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

1. Background

The conventions of mainstream Western harmony may be divided into conventions for writing individual chords (stacked thirds on a root; inversion, doubling, spacing), pairs of chords (harmonic relationship, voice leading, avoidance of parallels), and chord progressions (scales, keys, tonality). A range of scientific and pseudoscientific theories have been advanced to explain harmonic conventions, based on frequency ratios, the harmonic series, beats, combination tones, periodicity, cognitive structures, generative grammars, and mathematical groups. Terhardt [1974a, 1976] explained the perception of harmony in terms of perceptual familiarity with the pitch pattern of typical complex tones. Unlike other theories, Terhardt's is both empirically based and capable of explaining harmony's origins.

1.1 Music Theory

1.1.1 Introduction

Music theory comprises conventions ("rules") and guidelines for arranging, composing and analyzing music. The music theory of interest in this study is that of diatonic Western music: music in major and minor keys from the Baroque, Classical and Romantic periods. Of particular interest is the theory of chord progressions in three or more voices, i.e., with three or more notes per chord.

All rules can be broken. Perhaps the most successful composers were those who broke the so-called rules of music theory most effectively. In order to maintain stylistic unity, however, it is necessary most of the time to adhere to appropriate compositional conventions.

The conventions of music theory are generally formulated by theorists long after composers have begun to use them consistently. For example, the "rules" of harmony and form in eighteenth century music were largely developed in the nineteenth century. Successful composers in the eighteenth century tended to compose by intuition and direct imitation. In the nineteenth century, music-theoretical formulas were developed by which budding composers could imitate the harmonic styles of their predecessors in a more general and organized fashion. The time lag between practice and theory, and the tendency for the

theory to be created by theorists (rather than "great" composers), make music theory a somewhat idealized and simplified form of musical reality.

The conventions of diatonic music are explained in a variety of texts [e.g. Schoenberg 1911; MacPherson 1920; Hindemith 1940; Forte 1962; Goldman 1965; Aldwell and Schachter 1978; Piston 1978]. The most important of these conventions are introduced and summarized, along with brief accounts of their historical origins, in the following sections.

1.1.2 Single Chords

A chord is a simultaneity of three or more notes. In Western music theory, chords are constructed by superposing thirds on a note called the root (Fig. 1.1a). Superposing two thirds produces a triad of root, third and fifth. The major and minor triads are regarded as the most consonant triads, because their fifths are "perfect" (7 semitones). Other possible constructions are the diminished and augmented triads, named after the quality of their fifths, and the suspended triad, named after its suspended fourth (which normally resolves to a third). A seventh chord is formed by adding a third above the fifth

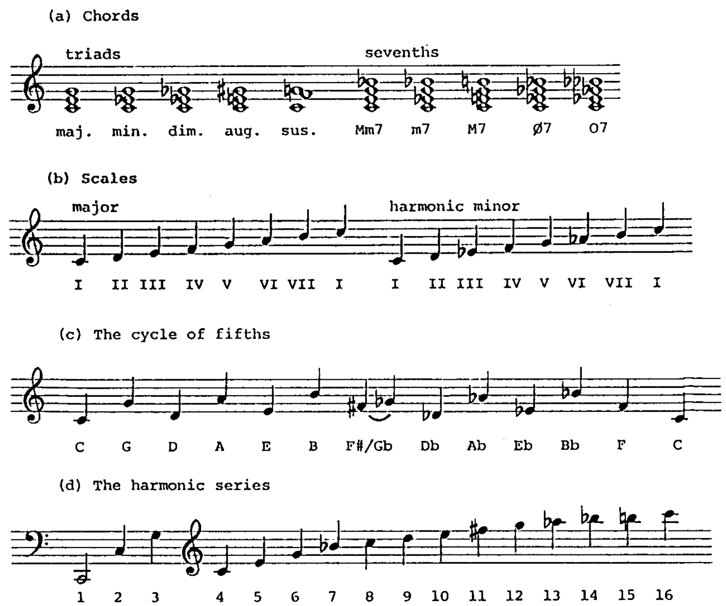

Fig. 1.1. Basic structures in music theory

of a triad. The most consonant seventh chords in Western music theory are those with perfect fifths above the root: the major-minor (or dominant) seventh chord (with a major third and a minor seventh above the root), the minor seventh chord (minor third, minor seventh) and the major seventh chord (major third, major seventh). Seventh chords with diminished fifths include the half-diminished seventh (minor third, diminished fifth, minor seventh) and the diminished seventh (minor third, diminished fifth, diminished seventh).

Music theory assumes *octave equivalence* (or octave generalization): notes one or more octaves apart are assumed to have the same harmonic function. Such notes are said (in music theory) to belong to the same *pitch class* [Babbitt 1955], or (in music psychology) to have the same *chroma.* Similarly, a *chord class* is specified by the pitch classes of its notes. Different *voicings* of a chord class have different bass notes (inversion), spacing and doubling.

Chords are *inverted* by transposing one of the upper notes down an octave into the bass (i.e. lowest) part. Major and minor triads most commonly appear in root position (i.e. not inverted); and first inversions (third in the bass) occur more commonly than second inversions (fifth in the bass). Diminished triads are most often heard in first inversion. Augmented triads (and also diminished seventh chords) are enharmonically equivalent in all inversions, and – partly because of this ambiguity – are relatively rare in diatonic music. Seventh chords appear in all inversions, but some inversions, like the root position of the major-minor seventh, minor seventh and major seventh chords, the first inversion of the minor seventh (producing a major added sixth), and the first inversion of the half-diminished seventh (producing a minor added sixth), are more usual.

The commonly taught conventions of *spacing* of chords are as follows. The intervals between neighbouring upper notes (in the treble staff) should not exceed one octave. Intervals between lower notes (in the bass staff) may exceed an octave, and should not be too small: thirds in the lower bass region are quite dissonant.

Four-part music largely comprises triads in which one note has been *doubled* at the octave, producing two notes with the same chroma in the one chord. It is generally recommended that the root of a major or minor triad be doubled, and that the third of a major triad, or the lowest note (the conventional root) of a diminished triad, should not. In any case, the leading note (scale degree VII, see Fig. 1.1 b) is not normally doubled in four-part writing. The most commonly doubled notes fall on tonally strong scale degrees such as I, V, IV and II (in order of decreasing importance). For example, the best note to double in an "augmented sixth" chord on ♭VI[1] is not the conventional root of the chord (♭VI) but the major third (I), because I is so much more stable than ♭VI (and also because parallels are avoided this way, see below).

Rules for writing single chords were in evidence, at least intuitively, in the Middle Ages. The number of pitch classes normally used in chords in homo-

[1] Flat sixth scale degree (flat submediant)

phonic music increased from one (octave-unison) through two (open fifths or fourths, as in some *organum* styles) to three (major and minor triads) during these centuries [Randel 1971]. Conventions of doubling and spacing also developed during this time. Chords of four pitch classes (tetrads, or sevenths) appeared in passing in contrapuntal Medieval and Renaissance styles, but were not recognized as harmonies (albeit as dissonant ones) until the early Baroque in the works of Monteverdi [Arnold and Fortune 1968].

1.1.3 Pairs of Chords

The contrapuntal techniques of the Middle Ages and Renaissance provided the basis from which conventions for writing sequential pairs of chords in the Renaissance and Baroque periods developed. For example, the emergence of what we now know as triads and perfect (V-I) cadences in the fifteenth century may be traced to medieval conventions for approaching perfect consonances, e.g. contrary motion from the major sixth to the octave; a shift from three- to four-part harmony; and increasingly strict prohibition of parallel fifths and octaves [Randel 1971]. These developments eventually led to the establishment of the following music-theoretical concepts.

Sequential chords sound related, or "progress" well, if they stand in a strong *harmonic relationship* with each other. In music theory, the nature of harmonic relationship is explained indirectly, in terms of the diatonic scales and the cycle of fifths (Fig. 1.1 b, c). The *diatonic scales* include major, harmonic minor and melodic minor forms. The *cycle of fifths* is a theoretical construction of chroma in which neighbours are a fifth (or fourth) apart. Chords are generally supposed to be harmonically related if all their notes belong to the same diatonic scale or if their roots are close together on the cycle of fifths.

Sequential chords generally progress well if they have notes in common or if their voices move by step (major or minor seconds) rather than by leap (thirds, fourths, etc.). This aspect of the relationship between sequential chords is called *voice leading,* or *melodic relationship.* Leaps which do occur preferably cover harmonically simple intervals such as the octave, perfect fourth or perfect fifth. The disruptive effect of a large leap is reduced if the note following the leap falls between the pitches of the last two notes – preferably a step away from the second note.

When two voices proceed from one chord to the next in parallel motion at a constant pitch distance, *parallels* (or *consecutives*) result. Parallel fifths and octaves were common in the music of the Middle Ages (e.g. *organum*) but were generally avoided in the three- and four-part writing of the Baroque, Classical and Romantic periods.

It is useful to think of *tonal relationships* as combinations of harmonic and melodic relationships (or harmonic and scalar relationships [Sloboda 1985, p. 22]). For example, a chromatic chord (e.g. the diminished seventh B–D–F–A♭) may not stand in a particularly strong *harmonic* relationship with its chord of resolution (e.g. C–E–G), but if the *melodic* relationship between the chords is strong (e.g. due to the semitone relationships B – C and A♭ – G) then their overall *tonal* relationship is strong.

1.1.4 Chord Progressions

The notes of chord progressions written in the Middle Ages and Renaissance largely conformed to *standard heptatonic scales* (corresponding to the white notes of the piano, or transpositions thereof). Depending on the *mode*, different notes in the scale were used more than others. The note which was used most often in a particular mode was called the "dominant", and the note on which melodies ended was the "final". The dominant was often a perfect fifth above the final [Scholes 1970].

As triadic harmony developed during the Renaissance, the modal system was gradually supplanted by the diatonic system in which only two scales (major and minor) were theoretically recognized. The fifth relationship usually found between dominant and final (now called the *tonic*) became standard. The triad formed by adding a major or minor third between the tonic and dominant was called the *tonic triad*. The tonic note was regarded as the *root* of this triad.

The tonic triad acted as a kind of home base – a point of departure, and of return – in diatonic chord progressions. A common feature of chord progressions written from the seventeenth century on was the maintenance of a strong tonal centre by the use of chords harmonically related to the tonic, i.e. chords whose roots belong to the prevailing diatonic scale and are close to the tonic on the cycle of fifths. The resultant *functional harmony* [Riemann 1893] had a kind of stylistic unity and balance which had not been present in earlier musics.

The tonal framework set up by such a chord progression was called a *key*. Keys take the same name as their tonic triad (e.g. E. minor, G major). The key of a progression can change by a process called *modulation*. The chord or chord at the point of modulation (pivot chords) normally belong or are closely related to both preceding and following keys. The original key and tonic normally return at the end of a diatonic piece.

1.1.5 A Scientific Basis?

Until the start of this century, Western musicians generally believed that the diatonic system was in some way "natural". Helmholtz [1863] gave this belief a scientific basis through his investigations of the physics, physiology and psychophysics of tone perception, including the effect of the beating of near-coincident pure tone components on the dissonance of tone simultaneities, and the effect of frequencies in common on pitch relationships between sequential sounds. Helmholtz did not, however, believe that the diatonic system was an inevitable consequence of the psychophysics of tone perception. On the contrary, he emphasized the role of culture and aesthetics in the development of musical style.

Diatonic harmony dominated Western art music until the early twentieth century. Gradually it became clear that most of the possibilities of the system had been exhausted. Composers, seeing the need to experiment with new styles, began to regard physically based arguments in favour of the diatonic system as inadequate or spurious. Consonance and dissonance, as investigated by Helmholtz, appeared to be specific to Western music. Inflexible approaches to consonance and dissonance had inhibited the progress of music on previous occasions – for example, unprepared seventh chords were not accepted until Monteverdi because of theoretical considerations – and composers such as Schoenberg did not want such theoretical ideas to prevent progress into new harmonic styles [Griffiths 1978].

Nondiatonic styles met with mixed success. The fact that many pieces have now been accepted into the concert repertoire supports the idea that any new style can be appreciated given sufficient exposure to that style, and provides evidence against the theory that diatonicism is in some way more "natural" than other ways of organizing pitch in music. But most atonal (or twelve-tone) styles have still not been absorbed into the mainstream of Western music – perhaps because of the high rate of information processing required of the listener [Klingenberg 1974]. Meanwhile, "old-fashioned" tonal styles continue to flourish. If the success of innovations is gauged by their long-term popularity, then the greatest progress in tonal structure this century has occurred in the genres of impressionism and jazz.

Composers of twelve-tone music generally agree that most of the procedures they use to organize their music are imperceptible, even to trained listeners [Lerdahl and Jackendoff 1983, p. 299], and musicologists and psychologists agree that twelve-tone music has no sensory basis [Bruner 1984]. At the time of this study, however, there is still widespread disagreement concerning what, if any, sensory ("natural"?) basis exists for diatonic harmony. The rest of this chapter is concerned with attempts to find such a basis.

1.2 Physically Based Theories

1.2.1 Introduction

Perhaps the most obvious place to look for a scientific explanation of music theory is in the "real" worlds of physics and biology: the acoustics of sound transmission, the physiology of sound detection by the ear and the neurophysiology of sound "interpretation" in the brain. A physically oriented approach lacks many of the conceptual problems of psychological approaches. More importantly, physical measurements are made in well-defined ways, and tend to be easier to control and to reproduce than more subjective, psychological measurements.

The following aspects of the physics of sound (as explained more fully in [Wood 1961; Backus 1969; Benade 1976; Roederer 1979; Hall 1980; Moore 1982; Rossing 1982; Pierce 1983]) seem relevant for the theory of musical intervals and harmony. (i) The *waveform* of a sound (a finite-duration function of air pressure against time) can be analyzed into sinusoidal (pure tone) *components* by spectral analysis; (ii) pairs of pure tones in simple, whole-number *frequency ratios* can be combined (by addition of their waveforms) to form *periodic* waveforms; (iii) pairs of periodic waveforms combine to form amplitude-modulated waveforms (*beats*) if their frequencies are close, but not identical; and (iv) pairs of tone components in distorting electroacoustical systems produce *combination tones*. At first sight, the above phenomena appear to explain aspects of music perception such as the perception of (i) pitch, (ii) intervals, (iii) dissonance, and (iv) the roots of chords.

1.2.2 Frequency Ratios

It has been held since Pythagoras that certain musical intervals are simple or basic because they correspond to simple frequency ratios. The "harmony of the spheres" concept [Schneider 1960; Schavernoch 1981; Haase 1986], originally developed by Pythagoras and Plato [Lippman 1963a], persisted through the Middle Ages (e.g. Boethius, around 500 A.D.) and into the Renaissance (e.g. Zarlino, see [Wienpahl 1959]). A more "naturalistic" approach to frequency ratios was adopted by Rameau [1721] (Sect. 1.4.1). Modern music theorists such as Boomsliter and Creel [1961], Tanner [1981] and Barlow [1987] appealed to psychology and neuropsychology to explain the supposed sensitivity of the auditory system to frequency ratios.

The frequency ratio concept has some serious shortcomings. Not the least of these is that the very name "interval", as used by musicians, implies a kind of *distance*, not a ratio. Musicians assume that the same interval covers the same pitch distance, independent of transposition into different registers.

In the frequency ratio approach, most intervals may be assigned two different ratios, one of which contains only powers of two and three (the *Pythagorean* version) and one of which may also include powers of five (the *just* or *pure* version). For example, the interval of a major second (two semitones) may be assigned the frequency ratios 9/8 ($3^2/2^3$) and 10/9 ($2\times5/3^2$). If intervals corresponded to specific frequency ratios, one would expect intonation to vary randomly about one of these. In fact, intonation of intervals with different just and Pythagorean versions normally varies over a range which encompasses both ratios [Burns and Ward 1982], making it impossible to establish experimentally which of these is "the" frequency ratio of the interval.

If intervals were perceived as frequency ratios, then different *enharmonic spellings* of the same interval might be expected to correspond to different ratios. Consider, for example, the augmented fifth and minor sixth. These intervals are enharmonically equivalent: both span eight semitones in the chromatic scale. The minor sixth may be regarded as an octave (2:1) minus a major third (5:4), so its frequency ratio should be $(2/1)/(5/4) = 8/5 = 1.60$. An augmented fifth may be regarded as two superposed major thirds: $(5/4)\times(5/4) = 25/16 = 1.56$. In the frequency ratio approach, then, an augmented fifth is smaller than a minor sixth. Musical experience suggests exactly the opposite: the upper note of an augmented fifth interval normally resolves by rising a semitone, and tends to be played slightly sharp (if the instrument allows) to emphasize its "leading" nature, while the upper note of a minor sixth normally resolves by falling a semitone, and is therefore more likely to be played flat than sharp. Similarly, augmented fourths are consistently played "sharper" than diminished fifths [Ward 1970], even though the simplest (just) frequency ratio of an augmented fourth is smaller than that of diminished fifth.

The frequency ratio theory leads to the idea that harmonic language could be expanded by introducing frequency ratios including factors not only of 2, 3 and 5, but also of 7, 11, 13, etc. [e.g. Fokker 1966]. But in spite of decades of experimentation, Western harmony shows no inclination to move in this direction (Sect. 3.3.2).

1.2.3 Harmonic Series

Music theory texts generally make some reference to the harmonic series: the series of frequencies of harmonics of a single musical note. The first sixteen harmonics of the note C_2 are notated musically in Fig. 1.1d. Of these, the seventh (B*b*), eleventh (F#) and thirteenth (A*b*) are noticeably out of tune by comparison to equal temperament.

Since Rameau [1721], the fourth, fifth and sixth harmonics have been regarded by many as the origin of the major triad, expressed in frequency ratio form as 4:5:6. The harmonic series seems to explain why we call the triad C–E–G a "C chord" as opposed to an "E chord" or a "G chord": the root of the chord is octave equivalent with the fundamental of the harmonic series from which it is supposed to be derived.

The harmonic series idea begins to break down as soon as we try to "explain" the second most common chord in Western music, the minor triad. The simplest conceivable frequency ratio representation of the minor triad is 10:12:15 (6:7:9 is out of tune). To suggest that the minor triad is derived from the 10th, 12th and 15th harmonics of a complex tone is far-fetched, on three counts. (i) Chords such as the diminished triad (5:6:7?) should, by this argument, be more consonant than the minor triad (10:12:15). (ii) The tenth harmonic is not octave-equivalent with the first, so the harmonic series model does not predict the root of the minor triad in the direct way that it predicts the root of the major triad. (iii) Harmonics above the tenth are generally inaudible in typical complex tones [Plomp 1964; Terhardt 1979a], so it is hard to see how they could play any direct role in music.

The present study describes a more sophisticated version of the harmonic series model of harmony, based not on the "naturalness" of the harmonic series as a physical entity (the frequency relationship between harmonics of complex tones) but on the *familiarity* of the auditory system with the harmonic series, via the pitch pattern of the audible harmonics of complex tones. In later sections, it will be shown how this approach solves problems which have traditionally plagued the harmonic series approach, such as the problem of tuning and frequency ratios (Sect. 3.3.3) and the problem of "nonharmonic" notes in musical chords, such as the third of the minor triad (Sect. 3.4.2).

1.2.4 Beats

Helmholtz [1863] was aware of the arbitrary nature of explanations of musical intervals and chords based on frequency ratios and the harmonic series. He developed an alternative scientific explanation of musical intervals in terms of *beats*: simultaneous pairs of periodic complex tones in simple frequency ratios sound consonant because some pairs of harmonics coincide and combine. In the case of the perfect fifth, for example, the third harmonic of the lower tone coincides with the second harmonic of the higher tone. If a musically important interval like a fifth is mistuned by a small amount, then pairs of near-coincident harmonics produce dissonant beats.

The Helmholtz theory of beats only directly applies to *simultaneous* tones. According to the theory, the musical meaning of intervals between *sequential* tones is somehow learned by exposure to simultaneous tones. History suggests, however, that musical intervals (octaves, fifths) between sequential tones existed long before people started singing or playing tones simultaneously in music [Grout 1960]. Further, melodies in which the most important notes are separated by octave, fifth and fourth intervals have evolved independently in different cultures [Burns and Ward 1982]. Wood [1961] points out that many melodies were originally sung in caves, where reverberation may have caused beats to influence interval sizes. However, an echo generally has much less physical amplitude than the original signal, so the amplitude modulation (beating) of near-coincident harmonics in the original and the echo would normally be imperceptible [Terhardt 1974b]. Thus, the theory of beats does not seem to be able to explain the melodic origins of musical intervals. (A more satisfactory explanation involves the *pitch commonality* of sequential tones, see Sect. 3.2.3.)

Beat frequencies exceeding 20 Hz evoke a sensation known as *roughness*, which is normally perceived as dissonant. The roughness of a chord does not indicate what musical intervals lie between the notes of the chord; it depends as much on the spectra and registers of the notes as it depends on the intervals between them. So it is questionable whether beats and roughness have anything to do with the "musical meaning" of simultaneous intervals and chords [Pierce 1966; Terhardt 1974a].

1.2.5 Combination Tones

Flute and recorder players are sometimes disturbed by low, buzzing sounds which seem to accompany their duet music. These "combination tones" result from non-linearities in the ear's frequency-encoding system (the cochlea and basilar membrane). The frequencies of combination tones depend on those of the actual pure tone components which produce them. One kind of combination tone is the *simple difference tone*. Its frequency is simply the difference between the frequencies of the two other tones. If these two tones are adjacent harmonics of a fundamental frequency, then the frequency of the simple difference tone is the same as that of the fundamental. A more commonly audible kind of combination tone is the *cubic difference tone*, equal to twice the lower frequency minus the higher frequency [Moore 1982]. For example, pure tone components at frequencies of 200 and 300 Hz produce a cubic difference tone at $2 \times 200 - 300 = 100$ Hz.

The pitch of a *residue tone* (a complex tone with no fundamental) normally corresponds to that of its "missing fundamental". Examples of residue tones in the modern environment are speech vowels heard over the telephone and musical tones produced by a pocket radio. Helmholtz [1863], and many of his successors up to and including Fletcher [1924], thought that the pitch of residue tones was due to combination tones, i.e. non-linear distortions in the cochlea. However, Schouten [1938] showed experimentally that the pitch at the "missing fundamental" is perceived even when no such distortion products are present.

Music theorists such as Tartini [1754], Krueger [1910] and Hindemith [1940] argued that the root of a chord (Sect. 1.1.2) is a combination tone. This theory was discredited by Stumpf [1909] and Cazden [1954], who pointed out that combination tones in music are generally inaudible (due to masking, see [Plomp 1965; Smoorenburg 1972], and, when they *are* audible, they sound dissonant (whereas chords with clear roots are consonant).

In summary, theories of musical intervals based directly on frequency ratios, the harmonic series and combination tones are implausible and unscientific. In the words of Cazden [1954, p. 289], "These doctrines represent in reality a species of mysticism dealing with magic numbers."

1.2.6 Periodicity

The above theories all take as their starting point the frequency spectrum of a sound, derived from its waveform by spectral analysis. Could an approach based directly on the waveform of a musical sound yield more feasible explanations of musical intervals?

The frequency spectrum of an exactly periodic waveform is exactly harmonic, i.e. its component frequencies are exact whole multiples of the fundamental (or waveform) frequency. So detection of harmonicity in the frequency domain (i.e. detection of a harmonic series of frequencies) is essentially equivalent to detection of periodicity in the time domain. Perhaps, then, the pitch of residue tones and the "meaning" of musical intervals can be explained in terms of the periodicity of their pressure waveforms, instead of the harmonicity of their frequency spectra. (For explanatory notes and illustrations, see the references cited in Sect. 1.2.1.)

The ear itself is not physiologically suited to periodicity detection. Theories of periodicity detection in hearing normally refer instead to the physiology of the brain. A starting point for such theories was the finding that neuron firing in the auditory nerve can, under certain conditions, by synchronized with the periodicity of the waveform of a tone [Wever and Bray 1937; Rose et al. 1967].

A possible explanation for the pitch of a residue tone based on periodicity is that neural signals from different parts of the basilar membrane, corresponding to different harmonics of the tone, are synchronized such that the combined signal (in the auditory nerve) has a periodicity corresponding to the frequency of the missing fundamental. Explanations of pitch based on periodicity of nerve impulse patterns date back to the *residue* theory of Schouten [1940]. The "musical meaning" of intervals corresponding to exact frequency ratios may be explained in a similar fashion [Boomsliter and Creel 1961]. The theory may be further refined by allowing for limited departures from periodicity to account for pitch shifts, octave stretch, and the pitch of slightly non-harmonic complex tones [Ohgushi 1983].

A theory of musical intervals based on periodicity of neuron firing is fraught with technical and philosophical problems. For the moment at least, it is impossible to establish a definite link between the sensation of pitch produced by a residue tone and the synchrony of the signals in the auditory nerve corresponding to its harmonics: any given sensation (or, for that matter, any experience) could be linked to any of countless simultaneous neural processes (Sect. 2.1.2). Until we understand some basic things about brain function, such as exactly how large ensembles of neurons behave under certain conditions and why, musically useful neural explanations of music perception will remain unattainable.

1.3 Psychologically Based Theories

1.3.1 Introduction

Purely physically based theories of harmony would have us believe that music is a physical phenomenon, heard by a frequency-analyzing ear connected to a neural network which is sensitive to periodicities of neural pulses. In a psychological approach, by contrast, music is inherently social; it is heard and appreciated by people, not organisms.

In music theory and analysis, there is little coordination between the activity of analyzing the score and that of analyzing the *experience* of the music [Clifton 1983]. Modern psychological models attack this problem by concentrating on *perceptual* (rather than physical) interactions between tones and groups of tones (chords, melodies, progressions, key areas, etc.). The cognitive approach to musical pitch perception, for example, looks for "structural relations within a set of perceived pitches independently of the correspondence that these structural relations bear to physical variables" [Shepard 1982, p. 306]. The same may be said for the linguistically based approach of generative grammars [Lerdahl and Jackendoff 1983] and the mathematically based approach of group theory [Balzano 1980]. These three modern, psychologically based approaches to harmony theory are briefly introduced in the following sections.

1.3.2 Cognitive Structures

In a cognitive approach, structural relationships between musical elements (tones, chords, keys . . .) may be represented by points in a multidimensional space, by means of the mathematical technique of *multidimensional scaling* [Kruskal 1964]. The input data to the multidimensional scaling algorithm consists of a matrix of experimental results, in which each element in a set is compared with each other element (e.g. for apparent similarity). The output of the algorithm is a multidimensional map of the elements in the set, in which the distances between elements are approximately proportional to their apparent "differentnesses".

Shepard [1964] presented listeners with sequential pairs of octave-spaced tones: complex tones whose pure tone components are spaced at octave intervals over most of the audible spectrum. The listeners' task was to decide which of the two complex tones was higher in pitch. A multidimensional scaling solution of the results [Shepard 1978] yielded an almost perfectly circular solution, corresponding to the *chroma cycle* in music theory (a construct where successive pitch classes C, C#, D, etc. are substituted for numbers on a clock face). This result inspired a revival of the helical model of pitch, proposed as early as 1846 by Drobisch [Ruckmick 1929], in which tones an octave apart are represented by corresponding points on successive curves of a helix. Recently, more complex structures have been developed to account for musical pitch relationships in greater detail [Shepard 1982].

Krumhansl and Shepard [1979] asked listeners to rate how well a major scale was completed by an octave-spaced probe tone, presenting probe tones at all 12 different chroma during the experiment. The result was a *tone profile* reflecting the tonal structure of a major key. Tone profiles may be used to calculate the strength of the relationship between musical keys. For example, the profile for C major shows high positive correlations with profiles for closely related keys such as G major and A minor, and high negative correlations with distant keys such as F# major and G# minor. Multidimensional scaling solutions of the 24 by 24 matrix of correlation coefficients between musical keys yield maps in which closely related keys in music theory (dominant, subdominant, relative minor, parallel minor, etc.) are close to each other [Krumhansl and Kessler 1982].

Multidimensional scaling solutions of experimental data primarily *summarize* that data in a convenient and attractive form. They do not necessarily *explain* it. For example, the multidimensional scaling approach to the analysis of pitch relationships in music does not distinguish between universal (or sensory) and cultural effects. So it is not clear from multidimensional scaling solutions of similarity ratings of musical chords [Bharucha and Krumhansl 1983] that the data are largely sensory in origin (Sect. 5.7.3). The same applies for tone profiles of musical keys [Krumhansl and Kessler 1982] (Sect. 6.2.5).

Krumhansl and Shepard regard psychoacoustical data as essentially irrelevant to musical relationships. For example, Shepard [1982, p. 307] states that "because of the unidimensionality of scales of pitch such as the mel scale, perceived similarity must decrease monotonically with increasing separation between tones on the scale. There is, therefore, no provision for the possibility than tones separated by a particularly significant interval, such as the octave, may be perceived as having more in common that tones separated by a somewhat smaller but musically less significant interval, such as the major seventh." These criticisms of the psychoacoustical approach are invalid on two counts. First, the mel scale applies only to pure tones, not to musical (complex) tones (Sect. 2.5.2). Second, in a psychoacoustical approach, musical (complex) tones an octave apart have pitches in common: a musical tone has several possible pitches, separated by musical intervals such as octaves and fifths [Terhardt 1972, 1974a]. The latter explains why the similarity of sequential complex tones does not decrease monotonically with increasing pitch separation, but has local maxima at musically important intervals (Sects. 5.4–5.6).

The sensory approach of Terhardt and the cognitive approach of Krumhansl and Shepard are not mutually exclusive. They may be synthesized into a single method, by applying correlation and multidimensional scaling techniques to tone profiles calculated by psychoacoustical modelling (Sect. 6.2.5). In this way, aspects of harmony theory may be explained purely psychoacoustically, without the need to postulate the existence of abstract internal representations of musical structure.

1.3.3 Generative Grammars

Perceived music may be regarded as a *hierarchical structure of perceptual elements* [Deutsch and Feroe 1981; Krumhansl and Castellano 1983; Kessler et al. 1984]. Elements at one level of the hierarchy which are perceived to be related or similar are grouped to form elements at a higher level. The functions of chords in a musical key may be assigned to a hierarchy in which the tonic triad is at the top, diatonic chords are near the top, and chromatic chords are further down. The rhythmic functions of different metrical positions (bars, downbeats, off-beats, etc.) are similarly hierarchical. Schenker's [1935] concepts of *foreground*, *middleground* and *background* are examples of hierarchical levels applied to the analysis of entire musical pieces.

In their approach to perceptual hierarchies in music, Lerdahl and Jackendoff [1977, 1983] interpreted the conventions of music theory as a special case of the conventions of linguistic grammar. Linguistic conventions ar intuitively understood by all speakers of a given language; musical conventions, by musically trained or experienced members of the corresponding musical culture. The theory of Lerdahl and Jackendoff provides a psycholinguistic basis for the communication of implicit or abstract ideas by music, while at the same time recognizing the pitfalls of making too close an analogy between music and language [cf. Feld 1974].

In general, a musical passage has a number of different possible hierarchical descriptions, corresponding to different analytical interpretations. These are illustrated by Lerdahl and Jackendoff by means of tree structures. Rules of "well-formedness" are used to decide which structural groupings are possible, and preferences for particular structural groupings over others are accounted for by means of "preference rules". Recent experimental work [Deliege 1987] has confirmed the validity of the grouping preference rules for musically trained listeners, and has shown that nonmusicians exhibit similar grouping behaviour.

Problems associated with the generative grammars approach to harmony include the failure to distinguish between compositional and perceptual structures, and the failure to account for the possibility of network structures and transient levels [Narmour 1983]. Furthermore, the theory does not explain *why* "... a triad in root position is more stable than its inversions; ... the relative stability of two chords can be determined by the relative closeness to the local tonic of their roots on the circle of fifths; conjunct linear connections are more stable than disjunct ones; ... and so forth" [Lerdahl and Jackendoff 1977, pp. 130–131]. In their book [1983], Lerdahl and Jackendoff discuss the problems associated with explanations of intervals, chords and scales based on the harmonic series, but provide no concrete alternative other than an appeal to cognition: "Tonality in music provides evidence for a cognitive organization with a logic all its own", and "musical idioms will tend to develop along lines that enable listeners to make use of their abilities to organize musical signals" (p. 293). This may be true, but it does not satisfactorily explain fundamental aspects of Western music such as the importance of the major and minor triads.

The methods of Lerdahl and Jackendoff are useful and practical teaching tools. The structure of tonal music becomes clearer to the student when it is expressed in terms of a strict formalism. Their work could be appropriately expanded by incorporating explanations of musical elements and preference rules based on Terhardt's [1974a, 1976] concept of pitch, consonance and harmony in music. The result would be a more comprehensive and satisfying theory of harmony than has previously been possible.

1.3.4 Mathematical Groups

Balzano [1980] contributed to the debate on the origins of the diatonic scales and triadically based harmony by an analysis of the mathematical properties of the set of twelve pitch classes in the chromatic scale. He treated this set as a cyclic group of order 12, i.e. as a group of 12 elements in which each element has an upper and a lower neighbour. This approach grew out of the application of set theory to the theory of atonal music [Forte 1964, 1973; Beach 1979]. Its most fundamental assumptions are that musical intervals are perceived categorically [Burns and Ward 1978] and that notes an octave apart are harmonically equivalent (Sect. 3.3.1).

Balzano generated the cyclic group of order 12 in various ways. The simplest method involves adding semitone intervals one at a time until all members of the group have been generated. The result is the chroma cycle $\{0,1,2,3,\ldots,11\}$, or "semitone group". Another method involves adding perfect fourth or fifth intervals, producing the cycle of fifths $\{0,7,2,9,\ldots,5\}$. The standard heptatonic scale consists of neighbouring elements in this set: 5, 0, 7, 2, 9, 4 and 11. No other intervals beside 1 (or 11) semitones and 7 (or 5) semitones can alone generate the entire cyclic group of order 12, a fact which Balzano interpreted as indicative of the importance of the semitone (or minor second) in melodic relationships and of the fourth and fifth in harmonic relationships.

A further method of generating the cyclic group of order 12 proposed by Balzano involved a two-dimensional generator. This idea was not entirely new. Schoenberg [1954] developed a two-dimensional map of musical keys, extending vertically by perfect fifths and horizontally by relative and parallel major/minor relationships, in which closely related keys were always close together. Longuet-Higgins [1979] developed a two-dimensional map of pitch classes, extending vertically by major thirds and horizontally by perfect fifths. Balzano's method was to express each musical interval as as a combination of major and minor thirds: in the terminology of set theory, he generated the cyclic group of order 12 by multiplying together cyclic groups of order 3 (there are 3 major thirds in an octave) and of order 4 (there are 4 minor thirds in an octave). The resultant two-dimensional space was similar to that of Longuet-Higgins in that the notes of common (major and minor) triads and of diatonic (major and minor) scales were next to each other, and tonic triads were centrally located relative to the notes of the diatonic scale.

In a later article, Balzano [1982] used the set-theoretical approach of Forte [1964] to demonstrate the uniqueness of the standard heptatonic scale. First, he defined a *coherent* scale to be one in which interval size increases monotonically with the number of scale steps, regardless of where one begins in the chroma cycle. Of the 66 different possible heptatonic scales (7-note subsets of the chromatic scale, not counting transpositions), only the standard heptatonic scale is coherent.

Balzano's theory has the following drawbacks. First, it ignores the sensory basis for the musical importance of the perfect fifth and fourth intervals

[Terhardt 1974a, 1976]. In order to "derive" the importance of those intervals, Balzano presupposes the existence of the chromatic scale; but this scale first emerged in Western music long after fifths and fourths came into widespread use (Sect. 3.3.2). Second, it fails to give a detailed account of the *process* by which the described group-theoretical properties may have influenced the historical development of music. Consequently, the *extent* of this influence is also unclear. Finally, the group-theoretical principles which Balzano invokes to demonstrate the usefulness of the 12-note scale also suggest divisions of the octave into 20, 30 or 42 equal parts [Balzano 1980]; but such scales have failed to generate viable new harmonic systems.

1.4 Towards a Psychophysical Theory

1.4.1 From Rameau to Terhardt

In his *Traité de l'harmonie* [1721], the composer and theorist Jean-Philippe Rameau adopted a physical entity, the *corps sonore*, as the basis for his theory of harmony [Christensen 1987]. By this he meant any vibrating system (string, pipe, vocal chords, etc.) whose sound spectrum comprises whole-number multiples of a fundamental frequency. Rameau hypothesized a link between the fundamental frequency of the *corps sonore* and the *basse fondamentale*, or root, of a chord. This led to a distinction between chords whose bass note is octave-equivalent to the root (root position chords such as G–B–D) and chords whose bass note and root are different (inversions such as B–D–G, D–G–B).

Rameau's seminal work on harmony inspired generations of music theorists, especially during the music-theoretical boom of nineteenth-century Europe [Watson 1982]. To this day, harmony textbooks begin with a discussion of the musical intervals of the harmonic series. But, as we have seen (Sect. 1.2.3), the harmonic series approach fails to explain satisfactorily such elementary structures as the minor triad and the diatonic and chromatic scales, in spite of valiant efforts to force these structures into the harmonic series mould by music theorists.

The trouble with Rameau's approach seems to be that the *corps sonore* is an objective, *physical* entity, but the musical harmony which it is supposed to explain is a subjective, psychological *experience*. The seeds for a resolution of this problem were sown early this century with the emergence of the scientific discipline of *psychophysics*, or – more specifically – *psychoacoustics*. Psychoacoustics is concerned with the relationship between physical and perceptual descriptions of sound, e.g. between frequency, intensity and spectral composition on the one hand, and pitch, loudness and timbre on the other [Fletcher 1934].

A turning point in the theory of harmony was Terhardt's [1974a, 1976] psychoacoustical reinterpretation of Rameau's theory. Instead of adopting a physical entity (the *corps sonore*, or, in modern parlance, the *harmonic complex tone*) as the ultimate basis of harmony, he adopted the principle of *familiarity of the auditory system with harmonic complex tones.* He hypothesized that our sensitivity to musical harmony and consonance is ultimately based on our sensitivity to the specific pattern of pitch which characterizes harmonic complex tones, a sensitivity which is fostered primarily by repeated exposure to this specific pattern in speech vowels, and which presumably influenced Western harmony at all stages of its development. Terhardt thus shifted the emphasis in Rameau's theory from the world of physics and "Nature" to the world of sensation and perception. In this way he was able to solve the main conceptual and (eventually) the main practical problems of Rameau's theory.

Rameau was not entirely oblivious of the role of perception in his theory, nor of the importance of speech vowels. He began his *Nouveau système de musique théorique* [1726] as follows:

> There is *actually in us* a germ of harmony which apparently has not been noticed until now. It is nonetheless easily perceived in a string or a pipe, etc. whose resonance produces three different sounds at once. Supposing this same effect in all sonorous bodies, one ought logically to suppose it in the *sound of our voice*, even if it is not evident. ([Christensen 1987, p. 26] my italics; the "three different sounds" to which Rameau refers are those of the octave, twelfth and seventeenth, i.e. of harmonics 2, 3 and 5.)

However, Rameau was working long before the major developments in the psychoacoustics of pitch perception of the nineteenth and twentieth centuries [Ohm 1843; Helmholtz 1863; Fletcher 1934; Schouten 1940; Plomp 1967; Terhardt 1972] and so was in no position to realize the implications of a perceptual approach to the relationship between harmony and the *corps sonore.* Nor was he in a position to identify the important role of the voice as the most important *corps sonore* underlying musical harmony (Sect. 3.1.2).

1.4.2 The Psychoacoustical Approach

The conventions of Western music theory are general statements about what "sounds right" in diatonic music. As we have seen, attempts to find a basis for music theory in *acoustics* run into difficulties because acoustics is based on physical measurements – what sound "really is", perhaps, rather than what it sounds like. Attempts to find a basis for music theory in *psychology* have so far run into difficulties because psychologists have tended to assume a simple one-to-one correspondence between the notes of a musical element (such as a chord or melody) and the tone sensations evoked by that element. This relationship is actually quite complex [Terhardt et al. 1982b].

Psychoacoustics is concerned with the relationship between the physical properties of stimuli and observers' descriptions of the sensations they evoke. On the basis of experimental results indicating the nature of this relationship in specific cases, general rules are developed by which sensory properties of

sounds may be calculated from their acoustical properties. This general understanding allows the sensory properties of sounds, as distinct from their physical properties or musical notation, to be accurately described [Terhardt 1978].

Psychoacoustical models allow the sensory nature of sounds to be specified before higher psychological processes such as hierarchical grouping are considered. A complete analysis of music perception involves appropriate acoustical description of musical sounds, psychoacoustical theories and models by which acoustical descriptions of sounds may be converted into appropriate experiental descriptions, and appropriate psychological theories by which the organization of (and selective attention to) musical sounds may be understood.

Psychoacoustics is normally restricted to isolated sounds, i.e. sounds out of context. Extrapolation of results to sounds in a musical context can be problematic. In the present study, limited extrapolation of this kind is achieved by means of quantitative estimates of the "pitch commonality" and "pitch proximity" of sequential sounds (Sect. 4.6), bridging the gap between existing psychoacoustical and psychological approaches.

In psychoacoustics, experiential (or psychological) variables such as "loudness" and "roughness" are measured experimentally, usually by averaging the results of a number of different listeners. Such averaging is only meaningful when results vary relatively little from one person to the next. Care must be taken when transferring psychoacoustical results across linguistic or cultural boundaries, as translation and cultural conditioning may have a considerable effect on the measurement and scaling of psychoacoustical variables.

A disadvantage of the psychoacoustical approach is its tendency toward excessive complexity. For example, "appropriate evaluation and prediction of the various pitch percepts of complex tonal signals on the basis of the physical parameters is far from being either simple or trivial" [Terhardt et al. 1982b, p. 679]. The complexity of psychoacoustical modelling may be reduced by isolating effects which are relatively unimportant in a given application, and neglecting them (as in Chap. 4 of the present study).

1.4.3 Outline of the Book

This study aims to explain some of the conventions of Western music theory (as outlined earlier in this chapter), using psychoacoustical data, theories and techniques to explore their sensory basis. Chapters 2 and 3 survey relevant literature in psychoacoustics and psychomusicology, and introduce some new concepts and approaches to music theory. Chapter 4 describes a mathematical model by which sensory (or universal) aspects of the perception of chords and chord sequences may be simulated. The model is then tested and fine-tuned by comparing its calulations against the results of experiments (Chap. 5): deviations between calculations and experimental data point to specific effects of musical conditioning. Finally, the model is applied in the theory, analysis and composition of music (Chap. 6). Technical terms from music, psychology and physics are explained in an extensive glossary.