) it can be induced by dichotic presentation of subsets of harmonics (Houtsma and Goldstein 1972), and (3) the synthetic pitch sometimes corresponds to neither the fundamental nor other distortion products (Hermann 1912; de Boer 1956; Schouten, Ritsma, and Cardozo 1962).

This last piece of evidence provides perhaps the most challenging empirical constraint on a theory of pitch perception. A tone consisting of sine waves at 800, 1000, and 1200 Hz has a predominant pitch at 200 Hz. Here 200 Hz is both the fundamental (the highest common divisor) and the difference frequency (a distortion product). However, a tone consisting of sine waves at 850, 1050, and 1250 Hz has no pitch at 50 Hz, the fundamental frequency, and no pitch at 200 Hz, the difference frequency. Its pitch is somewhat ambiguous but is most closely matched to around 210 Hz.

Patterson (1973) presented subjects with tones with constant difference frequencies of 200 Hz between adjacent partials, i.e., tones with frequencies $L + 200n$, where L is the lowest frequency ($100\,\text{Hz} < L < 2600\,\text{Hz}$) and n is an integer (ranging from 0 to 6). He found that the pitch is matched to 200 Hz whenever the fundamental is 200 Hz (i.e., whenever L is a harmonic of 200 Hz). As the spectrum shifts away from these points (while maintaining additive invariance), the pitch shifts as well – fairly linearly – even though the fundamental frequency jumps around and the difference frequency remains constant.

One promising solution to this puzzle consisted of computing the period from peak-to-peak distances when the component sine waves are added in cosine phase (Bilsen and Ritsma 1969). As the tone is shifted additively, peak-to-peak distances become ambiguous because of competing local peaks; the average of the candidate peak-to-peak distances seems to yield the empirically determined pitch of 210 Hz in the above example. This 'temporal fine-structure' model fails because the shape of the time-domain waveform depends upon the relative phases of the partials, whereas the pitch percept in Patterson's (1973) experiment does not (Wightman 1973a).

In their mathematical models of auditory pattern recognition, Goldstein (1973) and Wightman (1973b) account for these data by adding noise or 'smear' to the power spectrum. Goldstein's model represents peaks in the power spectrum as Gaussians and then finds the optimal estimate of the fundamental frequency. Wightman's model postulates a triangular smearing of the power spectrum and then a Fourier transformation, from which the most prominent pitch is derived. Harmonics of high order (above around 10) are thought not to be resolvable and are therefore weak determinants of pitch; hence the fundamental of 850, 1050, and 1250 Hz (the 17th, 21st, and 25th harmonics of 50) is not considered a strong candidate for pitch. The noisy or smeared representation permits these models to recognise that the closest match of <850, 1050, 1250> to low-order harmonics is approximately <840, 1050, 1260>, the 4th, 5th, and 6th harmonics of 210, thereby producing 210 Hz as the strongest pitch.

Although these models give a compelling quantitative account of aspects of pitch perception, they assume a mechanism for either finding the fundamental or performing Fourier transformations. From a cognitive point of view, both assume prior knowledge of the structure of harmonic signals and an expectation that incoming signals are harmonic.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

Terhardt's (1974, 1979) model is a substantial advance because it postulates that this knowledge is learned. Terhardt suggested that we develop harmonic templates from our exposure to speech (vowels, in particular). The hypothesis that harmonic templates are learned from speech receives support from the dramatic diminution of the missing fundamental effect as one goes much beyond the speech range (roughly 700 Hz).

According to Terhardt's theory, a signal produces pitches corresponding to the spectral frequencies ('spectral pitches') but is also matched to learned harmonic templates. The lowest component of each template that produces a match is heard as a 'virtual pitch' whose strength depends upon the strength of the match (which in turn depends upon the number of components that match and their strengths). The missing fundamental is heard as a virtual pitch because the upper harmonics are sufficient to produce a strong match to the template of the same fundamental. Virtual pitches other than the fundamental are also generated because of weaker matches between subsets of the spectrum and other templates. These additional virtual pitches tend to be subharmonics (integer fractions) of the fundamental.

Terhardt (1974) proposed a learning matrix model in which the lowest spectral component is extracted and feeds into a pathway that generates a virtual pitch at that frequency. With each spectral experience, the simultaneous occurrence of the lowest spectral component with higher components lowers resistances connecting the higher components to the same pathway. Eventually, these resistances are low enough that an individual spectral component can generate virtual pitch cues corresponding to the lowest components of all the spectra in which it typically occurs. Assuming that the most prevalent spectral structure in our experience is harmonic, each component of a signal produces virtual pitches at all the possible fundamentals of which it is a (low order) harmonic. When all the components of a tone are harmonics of some missing fundamental, the fundamental is the strongest element of the intersection of the sets of virtual pitches generated by the harmonics that are present.

Although Terhardt's (1974) learning model is an important advance because it does not presuppose a specialised mechanism tailor-made for harmonic spectra, it does presuppose a privileged role for the lowest component. Furthermore, it does not have cognitive generality beyond pitch perception.

The model of pitch sketched below is based on a place mechanism, i.e., codes frequency via the filtering characteristics of frequency sensitive neurons. Although recent interest has focused on timing mechanisms in pitch (e.g., Slaney 1990; Meddis and Hewitt 1991; Cariani and Delgutte 1996), place models continue to be viable (e.g., Cohen et al. 1995) because both timing and place mechanisms seem to exist in the brain. The objective of the section below is to show a possible parallelism between neural self-organisation from pitch class to key (as shown above) and from frequency to pitch, thereby showcasing the possibility that the same general principles may apply to a hierarchical system from frequency to key.

The central mathematical idea in self-organisation is disarmingly simple. At the input level, units tuned (either innately or through prior self-organisation) to features in the stimulus represent a vector of activations corresponding to the features present (or filled in). Units available at a higher level become tuned to consistent patterns of activations present in the input layer, creating a more abstract set of feature or pattern detectors. Learning consists of changing the weight vectors that feed into the abstract units so that they approach co-linearity with the input vectors. Co-linearity of input and weight vectors is the key computational idea, because it enables the network to serve as a template that recognises patterns that correlate with the internalised patterns of weights. On any given trial, the abstract unit selected for weight changes is the one whose weight vector is already more highly correlated with the input pattern, i.e., the winner. All the major models of self-organisation function in this way (i.e., Grossberg 1970, 1972; von der Malsberg 1973; Fukushima 1975; Grossberg 1976; Kohonen 1984; Rumelhart and Zipser 1985). The Kohonen variant extends to the neighbours of the winner the movement toward co-linearity, while gradually

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

reducing the neighbourhood size, thereby sorting the learned feature detectors so that they become topographic representations of the space of patterns being learned. Most of the complexity (and differences between) these models address measures to ensure stability and the granularity of categorisation.

Learning is essentially Hebbian (Hebb 1949): weights to the winning output unit (and perhaps its neighbours) are strengthened in proportion to the activations of the input units, and then the weight vector is renormalised. This increases the likelihood that the same output unit will win in response to this stimulus and similar stimuli, and reduces the likelihood that it will win in response to dissimilar stimuli.

In the model described here, the self-organisation of pitch begins with the extraction of pitch from an acoustic spectrum. Only units in the first layer, the spectral layer, of this hierarchy have innate tuning characteristics. Units in the spectral layer are modelled after frequency-tuned neurons found in the cochlear nerve. The filtering characteristics of cochlear neurons were based on physiological data. For cochlear neurons, which are essentially frequency filters, Q increases roughly linearly as a function of log characteristic frequency (Rose, Hind, Anderson, and Brugge 1971). A straight line fit to the Rose et al. data yields the following function relating Q to characteristic frequency, c:

$$Q = 3.04 \log c - 4.17.$$

The bandwidth, b, of a neuron with characteristic frequency, c, is then

$$b = c/Q.$$

One thousand spectral units were used in the model, with characteristic frequencies ranging from 10–10,000 Hz at 10 Hz intervals. This is a rough approximation of the range within which the first eight harmonics of the speech signal typically fall. (The first eight harmonics were deemed sufficient because energy drops off rapidly with harmonic number in natural signals and because only lower order harmonics are critical for pitch perception, the higher ones not being resolvable). The choice of a 10 Hz separation was designed to assure extensive overlap in the receptive fields of neighbouring units. Even though increments in characteristic frequency are likely logarithmic (Greenwood 1961a,b, 1990), the properties of the system in the present context do not change as a function of the density of characteristic frequencies, as long as there is substantial overlap between receptive fields, permitting course coding.

The second layer represents pitch. The spectral and pitch layers were fully connected. The input spectra to be learned were tones with the first eight harmonics with intensities inversely proportional to harmonic number, whose fundamental frequencies ranged from 10 to 1250 Hz in 10 Hz increments. This simulates roughly the environment of pitch-evoking stimuli (principally from speech and music) to which we are exposed – subtracting the formants (from speech) or resonances (from musical instruments), which are not pertinent to the perception of pitch per se.

The limit state of the learning process was simulated by setting the weight vectors to be co-linear with the activations of the input units. Specific algorithms for self-organisation vary in terms of stability conditions for learned categories, and variability due to iterative learning. Simulations using different self-organising algorithms, or using the same algorithm with different initial settings of weights, yield similar results, with noise around a central tendency. To understand the system mathematically in its most stable form, this central tendency is best approximated by setting a weight vector to be co-linear with the activation vector across the input units for each learned spectrum. Thus, for each tone, t, of the 125 tones learned, if $\mathbf{s}_t$ is its spectral vector, and $f(\mathbf{s}_t)$ is the filtered vector denoting the activations of the input units, a unit in the second layer has its weight vector $\mathbf{w}_t$ tuned to $f(\mathbf{s}_t)$:

$$\mathbf{w}_t = f(\mathbf{s}_t).$$

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

While spectral categories may be learned at different resolutions, we consider here only the finest resolution possible with this network, i.e., one category per stimulus. The units in the higher layer thus function more like coding units (Estes 1972) than like categories, relative to the stimulus set presented. However, their generalisation to novel stimuli interpolated between stimuli in the training set shows that they are indeed categories.

The model thus represents a template matching system that learns its templates through exposure to harmonic spectra, searches templates in parallel and does fuzzy matching. Learning through exposure is accomplished by Hebbian weight changes at the winning analytic pitch unit, parallel search is an intrinsic feature of the architecture, and fuzzy matching is a consequence of the overlapping receptive fields of the spectral units.

To test for the emergence of the missing fundamental, the network was tested with harmonic spectra minus their fundamentals. Figure 2 shows the activations of the spectral units following filtering of a harmonic tone missing its 250 Hz fundamental. Peaks are clear at multiples of 250 Hz. These activations are fed through the weight matrix and activate the pitch units as shown in Figure 3. A strong pitch is evident at the missing fundamental (250). Pitches are also evident at harmonics and subharmonics (integer fractions) of 250. Furthermore, because the input tone is also consistent with a fundamental of 125, the harmonics and subharmonics are also evident.

The behaviour of the network is consistent with Terhardt's claim that some of the virtual pitches we hear are subharmonics. Parncutt (1989) has obtained some evidence to support this. According to Terhardt's theory and according to the model we propose here, generalisation to subharmonics is precisely what causes the perception of the missing fundamental. When the fundamental is removed from a harmonic complex, the fundamental frequency is a subharmonic of all the harmonics present, hence is by far the strongest subharmonic to be induced. On this view, the classical problem of the missing fundamental does not warrant a tailor-made theory but is an emergent consequence of a more general pattern-learning-and-matching process.

In accord with the human data, testing the network with frequencies 850, 1050, and 1250 Hz yields a clear pitch at 210 Hz, which is the best template match the network can make. The network was also tested with the tones used in Patterson's experiment, i.e. tones with a constant difference frequency of 200 Hz. Figure 4 plots the highest pitch peaks within the range 150–250 Hz, following Patterson's analysis. Figure 4 shows the same scalloped patterns obtained by Patterson. When the

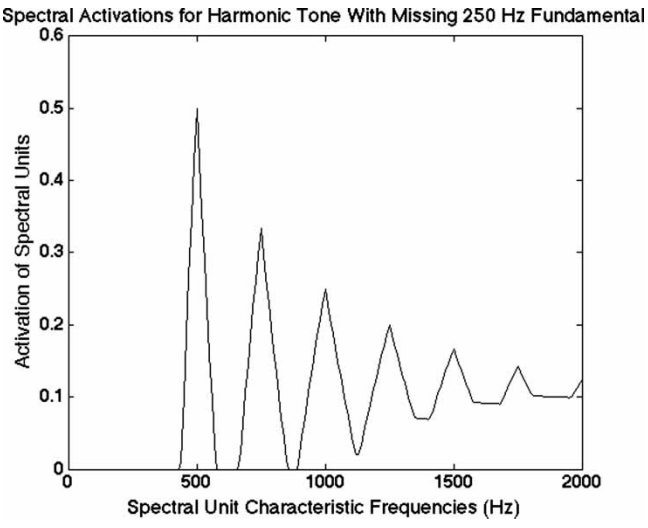

Figure 2. Activation of spectral units (after filtering) in response to a harmonic tone missing its 250 Hz fundamental frequency.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

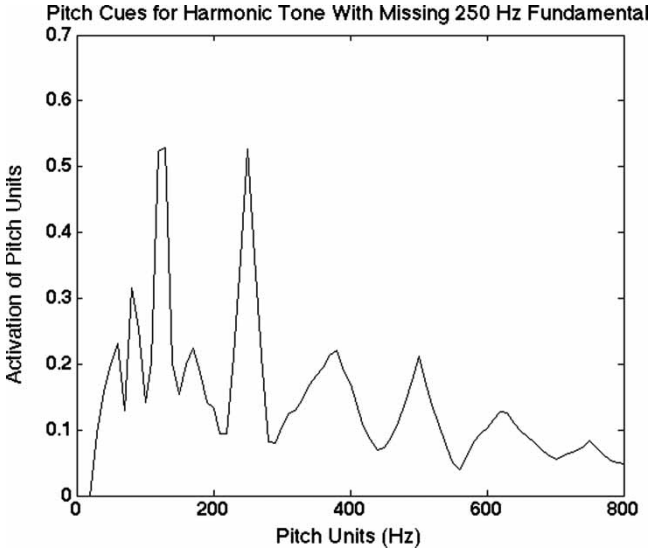

Figure 3. Activation of pitch units in response to a harmonic tone missing its 250 Hz fundamental frequency. At this level of the network, the fundamental is present, as are the other subharmonics of 250 Hz. Because the harmonics of 250 Hz are also harmonics of 125 Hz, the network shows activations corresponding to the harmonics and subharmonics of 125 Hz as well.

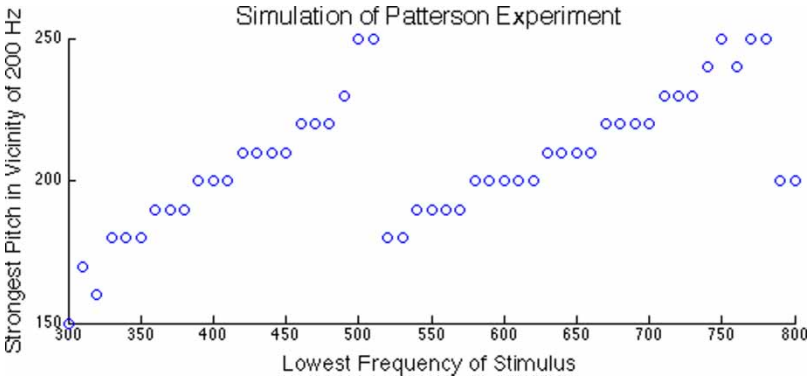

Figure 4. The network's performance in Patterson's (1973) experiment.

lowest frequency is a multiple of 200 Hz, the pitch within this range corresponds to 200 Hz. However, as the lowest frequency increases beyond a multiple of 200 Hz, the pitch rises above 200. Roughly half way between successive multiples of 200 Hz, the pitch drops to below 200 and starts rising again, until it reaches 200 again at the next multiple. Whereas Patterson found some additional scallops when the lowest frequency was a multiple of 100 but not of 200, the network does not. This can be attributed to Patterson's use of a complex tone as the reference pitch. Patterson acknowledges in his article that this may have created additional octave confusions.

The perceptual salience of a pitch in a complex tone reflects the activation of the pitch unit that responds maximally to the sinusoid to which this pitch is matched. Thus the attribution of pitch is always made with reference to the spectral layer. The fusion of pitch (i.e., synthetic pitch) of a complex tone is postulated to be the result of attending to the pitch peak. The listener's ability to 'hear out' the harmonics, often only after guidance, is thus a result of directing attention to

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

different peaks. That analytic and synthetic pitch percepts do not reflect fundamentally different structures or processes is supported by the novice's ability to 'hear out' an upper harmonic when attention is directed to it by alternating between tones that differ only in the presence or absence of that harmonic (Diufhuis 1970).

Note that in this model the pitch surface makes no distinction between spectral and virtual pitch as described by Terhardt: both appear as peaks and are not qualitatively distinguished. The ability to qualitatively distinguish between spectral and virtual pitch (i.e., pitch in the presence or absence of energy at the pitch-matched sinusoid) derives simply from the difference between the spectral representations in the two cases, since it is assumed that the listener can attend to all levels of organisation. (In principle one could postulate self-organisation based on vectors that include several layers, thereby achieving coding or category units for arbitrarily fine distinctions of sound that cut across levels of organisation. This may well happen, particularly in listeners who train their 'ears' by deliberately directing their attention in certain ways.) Terhardt's model requires two kinds of pitch – spectral and virtual – because only the lowest element of the matched template becomes a virtual pitch (see also Parncutt 1989). The present model gives no privileged status to the lowest pitch; all pitches are 'inferred' or 'virtual,' though some correspond to energy components at the spectral layer and others do not.

In conclusion, the model is not just an implementation of parallel fuzzy template matching. It is essentially a model of how this process may be learned. In comparing models of pitch that resort to internal harmonic templates, a model that can learn these templates through mere exposure is more parsimonious than ones that require the prior existence of these templates. The model proposed here is not the set of weights determined by the harmonic spectra with which it was trained. Rather, the model is the architecture of the constrained *tabula rasa* which adapts itself to whatever spectral regularities it may encounter. If the network is exposed to an inharmonic world with regularities, it should learn those. In contrast, Terhardt and Parncutt's theories assume rigid harmonic templates even while acknowledging that there are differences in pitch perception as a function of musical and cultural exposure (see Parncutt 1989). Having said that, I do not argue that pitch perception is necessarily learned. What the model demonstrates is that it *can* be learned. Whether or not the weights have been tuned to harmonic spectra phylogenetically, they can be tuned ontogenetically. Future research will have to determine how much, if any, of this structure is innate.

A pitch extraction model does not account for how the integrity of a frequency spectrum is maintained in relative terms. The relative strengths of partials in a spectrum contribute to the timbre. There are other determinants of timbre, including fixed resonances, onsets and offsets, and changes in the relative strengths of partials over time. Even with fixed resonances, small changes in pitch maintain the perceptual integrity of the timbre. Thus there must be a representation of spectra in a pitch-invariant format.

Figure 5 shows a scheme in which this can be accomplished (adapted from Bharucha and Mencl 1996). It parallels the system described earlier for establishing a pitch-invariant format for tones across keys. Units along the bottom (labelled 'Log Frequency') represent the frequency spectrum (the input to the pitch model). The units to the left (labelled 'Mapping Units') represent the most salient pitch extracted by the pitch model. Activation from the two sources is gated by the sigma–pi units, yielding a pattern of activation across the units at the top (labelled 'Pitch Invariant Representation').

Consider a signal that has frequencies at 100, 300 and 500 Hz. The pitch network extracts 100 Hz as the most salient pitch, and the corresponding gates enable the signal to project to the units at the top as shown in the figure. Another signal with frequencies 200, 600 and 1000 Hz has the same relative spectrum, because the ratios between the frequencies are the same as in the first signal. The pitch network extracts 200 Hz as the most salient pitch, and the corresponding gates project the signal to the same pitch-invariant representation as the first signal. In other words,

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

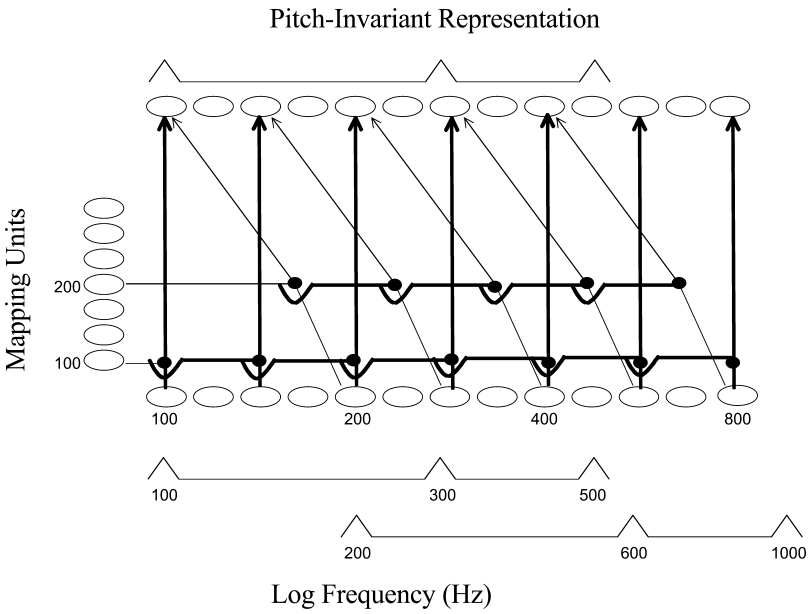

Figure 5. Mapping frequency spectra from an absolute frequency format to a pitch-invariant format. (Adapted from Bharucha and Mencl 1996).

the network projects both signals onto the same pitch-invariant representation. The extraction of pitch as a unitary percept may thus play a role in providing a reference point that enables a pitch-invariant format.

3. Conclusion

In conclusion, the fusion of tones into chords and key, as it occurs in harmony, and the fusion of frequencies into pitches, as it occurs in pitch, can be thought of as the result of a consistent process of hierarchical self-organisation. Furthermore, the extraction of keys or other abstract units may serve to transform representations based on absolute pitch into ones that are pitch-invariant. At a lower level, pitch may play a similar role, serving to transform spectra based on absolute frequencies into ones that are invariant across changes in pitch.

Acknowledgements

The author thanks W. Einar Mencl, Subhobrata Mitra, Mark McNellis, Barbara Tillman and two anonymous reviewers.

References

Anderson, J.R. (1976), *Language, Memory, and Thought*, Hillsdale, NJ: Erlbaum.

Bharucha, J.J. (1984), 'Event Hierarchies, Tonal Hierarchies, and Assimilation: A Reply to Deutsch and Dowling', *Journal of Experimental Psychology: General*, 113, 421–425.

Bharucha, J.J. (1987a), 'MUSACT: A Connectionist Model of Musical Harmony', in *Proceedings of Ninth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum, pp. 508–517.

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

Bharucha, J.J. (1991), 'Cognitive and Brain Mechanisms in Perceptual Learning', in *Music, Language, Speech and Brain*, eds. J. Sundberg, L. Nord, and R. Carlson, London: MacMillan.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

Bharucha, J.J. (1992), 'Tonality and Learnability', in *Cognitive Bases of Musical Communication*, eds. M.R. Jones, and S. Holleran, Washington, DC: American Psychological Association, pp. 213–223.
Bharucha, J.J. (1998), 'Neural Nets, Temporal Composites and Tonality', in *The Psychology of Music* (2nd ed.), ed. D. Deutsch, New York: Academic Press. [Reprinted in Levitin, D.L. (ed.) (2002), *Foundations of Cognitive Psychology*, Cambridge: MIT Press.]
Bharucha, J.J., and Krumhansl, C.L. (1983), 'The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context', *Cognition*, 13, 63–102.
Bharucha, J.J., and Mencl, W.E. (1996), 'Two Issues in Auditory Cognition: Self-Organization of Categories and Pitch-Invariant Pattern Recognition', *Psychological Science*, 7, 142–149.
Bharucha, J.J., and Olney, K.L. (1989), 'Tonal Cognition and Artificial Intelligence: Priming Studies and Connectionist Modelling', *Contemporary Music Review*, 4, 341–356.
Bharucha, J.J., and Stoeckig, K. (1986), 'Reaction Time and Musical Expectancy: Priming of Chords', *Journal of Experimental Psychology: Human Perception and Performance*, 12, 403–410.
Bharucha, J.J., and Stoeckig, K. (1987), 'Priming of Chords: Spreading Activation or Overlapping Frequency Spectra?', *Preception and Psychophysics*, 41, 519–524.
Bharucha, J.J., and Todd, P. (1991), 'Modeling the Perception of Tonal Structure with Neural Nets', *Computer Music Journal*, 13, 44–53. [Reprinted in Todd, P., and Loy, G. (eds.), (1991), *Music and Connectionism*, Cambridge: MIT Press, pp. 128–137.
Bilsen, F.A., and Ritsma, R.J. (1969), 'Repetition Pitch and its Implication for Hearing Theory', *Acustica*, 22, 63–73.
Cariani, P., and Delgutte, B. (1996), 'Neural Correlates of the Pitch of Complex Tones. I. Pitch and Pitch Salience', *Journal of Neurophysiology*, 76, 1698–1716.
Cohen, M.A., Grossberg, S., and Wyse, L. (1995), 'A Spectral Network Model of Pitch Perception', *The Journal of the Acoustical Society of America*, 98, 862–879.
de Boer, E. (1956), 'On the Residue in Hearing', Ph.D. thesis, University of Amsterdam.
Deutsch, D. (1969), 'Music Recognition', *Psychological Review*, 76, 300–307.
Dewitt, L.A., and Crowder, R.G. (1987), 'Tonal Fusion of Consonant Musical Intervals: The Oomph in Stumpf', *Perception & Psychophysics*, 41, 73–84.
Diufhuis, H. (1970), 'Audibility of High Harmonics in a Periodic Pulse', *Journal of the Acoustical Society of America*, 48, 888–893.
Dowling, W.J. (1978), 'Scale and Contour: Two Components of a Theory of Memory for Melodies', *Psychological Review*, 85, 341–354.
Estes, W.K. (1972), 'An Associative Basis for Coding and Organization in Memory', in *Coding Processes in Human Memory*, eds. A.W. Melton and E. Martin, New York: Winston.
Fukushima, K. (1975), 'Cognitron: A Self-Organizing Multilayered Neural Network', *Biological Cybernetics*, 20, 121–136.
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.
Greenwood, D.D. (1961a), 'Auditory Masking and the Critical Band', *Journal of the Acoustical Society of America*, 33, 484–502.
Greenwood, D.D. (1961b), 'Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane', *Journal of the Acoustical Society of America*, 33, 1344–1356.
Greenwood, D.D. (1990), 'A Cochlear Frequency-Position Function for Several Species – 29 Years Later', *Journal of the Acoustical Society of America*, 87, 2592–2605.
Griffith, N. (1994), 'Development of Tonal Centers and Abstract Pitch as Categorizations of Pitch Use', in *Musical Networks: Parallel Distributed Perception and Performance*, eds. N. Griffith, and P. Todd, Cambridge: MIT Press.
Grossberg, S. (1970), 'Some Networks that Can Learn, Remember, and Reproduce any Number of Complicated Space-Time Patterns', *Studies in Applied Mathematics*, 49, 135–166.
Grossberg, S. (1972), 'Neural Expectation: Cerebellar and Retinal Analogs of Cells Fired by Learnable or Unlearned Pattern Classes', *Kybernetic*, 10, 49–57.
Grossberg, S. (1976), 'Adaptive Pattern Classification and Universal Recoding, I: Parallel Development and Coding of Neural Feature Detectors', *Biological Cybernetics*, 23, 121–134.
Grossberg, S. (1997), 'Pitch-Based Streaming in Auditory Perception', in *Musical Networks: Parallel Distributed Perception and Performance*, eds. N. Griffith, and P. Todd, Cambridge: MIT Press.
Hebb, D.O. (1949), *The Organization of Behavior,* New York: Wiley.
Helmholtz, H. (1863/1954), *On the Sensation of Tone*, New York: Dover (Originally published in German in 1863).
Hermann, L. (1912), 'Neue Versuche zur Frage der Unterbrechungstone', *Archiv für die gesamte Psychologie*, 146, 249–294.
Houtsma, A.J.M., and Goldstein, J.L. (1972), 'The Central Origin of the Pitch of Complex Tones: Evidence from Music Interval Recognition', *Journal of the Acoustical Society of America*, 51, 520–529.
Janata, P., Birk, J., Van Horn, J. D., Leman, M. Tillmann, B., and Bharucha, J.J. (2003), 'The Cortical Topography of Tonal Structures Underlying Western Music', *Science*, 298, 2167–2170.
Justus, T.C., and Bharucha, J.J. (2001), 'Modularity in Musical Processing: The Automaticity of Harmonic Priming', *Journal of Experimental Psychology*: *Human Perception and Performance*, 27, 1000–1011.
Kohonen, T. (1984), *Self-Organization and Associative Memory*, Berlin: Springer-Verlag.
Krumhansl, C.L. (1990), *The Cognitive Foundations of Musical Pitch*, Oxford: Oxford University Press.
Leman, M. (1995), *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*, Springer.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015

Licklider, J.C.R. (1954), 'Periodicity Pitch and Place Pitch', *Journal of the Acoustical Society of America*, 26, 945.
McClelland, J.L., and Rumelhart, D.E. (1981), 'An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings', *Psychological Review*, 88, 375–407.
Meddis, R., and Hewitt, M.J. (1991), 'Virtual Pitch and Phase Sensitivity of a Computer Model of the Auditory Periphery. I: Pitch Identification', *Journal of the Acoustical Society of America*, 89, 2866–2882.
Parncutt, R. (1989), *Harmony: A Psychoacoustical Perspective*, New York: Springer-Verlag.
Patterson, R.D. (1973), 'The Effects of Phase and the Number of Components on Residue Pitch', *Journal of the Acoustical Society of America*, 53, 1565–1572.
Pitts, W., and McCulloch, W.S. (1947), 'How We Know Universals: The Perception of Auditory and Visual Forms', *Bulletin of Mathematical Biophysics*, 9, 127–147.
Rose, J.E., Hind, J.E., Anderson, D.J., and Brugge, J.F. (1971), 'Some Effects of Stimulus Intensity on Response of Auditory Nerve Fibers in the Squirrel Monkey', *Journal of Neurophysiology*, 34, 685–699.
Rumelhart, D.F., and Zipser, D. (1985), 'Feature Discovery by Competitive Learning', *Cognitive Science*, 9, 75–112.
Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962), 'Pitch of the Residue', *Journal of the Acoustical Society of America*, 34, 1418–1424.
Slaney, M. (1990), 'A Perceptual Pitch Detector', in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Albuquerque.
Stumpf, C. (1898), 'Konsonanz und dissonanz', *Beiträge zur Akustik und Musikwissenschaft*, 1, 1–108.
Taylor, I., and Greenhough, M. (1994), 'Modelling Pitch Perception with Adaptive Resonance Theory Artificial Neural Networks', in *Musical Networks: Parallel Distributed Perception and Performance*, eds. N. Griffith, and P. Todd, Cambridge: MIT Press.
Tekman, H.G., and Bharucha, J.J. (1998), 'Implicit Knowledge Versus Psychoacoustic Similarity in Priming of Chords', *Journal of Experimental Psychology: Human Perception and Performance*, 24, 252–260.
Terhardt, E. (1972), 'Zur Tonhohenwahrnehmung von Klangen', *Acoustica*, 26, 173–199.
Terhardt, E. (1974), 'Pitch, Consonance, and Harmony', *Journal of the Acoustical Society of America*, 55, 1061–1069.
Terhardt, E. (1979), 'Calculating Virtual Pitch', *Hearing Research*, 1, 155–182.
Terhardt, E. (1982), 'Pitch of Complex Signals According to Virtual Pitch Theory: Tests, Examples, and Predictions', *Journal of the Acoustical Society of America*, 71, 671–678.
Terhardt, E. (1986), 'Gestalt Principles and Music Perception', in *Auditory Processing of Complex Sounds*, eds. W.A. Yost, and C.S. Watson, Hillsdale, NJ: Erlbaum.
Terhardt, E., Stoll, G., and Seewann, M. (1982), 'Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals', *Journal of the Acoustical Society of America*, 71, 679–688.
Thurlow, W.R., and Rawling, L.L. (1959), 'Discrimination of Number of Simultaneously Sounding Tones', *Journal of the Acoustical Society of America*, 31, 1332–1336.
Tillmann, B., Bharucha, J.J., and Bigand, E. (2000), 'Implicit Learning of Tonality: A Self-Organizing Approach', *Psychological Review*, 107, 885–913.
Tillmann, B., Bharucha, J.J., and Bigand, E. (2003), 'Learning and Perceiving Musical Structures: Further Insights from Artificial Neural Networks', in *The Cognitive Neuroscience of Music*, eds. R. Zatorre, and I. Peretz, Oxford: Oxford University Press.
Tillmann, B., Janata, P., Birk, J., and Bharucha, J.J. (2003), 'The Costs and Benefits of Tonal Centers for Chord Processing', *Journal of Experimental Psychology: Human Perception and Performance*, 29, 470–484.
Tillmann, B., Janata, P., Birk, J. and Bharucha, J.J. (2008), 'Tonal Centers and Expectancy: Facilitation or Inhibition of Chords at the Top of the Harmonic Hierarchy?', *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1031–1043.
von der Malsberg, C. (1973), 'Self-Organizing of Orientation Sensitive Cells in the Striate Cortex', *Kybernetic*, 14, 85–100.
Wightman, F.L. (1973a), 'Pitch and Stimulus Fine Structure', *Journal of the Acoustical Society of America*, 54, 397–406.
Wightman, F.L. (1973b), 'The Pattern-transformation Model of Pitch', *Journal of the Acoustical Society of America*, 54, 407–416.

Downloaded by [Jamshed Bharucha] at 13:51 04 November 2015