APPLYING PSYCHOACOUSTICS IN COMPOSITION: "HARMONIC" PROGRESSIONS OF "NONHARMONIC" SONORITIES

RICHARD PARNCUTT
AND
HANS STRASBURGER

INTRODUCTION

DURING THE LAST TWO DECADES, a psychoacoustic theory of pitch, consonance, and harmony has been developed by Ernst Terhardt and his colleagues at the Institute of Electroacoustics, Technical University of Munich. Applications of the model in music theory (Terhardt 1974a and 1976) include estimation of the perceptual importance (salience) of specific pitches in musical chords (Terhardt, Stoll, and Seewann 1982) and estimation of the strength of the harmonic relationship perceived between successive pairs of tones or chords (Parncutt 1989 and 1993).

Terhardt's model, as formulated and extended by Parncutt, would appear to have considerable potential as a basis for algorithmic composition. The aim of the present article is to explore ways in which this potential might be tapped. In other words, we would like to bridge the

gap between compositional and perceptual theory by making the results of psychoacoustic research more accessible to composers.

Like other compositional models, the present model aims to provide tonal materials for inclusion in musical works. The model differs from mainstream posttonal theory in several fundamental respects.

(i) The model takes as its starting point the spectrum of a sonority, rather than its musical notation. This allows one to work with individual pure-tone components, rather than notes with unspecified spectral content. Complex tones input to the model need not have harmonic spectra.

(ii) The model is not octave generalized; it is based on pitch height (measured on a one-dimensional scale, from low to high) rather than pitch class. Effects of voicing are not dealt with *post hoc*, but are intrinsic to the model. For example, octave transposition of components within a sonority affects the output of the model—although to a much smaller extent than transposition through other intervals.

(iii) In its present form, the model is capable of dealing only with individual sonorities—that is, tone *simultaneities*—and the harmonic and melodic relationships perceived between them. Compositional theories based on pitch-class sets tend to be more flexible, with relationships between sets applying regardless of whether pitch classes are sounded simultaneously or successively.

(iv) The model arises from a scientific tradition primarily concerned with the perceptual properties of individual, nonmusical sounds. By contrast, pitch-class theory has generally been driven by the requirements of music theory, analysis, and composition. Both approaches have specific advantages and disadvantages. An advantage of the psychoacoustic approach is that predictions of the model are based on experimental results, and may be expected to have a considerable degree of generality. A disadvantage is that such experiments have, for the most part, been performed on isolated, nonmusical sounds, making it difficult to assess their validity in a musical context.

In this article, we propose techniques for composing harmonic progressions of nonharmonic sonorities by calculating and manipulating relevant psychoacoustic parameters. The suggested techniques are intended only as examples; composers may apply the theory in many different ways and achieve a wide variety of musical results. Aspects of composition such as form, aesthetics, and cultural meaning are beyond the scope of the present article.

The ideas set out here have yet to be applied in composition. Foreseeable compositional applications promise to be rewarding, but will also require considerable programming expertise on the part of the composer, due to the complexity of the software needed to perform the various calculations and to generate the resultant sounds and progressions.[1] In any case, the present model touches only on one local aspect of the composition of a piece.

We begin by briefly examining relevant aspects of Terhardt's theory. We then provide a complete mathematical statement of the present model, and some examples of its application. Finally, we suggest tentative ways in which the theory may be applied in algorithmic composition.

THEORETICAL BACKGROUND

The perception of the pitch of a complex tone such as a musical tone (piano, violin, voice, and so on) involves pattern recognition (Goldstein 1973, Terhardt 1972 and 1974). The pattern in question is the pitch pattern formed by the audible harmonics of the tone, corresponding to the lower echelons of the harmonic series. The intervals between the elements of the pattern correspond to octaves, fifths, fourths, thirds, and so on.

According to Terhardt, this pattern becomes familiar to the auditory system during early life, by repeated exposure to harmonic-complex tones, especially in speech (i.e., vowels). Complex tones are subsequently perceived by matching this familiar pattern against real-time auditory spectra. Whenever a match occurs, a tone may be perceived at the lowest element of the pattern, with a pitch corresponding to the fundamental frequency. This may occur even if important elements of the pattern are missing, including (in the case of residue tones) the fundamental.

Terhardt calls pitches formed in this way *virtual pitches*,[2] and pitches corresponding to individual, audible pure-tone components (i.e., the audible partials) *spectral pitches*. Parncutt (1989, 33) instead focuses on the concept of *tone sensation* (or perceived tone), defining a *pure tone sensation* as a single tone sensation evoked by a pure tone or pure-tone component (partial), and a *complex-tone sensation* as a single tone sensation evoked by a complex tone or complex-tone component (e.g., a note in a chord). The definitions of Terhardt and Parncutt are closely related: A spectral pitch is the pitch of a pure-tone sensation, and a virtual pitch is the pitch of a complex-tone sensation. Parncutt's definition has the advantage that *attributes* of tone sensation, such as pitch and timbre, are clearly distinguished from the tone sensations themselves.

Tone sensations and their pitches differ in their perceptual promi-
nence, or *salience*. The salience of a spectral pitch or pure-tone sensation
depends on the audibility of the corresponding pure-tone component,
which in turn is affected by masking and by spectral dominance
(explained below). The salience of a virtual pitch or complex-tone sensa-
tion depends on how well the corresponding harmonic pitch pattern
(which may or may not include the fundamental) is represented among
the spectral pitches.

An important prerequisite for the perception of music is familiarity, or
conditioning. Listeners become conditioned to particular styles, and sub-
sequently "understand" them better, the more familiar stylistic elements
they contain. In Terhardt's approach, conditioning is a prerequisite not
only for musical understanding but also, at a more basic level, for the per-
ception of complex tones as individual entities, that is, for the perception
of virtual pitch. This in turn is necessary for sound source-identification
and speech understanding.

Musical conditioning and conditioning by exposure to individual com-
plex tones may be regarded as a *passive* process, in that consciousness is
not a prerequisite for perceptual learning, or an *active* process, in that it
generally involves physical interaction with the environment (Gibson
1979). For example, the acquisition of speech involves both passive expo-
sure to speech and active experimentation with the sounds that can be
produced by the vocal tract (Papousek and Papousek 1989). Aspects of
auditory conditioning may also be classified as predominantly universal,
applying to most or all of the human race, or as predominantly culture-
specific. For example, the ability to extract a single pitch from the audible
spectrum of a harmonic complex tone is presumably universal, as pitch
perception is a basic prerequisite for the perception of speech. However,
the perception of musical pitch structures may vary considerably from
one culture to another.

Both universal and culture-specific aspects of conditioning, as well as
active and passive aspects, may be modeled by neural nets (Laden and
Keefe 1989, Bharucha 1991). Neural nets can "learn" either by passive
exposure or by instruction, and then apply the results of that learning to
the processing of new inputs. In composition, original musical outputs
may be generated that are similar to musical structures to which a net has
previously been exposed (e.g., Todd 1989). However, the exact proce-
dure by which a net arrives at its results typically remains hidden from the
user. This is no problem, provided that only the output is of interest. The
present approach to algorithmic composition has the advantage that the
procedure linking input to output is explicit and under the composer's
direct control.

CONSONANCE OF INDIVIDUAL SONORITIES

A salient parameter in the delineation of musical structure is consonance-dissonance. This concept has been interpreted in many different ways during the history of western music, depending on whether it has been applied to successive or simultaneous dyads or chords, or to individual tones in chords (Tenney 1988). Here, we apply the concept to sonorities as a whole (cf. Tenney's "CDC-5"). In addition, we consider relationships between successive sonorities. These include harmonic, voice-leading (melodic), and tonal (key) relationships. Like most psychological parameters, perceived dissonance depends on experience and context: the cultural experience of the listener, and the musical context in which a particular sonority or progression is presented.

Of course, consonant sounds are not necessarily preferred to dissonant sounds. Instead, listeners tend to prefer a certain optimal amount of dissonance, complexity, or information flow (Smith and Cuddy 1986). Optimal dissonance gradually increased during much of the history of western music, and depends to a large extent on the musical experience and taste of individual listeners. Optimal dissonance may be conceived relative to an underlying "objective" scale of dissonance, that depends primarily on the acoustic signal and is relatively independent of "subjective" aspects such as musical experience and taste.

Various aspects of consonance-dissonance have been quantified for compositional purposes (e.g. Hiller and Isaacson 1958, Barlow 1987). Barlow, in a heuristic approach for use in his compositional work, defined the harmonicity of an interval as a function of the prime factors of the two integers making up its frequency ratio. He then derived a vocabulary of intervals for which harmonicity exceeded a certain value. The calculated harmonicity values of the intervals corresponded closely to their consonance in music theory, and vocabularies derived from different critical (minimum) harmonicity values corresponded closely to vocabularies of intervals actually used in tonal Western music.

The theory underlying Barlow's algorithm is, however, at conflict with recent research in perception. The role played by frequency ratios per se in interval perception is doubtful (Burns and Ward 1982); psychoacoustic research suggests that intervals are perceived and remembered not as frequency ratios but as pitch distances (Terhardt 1976). As a consequence, Barlow's algorithm, though compositionally useful in the intended scope of application, does not necessarily apply to more general cases of musical interest.

The consonance of a chord, according to Terhardt (1974a and 1976), has two main psychoacoustic aspects. One is called *sensory consonance*,

and is relatively independent of cultural conditioning. The most important component of sensory consonance in tonal music is smoothness, or absence of roughness. Roughness depends on the amplitude modulation of the overall waveform of a sonority, that is, on beating between pure-tone components. The roughness of a pair of simultaneous pure tones is maximal when the interval between them corresponds to about one quarter of a critical bandwidth (Plomp and Levelt 1965). Contributions to the overall roughness of a complex sonority may be modeled by summing contributions from different critical bands (Terhardt 1974b, Aures 1985a). Sensory consonance also depends on *sonorousness* (or *Klanghaftigkeit*, the perceptual clarity of the pitch or pitches evoked by a sonority), *sharpness* (depending on the amount of high-frequency energy in the sonority), and *loudness* (loud sounds tend to be more dissonant than quiet sounds) (Aures 1985b and 1985c).

The second main psychoacoustic correlate of consonance, according to Terhardt, is "harmony." This concept of harmony may be interpreted as involving familiarity with patterns of pitch within musical chords or sonorities, on various cognitive levels. The lowest level is not specific to music: Complex tones are perceived within a chord by harmonic pitch-pattern recognition, just as pitch is perceived in speech. In general, the easier it is for a complex tone to be perceived in this way within a chord, the greater is the consonance of the chord (Parncutt 1988 and 1989). At higher, more musical levels of familiarity, specific chords or chord progressions may be perceived as consonant if they occur often in music, or have particular musical functions.

Terhardt has explained the origins of the roots of chords in western music in terms of the perception of complex tones. The spectrum of a musical chord typically contains many incomplete harmonic patterns, suggesting the presence of incomplete (or partially masked) complex tones. The chord therefore evokes many different virtual pitches, corresponding (approximately) to the fundamental frequencies of these patterns. In general, the most prominent or salient of these tone sensations correspond to the actual notes in the chord, while less prominent tone sensations are octave equivalent or harmonically related to the notes.

The components of a given harmonic pattern may originate from different complex tones (notes). The perceptual integration of components from different notes is called blending. In general, the better a chord blends, the more likely it is to be perceived as a single sonority, with a single, well-defined root; and the more consonant it sounds. In musical performances, the auditory system is often able to parse which components in a musical chord belong to which notes, since different notes generally start at slightly different times (Rasch 1978), and the amplitude envelopes of the pure-tone components of a note normally rise and fall in

synchrony (McAdams 1984). However, these cues may be missing or weak in some cases, e.g. in a choir or string orchestra (ill-defined onsets of parts) or in an organ (without vibrato) in a reverberant church. In computer-generated composition, these cues may be manipulated by the composer. In any case it is clear that the perceived consonance of a chord depends not only on its notation but also on its physical realization.

All aspects of music perception may be influenced by cultural factors. Terhardt's theory is primarily concerned with psychophysical relationships as they obtain in a Western cultural context. According to the theory of virtual pitch, some of these relationships may have a universal basis, due to the universal exposure of the auditory system to complex tones in speech. Of course psychophysical relationships in non-Western contexts are also expected to have universal substrates; however, such relationships lie beyond the scope of the theory.

EXTENSIONS TO TERHARDT'S THEORY

The harmonic relationship between two chords is held in classical western music theory to depend on the degree to which the notes of the chords conform to the same diatonic scale, and on the distance between the chord roots on the cycle of fifths. These two factors may be accounted for by a single, more general parameter, called *pitch commonality* (Parncutt 1989 and 1993). Pitch commonality explains successive harmonic relationships more directly than diatonic scale conformity and distance on the cycle of fifths, as it depends to a smaller degree on musical conditioning and culture-bound music-theoretic concepts. Instead, it is directly linked to the acoustic makeup of the sounds themselves. It may appropriately be described as a *sensory basis* of successive harmonic relationships.

The pitch commonality of two sonorities may be defined as the degree to which they evoke, or are perceived to have, common pitches. As described above, perceived pitches normally correspond (roughly) to fundamentals of harmonic patterns—complete or incomplete—of audible pure-tone components. In the calculation of pitch commonality, common perceived pitches are weighted according to their calculated salience, or perceptual importance. Pitch commonality is greater if common pitches have greater salience, or if the number of coincidences in a given template match is high by comparison to the number of noncoincidences.

Perceived pitches are deemed "common" in the present definition if they both lie well within the same chromatic pitch category, or within

about a quarter of a semitone (say, twenty-five cents) of each other.[3] This margin of error may seem surprisingly wide; it is nevertheless consistent with variations (often imperceptible) in tuning observed in excellent performances by professional musicians (Fransson, Sundberg, and Tjernlund 1974).

As an example of pitch commonality, consider the chords of C major and D minor. These have no notes in common, yet they are perceived as harmonically related. Three possible explanations of this phenomenon come to mind. First, in music theory, the phenomenon may be explained in terms of proximity of the two chord roots on the cycle of fifths. This explanation is hard to generalize, as the root of a chord is often ambiguous. A second, psychological explanation is that both chords belong to the same "overlearned" diatonic scale (C major). A problem with this approach is that chords need not belong to the same scale to be perceived as harmonically related.

A third, psychoacoustic explanation is that the two chords have perceived pitches in common. The main pitches evoked by the C chord are, of course, C, E, and G. The chord also evokes other, perceptually weaker pitches, at harmonically related intervals. For example, A may be evoked, as the note E corresponds to the third harmonic of A, and the note G to the seventh harmonic of A. Similarly, the D minor chord evokes pitches not only at D, F, and A, but also at G, whose third harmonic is D, seventh harmonic is F, and ninth harmonic is A. So C major and D minor chords typically have at least two perceived pitches—A and G—in common.

Pitch commonality is related to various other similarity functions that have been defined in the music-theoretic literature. As noted by Forte (1973) and Morris (1987), there are two distinct ways in which pitch-class sets may be similar. They may have either interval classes or pitch classes in common.

The similarity of different transpositions and inversions of the same prime form may be related to the *interval* classes that they have in common. For example, a C major triad is similar to a Db major triad (by transposition), and major triads are similar to minor triads (by inversion). Morris (1979) proposed a more general measure of the (dis)similarity of two pitch-class sets—the difference between their interval-class vectors, or the number of interval classes that are present in one set but not in the other (called SIM). Morris derived a mathematical relationship between SIM, the number of interval classes common to two sets, and the respective cardinalities (number of pitch classes) of the sets.

Similarity due to common *pitch* classes was investigated by Lewin (1977 and 1979) and John Rahn (1979). Lewin defined a measure of

similarity called EMB, the number of distinct forms of one set that are embedded in, or subcollections of, the other set. Two sets may be regarded as similar if one is embedded in the other, or if they are both embedded in a third, larger set (such as a prevailing scale). For example, the major and minor triads are each embedded three times in the major scale. Lewin then normalized (or generalized) EMB by dividing it by its hypothetical maximum value (dependent on the cardinalities of the two sets in question). The result was a *probability* function (PROB) confined to the range 0 to 1. Rahn developed a measure of the similarity of two sets by counting the number of subsets of a specified size that are embedded mutually in the two sets (MEMB), summing the results over all possible subset sizes (TMEMB), and again normalizing the result so that its maximum value is one for equal sets.

The above similarity indices are *context-free* in that they do not discriminate among pitch class sets of the same type (Rahn 1979). Such sets may be discriminated by Regener's (1974) *common-note function*, defined as the number of pitch classes common to two pitch-class sets, for each of twelve transpositions of the second set. Regener also defined a partition function comprising the number of possible configurations of two sets producing a given number of common pitch classes.

In the present approach, we are primarily interested in relationships perceived between successive sonorities. In this regard, pitch commonality is of primary importance, as pitches common to successive sonorities are perceived more directly than intervals in common. On the other hand, sonorities containing similar patterns of intervals may have similar *timbres*. Timbral similarity is of central importance in electroacoustic composition; however, the problem is beyond the scope of the present article.

The present model goes beyond existing music-theoretic approaches to pitch relationships by taking into account the perceptual salience of each pitch. Pitch salience is understood as an idealized measure of the probability of consciously perceiving (or noticing) that pitch, when the sonority in which the pitch is embedded is heard in isolation, or in a randomly selected context. As we have seen, perceived pitches are generally fundamentals of incomplete harmonic series, that is, virtual pitches. The salience of a virtual pitch depends on how well the harmonic series is represented by audible pure-tone components (spectral pitches) at and above that pitch.

Relationships perceived between successive sonorities are also affected by voice leading. Melodic or voice-leading relationships are quantified in the present model by a parameter called *pitch distance*, or by its inverse, *pitch proximity*. The overall pitch distance perceived between two sonorities is assumed to depend on distances between *all* successive perceived

pitches within the sonorities, and on the perceptual salience of each pitch.

Conventional tonal theory generally requires that chromaticisms be resolved by step. For example, in an augmented sixth chord (such as Ab–C–F♯ in the key of C), the dissonances normally resolve by semitone steps. This suggests that, if a smooth connection is desirable, low pitch commonality may be compensated for by high pitch proximity. Conversely, relatively large leaps are often permitted between chords in close harmonic relationship, such as different inversions of the same chord, implying that low pitch proximity may be compensated for by high pitch commonality. In sum, chords may be perceived to progress smoothly due to *either* high pitch commonality (strong harmonic relationship), *or* high pitch proximity (good voice leading), or both.

A general treatment of pitch relationship (e.g., Jay Rahn 1992) might begin with the fundamental observation that two pitches can be either the same or different—that is, two tone sensations may lie in the same or in different pitch categories. The distinction between pitch commonality and pitch distance, and by inference between harmonic and melodic aspects of successive pitch relationships, may be traced to this fundamental dichotomy. In the model presented below, pitch commonality depends on successive tone sensations falling in the same pitch categories, whereas pitch distance depends on successive tone sensations falling in different pitch categories.

Roeder (1989) emphasized the dichotomy between harmonic and melodic relationships with reference to examples from Schönberg, Webern, and Berg, and demonstrated that voice-leading constraints, such as the prohibition of parallel octaves and common tones, can favor certain chord types. In other words, the consonance of a chord is dependent not only on its internal structure but also on the context in which it appears. The present approach to algorithmic composition is primarily concerned with nonharmonic sonorities. Voice-leading constraints of the kind considered by Roeder (defined in terms of musical notes, or harmonic-complex tones) are assumed here to have a relatively small influence on the consonance or prevalence of nonharmonic sonorities containing many pure-tone components. Sensory consonance is assumed to be largely independent of successive relationships, and vice-versa.

DESCRIPTION OF THE MODEL

A model incorporating algorithms for pitch commonality and pitch distance was presented by Parncutt (1989). The model was tested and

fine-tuned by comparing its predictions with the results of various experiments. The experiments involved auditory pitch analyses of tones and chords, and similarity ratings of successive pairs of tones and successive pairs of chords.

Here, we summarize aspects of that model that are relevant for composition. The summary provides sufficient information for the model to be implemented. The present account emphasizes compositional applications, whereas the 1989 version emphasized experimental testing of the model and its application in tonal theory. The present version differs from its predecessor primarily in the formulation for pure-tone height (Equation 3 below). This change necessitated adjustment of the free parameter k_M in Equation 4, and the constants k_p and k_c in Equations 12 and 13. Software implementations in Fortran and C are available from the first author.

INPUT

The input to the model is a progression of quasi-steady-state sonorities, each composed of pure-tone components in arbitrary frequency relationships. The physical spectrum of each sonority is determined by the *pitch categories P* (in semitones) and *auditory levels YL* (in dB above the threshold in quiet) of its pure-tone components.[4] Phase relationships seldom affect pitch perception, and are not considered in the model.

A *pitch category* is defined as a pitch range of width approximately ±25 cents about a step in the (stretched) equally tempered scale. The variable *P* gives the approximate position of the center of this category in semitones above 16.35 Hz (C_0), such that the pitch category of middle C (C_4, 262 Hz) is $P = 48$. In general, the pitch category of a given musical note may be calculated by multiplying its octave register by 12 and adding its pitch-class. For example, C_4 lies in register 4 and has a pitch-class of 0, so $P = (4 \times 12) + 0 = 48$. Similarly, A_4 has $P = 48 + 9 = 57$. Pitch-category values are limited to the range 0 to 120, or C_0 to C_{10} (corresponding to the range of hearing).[5]

The nominal center frequency *f* in kHz of each pitch category *P* may be calculated as follows:

$$f = 0.44 \times 2^{(P-57)/12} \tag{1}$$

since A_4 ($P = 57$) has a frequency of 0.44 kHz. A stretch of approximately ten cents per octave may be produced by replacing the value 12 in Equation 1 by 11.9.

Auditory level is defined as the level of a pure tone or pure-tone component relative to the free-field threshold of hearing in quiet. According to Terhardt et al. (1982), that threshold may be formulated

$$L_{TH} = 3.64\,f^{-0.8} - 6.5\,exp\,\{-0.6\,(f-3.3)^2\} + 10^{-3}\,f^4, \qquad (2)$$

where L_{TH} is threshold level in dB SPL, f is the frequency of the pure tone in kHz, and the function *exp* (exponential) means "e = 2.72 to the power."

Exact intonation within the twelve pitch categories is not specified in the model. However, certain limitations on tuning are recommended. First, octaves should be stretched (Ward 1954, Terhardt 1974a)—typically, by about ten cents per octave. Second, intervals *between* pitches should be tuned to within ±50 cents of stretched equal temperament. This latter condition will always be fulfilled if individual pitches are tuned to within ±25 cents of stretched equal temperament—that is, if they are confined to categories of width fifty cents. Composers may adjust tunings within these categories according to personal taste, or the requirements of a theoretical tuning system.

MASKING

The next stage of the model simulates masking between simultaneous pure-tone components. First, *pure-tone height* (or critical-band rate) is defined as the pitch-height of a pure tone or pure-tone component, expressed in units of critical bandwidth. Moore and Glasberg (1983), drawing on a range of experimental data on critical bandwidth and the shape of the auditory filter, proposed an analytic formula for *equivalent rectangular bandwidth ERB*, and derived from this a formula for *ERB*-rate (cf. Zwicker and Terhardt 1980). They formulated *ERB*-rate as follows:

$$H_p\,(f) = H_1\,\log_e\left(\frac{f+f_1}{f+f_2}\right) + H_0, \qquad (3)$$

where Hp is pure-tone height (or *ERB*-rate) expressed in units of equivalent rectangular bandwidth (erb) and f is frequency in kHz (see Equation 1). Moore and Glasberg chose the parameters H_1, H_0, f_1, and f_2 by nonlinear regression to experimental ERB estimates. A good fit was obtained with $H_1 = 11.17$ erb, $H_0 = 43.0$ erb, $f_1 = 0.312$ kHz, and $f_2 = 14.675$

kHz. H_p varies between about 0 erb at $f = 0$ kHz and about 36 erb at $f = 16$ kHz. The equation provides a good fit to experimental data in the approximate range from $f = 0.1$ to $f = 6.5$ kHz.

The degree to which pure-tone components mask each other depends on their distance from each other on the pure-tone height scale: Components which are closer to each other mask each other more. The degree to which a pure-tone masker of frequency f' and auditory level $TL(f')$ masks another component f (the maskee) is expressed in the model as a *partial masking level* $ml(f, f')$—the effective reduction in dB of the audible level of the maskee by the masker:

$$ml(f, f') = TL(f') - k_M |H_p(f') - H_p(f)|. \qquad (4)$$

The expression within vertical bars (absolute value symbols) is the distance between the two pure-tone components on the pure-tone height scale. This formulation assumes that the masking pattern is approximately symmetrical, which is true for low to medium sound levels (Zwicker and Jaroszewski 1982). The parameter k_M represents the gradient of the masking pattern of a single pure tone, and may be determined by comparing predictions of the entire masking algorithm with experimental data on the number of audible harmonics in typical complex tones. When k_M lies between 12 and 18 dB per critical band, the algorithm predicts that typical harmonic-complex tones in the central pitch range have about ten audible harmonics.[6] Note that if masking is negligible, ml is not zero, but large and negative.

For a calculation example, consider a simultaneous dyad of pure tones with frequencies $f = 0.4$ and $f' = 0.5$ Hz (a major third) and TLs (levels above the threshold in quiet) of 50 and 60 dB respectively. The above formulations may be used to predict the extent to which the components mask each other. The first step is to convert the two frequencies to pure-tone heights Hp according to Equation 3, producing the values 8.90 and 10.29 erb respectively. The interval between the two tones is thus 1.39 equivalent rectangular bandwidths. According to Equation 4, the higher component f' masks the lower f by $ml = 60 - (12 \times 1.39) \approx 43$ dB. The audible level AL (Equation 6 below) of the lower component is therefore $50 - 43 = 7$ dB—the component is only weakly audible. Conversely, the degree to which the lower component masks the upper component is $ml = 50 - (12 \times 1.39) = 33$ dB. The audible level of the higher component is thus $60 - 33 = 27$ dB—considerably greater than that of the lower one.

The *overall masking level ML* (in dB) at a given frequency f due to maskers at frequencies f' is estimated by addition of masking amplitudes.

Specifically, partial masking levels are converted to amplitudes,[7] added up, and converted back to levels:

$$ML(f) = \max \{20 \log_{10} \sum_{P \ne P'} 10^{*l(f,f')/20} \; ; 0\}, \tag{5}$$

where the max function prevents *ML* from becoming negative (in the case of no maskers). The *audible level AL* (in dB) of each pure-tone component (Terhardt: *SPL excess*) is defined as its level above masked threshold, and calculated as follows:

$$AL(P) = \max \{YL(P) - ML(P); 0\}. \tag{6}$$

Finally, the *audibility* A_p of each pure-tone component (Terhardt: spectral pitch weight) is made to saturate with increasing audible level:

$$A_p(P) = 1 - exp \; \{\frac{-AL(P)}{AL_\theta}\}. \tag{7}$$

The constant AL_0 was estimated experimentally by Hesse (1985) at about 15 dB. The output of this stage of the model thus consists of audibility estimates of the pure-tone components present in the input sonority. Components are assigned zero audibility if, according to the model, they completely mask each other. Only audible components are further processed.

RECOGNITION OF HARMONIC PITCH PATTERNS

In the next stage, the model simulates the recognition of harmonic pitch patterns among the audible pure-tone components of each sonority. A template is set up for this purpose, the components of which correspond to the audible harmonics of a typical complex tone (see Example 1).

The use of template matching to model pitch perception was advocated by Goldstein (1973). His pattern recognition procedure differed from the present model in that: (i) masking was not taken into account; (ii) perceived pitch was predicted by means of statistical estimation theory; and (iii) harmonics were assumed to be neighboring. Goldstein's mathematical formulations embody certain important aspects of pitch perception, but they are generally too sophisticated for musical purposes.

Specifically, partial masking levels are converted to amplitudes,[7] added up, and converted back to levels:

$$ML(f) = \max \{20 \log_{10} \sum_{P \neq P'} 10^{ml(f,f')/20} ; 0\}, \qquad (5)$$

where the max function prevents ML from becoming negative (in the case of no maskers). The *audible level AL* (in dB) of each pure-tone component (Terhardt: *SPL excess*) is defined as its level above masked threshold, and calculated as follows:

$$AL(P) = \max \{ YL(P) - ML(P); 0\}. \qquad (6)$$

Finally, the *audibility A_p* of each pure-tone component (Terhardt: spectral pitch weight) is made to saturate with increasing audible level:

$$A_p(P) = 1 - exp \{\frac{-AL(P)}{AL_o}\} . \qquad (7)$$

The constant AL_0 was estimated experimentally by Hesse (1985) at about 15 dB. The output of this stage of the model thus consists of audibility estimates of the pure-tone components present in the input sonority. Components are assigned zero audibility if, according to the model, they completely mask each other. Only audible components are further processed.

RECOGNITION OF HARMONIC PITCH PATTERNS

In the next stage, the model simulates the recognition of harmonic pitch patterns among the audible pure-tone components of each sonority. A template is set up for this purpose, the components of which correspond to the audible harmonics of a typical complex tone (see Example 1).

The use of template matching to model pitch perception was advocated by Goldstein (1973). His pattern recognition procedure differed from the present model in that: (i) masking was not taken into account; (ii) perceived pitch was predicted by means of statistical estimation theory; and (iii) harmonics were assumed to be neighboring. Goldstein's mathematical formulations embody certain important aspects of pitch perception, but they are generally too sophisticated for musical purposes.

In Terhardt's model (1972 and 1974a), virtual pitches are generated by a bottom-up procedure called *subharmonic coincidence detection*. The procedure aims to predict pitch properties of complex tonal signals as simply, directly, and accurately as possible, on the basis of available experimental data. By contrast, Goldstein's model (1973) is based on a top-down process of pattern matching that aims for an optimal estimate of fundamental frequency. At first sight, the two models seem quite different. Closer examination reveals that they have a number of basic features in common. Both essentially involve matching a harmonic-series template to a real-time stimulus. The outcome of the two procedures, when applied to most everyday environmental and musical sounds (including the musical examples given below), is quite similar. The procedures invoked (subharmonic coincidence detection, pattern matching) depend in both cases on assumptions about the prevalence or probability of specific configurations of spectral frequencies in the auditory environment. Neither model addresses the issue of temporal versus spectral cues in the determination of spectral pitches, and neither model attempts to emulate neurophysiological aspects of pitch perception.
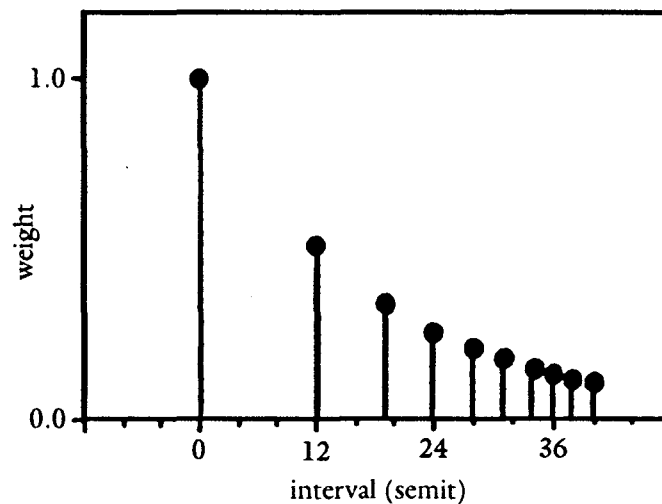
EXAMPLE 1: TEMPLATE USED TO SIMULATE THE RECOGNITION OF
HARMONIC PITCH PATTERNS (EQUATIONS 8, 9). THE LEFTMOST
(HIGHEST) COMPONENT CORRESPONDS TO THE FUNDAMENTAL,
THE SECOND TO THE SECOND HARMONIC, ETC.

In the present model, the size of the harmonic template is limited to ten components, for the following reasons. First, the eleventh harmonic is usually inaudible in typical harmonic-complex tones (Plomp 1964), and so—in Terhardt's approach—presumably plays little or no role in pitch perception. Second, the eleventh harmonic is hard to categorize relative to the conventional chromatic scale, lying 5.5 semitones above the eighth harmonic.

Each component is labeled according to its harmonic number $n = 1$, 10. The pitch of each element is given by its pitch category P_n relative to the lowest element P_1:

$$P_n = P_1 + \text{int}\{12 \log_2(n) + 0.5\}, \qquad (8)$$

where int denotes integer part. The pitch of the third harmonic, for example, is $P_3 = P_1 + 19$.

The components are also weighted relative to each other. The higher the harmonic number n, the lower the weight W_n, as higher harmonics are less often audible, and hence less important for complex tone perception:

$$W_n = \frac{1}{n} \qquad (9)$$

For each sonority, the template is shifted in steps of one semitone through the entire pitch range. In each position, matches are sought between the template and the spectrum of audible components. Wherever a match is found, the pitch category at the template's fundamental is assigned a template-match weight called *complex-tone audibility* A_c. Its value depends on both the audibilities of matching pure-tone components and the weights of corresponding template components. Complex-tone audibility A_c may be regarded as a measure of the degree to which the harmonic series, or part thereof, is embedded in the audible spectrum of a sonority at a given pitch. We define it as a nonlinear sum of pure-tone audibilities, where each audibility value is weighted according to its effective harmonic number:

$$A_c(P_1) = \frac{1}{k_T}\left(\sum_n \sqrt{W_n A_p(P_n)}\right)^2, \qquad (10)$$

where k_T is a free parameter, chosen such that resulting values of complex-tone audibility A_c are correctly scaled relative to pure-tone audibility

values A_p (Parncutt 1989). The parameter k_T typically takes a value of about 3. In analytic listening where pure-tone components tend to be heard out, k_T generally exceeds 3 (approaching roughly 10). In holistic listening, k_T may be smaller than 3 (say, 1).

Where pure- and complex-tone components share the same pitch category, the one with the higher salience prevails:

$$A(P) = \max \{A_p(P) ; (A_c(P))\}, \tag{11}$$

where A denotes the audibility of a tone component, regardless of whether it is pure or complex. This equation, borrowed from Terhardt et al. 1982, is consistent with the assumption that it is impossible to hear both a pure and a complex-tone sensation simultaneously at the same pitch. Calculations according to Equation 11 (with the max function) agree more closely with experimental results than calculations based on the sum of A_p and A_c (Parncutt 1989).

SENSORY CONSONANCE

As described above, consonance may act as a parameter according to which sonorities are selected for use in composition. The consonance of a sonority is modeled here in terms of the variables *roughness* (*Rauhigkeit*) and *sonorousness* (*Klanghaftigkeit*, often translated as *tonalness*)—the main aspects of Terhardt's sensory consonance. We consider the general case of non-harmonic sonorities (rather than just chords), with the aim of composing progressions of non-harmonic sonorities from psychoacoustic principles.

Following Aures 1985c (Equation 11), we define the *pure sonorousness* T_p of a sonority as the normalized quadratic sum of the audibilities of its pure-tone components:[8]

$$T_p = K_p \sqrt{\sum_P A_p(P)^2} \tag{12}$$

where K_p is a constant. The *complex sonorousness* T_c of a sonority is given by the audibility of its most audible complex-tone component—that is, the extent to which the harmonic series is audibly embedded in the sonority:

$$T_c = K_c \max_P \{A_c(P)\}. \tag{13}$$

For typical musical tones, both these formulations of sonorousness are maximal near the center of the musical pitch range (the lower part of register 4). The main reason for this is spectral dominance (Plomp 1967, Ritsma 1967): spectral pitches are more salient in a broad region centered on about 700 Hz (Terhardt et al. 1982), and virtual pitches, on average, are usually lower than the spectral pitches that produce them. Moreover, lower-pitched complex tones are less sonorous due to increased masking between neighboring pure-tone components. The maximum degree of sonorousness may be normalized to a value of approximately one by setting $K_p = 0.5$ in Equation 12 and $K_c = 0.2$ in Equation 13.[9]

The *roughness* of a sonority may be estimated along the lines of Plomp and Levelt 1965, Kameoka and Kuriyagawa 1969, Terhardt 1974b, Hutchinson and Knopoff 1978,[10] Danner 1985, and/or Aures 1985a. The result may then be combined with sonorousness, to produce an estimate of overall consonance-dissonance. Rough sounds generally have low (pure or complex) sonorousness, since roughness is produced by beating between nearby pure-tone components, and nearby components mask each other, reducing sonorousness. Similarly, smooth sounds tend to have high sonorousness. For this reason it is often sufficient to evaluate only sonorousness and to neglect roughness, or vice-versa.[11]

MULTIPLICITY AND SALIENCE

We define the multiplicity of a sonority as the number of tones consciously perceived (or noticed) in that sonority. Multiplicity (otherwise known in the psychological literature as *numerosity*) may be measured simply by presenting sonorities to listeners and asking them how many tones they hear. Results depend considerably on listener and context (Parncutt 1989 and 1993). In the following, multiplicity values are assumed to be averaged over a large numbers of listeners and contexts. (The same applies to most parameters in the model—notably, the audibility and salience of tone components or sensations.)

An initial estimate of multiplicity is made on the basis of tone audibility values from Equation 11:

$$M' = \frac{1}{A_{max}} \sum_P A(P), \tag{14}$$

where $A_{max} = \max_P \{A(P)\}$. This initial estimate is then adjusted, or scaled, as follows:

$$M = (M')^{k_S}, \tag{15}$$

where k_S depends on the mode of listening. A typical value of k_S is 0.5. In analytic listening, k_S is higher (approaching 1), while in holistic listening, it is lower (approaching 0).

Tone salience S is formulated as a probability of consciously perceiving (or noticing) a given pitch, as follows:

$$S(P) = \frac{A(P)}{A_{max}} \frac{M}{M'}. \tag{16}$$

The interpretation S as a probability of noticing may be confirmed by summing the right-hand side of Equation 16 over all 120 values of P, and substituting Equation 14 for M'. This procedure yields the definitive equation

$$M = \sum_P S(P).$$

In other words, the sum of the probabilities of noticing all tone saliences in a sonority is equal to the average number of simultaneously perceived tones in that sonority.

SUCCESSIVE PITCH RELATIONSHIPS

The pitch commonality C of two successive sonorities may be evaluated as a Pearson correlation coefficient r between the two tone salience profiles S (arrays of up to 120 elements).[12] When formulated in this way, pitch commonality satisfies the following conditions:

(1) It increases with the number of perceived pitches common to two sonorities, taking into account their calculated saliences (probabilities of noticing).

(2) It is equal to 1 in the special case of two identical sonorities, and has a hypothetical minimum value of –1 (for perfectly complementary sonorities).

Regarding the overall pitch distance between two sonorities, consider a particular pitch P in the first sonority and another pitch P' in the second. Let the probability of noticing a pitch P in the first sonority be $S_1(P)$, and the probability of noticing pitch P' in the second sonority be $S_2(P')$. Assuming independent probabilities, the probability of noticing *both* pitch P in the first sonority and P' in the second sonority is $S_1(P)\,S_2(P')$. If both these pitches are perceived, then it may be assumed that the interval between them is perceived—not necessarily recognized as containing a certain number of semitones, but at least perceived as relatively big or small on a continuous scale.[13] The size of the interval between the two pitches in semitones is $|P' - P|$. On this basis, we define the overall pitch distance between the two sonorities as:

$$D = \sum_P \sum_{P'} S_1(P)\,S_2(P')\,|P' - P|$$

$$- \sqrt{\sum_P \sum_{P'} S_1(P)\,S_1(P')\,|P' - P| \sum_P \sum_{P'} S_2(P)\,S_2(P')\,|P' - P|}, \qquad (17)$$

where both P and P' vary from 0 to 120. This formula takes into account all possible intervals between pitches perceived in two sonorities. In this respect it is reminiscent of Lewin's (1977) interval function, which calculates the probability of hearing a particular interval class between two sets of pitch classes. The present function is more sophisticated in that it is not octave-generalized, and additionally takes into account the probability of noticing each pitch.

The above formula for pitch distance satisfies the following requirements:

(1) In the case of two identical sonorities, pitch distance is zero.

(2) In the case of two pure tones, pitch distance is equal to the interval between them in semitones.

(3) Pitch distance always exceeds zero.

The above formulations of pitch commonality and pitch distance are based on the assumption that pitch saliences represent independent probabilities. Research in melodic stream segregation (Bregman 1990) has shown that this is not entirely true. The probability of noticing a given pitch generally depends to some extent on other pitches sounding simultaneously, as well as the temporal context of the pitch. The formulae have nevertheless been found to account for the results of several experiments

involving comparisons between quasi steady-state sonorities (Parncutt 1989) and are probably sufficiently accurate for music-theoretic applications involving homorhythmic progressions.

Like roughness and sonorousness, pitch commonality and pitch proximity (the inverse of pitch distance) may be combined to produce a measure of the overall tonal relationship between any two sonorities in a composition. However, the proportions in which they are combined depend on context, style, and aesthetics.

EXAMPLES

To illustrate the model's operation, three examples are given in Examples 2 and 3. Example 2 shows three sonorities in common music notation. The first sonority (Example 2a) is a chord familiar from mainstream jazz. The second and third sonorities (Examples 2b and c) are not made up of regular harmonic-complex tones, so they are notated as simultaneities of pure-tone components.
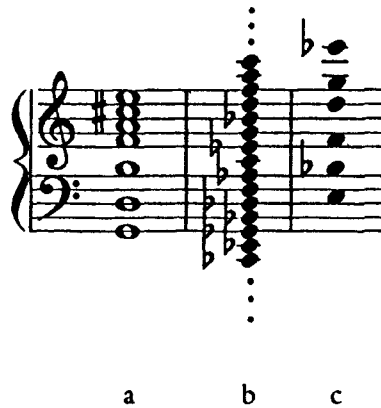
EXAMPLE 2: MUSIC NOTATION OF THE THREE SONORITIES ANALYZED IN
EXAMPLE 3. THE OPEN NOTEHEADS IN (a) REPRESENT ORDINARY
MUSICAL NOTES (HARMONIC-COMPLEX TONES). THE BLACK
NOTEHEADS IN (b) AND (c) REPRESENT PURE-TONE COMPONENTS.
THE COMPONENTS IN (b) CONTINUE UPWARD AND DOWNWARD
TO THE THRESHOLD OF HEARING.

Example 2b may be described as a *simultaneous dyad of fifth-complex tones*. Octave-complex tones—complex tones whose pure-tone components span octave intervals across a wide frequency band—are well-known from Shepard's (1964) investigations. By analogy, fifth-complex tones may be defined as complex tones whose pure-tone components form chains of perfect fifth intervals across a wide range of frequencies. Here, for the purpose of argument, the tone components are allowed to cover the entire audible range. In practice, the roughness of this sonority may be reduced by attenuating the low-frequency components. Example 2b shows a simultaneous dyad of such tones, one including a pure-tone component at C_4, the other at $E\flat_4$.

The third example is a sonority constructed according to the following requirements: (i) The sonority comprises six pure-tone components; (ii) intervals between neighboring components vary randomly between five and ten semitones; (iii) no pitch class is repeated; (iv) the mean pitch of the components is approximately A_4 (440 Hz). Specifications of this kind may be used to generate a vocabulary of sonorities for use in composition. A more sophisticated procedure could include limitations on sensory consonance (roughness, sonorousness) or other requirements stipulated by a composer.

Example 3 shows (i) the physical spectra of the above three examples, and (ii) corresponding "perceptual spectra" according to the model. The physical spectra are expressed in terms of *auditory level*, defined as level relative to the threshold of hearing. In the figure, levels have been chosen to be musically typical; in general, a composer is free to determine and adjust the levels of components according to musical requirements.

By definition, a pure-tone component with a positive auditory level will always be audible when presented in isolation. However, masking by other components may render it inaudible. In Example 3a, masking is more severe at higher frequencies, where components are closer together in pitch. In Example 3b, pure-tone components are spaced at roughly equal intervals (major and minor thirds) across the audible range; masking is most severe at lower frequencies, where critical bandwidth (expressed in semitones) is greater. The intervals between pure-tone components in Example 3c are relatively large, so little masking occurs.

Parts (ii) of Example 3a and b include complex-tone sensations or virtual pitches (white bars). These are derived from pure-tone sensations or spectral pitches (black bars).

The most salient complex-tone sensations in the bass region of Example 3a correspond to pitch-class G, the root of the chord. Another prominent complex-tone sensation occurs at pitch-class A in the central pitch region. The A may be regarded as a subsidiary root and suggests

that the chord may be interpreted as bitonal. The model treats the C$\sharp$ in this chord as a major third (or fifth harmonic) above A, but not as the eleventh harmonic of G, as the harmonic template (Example 1) is limited to ten components.

The described model may produce a range of different qualities of chord progression. For example, if individual chords are constrained to have high sensory consonance, and harmonic and melodic relationships between chords are constrained to be strong (high pitch commonality, low overall pitch distance), but if there is still melodic movement in most or all parts (so that successive chords have few—or, as is often the case at cadences, no—actual tones in common), then modal-sounding chord progressions may be expected.[14] If chords are allowed to have greater dissonance, and harmonic relationships are less strong, but pitch distance is constrained to small values, then a kind of chromatic harmony might result (consider, for example, the progression from French augmented-sixth to cadential six-four).

The sonority depicted in Example 3b is highly ambiguous regarding its most salient pitch. The strongest tone sensations, according to the model, correspond to the pure-tone components of the fifth-complex tone on C_4; the complex-tone sensations corresponding to the fifth-complex tone on $E\flat_4$ are much weaker. This difference may be understood by looking at which high components correspond to harmonics of which low components. For example, the component at C_4 is supported by "pseudo-harmonics" at C_6, E_6, G_6, and D_7. However, the component at $E\flat_4$ is supported only by G_6.

The pitch circularity of octave-complex tones, as demonstrated by Shepard (1964), applies equally well to any complex tone whose pure-tone components are equally-spaced in the logarithm of frequency (Burns 1981), including fifth-complex tones. Given a constant spectral envelope, there exist only seven different fifth-complex tones in the chromatic scale (pitch classes 0 to 6 in a modulo-7 system), and only three different interval-classes between them (one to three semitones). The described model makes specific predictions concerning harmonic relationships (pitch commonality) between successive pairs of fifth-complex tones, and between successive pairs of simultaneous dyads of fifth-complex tones, and may therefore aid in creating meaningful harmonic progressions from such sonorities. Such progressions could conceivably become *functional* (Riemann 1893) in that one sonority (analogous to the tonic in the major-minor system) functions as a goal of harmonic motion, another sonority (analogous to the dominant) functions as a harmonic preparation for the tonic, and other sonorities (analogous to subdominant or pre-dominant functions) prepare the dominant.
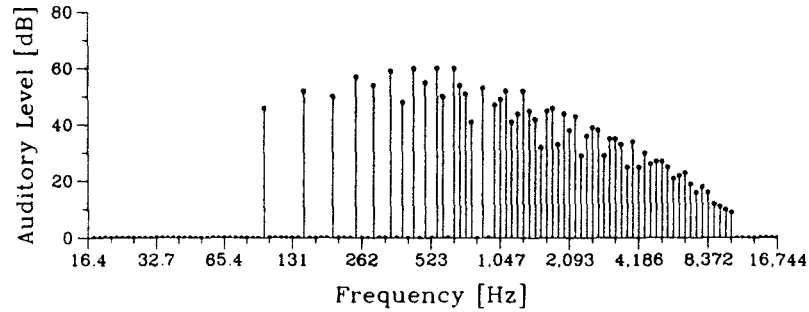
a(i)
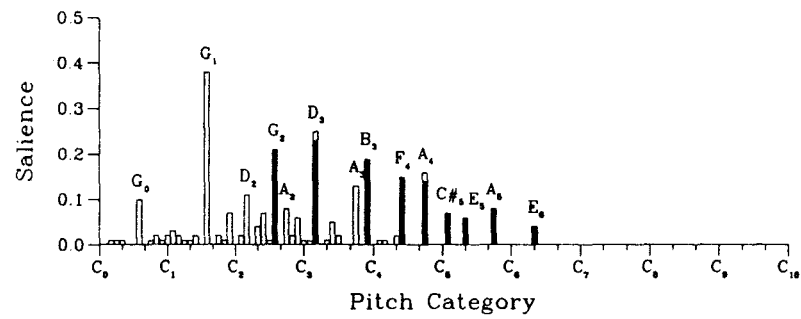
a(ii)

EXAMPLE 3: ANALYSES OF THE THREE SONORITIES IN EXAMPLE 2.
(i) SPECTRA SHOWING AUDITORY LEVEL (ΥL, LEVEL ABOVE THRESHOLD
IN QUIET) AGAINST LOG FREQUENCY FOR EACH PURE-TONE COMPONENT.
(ii) CALCULATED SALIENCES OF ALL EVOKED TONE SENSATIONS
ACCORDING TO EQUATION 16.
BLACK BARS: PURE-TONE SENSATIONS (SPECTRAL PITCHES).
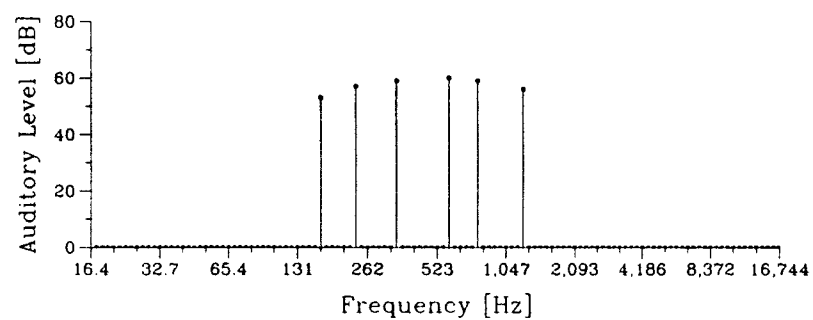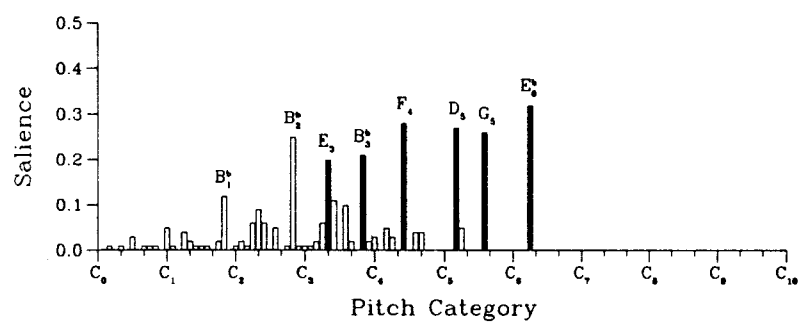WHITE BARS: COMPLEX-TONE SENSATIONS (VIRTUAL PITCHES).
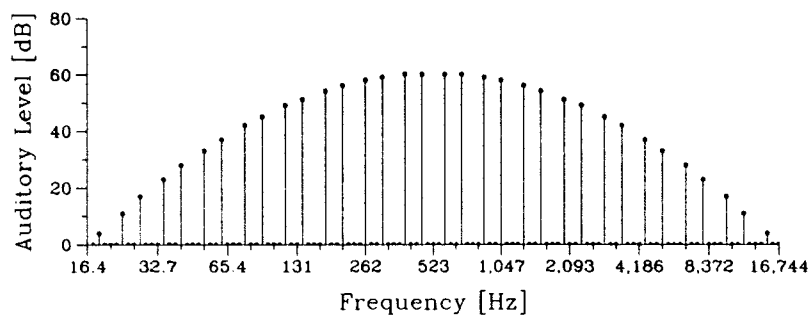
b(i)
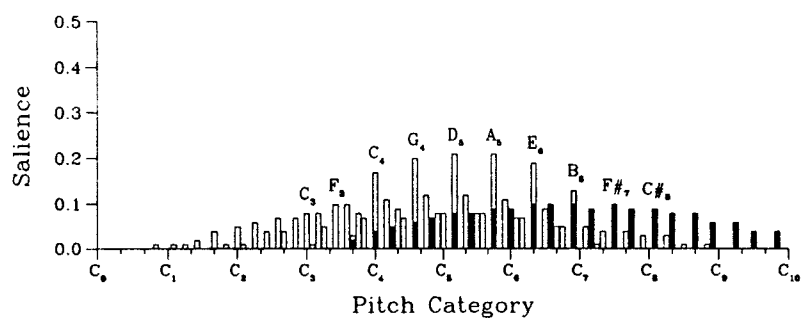
b(ii)

EXAMPLE 3: (CONT.)

c(i)

c(ii)

EXAMPLE 3: (CONT.)

Example 3c suggests that sonorities with relatively few components may still evoke pitches that do not correspond to frequencies. The stochastic procedure used to "compose" this sonority resulted in a spectrum containing part of a harmonic series: the embedded pitch set [Bb_3, F_4, D_5]. If these pitches are interpreted as harmonics 2, 3, and 5, the implied fundamental is Bb_2; if 4, 6, and 10, Bb_1. For the present purpose, this embedded harmonic pitch set is the most interesting aspect of this sonority. In composition, a family of sonorities containing embedded harmonic fragments of this kind might be produced by generating all possible sonorities whose sonorousness lies in a given range, where sonorousness is formulated as maximum complex-tone audibility (see Equation 10).

APPLICATION OF THE THEORY IN COMPOSITION

The described theory may be applied in composition in various ways. One way is to use the model to produce homogeneity of consonance, that is, to limit the consonance of sonorities and the strength of successive relationships to specific ranges deemed typical of a particular style. More generally, consonance and successive relationships may be allowed to follow a predetermined course during a progression, or an entire piece. Composers may thus adapt the model to suit their own purposes.

SELECTING A VOCABULARY OF TONAL SONORITIES

Historically, the chord vocabulary of Western tonal music has been determined by a number of factors, of which the most important is probably the consonance of individual chords. The well-known gradual expansion of harmonic vocabulary over several centuries to include intervals of thirds and sixths, major and minor triads, and seventh chords as harmonic entities reflects a gradual change in preferred degrees of consonance and dissonance. Different styles exploit different sections of the consonance-dissonance spectrum. Harmonic progressions in Renaissance music cover a relatively small range of consonance-dissonance, placed toward the consonant end of the spectrum—at least from a modern viewpoint. Romantic music covers a wider, more central range of consonance-dissonance. Many twentieth-century serial and jazz styles use a relatively small range of consonance-dissonance, placed toward the dissonant end of the spectrum.

Following this tradition, progressions of nonharmonic tonal sonorities may be composed by first imposing specific limits on the consonance/

dissonance of each sonority according to psychoacoustically based indices such as roughness and sonorousness. A kind of tonal vocabulary may then be established by identifying all possible sonorities within these limits (cf. Barlow 1987). In the process one may also specify frequency ranges, both overall and for each voice (if appropriate), and other limitations appropriate to the intended style.

COMPOSING PROGRESSIONS

Once a manageable vocabulary of sonorities has been established, progressions of sonorities may be composed by specifying homogeneous, increasing, decreasing, or irregular levels of consonance. For example, a homogeneous progression could be created by limiting pitch commonality and pitch distance values to specific ranges. Alternatively, these ranges could be allowed to evolve in a controlled fashion. A kind of tonality could be established by defining one sonority, which need not actually appear in the progression, as a kind of tonic, and requiring other sonorities to conform to a certain range of pitch commonality with the tonic sonority. Progressions composed according to restrictions such as these may then be used as raw material for a piece.

Tonal progressions may also be composed stochastically, that is, according to prescribed probabilities; see Jones (1981) and Dodge and Jerse (1985) for practical algorithms. At the simplest level, sonorities may be made to appear with probabilities determined by their calculated consonance. For example, a bell-shaped probability distribution may be chosen such that sounds with consonances near the peak of the distribution are most likely to appear. A more sophisticated stochastic algorithm might incorporate a Markov chain in which the probability of a sonority's appearance depends on the sonorities that precede it. Pitch commonality and proximity values may be used to formulate the required probability matrix. Stochastic algorithms may be implemented so as to "compose" original progressions in real time.

Once progressions have been created, it may be possible to relax the constraint that the pure-tone components of each sonority be sounded simultaneously. Harmonies could be prolonged by staggering onsets of pure-tone components, either individually or in groups, or by varying amplitudes incoherently so as to encourage stream segregation within harmonies. In this way it may be possible to introduce contrapuntal elements to a harmonic progression of nonharmonic sonorities.

CONCLUSION

The proposed techniques allow the control of low-level psychoacoustic properties of progressions of tonal sonorities, with the aim of creating raw materials for composition. Higher-level cognitive organization of tonal progressions is not considered here. Organizational principles on the basis of generative grammars, for example, have been implemented by Roads (1978). Incidentally, there is no reason why two quite different compositional algorithms cannot be used in the one piece: one, such as the model in this paper, for composing musical elements, and another for organizing them.

An important and useful technique in computer music is the gradual deviation from realistic instrumental *timbres*. This may be achieved by physical modeling of musical instruments (e.g., Karplus and Strong 1983, Smith 1987), by gradually modifying model parameters in the direction of physically impossible values. The procedures we have described may enable an analogous procedure in the domain of tonality, that is, gradual deviations from familiar harmonic progressions. The model would first be adjusted to produce conventional-sounding chord progressions, by confining values of roughness, pitch commonality, and so on to ranges typical of western music. After that, parameter values would be allowed to deviate gradually from familiar values, or complex-tone components would be allowed to deviate gradually from harmonicity.

Another central issue in computer music is that of *form*. Wessel 1979, Lerdahl 1987, and McAdams and Saariaho 1991 have described how timbre may be used as a basis for formal structures, taking over the role normally played by pitch in tonal Western music. Similar approaches could be applied to the formal organization of progressions of non-harmonic sonorities, in which the boundary between pitch structure and timbral structure is deliberately blurred.

ACKNOWLEDGMENTS

APPENDIX

ON SUBDIVISION OF THE OCTAVE

The limitation to twelve pitch categories per octave is clearly culture-specific, and is adopted here primarily as a matter of convenience. The model may equally well be applied to other equal subdivisions (such as 19, 31, or 53: see e.g. Blackwood 1991) by adopting a different categorization scheme for the pitches of pure-tone components as well as the model's "harmonic template." This may be done simply by replacing the number 12 in Equations 1 and 8 by another number. Alternatively, pitch categorization could be removed from the procedure altogether, and composers could work with pitch on a continuous, one-dimensional scale (as in Terhardt et al. 1982).

Several possible reasons can be cited for the predominance of the twelve-note chromatic scale in Western music. From the performer's viewpoint, the minimum spacing of scale steps may be traced to physical performance limitations such as the width of the violinists' fingertips, typical ranges of singing vibrato, and so on. From the listener's viewpoint, there are limits to the amount of information that may be processed by the cognitive system.

Psychoacoustic experiments have demonstrated the existence of an additional perceptual constraint. Moore, Glasberg, and Peters (1985) presented listeners with steady-state harmonic-complex tones in which one component had been shifted in frequency. They found that, for a frequency shift of up to about 50 cents (3% in frequency), the virtual pitch near the lowest component shifted in the same direction as the harmonic, suggesting that such a harmonic continues to make a full contribution to the salience of the virtual pitch. Darwin and Hill (1993) confirmed this result and additionally found that a harmonic mistuned by up to 50 cents continues to make a full contribution to the vowel quality (timbre) of the complex tone. The two studies further demonstrated that, as the mistuning of a harmonic is gradually increased to about 130 cents (8%), virtual pitch gradually returns to its original position (i.e., the mistuned harmonic gradually stops contributing to the virtual pitch), and the vowel quality of the complex ceases to be influenced by the mistuned harmonic. These results were relatively independent of harmonic number, depending much more on the size of the frequency shift.

A match between a pitch-perception template component and a real-time spectral pitch may thus be out of tune by up to 130 cents and still contribute (albeit increasingly weakly, as mistuning increases from 50 to 130 cents) to the perception of the pitch and timbre of a complex tone.

This idea is incorporated into the pitch algorithm of Terhardt et al. (1982). In that model, the contribution of a pair of spectral pitches to a virtual pitch decreases linearly in proportion to the mistuning of the interval between them. The contribution falls to zero when the deviation reaches 8% of frequency.

If, as Terhardt (1974 and 1976) suggests, the root of a chord is a virtual pitch, then the results of Moore, Darwin, and colleagues may have important consequences for the width of musical pitch categories. Consider, for the purpose of argument, the chord $C_4E_4G_4$, where C_4 is middle C. The many pitches evoked by this chord include the pitch-class C (the conventional root) in several different octave registers (for experimental data see Parncutt 1989, Section 5.3). Let us concentrate on the perceived (residue) pitch at C_3. The above data suggest that, if the tone C_4 is shifted in frequency by up to about 50 cents, the perceived pitch at C_3 will shift in the same direction, but to a smaller degree. If the tone is shifted by more than 50 cents, the perceived pitch at C_3 will gradually move back to its original position, until at a shift of 130 cents the tone near C_4 no longer contributes to the salience of the pitch at C_3. The pitch perceived at C_3 will thus become progressively weaker. Meanwhile, other, nearby pitches will have become stronger. For example, if the note C_4 in the chord $C_4E_4G_4$ is shifted downward in the direction of B_3, the perceived pitch at E_3 will likely become more prominent than that at C_3; and if the note C_4 is shifted upward in the direction of $C\sharp_4$, the perceived pitch at A_2 will become more prominent than C_3. The above observations are consistent with both psychoacoustics and musical experience, suggesting a causal link between the two.

Effects of mistuning are also influenced by musical experience and conditioning. In the above example, the degree of mistuning of the note C_4 required to change the root of the chord $C_4E_4G_4$ from C to E or from C to A depends on musical context. If the context is consistent with the interpretation of the mistuned chord as E minor or A dominant-seventh, then less mistuning will be needed to induce these interpretations. If the context is inconsistent with these interpretations, then more mistuning will be required to induce them. The degree of mistuning required to change the root probably also depends to a considerable extent on the listener.

The findings of Moore and of Darwin, together with Terhardt's hypothesis that the root of a chord is a virtual pitch, suggest that there may be a psychoacoustically determined limit to the density of harmonically distinct steps in a scale (Parncutt 1989). By "harmonically distinct" we mean contributing to different virtual pitches, which—according to Terhardt—is essentially the same thing as contributing to different roots.

The psychoacoustic data suggest that the maximum possible density of harmonically distinct scale steps lies somewhere in the range nine to twenty-four per octave, corresponding to the mistuning limits of 130 cents and 50 cents respectively. As the density of scale steps is increased, there comes a point where neighboring scale steps can contribute to the same virtual pitch—that is, act as different tunings of the same harmonic function. Such scale steps would no longer be harmonically distinct. If, to be on the safe side, scale-step mistunings of about ±50 cents are allowed, then the maximum density of scale steps is approximately twelve per octave.[15]

The effect of the mistuning of a single chord tone on the exact pitch of its root has not been investigated experimentally. A similar effect has nevertheless been observed in the case of melodic musical intervals. Siegel and Siegel (1977) and Burns and Ward (1978) observed that melodic intervals could be mistuned by up to 50 cents and still be quickly recognized by musicians as belonging to the nearest chromatic interval category (minor second, major second, minor third, and so forth). These data are apparently determined by musical experience, and are probably not directly related to the psychoacoustic measurements of Moore and of Darwin.

What could determine perceptual tuning tolerances for components of complex tones? Terhardt has suggested that pattern recognition in pitch perception is primarily learned from exposure to complex tones, especially speech vowels. If this is true, then the auditory system's tolerance to the mistuning of harmonics in complex tones may, like tolerance to the mistuning of musical intervals, be determined by auditory experience. The pure-tone components of everyday complex tones may be either physically mistuned (as in the stretched pure-tone components of piano tones: Schuck and Young 1943) or "perceptually mistuned" due to pitch shifts (Terhardt 1972 and 1974a). Both these kinds of "mistuning" regularly approach or exceed ±50 cents. Clearly, the auditory system needs to be able to accommodate mistunings of this magnitude if it is to recognize mistuned harmonics as belonging to a single sound source.

Is the width of perceptual pitch categories immutable? Terhardt's theory of pitch suggests not. If the width of pitch categories are indeed dependent on learning, then it should be possible to reduce their width by repeated exposure to appropriate patterns of sound—for example, microtonal music. However, psychological literature on perceptual imprinting suggests that such exposure would only be effective if it occurred quite early in life. In the visual system, for example, the spatial frequency bandwidth of neurons is learned by imprinting at an early age.

In summary, Terhardt's concept of pitch perception is consistent with the view that tuning preferences and limitations are largely a matter of familiarity and conditioning.

Notes

1. An implementation of the basic algorithm described in this paper (written in the C programming language) may be obtained from the first author, email parncutt@music.mcgill.ca.

2. Note that virtual pitch is not the same as residue pitch. A residue pitch is a virtual pitch at a missing fundamental. The fundamental is present and audible in most harmonic complex tones, so most virtual pitches in everyday sounds correspond to spectral components. However the exact pitch of a complex tone is generally determined by higher components (Plomp 1967, Ritsma 1967).

3. The range within which pitches are considered to coincide decreases as the number of successive pitches increases. Consider three successive pitches x, y, and z, each embedded in a separate sonority, and let d be the margin of error within which pitches are deemed to coincide: Pitches x and y coincide if $|x - y| < d$. Jay Rahn (1992) has pointed out that this relationship is not transitive, i.e., that if both $|x - y| < d$ and $|y - z| < d$ then it is not necessarily true that $|x - z| < d$. In other words, x may be perceived (categorically) equal to y in one context, and y equal to z in another, but when x, y, and z follow each other in succession, it is possible that x will be perceived to be different from z. This effect has been observed in experiments on the perception of microtonal scales (Parncutt and Cohen, in preparation): In rising scales of microtonal intervals, the first two notes may be perceived to be identical, while the third is perceived to be different from both its predecessors—even though all notes are equally spaced in log frequency.

4. The symbol YL distinguishes audito*ry* l*e*vel YL from *a*udible l*e*vel AL. YL is level above the threshold in quiet; AL is level above masked threshold, that is, threshold in the presence of a masking sound. Auditory level YL is not to be confused with sensation level SL: Both mean level above the threshold in quiet, but SL generally refers to the overall level of a sonority, whereas YL refers to an individual pure-tone component.

5. Corresponding MIDI numbers exceed pitch category labels by 12; for example, middle C (C_4) is called 48 here but 60 in MIDI.

6. Plomp (1964) presented listeners with complex tones in the central pitch range comprising twelve harmonics of equal amplitude. He

found that listeners could, on average, only hear about six to eight harmonics in such tones. However, the error bars in Plomp's data suggest that some of his listeners heard as many as ten harmonics. The pitch model of Terhardt et al. 1982 predicts that the tones investigated by Plomp have roughly ten audible harmonics.

7. In elementary acoustics, the intensity of a tone of level L is given by $10^{L/10}$ and the amplitude by $10^{L/20}$ (relative to specified reference points).

8. Pure sonorousness was called "pure tonalness" by Parncutt (1989).

9. The difference between the present values of K_p and K_c and corresponding values in Parncutt 1989 are due to revision of the masking algorithm (specifically, Equation 3).

10. An implementation in the C programming language of the algorithm described by Hutchinson and Knopoff is available from the first author.

11. These observations apply to pure-tone components separated by more than about one quarter of a critical bandwidth. For smaller intervals, roughness decreases as interval size decreases (Plomp and Levelt 1965), while masking continues to increase. This situation rarely affects the present model, in which the frequencies of pure-tone components are confined to the chromatic scale, since the interval of one semitone is approximately equal to or greater than one-quarter of a critical bandwidth across most of the musically important spectral frequency range. The rule only breaks down at low spectral frequencies (below about middle C) for partials separated by small intervals. However, contributions to roughness and sonorousness from frequencies in this range are relatively small, due to the phenomenon of spectral dominance (Plomp 1967, Ritsma 1967, Terhardt 1974b).

12. Parncutt (1989) developed a different formula for pitch commonality (Equation 4.29). In later work (1993), calculations according to that formula were found to fit experimental results no more or less closely than correlation coefficients between tone salience profiles. In the present version of the model, therefore, pitch commonality is calculated as a correlation coefficient, primarily because the correlation coefficient is a well-known mathematical function. Krumhansl (1990) often uses correlation coefficients to quantify pitch relationships between tone profiles; the use of the correlation coefficient here facilitates comparison of her approach to the present model.

13. This concept of pitch distance corresponds to the *mel* scale of Stevens and Volkmann 1940. The *mel* scale is not used as a metric for computing pitch distance in the present model, because it applies only to pure tones, whereas most of the tone sensations evoked by complex sonorities are of the complex variety (virtual rather than spectral pitches). The most appropriate metric for pitch distance between complex tone sensations in music would appear to be simply the logarithm of frequency. (Intervals in the deep bass do tend to sound smaller than their size in semitones: however, the pitches that bound such intervals are seldom salient.)

14. For an overview of the history of harmonic progression, with an emphasis on the history of cadences, see Eberlein and Fricke 1992.

15. The argument upon which this estimate is based involves chord roots, and only applies to musics in which chord roots play an important role. Some non-Western styles are known to incorporate finer subdivisions of the octave. The above arguments do not apply to such styles.

REFERENCES

Aures, W. 1985a. "Ein Berechnungsverfahren der Rauhigkeit" (A Procedure for Calculating Roughness). *Acustica* 58:268–81.

————. 1985b. "Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen" (Sensory Consonance as a Function of Psychoacoustic Parameters). *Acustica* 58:282–90.

————. 1985c. "Ein Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale" (A Procedure for Calculating the Sensory Consonance of Any Sound). *Acustica* 59:130–41.

Bregman, A. S. 1990. *Auditory Stream Segregation.* Cambridge: MIT Press.

Barlow, Clarence. 1987. "Two Essays on Theory." *Computer Music Journal* 11, no. 1:44–60.

Bharucha, J. J. 1991. "Pitch, Harmony, and Neural Nets: A Psychological Perspective." In *Music and Connectionism*, ed. P. M. Todd and David G. Loy. Cambridge: MIT Press.

Blackwood, Easley. 1991. "Modes and Chord Progressions in Equal Tunings." *Perspectives of New Music* 29, no. 2 (Summer): 166–200.

Burns, E. M. 1981. "Circularity in Relative Pitch Judgments for Inharmonic Complex Tones: The Shepard Demonstration Revisited, Again." *Perception and Psychophysics* 30:467–72.

Burns, E. M., and W. D. Ward. 1978. "Categorical Perception—Phenomenon or Epiphenomenon: Evidence from Experiments in the Perception of Melodic Musical Intervals." *Journal of the Acoustical Society of America* 63:456–68.

————. 1982. "Intervals, Scales and Tuning." In *The Psychology of Music*, ed. Diana Deutsch. New York: Academic Press.

Danner, G. 1985. "The Use of Acoustic Measures of Dissonance to Characterize Pitch-Class Sets." *Music Perception* 3:103–22.

Darwin, C. J., and N. I. Hill. 1993. "Auditory Grouping in Lateralization, Pitch, and Vowel Perception: A Comparison of the Effects of Onset Asynchrony and of Mistuning." *Journal of the Acoustical Society of America* 93:2307 (abstract).

Dodge, Charles, and Thomas A. Jerse. 1985. *Computer Music: Synthesis, Composition and Performance.* New York: Schirmer Books.

Eberlein, R., and J. P. Fricke. 1992. "Kadenzwahrnehmung und Kaden-zgeschichte—ein Beitrag zu einer Grammatik der Musik" (with abridged version in English). Frankfurt am Main: Lang.

Forte, Alan. 1973. *The Structure of Atonal Music*. New Haven: Yale University Press.

Fransson, F., Johan Sundberg, and P. Tjernlund. 1974. "The Scale in Played Music." *Swedish Journal of Musicology* 56:49–54.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Miffin.

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.

Hesse, A. 1985. "Zur Ausgeprägtheit der Tonhöhe gedrosselter Sinustöne." *Fortschritte der Akustik*, 535–38. Proceedings of the Deutsche Arbeitsgemeinschaft für Akustik (DAGA).

Hiller, Lejaren, and Leonard Isaacson. 1958. *Experimental Music*. New York: McGraw-Hill.

Hutchinson, W., and L. Knopoff. 1978. "The Acoustic Component of Western Consonance." *Interface* 7:1–29.

Jones, K. 1981. "Compositional Applications of Stochastic Processes." *Computer Music Journal* 5, no. 2:45–61.

Kameoka, A., and M. Kuriyagawa. 1969. "Consonance Theory Part II: Consonance of Complex Tones and its Calculation Method." *Journal of the Acoustical Society of America* 45:1460–69.

Karplus, K., and A. Strong. 1983. "Digital Synthesis of Plucked String and Drum Timbres." *Computer Music Journal* 7, no. 2:43–55.

Laden, Bernice, and Douglas H. Keefe. 1989. "The Representation of Pitch in a Neural Net Model of Chord Classification." *Computer Music Journal* 13, no. 4:12–26.

Lerdahl, Fred. 1987. "Timbral Hierarchies." *Contemporary Music Review* 2, no. 1:135–60.

Lewin, David. 1977. "Forte's Interval Vector, My Interval Function, and Regener's Common-Note Function." *Journal of Music Theory* 21: 194–237.

———. 1979–80. "Some New Constructs Involving Abstract Pcsets, and Probabilistic Applications." *Perspectives of New Music* 18:433–44.

McAdams, S. 1984. "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organisation." In *Cognitive Processes in the Perception of Art*, ed. W. R. Crozie and A. J. Chapman. Amsterdam: Elsevier.

McAdams, S., and Kaija Saariaho. 1991. "Qualités et fonctions du timbre musical." In *Le Timbre: Métaphore pour la composition*, ed. Jean-Baptiste Barrière, 164–81. Paris: IRCAM.

Moore, B. C. J., and B. R. Glasberg. 1983. "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns." *Journal of the Acoustical Society of America* 74:750–53.

Moore, B. C. J., B. R. Glasberg, and R. W. Peters. 1985. "Relative Dominance of Individual Partials in Determining the Pitch of Complex Tones." *Journal of the Acoustical Society of America* 77:1853–60.

Morris, Robert D. 1979. "A Similarity Index for Pitch-Class Sets." *Perspectives of New Music* 18:445–60.

———. 1987. *Composition With Pitch-Classes: A Theory of Compositional Design*. New Haven: Yale University Press.

Papousek, M., and H. Papousek. 1989. "Forms and Functions of Vocal Matching in Interactions between Mothers and their Precanonical Infants." *First Language* 9:137–58.

Parncutt, Richard. 1988. "Revision of Terhardt's Psychoacoustical Model of the Root(s) of a Musical Chord." *Music Perception* 6:65–94.

———. 1989. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.

———. 1993. "Pitch Properties of Chords of Octave-Spaced Tones." *Contemporary Music Review*, in press.

Parncutt, Richard, and A. J. Cohen. "Identification of Melodic Patterns in Microtonal Scales: Effects of Serial Order, Scale Step Size, and Training." In preparation.

Plomp, R. 1964. "The Ear As Frequency Analyser." *Journal of the Acoustical Society of America* 36:1628–36.

———. 1967. "Pitch of Complex Tones." *Journal of the Acoustical Society of America* 41:1526–33.

Plomp, R., and W. J. M. Levelt. 1965. "Tonal Consonance and Critical Bandwidth." *Journal of the Acoustical Society of America* 38:548–60.

Rahn, Jay. 1992. "An Advance on *A Theory for All Music*: At-Least-As Predicates for Pitch, Time, and Loudness." *Perspectives of New Music* 30, no. 1:158–83.

Rahn, John. 1979. "Relating Sets." *Perspectives of New Music* 18:483–98.

Rasch, Rudi A. 1978. "The Perception of Simultaneous Notes Such as in Polyphonic Music." *Acustica* 40:21–33.

Regener, Eric. 1974. "On Allen Forte's Theory of Chords." *Perspectives of New Music* 13, no. 1 (Fall-Winter): 191–212.

Riemann, Hugo. 1893. *Vereinfachte Harmonielehre*. London: Augener.

Ritsma, R. J. 1967. "Frequencies Dominant in the Perception of the Pitch of Complex Tones." *Journal of the Acoustical Society of America* 42:191–98.

Roads, Curtis. 1978. *Composing Grammars*. San Francisco: Computer Music Association.

Roeder, John. 1989. "Harmonic Implications of Schoenberg's Observations of Atonal Voice Leading." *Journal of Music Theory* 33, no. 1:27–62.

Schuck, O. H., and R. W. Young. 1943. "Observations on the Vibrations of Piano Strings." *Journal of the Acoustical Society of America* 15:1–11.

Shepard, R. N. 1964. "Circularity in Judgments of Relative Pitch." *Journal of the Acoustical Society of America* 36:2346–53.

Siegel, J. A., and W. Siegel. 1977. "Categorical Perception of Tonal Intervals: Musicians Can't Tell Sharp from Flat." *Perception and Psychophysics* 21:399–407.

Smith, J. O. 1987. "Waveguide Filter Tutorial." *Proceedings of the International Computer Music Conference, Illinois*, 9–16. San Francisco: Computer Music Association.

Smith, K. C., and L. L. Cuddy. 1986. "The Pleasingness of Melodic Sequences: Contrasting Effects of Repetition and Rule-Familiarity." *Psychology of Music* 14:17–32.

Stevens, S. S., and J. Volkman. 1940. "The Relation of Pitch to Frequency." *American Journal of Psychology* 53:329–53.

Tenney, James. 1988. *A History of 'Consonance' and 'Dissonance'*. New York: Excelsior.

Terhardt, Ernst. 1972. "Zur Tonhöhenwahrnehmung von Klängen" (Perception of the Pitch of Complex Tones). *Acustica* 26:173–99.

————. 1974a. "Pitch, Consonance, and Harmony." *Journal of the Acoustical Society of America* 55:1061–69.

————. 1974b. "On the Perception of Periodic Sound Fluctuations (Roughness)." *Acustica* 30:201–13.

————. 1976. "Ein psychoakustisch begründetes Konzept der musikalischen Konsonanz." *Acustica* 36:121–37.

Terhardt, Ernst, G. Stoll, and M. Seewann. 1982. "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals." *Journal of the Acoustical Society of America* 71:671–78.

Todd, P. M. 1989. "A Connectionist Approach to Algorithmic Composition." *Computer Music Journal* 13, no. 4:27–43.

Ward, W. D. 1954. "Subjective Musical Pitch." *Journal of the Acoustical Society of America* 26:369–80.

Wessel, David. 1979. "Timbre Space as a Musical Control Structure." *Computer Music Journal* 3, no. 2:45–52.

Zwicker, E., and Ernst Terhardt. 1980. "Analytical Expressions for Critical-Band Rate and Critical Bandwidth as a Function of Frequency." *Journal of the Acoustical Society of America* 68:1523–25.

Zwicker, E., and A. Jaroszewski. 1982. "Inverse Frequency Dependence of Simultaneous Tone-on-Tone Masking Patterns at Low Levels." *Journal of the Acoustical Society of America* 71:1508–12.