Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

4. Model

The quantitative relationship between physical and experiential properties of tones, chords and progressions is simulated by means of a model. The input to the model comprises either the amplitude spectra or the musical notes of each tone or chord. Evaluation of masking effects leads to an estimate of the audibility of each pure tone component. The perception of complex tones is simulated by matching the pitches of audible pure tone components against those of a template, representing the audible components of a typical complex tone. Calculated audibilities of pure and complex tone components are used to estimate the "tonalness" and "multiplicity" of a simultaneity, and the salience of each tone sensation. Calculated salience values enable the evaluation of pitch commonality and pitch distance between sequential sounds. These are combined to model the results of pitch analysis and tone/chord similarity experiments.

4.1 General Aspects

4.1.1 Aim, Form and Implementation

The prediction of the pitch properties of a tone simultaneity on the basis of its waveform or frequency spectrum entails extensive understanding of psychoacoustical data and theory [Terhardt et al. 1982b]. Similarly, the determination of the musical function of a chord from the notes of a musical score requires considerable familiarity with music and music theory.

The model described in this chapter aims to facilitate the prediction and understanding of these relationships. It formalizes experimental data and psychoacoustical theory on the perception of tone simultaneities in such a way that: the data and theory are logically, concisely and conveniently expressed; the theory may be quantitatively tested (Chap. 5); and the theory may be applied in music theory, analysis and composition (Chap. 6).

Mathematics itself cannot *explain* music, it can only *describe* it [cf. De la Motte-Haber 1985]. In the model, mathematics is used to describe psychoacoustical data and theory concerning the perception of tone simultaneities. The mathematics is drawn together into a single *algorithm*: a mathematical procedure which produces a unique output for a given set of input parameters.

Existing models of music theory and analysis such as those by Alphonce [1980], Rahn [1980] and Smoliar [1980] are based primarily on the experience and intuition of music theorists such as Schenker [1906, 1935]. The present model is similar in that it is also largely based on experience and intuition, in this case of psychoacousticians such as Terhardt. It has the advantage that it draws upon a higher proportion of experimental data and scientifically testable theory than its music-theoretical relatives, and the disadvantage that it fails to account for cultural or cognitive aspects of music perception such as the perceptual organization of relatively long time spans (compare Schenker's *middleground* and *background*).

The model may be regarded simply as a *black box*. Of primary interest is the relationship between the input and the output [cf. West et al. 1987]. When suitably implemented in the form of a computer program, the model may be used without prior knowledge of how and why it works. Its use may be justified solely in terms of the success with which it models the results of experiments and the conventions of music theory.

A model such as this would be impractical without modern computer technology. In this study, computers were used to perform the long strings of calculations required by the model, to automate experiments by which the model was tested, to analyze and reduce experimental data, and to compare calculations of the model with experimental results so as to allow objective judgment of the model's predictive power. By means of computers, various versions of the model could be quickly compared to see how well they modelled the results of particular experiments. Inferior versions and versions which did not improve modelling performance despite increased complexity and computation time could then be discarded. Finally, computers enabled the model to be applied in music theory and analysis (Chap. 6).

4.1.2 Formulation and Assessment

A variety of mathematical operations and functions were drawn upon to create the model. Among the more commonly used functions were logarithms and exponentials, functions which have traditionally played important roles in psychophysical modelling [Fechner 1860; Stevens 1957]. Another common operation in the model is summation (e.g. summations of functions of pitch over the range of hearing). Summations are used to simulate relationships between relatively analytical and relatively holistic sensations (Sect. 2.3.3). Mathematical forms were regarded as appropriate if they logically and non-arbitrarily embodied appropriate theoretical concepts, and so allowed theoretical ideas (e.g. concerning the supposed sensory basis of Western harmony) to be tested as objectively as possible. A further criterion was efficiency: mathematical forms were preferred if they were relatively simple to understand and to calculate, and enabled the results of appropriate experiments to be modelled within a reasonable margin of error.

The model, as it stands, is neither perfect nor final. In its present state it is sufficiently logical, simple and accurate for a range of music-theoretical applications. It could, however, be changed in innumerable ways to satisfy future requirements, or as new theoretical possibilities emerge.

The model was tested against the results of psychoacoustical experiments (Chap. 5) based on subjective ratings of, and comparisons between, tones and chords. Responses were influenced by interindividual variations (such as listener's personalities and abilities, how well they were able to concentrate at the time of the experiment, etc.) and the context of each experimental trial. The model accounts for certain interindividual variations and contextual effects by means of four *free parameters*. Each parameter is supposed to reflect *how analytically sound is perceived* by the listener at a certain level. The first level concerns the analysis of a simultaneity into pure tone components: the more analytically one listens, the more clearly pure tone components can be resolved, and so the greater is the number of audible pure tone components, see eq. (4.14) below [E. J. Gibson 1953]. The second level concerns the perception of individual tones within a simultaneity: the more analytically one listens, the more likely one is to notice pure tone sensations by comparison with complex tone sensations (4.19) [Terhardt 1972, 1974a]. The third level concerns the simultaneous perception of tones: the more analytically one listens to a simultaneity (such as a musical chord), the more tones one notices at the one time (4.24). The fourth level concerns perceived relationships between sequential sounds: the more analytically such a relationship is heard, the more it is supposed to be influenced by pitch commonality as opposed to pitch proximity (4.34, 37).

4.1.3 Culture-Specific Aspects

Listeners' musical experience – in particular, their experience of sounds similar to those heard in the experiments – has a considerable effect on experimental results. This was not accounted for at all in the model. Instead, specific cultural conditioning effects were explored by looking at systematic discrepancies between calculations and averaged experimental responses. The model is nonetheless culture-specific in several respects. The aim of the model is primarily to explain aspects of the perception of Western chord progressions. Consequently, it is formulated in such a way that it can only be applied to the theory and analysis of Western music.

The following list gives some idea of the extent of the model's ethnocentricity.

i) Pitches and frequencies are limited to the chromatic scale throughout the model. This limits application of the model to musical cultures whose scales may adequately be represented as subsets of the chromatic scale (Sect. 3.3.2). Note that exact tuning of the chromatic scale is not specified in the model.

ii) Emphasis is placed on how analytically sounds are heard (the four free parameters). This may reflect a kind of Western analytical bias.
iii) The importance of pitch for musical structure is emphasized. The relative importance of pitch (as opposed to rhythm, timbre, dynamics, etc.) differs across musical cultures.
iv) The importance of sounds with harmonic spectra is emphasized, and perceptual conditioning by sounds with nonharmonic spectra (important in many world musics) is ignored.
v) The importance of consonance – and of tonalness, pitch commonality and pitch distance as sensory components of consonance – is emphasized.

The psychoacoustics of hearing has many aspects, some of which are well understood, some yet to be explored. There presumably exist many universal properties of human hearing which are not exploited by Western music, but which influence non-Western musics in important ways and may therefore be described as sensory bases of these musics. The identification and investigation of such properties would require intimate knowledge of the music(s) in question, and is beyond the present scope.

4.1.4 Comparison with Terhardt's Model

The present model is based on and inspired by the *Algorithm for the extraction of pitch and pitch salience from complex tonal signals* of Terhardt, Stoll and Seewann [1982b], hereafter referred to as "Terhardt's model". The present model differs from Terhardt's in several important respects.

Terhardt's model operates directly on an amplitude spectrum, created by the fast Fourier transform (FFT) of a segment of the waveform of a sound, and includes a procedure for the extraction of tonal components from the spectrum. In the cases of interest for the current study, spectral analysis was unnecessary. In experimental testing of the model (Chap. 5), the spectra from which the sounds were synthesized were input directly to the model. In music theoretical applications of the model, the spectra of individual complex tones are specified by the model, so that only note names need to be supplied.

Terhardt's model is primarily intended to predict the *exact pitch* of the most prominent tone sensation in a simultaneity. It therefore deals with *pitch shifts* in considerable detail. Music theory is primarily concerned with relationships between pitch *categories*, i.e. notes (Sect. 2.5.3). It turns out that musical pitch relationships may be modelled without reference to pitch shifts. Consequently, pitch shifts are neglected altogether in the present model.

Tone saliences (Terhardt: *pitch weights*) are expressed as absolute values in the new model. Tone salience is defined as the probability of noticing a tone (i.e. the probability of experiencing the corresponding tone sensation). In Terhardt's model, pitch weights are used primarily to determine which pitch (i.e. which tone sensation) in a simultaneity is most prominent; they are not regarded as absolute values, nor are they given a definite psychoacoustical meaning.

Finally, the new model proceeds further in the direction of music theory than Terhardt's, by estimating the strength of the pitch relationship (pitch commonality, pitch proximity) between sequential musical tones or chords.

Because the two models are so different, the degree to which they may usefully be compared is limited. Ultimately, the models should be judged separately on their individual merits: how logically they embody the theories on which they are based, how simple they are, and how closely they predict experimental results corresponding to their separate aims.

4.2 Input

4.2.1 Pitch Category

Following a notation adopted by the American Standards Association [1960], the *register* of a note is written as a subscript to the note's letter name. For example, middle C is called C_4. Each register runs from C up to B, so the B next to middle C is B_3. The lowest pitch register (corresponding roughly to the range 16–32 Hz) is "register 0"; the highest (roughly 8–16 kHz), "register 9". The modern piano keyboard includes 7 complete registers (1–7).

The *chroma* (or pitch class) of a note is defined in the model as its distance in semitones from the nearest C below. So the chroma of any C is 0, of any G is 7, and so on. No distinction is made between enharmonically equivalent note names when they are expressed as chromas. For example, E sharp, F natural and G double flat all lie 5 semitones above C, and so have chroma 5.

The *pitch category* of a musical note in the model is obtained by multiplying its register by twelve and adding its chroma. For example, the pitch category of the note A_4 (440 Hz) is $12\times4+9=57$. The range of the modern piano (A_0 to C_8) may be expressed as a pitch category range of 9–96.

4.2.2 Experiments

The sounds presented in the modelled experiments (Sects. 5.2, 5.3, 5.6, 5.7) were initially specified by the pitch categories (P) and sound pressure levels (SPL) of their tone components. When realizing the sounds, the frequency levels FL [Fletcher 1934; Young 1939] of tone components were set equal to their pitch categories P, i.e. tone components were tuned to equal temperament:

$$\mathrm{FL}(P) = P \ . \tag{4.1}$$

It was unnecessary to account for octave stretch (Sect. 3.3.3), as sounds in each experimental trial covered relatively small frequency ranges.

In the experiments on the similarity of synthetic tones (Sects. 5.5, 5.6), fundamentals were tuned according to (4.1), and overtones were exact multiples

of fundamental frequencies. In other experiments using synthetic sounds (Sects. 5.2, 5.3 and 5.7), (4.1) was applied to each pure tone component individually, so that spectra were entirely equally tempered.

When synthesizing sounds for the experiments, frequencies f [Hz] of pure tone components were obtained from their frequency levels FL by the equation:

$$f(\mathrm{FL}) = 2^{(\mathrm{FL}-57)/12}\times 440 \ . \tag{4.2}$$

Similarly, the pressure amplitude p of each pure tone component was calculated from its specified SPL by

$$p(\mathrm{SPL}) = k\times 10^{\mathrm{SPL}/20} \ , \tag{4.3}$$

where the factor k depends on sound amplification.

The frequency spectrum of each sound heard in the modelled experiments was input to the model as an array of real numbers in the form SPL(P), where pitch category P ranged from 19 to 96. Empty pitch categories were assigned large negative SPLs (i.e. zero amplitudes).

4.2.3 Auditory Level

The threshold of audibility in free field for pure tones is expressed in the model as a *threshold level* (TL), in dB (SPL), as a function of pitch category P:

$$\mathrm{TL}(P) = 91-49\ \log_{10}(P-7) \ . \tag{4.4}$$

Over the pitch range of the modern piano (P from 9 to 96), threshold level values calculated in this way differ by 1 dB (average) to 2 dB (maximum) from values calculated by Terhardt's [1979a] formula for threshold level. Terhardt's formula had been fitted to data averaged over a number of people with good hearing [Zwicker and Feldtkeller 1967] over the entire audible range. The two formulas are compared in Fig. 4.1. Beyond $P = 96$ (4.2 kHz), the two formulations diverge rapidly and (4.4) becomes invalid. Frequencies in this range were not used in the experiments of this study.

The *auditory level* (YL) of a pure tone component is defined here as its level in decibels above the threshold of audibility:

$$\mathrm{YL}(P) = \max\{\mathrm{SPL}(P)-\mathrm{TL}(P), 0\} \ , \tag{4.5}$$

where the operator "max" selects the maximum of the two values in curly brackets.

A pure tone component with positive auditory level YL may be inaudible due to masking by other simultaneous components. In this case the component is said to have positive auditory level YL but zero *audible* level AL, see (4.16). Both YL and AL are supposed to apply to an idealized average listener with good hearing.

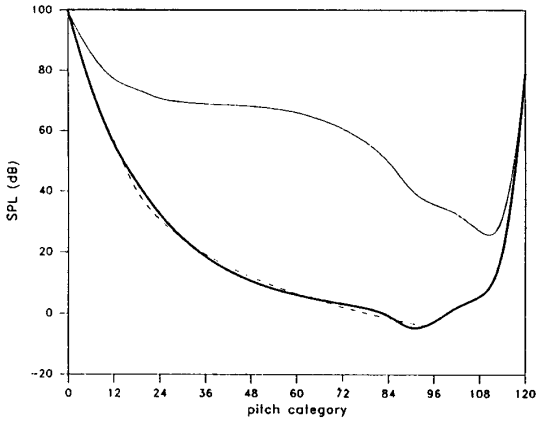

Fig. 4.1. Specified sound levels of musical tones. *Bold line*: threshold of audibility according to Terhardt [1979a]. *Broken line*: threshold of audibility according to (4.4). *Upper curve*: specified SPLs of fundamentals of musical tones according to (4.8), relative to the bold line

4.2.4 Applications

In most applications of the model (Chap. 6), the user inputs only the type and pitch category of each tone (component). Tone type may be "pure", "full complex" or "octave-spaced". In each case, the model specifies the auditory levels of pure tone components in a way which is consistent with the amplitude spectra of typical musical sounds.

The pure tone option is useful in the case of non-harmonic complex sounds. Only the pitch categories P of pure tone components are supplied by the user. Each component is then assigned an auditory level yl as follows:

$$\mathrm{yl}(P) = \frac{P(120-P)}{60} \; . \tag{4.6}$$

According to this function (which is just an arc of a circle), assigned auditory levels of pure tone components approach 60 dB over the central pitch range (roughly corresponding to the spectral dominance region [Fletcher and Galt 1950]) and fall to zero at the upper and lower thresholds of pitch.

Normally, musical sounds are made up of complex tones, specified by musical note names. So for music-theoretical or music-analytical work it is convenient for the model to accept musical note names as input. In the present model, each note is assumed to be realized as a full complex tone with all harmonics from the first to the sixteenth. Harmonics above the tenth are included even though they are rarely audible [Plomp 1964; Terhardt 1979a] as they can

still contribute to the masking of other components, and thereby to the overall sound.

If the fundamental (first harmonic) of a musical tone belongs to pitch category P_1, then the nth harmonic belongs to category P_n, as follows:

$$P_n = P_1 + \mathrm{int}\,\{12\,\log_2(n) + 0.5\} \; . \tag{4.7}$$

In this equation, the operator "int" (integer part) converts the frequency intervals between the components of a harmonic series into whole numbers of semitones, and the "0.5" ensures that values are correctly rounded. The values of $(P_n - P_1)$ generated by this equation, i.e. pitch distances above the fundamental in semitones for $n = 1$ to 16, are 0, 12, 19, 24, 28, 31, 34, 36, 38, 40, 42, 43, 45, 46, 47 and 48: the harmonic series in whole semitones (Fig. 1.1).

The auditory level yl of the nth harmonic is specified in the model by

$$\mathrm{yl}(n) = \frac{P_n(120-P_n)}{60}\left(1 - \frac{(P_n - P_1)}{120}\right) \; . \tag{4.8}$$

This expression satisfies the following requirements:

- On average, SPLs of harmonics of musical tones exceed threshold level by around 50 dB.
- Auditory levels fall to zero at the upper and lower thresholds of pitch ($P_n = 0$ and 120).
- Auditory levels are normally low in the lowest two pitch registers (below C_2, $P = 24$) due to the high threshold of audibility in these registers (Fig. 4.1).
- Pure tone components are seldom important for pitch perception in the highest two registers (above C_8, $P = 96$) due to spectral dominance (Sect. 6.1.2).
- The levels of coinciding pure tone components belonging to different complex tones in well-balanced musical chords are normally such that a component of higher harmonic number has a lower level than a coincident harmonic of lower harmonic number. For example, the 3rd harmonic of C_4 ($P_3 = 48+19 = 67$) is specified by (4.8) to be about 10 dB lower in level than the 1st harmonic of G_5 ($P_1 = 67$).
- Average SPL gradients for the harmonics of complex tones are shallower for lower tones (-8 to -10 dB per octave for fundamentals in the range C_1 to C_3) and steeper for higher tones (-12 to -17 dB per octave for C_4 to C_6) according to this equation [Parncutt 1987a].

In music-theoretical applications of the model, it is sometimes useful to analyze the pitch properties of chords composed of octave-spaced tones (Sects. 6.1.6, 6.2.1, 6.2.3–5). An octave-spaced tone at chroma c has pure tone components in every register $r = 0$ to 9 with pitch categories P_r given by

$$P_r = c + 12r \ . \tag{4.9}$$

In the model, these components are assigned auditory levels yl according to

$$\mathrm{yl}(r) = \frac{P_r(120 - P_r)}{60} \ . \tag{4.10}$$

In musical chords, pure tone components which fall into the same pitch category P are *incoherent*, i.e. their phase relationship is random. The auditory levels $\mathrm{YL}(P)$ of such components are combined according to elementary acoustics by adding their effective intensities (which are proportional to the squares of their sound pressures) by the formula

$$\mathrm{YL}(P) = 10 \log_{10} \sum 10^{\mathrm{yl}(P)/10} \ , \tag{4.11}$$

where $\sum$ denotes summation over all components falling in the same pitch category P, and the $\mathrm{yl}(P)$ are the individual auditory levels of each of the coinciding pure tone components.

4.3 Masking and Audibility

4.3.1 Critical Bandwidth

Fletcher [1940] suggested that the audible frequency range is divided into separate *critical bands*, as if the auditory system contained filters with variable centre frequencies. The loudness of a band of noise of constant power is constant for bandwidths less than a critical bandwidth, and increases for larger bandwidths [Zwicker et al. 1957]. Critical bandwidth, according to this definition, equals about 3 semitones above 500 Hz (C_5), and approaches a linear relationship with frequency at lower frequencies. Listener's responses in a variety of other experiments also change abruptly as bandwidth is increased beyond one critical bandwidth [Zwicker 1961; Scharf 1970].

There is some disagreement in the literature on the size of critical bandwidth at low frequencies. Zwicker et al. [1957] found that critical bandwidth below 500 Hz is approximately constant at 100 Hz [see also Zwicker and Terhardt 1980]. More recent measurements by other authors, using various methods, generally yielded lower estimates of critical bandwidth at low frequencies. Some published results are set out in Table 4.1.

In the present model, critical bandwidth W_{cb} is expressed in semitones as a function of the pitch category P at the centre of a band as

$$W_{\mathrm{cb}}(P) = \frac{5}{1 + x/\sqrt{x^2 + 44}} \ , \quad \text{where} \quad x = P/5 - 10 \ . \tag{4.12}$$

Table 4.1. Critical bandwidth [Hz] at low frequencies

Centre frequency [Hz]	125	200	250	500
Zwicker et al. [1957]	100	100	100	110
Patterson [1976]				80
Houtgast [1977]			50	80
Fidell et al. [1983]	40	50		
Calculated values (text)	60	70	80	110

This formulation was obtained by comparing calculations based on (4.13) below with experimental data (Table 4.1).

In the simulation of masking between pure tone components, it is useful to transform their frequencies or pitch categories onto a scale in which equal distances correspond to equal numbers of critical bandwidths. The accepted name for such a scale, *critical-band rate* [Zwicker 1961], is confusing because it implies that the scale is found by differentiation. In fact it is found by *integration* of the reciprocal of critical bandwidth. This conceptual difficulty may be avoided by anglicizing Zwicker's term *Tonheit* (literally: "tone-ness"; translated "tonalness" by Scharf [1970]) and renaming the critical-band rate scale *pure tone height*. The new name makes it clear that the scale is applicable only to the (spectral) pitch of pure tones and pure tone components – not to the (virtual) pitch of complex tones, such as in music. "Height" is used in

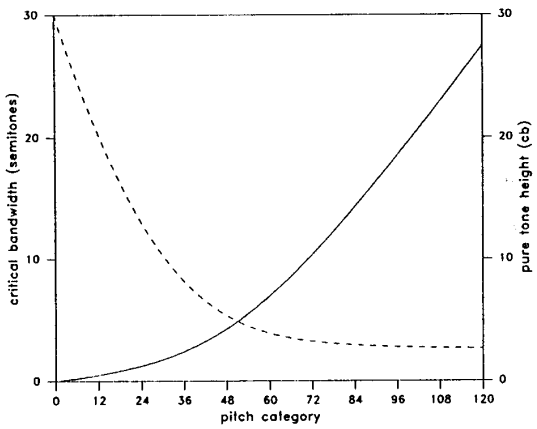

Fig. 4.2. Critical bandwidth and pure tone height. *Broken line, left scale*: critical bandwidth in semitones according to (4.12). *Full line, right scale*: pure tone height in critical bandwidths [cb] according to (4.13)

preference to "pitch", as "pitch" has musical connotations. The pure tone height scale is equivalent to the mel scale of pitch [Zwicker and Feldtkeller 1967].

Pure tone height H_p in critical bands [cb] may be derived from (4.12) by integrating its reciprocal and setting $H_p = 0$ at $P = 0$. The result is

$$H_p(P) = \sqrt{x^2+44}+x-2 \ , \quad \text{where} \quad x = P/5-10 \ . \tag{4.13}$$

Equations (4.12) and (4.13) are graphed in Fig. 4.2.

4.3.2 Masking

Masking may be described as the mutual drowning out (or inhibition, see [Moore 1982]) of one sound by another. Every pure tone component in a moderately loud sound partially masks every other component lying within a maximum distance of about 3 critical bands [Terhardt et al. 1982b]. The degree to which a component of auditory level YL in pitch category P' masks a component in a different category P is expressed in the model in terms of the effective reduction (in dB) of the audible level of P due P', as follows:

$$\mathrm{ml}(P,P') = \mathrm{YL}(P')-k_M\,|H_p(P')-H_p(P)| \ , \tag{4.14}$$

where ml stands for masking level and the expression within vertical bars (absolute value symbols) is the distance between P and P' in critical bandwidths. The "masking parameter" k_M is the first free parameter of the model. It has a typical value of about 25 dB/cb (Sect. 5.8.1). Note that ml in (4.14) may be negative; this is rectified in (4.15) below.

According to (4.14), the masking pattern (or audiogram) of a pure tone, when depicted as a graph of SPL against pure tone height, is triangular, with equal upper and lower gradients of k_M dB/cb (Sect. 2.2.2). In fact, the upper and lower gradients of masking patterns generally differ, and depend on level and frequency [Terhardt 1979a]. The masking parameter's optimal value of 25 mostly exceeds values used by Terhardt [1979a] to calculate masking, which are centred in the range 15–25 dB/cb. A possible explanation is that slightly subthreshold pure tone sensations can contribute to the formation of complex tone sensations. This would be consistent with the finding of Moore and Rosen [1979] that a residue tone can evoke a pitch at the missing fundamental even if its harmonics are (apparently) completely masked.

The interaction of several maskers has been investigated extensively [Zwicker and Herla 1975; Lutfi 1985; Moore 1985], but no rule has been found by which the effects of simultaneous maskers may be combined in a way which is accurate, yet appropriately simple for music-theoretical purposes. Terhardt's [1979a] solution was to combine contributions to the masking of a particular tone component by more than one other component by adding equivalent sound pressure *amplitudes*. This method is relatively simple, and makes plausi-

ble predictions of the number of audible harmonics in typical complex tones [Terhardt 1979a]. It is therefore retained in the present model:

$$\mathrm{ML}(P) = \max\left\{20\,\log_{10}\left(\sum_{P'\neq P} 10^{\mathrm{ml}(P,P')/20}\right), 0\right\} \ , \tag{4.15}$$

where the summation is carried out over all values of P' not equal to P, and the "max" function prevents ML from becoming negative (in the case of no maskers). This formation corresponds to that proposed by Lufti [1985] for the case of "appreciably overlapping maskers" (such as the pure tone components of full complex tones and musical chords).

The masking algorithm described in this section is only an approximation to that described by Terhardt [1979a], which itself will become obsolete as the Fourier-t-transform with appropriate frequency and amplitude dependencies is introduced to account for auditory masking [Terhardt 1985]. However, it is accurate enough for music-theoretical work, in which rough estimates of masking levels are sufficient (Sect. 6.1.1).

4.3.3 Audibility

The *audible level* (AL) of a pure tone component (Terhardt: *SPL excess*) is defined as its level above masked threshold. It is calculated in the present model by subtracting the masking level ML at the pitch category P of the component from the component's auditory level YL:

$$\mathrm{AL}(P) = \max\{\mathrm{YL}(P)-\mathrm{ML}(P), 0\} \ . \tag{4.16}$$

According to Hesse [1985], the perceptual prominence (*Ausgeprägtheit*) of the pitch of a partially masked pure tone relative to that of a clearly audible reference tone is equal to about 1/20 of its audible level in decibels at low audible levels, and approaches 1 (saturates) at high audible levels. Terhardt et al. [1982b] defined *spectral pitch weight* in a way which closely fits Hesse's data on the pitch prominence of pure tones. They allowed the audibility of each pure tone component to saturate with increasing audible level as follows:

$$A_p(P) = 1-\exp\{-\mathrm{AL}(P)/15\} \ , \tag{4.17}$$

where A stands for audibility, and the subscript p stands for pure tone. The saturation of audibility with increasing level has the effect that the tonalness (Sect. 4.4.3) of a clearly audible simultaneity is roughly independent of level.

4.4 Recognition of Harmonic Pitch Patterns

4.4.1 Harmonic Template

Complex tone perception may be regarded as a direct process: complex tone sensations are merely experiences accompanying the perception of tone sources [Gibson 1966] (Sect. 2.1.3). This state of affairs is not reflected in a psychoacoustical model, in which analytical sensations are regarded as basic. In the present model, the perception of complex tones is simulated as if it were mediated by the formation of pure tone sensations (Sect. 2.4.3).

According to the theory of virtual pitch [Terhardt 1972, 1974a] complex tone sensations result from the spontaneous recognition of (normally unnoticed) harmonic patterns among the (spectral) pitches of pure tone sensations. The present model stimulates the relationship between pure and complex tone sensations by searching for harmonic pitch patterns among pure tone sensations. A harmonic template, whose components form an idealized harmonic pattern, is used to estimate the importance of each pattern found (Sect. 2.4.3). The form of the template resembles the pitch pattern of the audible harmonics of a typical harmonic complex tone (Fig. 4.3).

The pure tone sensations contributing to the formation of a complex tone sensation at pitch category P_1 (the 1 denotes first harmonic) fall in pitch categories P_n, $n = 1, \ldots, 10$, as specified in (4.7). The degree to which pure tone sensations contribute to the formation of complex tone sensations becomes progressively less for higher harmonic numbers n [Ritsma 1967]. This is accounted for by assigning a weight (W) to each element of the template:

$$W_n = 1/n \ . \tag{4.18}$$

These weights resemble the audibilities of harmonics of typical complex tones, such as speech vowels [Terhardt 1979a]. Normally, no more than ten components are audible. The template is therefore limited to ten components, i.e. $n = 1, \ldots, 10$ in both (4.7) and (4.18). This limitation has the advantage that no decision has to be made about the pitch category of the eleventh component, which lies 41.5 semitones above the fundamental, on the borderline between two pitch categories. Note that the first component of the template has twice the weight of the second, reflecting the large difference between the pitch saliences of full complex tones and residue tones [Stoll 1983].

4.4.2 Complex Tone Sensations

The formation of complex tone sensations is simulated by shifting the harmonic template (Fig. 4.3) through the musical pitch range in steps of one semitone. At each step, pitch matches are sought between the components of the template and the pure tone sensations of the sound. Whenever one or more

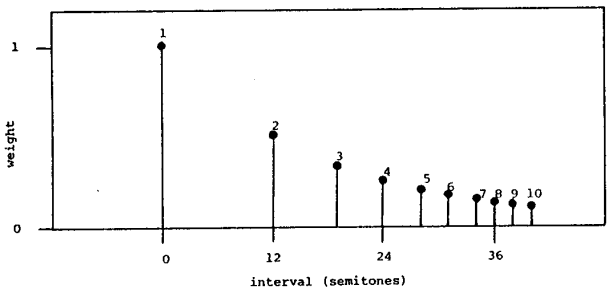

Fig. 4.3. Template for the simulation of harmonic pitch pattern recognition. Intervals are given by (4.7), weights by (4.18)

matches are found, a complex tone sensation is generated, whose pitch category (P_1) is that of the template's lowest component, and whose audibility (A_c) depends on the numbers and audibilities (A_p) of pure tone components matching template components as follows:

$$A_c(P_1) = \left(\sum_n \sqrt{W_n A_p(P_n)} \right)^2 / k_T \ , \tag{4.19}$$

where P_n is given by (4.7) and W_n by (4.18).

The "tone perception parameter" k_T is the second free parameter of the model. It is supposed to indicate how analytically tones are perceived, depending on listener and context. If k_T is low (about 1), the pure tone components of a complex sound are much less audible than complex tone components, and so are noticed only rarely. If k_T is high (about 10), pure and complex tone components compete with each other for the listener's attention on roughly equal terms. The value $k_T = 3$ was found to be typical in the experiments of this study (Chap. 5) and so was adopted for music-theoretical applications (Chap. 6).

Equation (4.19) may be regarded as a mathematical formulation of Houtgast's [1976] statement that "the potential of a multicomponent signal in evoking a particular low pitch [complex tone sensation] can be understood as a simple combination of the potentials of the individual components in evoking a particular subharmonic pitch" (p. 409). The "simple combination" chosen in the current model is not a simple sum but the square of a sum of square roots. In other words, complex tone audibility A_c is assumed to be proportional to the *square* of the number of template matches with audible pure tone components. This ensures that the calculated saliences of the subharmonic tone sensations of a pure tone are appropriately weak (Sects. 2.4.5, 6.1.5). Note that the value of A_c does not increase exponentially as harmonics

are added to a complex tone, because mutual masking of harmonics causes their audibilities A_p to decrease markedly.

In Terhardt's model, contributions to the salience of a virtual pitch are added (linearly) from *pairs* of spectral pitches. Consequently, pure tones are assigned single, unambiguous pitches. Equation (4.19) was chosen here in favour of Terhardt's formulation, as it is easier to calculate (requiring one loop instead of two nested loops), and it simulates the concept of template matching in a more straightforward manner. Also, spectral and virtual pitch weight are defined in Terhardt's model to depend on the absolute frequency of tone components (spectral dominance). Various versions of these dependencies were tried out in in (4.17, 19), but none improved the correlation between calculations and the results of the experiments reported in Chap. 5. Some consequences of neglecting spectral dominance in the model are investigated in Sect. 6.1.2.

Isolated pure tones of moderate loudness in the middle pitch range have audible levels of about 60 dB. According to (4.17), such tones have audibilities approaching one. Isolated complex tones of moderate loudness in the middle pitch range, according to the model [with $k_T = 3$ in (4.19)], have audibilities five to six times greater than this. (The same applies for spectral and virtual pitch weights in Terhardt's model.) This difference is understandable in terms of the number of sources of information about the pitch of a tone – pure tones have only one such source, while complex tones have several (the harmonics) – and in terms of the relative importances of pure and complex tones in the auditory environment.

In analytical listening, it is possible to switch attention from a complex tone as a whole to the pure tone component at its fundamental [Schouten 1940]. But is is not possible to attend to both at once, as their pitches are the same (or very close). Following Terhardt's model, this implies that the overall audibility $A(P)$ of a tone component (i.e. audibility regardless of whether that component is pure of complex) is given by the maximum of the pure and complex tone audibilities in that pitch category:

$$A(P) = \max\{A_p(P), A_c(P)\} \ . \tag{4.20}$$

4.4.3 Tonalness

The *pure tonalness* of a simultaneity is defined here to depend on the number and audibilities of pure tone components (Sect. 3.2.2). Following Aures [1984], pure tonalness may be modelled by quadratic addition of the audibilities of pure tone components:

$$T_p = \sqrt{\frac{\sum_P A_p(P)^2}{5.2}} \ . \tag{4.21}$$

The factor 5.2 scales pure tonalness values so that the calculated pure tonalness of a complex tone at middle C [specified by (4.7, 8)] is 1. Examples of calculated pure tonalness values are given in Sect. 6.1.4.

The *complex tonalness* of a simultaneity is assumed here to be proportional to the maximum virtual pitch weight in Terhardt's model. In the terminology of the present study, it is proportional to the audibility of a simultaneity's most audible complex tone component (Sect. 3.2.2):

$$T_c = \max_P\{A_c(P)\}/6.2 \ . \tag{4.22}$$

As before, the factor 6.2 scales complex tonalness values so that the calculated complex tonalness of a complex tone at middle C [specified by (4.7, 8)] is 1.

4.5 Salience

4.5.1 Multiplicity

In this study, the number of tones simultaneously noticed in a sound (e.g. in a musical chord) is called its *multiplicity* (Sect. 2.3.4). Multiplicity is assumed in the model to depend partly on a sound's pitch configuration – its configuration of tone audibility as a function of pitch category – and partly on how analytically the sound is perceived.

An initial, unscaled estimate M' of the number of tones noticed simultaneously in a sound may be made by assuming that the sound's most audible (pure or complex) tone component is noticed with a probability of 100%, while other, less audible tone components are noticed with probabilities proportional to their calculated audibilities:

$$M' = \frac{\sum_P A(P)}{A_{max}} \ , \tag{4.23}$$

where A_{max} is the maximum audibility in the sound, i.e. the audibility of the most audible (pure or complex) tone component.

When perceived holistically, a tonal sound evokes a single tone sensation. For example, a holistically perceived chord built from octave-spaced tones evokes a single tone sensation which usually corresponds to its root (Sect. 6.1.6). In this case, the actual number of simultaneously noticed tones M equals one, regardless of the value of M' in (4.23). A *power law* relationship [Stevens 1957] between M and M' allows for this: any value of M' may be scaled to a value of $M = 1$ by raising it to the power zero. In general, M' may be appropriately scaled by raising it to some power k_S:

$$M = (M')^{k_S} \ . \tag{4.24}$$

The "simultaneity perception parameter" k_S is the third free parameter of the model. It may take any positive or zero value; the higher k_S, the higher M, and the more analytically simultaneities are perceived. Calculations according to (4.23, 24) fitted the results of the multiplicity experiment (Sect. 5.2.3) most closely when the parameter was set to a value of about 0.5. So M is set equal to the square root of M' in music-theoretical applications of the model (Chap. 6).

Equations (4.23, 24) disregard the important effects of relative onset times of the notes of a chord [Rasch 1978] and coordination of component amplitude envelopes [McAdams 1984] on multiplicity. The auditory system is remarkably sensitive to these subtle effects, using them to establish the actual number of sound sources contributing to a particular sound (e.g. a musical chord played by several instruments). These cues were absent from the sounds presented in the multiplicity experiment (Sect. 5.2.2), allowing the formulation presented here to be tested relatively directly. In performed music, where onsets and amplitude envelopes are asynchronous and amplitude envelopes independent, the number of simultaneously perceived tones corresponds more closely to the actual number of simultaneous notes, provided this number remains small (say, no more than four).

When sounds (such as musical tones or chords) are perceived holistically (i.e. when they are assigned only one pitch at a time), the variable M may be interpreted as a measure of *pitch ambiguity*: an estimate of the number of different pitches a sound *could* have, where each pitch is weighted according to the salience of the corresponding tone sensation. The greater the pitch ambiguity of a sound, the greater the number of different pitches which could be assigned to the sound when it is perceived on different occasions and in different contexts.

4.5.2 Tone Salience

The salience (S) of an individual tone component is defined as its probability of being noticed. Assuming independent probabilities, and given that the tone simultaneity itself has been noticed, it follows that the sum of all the tone saliences $S(P)$ in a sound equals the number of simultaneously noticed tones M. In addition, tone salience is assumed to be proportional to tone audibility A. The following expression satisfies these criteria:

$$S(P) = \frac{A(P)}{A_{\max}} \frac{M}{M'} \,. \tag{4.25}$$

Assuming that only one tone is perceived at a given time in each pitch category of the chromatic scale, the variable S may also be interpreted as the "salience of a pitch category".

In music-theoretical applications, tone salience may be expressed in terms only of the audibilities A of other simultaneous tone sensations, by substituting $k_S = 0.5$ into (4.24) and combining (4.23–25):

$$S(P) = \frac{A(P)}{\sqrt{A_{\max} \sum_P A(P)}} \,. \tag{4.26}$$

4.5.3 Chroma Salience

In music-theoretical applications such as the determination of the root of a chord (Sect. 6.1.6) and the tonality of a progression (Sect. 6.2.3–5), it is useful to evaluate the saliences of the twelve chroma (pitch classes) in the chromatic scale. Two possible measures of chroma salience are considered here: *chroma tally* and *chroma probability*.

Chroma tally is defined as the average number of times a chroma is noticed in a musical element or passage. Assuming that tone saliences are independent probabilities, chroma tally $S_t(c)$ is given by

$$S_t(c) = \sum_r \sum_s S(c+12r) \,, \tag{4.27}$$

where the summations are carried out over all pitch registers r and simultaneities s in a homorhythmic (tone or chord) progression.

Chroma probability $S_p(c)$ is defined as the probability that a chroma is noticed (at least once) in a musical element or passage. Again assuming independent tone saliences,

$$S_p(c) = 1 - \prod_r \prod_s [1 - S(c+12r)] \,, \tag{4.28}$$

where $\prod$ denotes a product over all pitch registers r and simultaneities s in a homorhythmic progression.

4.6 Sequential Pitch Relationship

4.6.1 Pitch Commonality

The *pitch commonality* of a pair of sounds is assumed to be proportional to the number of pitches the sounds have in common (Sect. 3.2.3), i.e. the number of pitch categories containing noticed tones in both sounds. In the special case that the sounds are identical, pitch commonality is defined to take a maximum value of 1. The following formulation of pitch commonality satisfies these two criteria:

$$C = \frac{\sum_P \sqrt{S_1(P) S_2(P)}}{\sqrt{\sum_P S_1(P) \sum_P S_2(P)}} \,. \tag{4.29}$$

Here, S_1 is the array of tone saliences in the first sound, S_2 in the second; and pitch category P is varied over the range of hearing. In the case of two identical sounds, both the numerator and the denominator of the equation are equal to the sum of the tone saliences, i.e. the multiplicity, of each sound.

4.6.2 Pitch Distance

The average apparent *pitch distance* between a pair of sounds is formulated according to the following criteria. (i) The pitch distance between two identical sounds is zero. (ii) The pitch distance between two pure tones (with no subsidiary pitches) is equal to the difference between their frequency levels in semitones. (iii) The pitch distance between different sounds is always greater than zero. A formulation satisfying these criteria is

$$D = \sum_P \sum_{P'} S_1(P) S_2(P') |P'-P| - \sqrt{\sum_P \sum_{P'} S_1(P) S_1(P') |P'-P| \sum_P \sum_{P'} S_2(P) S_2(P') |P'-P|} , \quad (4.30)$$

where pitch categories P and P' are varied over the range of hearing.

In a first approximation, the saliences of (simultaneous or sequential) tones falling in different pitch categories may be assumed to be independent. So the product of $S(P)$ and $S(P')$ may be interpreted as the probability that tone sensations at P and P' are both noticed in a particular presentation, i.e. the probability that the interval of size $|P'-P|$ between them is noticed. This applies whether the two tone sensations in question are simultaneous or sequential.

The first term on the right-hand side of (4.30) is a weighted sum of pitch distances between *sequential* tone sensations in a pair of sounds; the second is the geometric mean of weighted sums of pitch distances between *simultaneous* tone sensations in each sound. It is easy to show that the equation satisfies criteria (i) and (ii) above. Verification that it satifies (iii) requires a long mathematical proof [Morris 1986]. Incidentally, criterion (iii) is not satisfied if the second right-hand term is formulated as an arithmetic mean rather than a geometric mean.

Equation (4.30) turns out to be unsuitable for use in music theory. The reason involves the distinction between actual tones (notes) and implied tones (such as Rameau's *basse fondamentale*, Sect. 1.4.1). In music theory, implied tone sensations are invoked to explain harmonic progression (in particular, through the concept of root progression), but they play no role in the theory of voice leading. Voice leading depends on *actual* notes, delineated for the listener by temporal cues such as timbre and onset asynchrony. Good voice leading involves making each melodic line within a chord progression coherent, so that the progression as a whole coheres regardless of which voice dominates (either accidentally or deliberately) in performance.

A more suitable equation than (4.30) for music-theoretical use is obtained by setting $S(P) = 1$ for the actual notes in a pair of chords and $S(P) = 0$ for all the other pitches:

$$D' = \sum_i \sum_j |P_j - P_i| - \sqrt{\sum_i \sum_{i'} |P_{i'} - P_i| \sum_j \sum_{j'} |P_{j'} - P_j|} , \quad (4.31)$$

where i labels the notes in the first chord ($i = 1, \ldots, I$, for a chord with I notes), j labels the notes in the second chord (similarly), and P_i and P_j are the pitch categories of notes in the first and second chords, respectively. This approximation is only valid for small numbers of simultaneous notes: it is difficult to hear more than three tones at once (Sect. 5.2.3), so tones in chords with more than three voices have saliences considerably less than one.

4.6.3 Pitch Analysis Experiment

The results of the pitch analysis experiment were modelled (Sect. 5.3.4) by a combination of pitch commonality and pitch distance, in the special case where one of the two sounds is a pure tone (the probe). Experimental evidence (Sect. 2.4.5) suggests that pure tones have weak subharmonic pitches. Inclusion of these in the model [according to (4.19)] greatly increased calculation time but made very little difference to the fit between results and calculations, so each pure probe tone was assumed to evoke a single tone sensation (with salience $S = 1$) in the pitch category corresponding to its frequency level. Consequently, the pitch commonality (C) of target sound and a probe tone was assumed to be simply

$$C = \sqrt{S(P)} , \quad (4.32)$$

i.e. the square root of the tone salience in the target sound at the pitch category P as the probe tone. The square root in (4.32) comes from the more general formulation of C in (4.29).

The average apparent pitch distance (D) between target and probe in the experiment was assumed to be

$$D = \sum_{P'} S(P') |P'-P| , \quad (4.33)$$

where P is the pitch category of the probe tone, P' is the pitch category of any tone sensation in the target sound, $S(P')$ is its calculated salience, and $|P'-P|$ – the absolute value of the difference between P' and P – is the interval between the two pitches in semitones. In the experiment P' varied from 9 to 96 (the range of the modern piano).

The theoretical or calculated outcome of each trial in the experiment was given in relative (unscaled) form by

$$R_t = k_R \frac{C}{\sigma_C} - (1-k_R)\frac{D}{\sigma_D} , \tag{4.34}$$

where k_R – the "relationship perception parameter" – is the fourth and final free parameter in the model, and σ_C and σ_D denote standard deviations over calculated values of C and D for all trials. The parameter k_R reflected the extent to which pitch commonality C, as opposed to pitch distance D, determined the theoretical response R_t. Its ideal value, for those listeners who were able to answer only the question asked of them in the pitch analysis experiment and ignore pitch distance altogether, was $k_R = 1$. Few participants in the experiment were able to listen this analytically, so the optimal value of the parameter was normally considerably less than one.

The theoretical (i.e. calculated) results R_t were scaled against the mean experimental responses (R_e) to each trial as follows:

$$R'_t = \frac{\sigma_e}{\sigma_t}(R_t - \bar{R}_t) + \bar{R}_e , \tag{4.35}$$

where $\bar{R}$ denotes the mean result, and σ denotes standard deviation of results about this mean, calculated over all 158 trials of the experiment. This equation performs a linear transformation on the theoretical responses R_t, setting their mean and standard deviation equal to the mean and standard deviation (respectively) of the mean experimental responses R_e.

4.6.4 Similarity Experiments

In the experiments described in Sects. 5.6 and 5.7, listeners judged the similarity of pairs of tones and chords. The results were modelled by means of the general formulations of pitch commonality and pitch distance in (4.29 and 4.30) above.

In addition, the pitch proximity of two sounds was defined to have a maximum value of 1 for identical sounds, and to approach 0 for sounds which are far apart in pitch relative to pitch distances typical of the prevailing context:

$$X = \exp(-D/\bar{D}) . \tag{4.36}$$

Here, $\bar{D}$ is the mean pitch distance between pairs of sounds, calculated over all trials of the experiment. Note that the effect of context was randomized in these experiments by presenting trials to each listener in a different (random) order.

The theoretical results R_t of the above two experiments were calculated by linear combination of pitch commonality and pitch proximity:

$$R_t = k_R \frac{C}{\sigma_C} + (1-k_R)\frac{X}{\sigma_X} \tag{4.37}$$

where k_R (the relationship perception parameter) is the fourth free parameter in the model [see also (4.34)], and σ denotes standard deviation. Final calculated responses R'_t were scaled against actual mean responses R_e in the same way as for the pitch analysis experiment (4.35).