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

3. Psychomusicology

The perception of music, and hence its performance, composition and historical development, is conditioned by the repeated exposure of musicians and audiences to auditory patterns and regularities. Such patterns occur in music itself, in everyday social and physical environments, and before birth. Musical consonance depends on the smoothness and tonalness of tone simultaneities, the pitch commonality and pitch proximity of sequences, and cultural conditioning. Pitch in tonal music is specified relative to diatonic and chromatic scales, in which notes an octave apart are harmonically equivalent; intonation within chromatic pitch categories involves a compromise between minimum roughness and maximum pitch commonality. Diatonic (major/minor) tonality depends on various aspects of consonance and musical pitch, but especially on the roots of chords and broken chords.

3.1 Conditioning

3.1.1 Sensory Versus Cultural

Ever since Pythagoras's conflict with Aristoxenes [Cazden 1958], music theorists have been trying to separate "nature" from "nurture." For example, Rameau [1721] and Schenker [1906] invoked the supposed "naturalness" of simple frequency ratios and the harmonic series to explain harmony. Cazden [1954], at the start of his criticism of Hindemith [1940], divided contemporary discussion of the fundamental questions of music theory into "three main types: those claiming to be founded on the laws of Nature, those claiming the priority of the arbitrary instinct of the composer, and those which rely on the observed practice of the art of music . . . the natural, the subjective and the empirical theories". More recently, DeWitt and Crowder [1987] distinguished between "rational" musical theories with an acoustical basis in "natural law" (simple integer ratios, harmonic series, etc.) and "empirical" theories based on context and learning.

In a psychoacoustical approach, the "natural" approach to the origin of harmony is rejected, as it fails to account for the difference between physical reality and the way it is perceived (Sect. 1.4.1). The nature/nurture distinction becomes a distinction between *innate* and *acquired.* Innate aspects of perception are supposedly dependent on the physiology of the sensory organs and the nervous system; acquired (or "learned", or "conditioned") aspects, on familiarity with regularities in the human environment. The innate/acquired distinction is similar to the distinction between "phylogenetic" (i.e. related to evolution) and "ontogenetic" (related to individual development).

The auditory system is conditioned by different patterns of sound at different developmental stages and in different situations. The foetus is repeatedly exposed to the internal sounds of the mother's body, of which the heartbeat, walking sounds and the voice appear to be the most relevant for music (Sect. 3.1.2). After birth, infants, children and adults are repeatedly exposed to sound patterns (in pitch and time) which indicate the presence of particular objects (sound sources) in the environment. In a particular musical culture, individuals are repeatedly exposed to the particular, arbitrary patterns of sound characteristic of the culture's musical styles. Thus, *music may be regarded as a multilayered structure of more or less familiar patterns, and its perception seen as a multilayered process of pattern recognition, based on familiarity.* In the words of Davidson et al. [1987, p. 606], "it would seem that, in the experiencing of music, familiarity is the most important variable."

From a musical viewpoint, the most important aspect of the ear's physiology is the spectral analysis of complex sounds into pure tone components enabled by the cochlea and basilar membrane (Sect. 2.2.2). The process of becoming familiar with and learning to recognize environmental sounds in early life is largely a process by which the central nervous system becomes attuned to auditory sound spectra as functions of time, output from the cochlea as patterns of neural activity.

The nature of these spectra depends as much on the physical nature of environmental sounds (e.g. the periodicity of the waveform of a tone; Sect. 1.2.6) as it depends on the physiology of the ear. In Gibson's [1966, 1979] approach, both of these belong to the same "environment"; both innate and acquired characteristics originate from a single integrated system. This is particularly true in the case of prenatal conditioning (Sect. 3.1.2).

In the present study, aspects of perception dependent on physiological (innate) properties and limitations and on experience of universal aspects of the human environment are brought together under the one umbrella and called *sensory.* Aspects of perception associated with consistent and systematic differences in human environments, especially social and musical environments, are called *cultural.*

The boundary between sensory and cultural, like the boundary between innate and acquired, is not cut and dried. Both sensory and cultural aspects of perception involve conditioning through perception of and interaction with physical and social environments. In general, it is hard to know where to draw the line between aspects of the perceptible world which are universal and aspects which are culture-specific, and hence between aspects of perception which are conditioned by one or the other of these. In the present study, the only universal environmental sound taken into consideration is the *harmonic complex tone*, which is fundamental to both speech and (Western) music. Experimental effects (Chap. 5) and aspects of music theory (Chap. 6) are described as "sensory" if they appear to be a result of exposure to complex tones.

Both sensory and cultural forces significantly influence the nature and development of music. Different musical styles are constrained by the same universal (sensory) properties of human hearing. These lead to cross-cultural similarities, such as the widespread use of the octave, fourth and fifth intervals in world musics, and the independent evolution of the chromatic scale in different musical cultures [Burns and Ward 1982]. On the other hand, cultural differences (different technologies, social conventions, attitudes to music, etc.) and the flexibility and adaptability of human perception combine to produce large intercultural differences in musical style. Thus, Westerners are conditioned to Western tonal styles [Francès 1972; Cuddy et al. 1979], and members of different cultures perceive the same music (or musical fragments) in different ways [Castellano et al. 1984; Kessler et al. 1984; Davidson et al. 1987].

The importance of both sensory and cultural effects in music perception is repeatedly acknowledged, although the "sensory" aspect is given different shades of meaning by different authors. Helmholtz [1863] recognized the important roles of both "physical and physiological acoustics" and "musicology and aesthetics" for the structure of scales and chords [Landau 1961]. Lerdahl and Jackendoff [1983] explicitly acknowledged the separate roles of sensory and cultural effects in the development of music-theoretical conventions by dividing their well-formedness and preference rules for the analysis of tonal music into those apparently reflecting innate perceptual or cognitive capacities and those apparently acquired by exposure to Western music. More recently, Roberts [1986] described experimental results which "support both psychoacoustic and relativistic theories of consonance and dissonance" (p. 170).

Cultural conditioning is often thought of as a kind of *brainwashing* – a process by which listeners' perception of music becomes biased toward the arbitrary structural characteristics of particular musical styles. Another aspect of cultural conditioning is *sensitization*. Repeated exposure to music makes listeners more sensitive to sensory aspects of the structure of particular kinds of sound, such as those modelled in Chap. 4 (see Sect. 5.8.2). For example, Westerners may be insensitive to the intricacies of African drum rhythms, while Africans may be insensitive to the intricacies of Western pitch structures. Aspects of African drum rhythms may nevertheless derive from universal aspects of pulse perception [Parncutt 1987b], and aspects of Western harmony may be based on universal aspects of the perception of complex tones [Terhardt 1974a, 1976].

3.1.2 Prenatal Conditioning

Perceptual learning (or conditioning) by repeated exposure to particular sound patterns occurs throughout life. Presumably, the process begins as soon as perceptual systems begin to function in the foetus. The process then accelerates until, at birth, it has gained enough momentum to enable the infant to cope with the suddenness and surprises of the postnatal world. An investigation of the first, prenatal stage of auditory conditioning suggests that it may have a significant influence on the perception of music by individuals, and on the historical development of music.

There exists considerable indirect evidence that both sensory and cultural aspects of musical conditioning begin before birth. In the case of cultural conditioning, there is anecdotal evidence from musicians whose mothers listened to or played specific pieces of music during pregnancy [Verny 1981]. The present study is concerned only with prenatal conditioning at the sensory (or universal) level: conditioning of the auditory system of the foetus by the internal sounds of the mother's body [Parncutt 1987a]. If the prenatal conditioning hypothesis is correct, these sounds influence auditory perception, and especially music perception, across cultures.

Recent experimental evidence [Trehub 1987 and references therein] has demonstrated that infants are remarkably sensitive to elementary musical structures. Specifically, they are sensitive to pitch contour and streaming (elements of melody), musical intervals such as the octave, fifth and third (elements of harmony), and long and short events in rhythmic patterns across tempo variations (elements of rhythm). This sensitivity is unlikely to be innate, as it doesn't directly improve the infant's chances of survival. Nor is it likely to be the result of musical conditioning – infants are presumably too busy learning how to interact with their environment (people, food, . . .) to pay much attention to the musical sounds to which they may be exposed.

Prenatal conditioning provides a more straightforward and tangible explanation of infants' sensitivity to musical elements. Sensitivity to melody and harmony may be a simple and direct consequence of repeated prenatal exposure to the mother's voice; sensitivity to rhythm, a consequence of repeated prenatal exposure to the mother's footsteps and heartbeat.

The ability of the foetus to hear is well documented. Unborn sheep respond to sounds from outside the mother from 6–7 weeks before term [Bernard et al. 1959]. In humans, both the inner and outer ear are completely developed about twenty weeks before term; from this point on, the foetus responds to sounds [see e.g. Verny 1981].

The prenatal auditory environment is loud and varied [Armitage et al. 1980]. In sheep, Bench [1968] referred to internal background sounds of around 72 dB. Sounds audible to foetal sheep are associated with drinking, eating, breathing, vocalizations, digestion, heartbeat and movement of the mother; sometimes the foetal heart is also audible to the foetus [Vince et al. 1982b]. Presumably, the human foetus is exposed to a similar range of sounds. Those associated with breathing, vocalizations (speech), heartbeat and movement of the mother include regularly repeating, and therefore easily recognizable, patterns (as further described below). Sounds associated with drinking, eating and digestion appear to be too irregular to influence mainstream Western music in obvious ways (although they could, perhaps, be exploited in electronic or non-Western musics).

Prenatal auditory stimulation affects postnatal behaviour in measurable ways. In the case of sheep, prenatally familiar sounds can cause sigh-like changes in respiratory patterns; they also produce less heartbeat acceleration than unfamiliar sounds [Vince et al. 1982a]. Sounds heard by birds before hatching affect their posthatching behavior [Vince 1980]. Presumably, human auditory perception is similarly affected by prenatal conditioning.

According to Gibson [1966, p. 5] "the infant does not have sensations at birth but starts at once to pick up information from the world". In other words, infants (like animals) do not have to be conscious of their actions and perceptions in order to learn to interact with their environment, and to become sensitive to perceptual invariances. The same is true for the foetus.

Of all the sounds heard by humans before birth, probably the loudest and/or most consistent is produced by pulsations in uterine blood vessels associated with the mother's heartbeat [Bench 1968; Grimwade et al. 1970], although this sound often falls below the threshold of audibility, which is quite high at low frequencies [Vince et al. 1982b]. The maternal heartbeat may be regarded as an imprinting stimulus on the foetus [Salk 1962]. Prenatal conditioning by heartbeat sounds explains why babies are calmer when fed from the left breast [Lockard et al. 1979]; it may also underlie the emotional connotation of musical *rubato* (tempo fluctuations) [Parncutt 1987b].

Prenatal sounds associated with walking are presumably at least as loud as those of general body movements [Vince et al. 1982b]. Such sounds may underlie strict rhythms, and their association with dance [Parncutt 1987b]. Music therapists [Thaut 1985 and references therein] have exploited auditory rhythms to help disabled children develop fundamental motor skills, supporting the hypothesis that auditory rhythms and body movements are linked through prenatal conditioning. A link between musical tempo fluctuations and prenatally experienced changes in the walking speed of the mother is suggested by experimentally measured timings of musical *ritardandi* [Kronman and Sundberg 1987]. Lerdahl and Jackendoff's [1983] *Metrical Preference Rule 10* – "prefer metrical structures in which at each level every other beat is strong" – is consistent with the theory that rhythm is conditioned by prenatal conditioning by footstep sounds: in asymmetrical positions, the foetus presumably hears one of the mother's feet louder than the other.

The prenatal conditioning hypothesis suggests that rhythmic sensitivity should be associated only with rhythmic *sounds*, and gross bodily movements; pulse-like patterns of visual and tactile sensations should not evoke rhythmic responses. The experimental results of Grant and LeCroy [1986] are consistent with this expectation. They found that performance at rhythmic perception tasks by intellectually disabled people was significantly worse when rhythms were presented in the form of silent taps on the shoulder (i.e. purely tactile stimuli) than when they were presented as drum strokes (purely auditory stimuli), and that the addition of a visual stimulus (allowing listeners to see a drum, or their knee being tapped) did not improve performance.

There is a tendency in music for lower-pitched sounds to be played on the beat, and higher-pitched sounds off the beat. Familiar examples from Western music are the bass/chord alternations of ragtime and waltz accompaniments. In Ghanaian drum music, low-pitched bells are often used at the start of a cycle, and high-pitched bells often sound on offbeats and syncopations. Similarly, in Indian *tabla* music, downbeats are more often played with lower-pitched than higher-pitched drums. Higher pitches *can* also be used to delineate the beat, with low sounds off the beat, but this normally produces a syncopated feel. These observations are consistent with the idea that rhythm is somehow related to heartbeat and walking sounds, as such sounds are concentrated at low frequencies by comparison to other sounds such as the mother's voice. Vince et al. [1982b] reported that intrauterine cardiovascular sounds are concentrated in the range 40–80 Hz.

The foetus is regularly exposed to the mother's speech. A feature of speech to which the foetus might be sensitive is its *intonation*, i.e. rising and falling of pitch. Intonation is developed and mastered by infants much earlier than individual phonemes and words [Tonkova-Yampol'skaya 1973]. The specific and highly structured patterns of Western speech intonation carry much of the meaning of speech [O'Connor and Arnold 1973], in particular, its emotional meaning [Nilsonne and Sundberg 1984].

The musical correlate of speech intonation is melodic *contour.* Apart from rhythm [Jones 1987], contour is the most important aspect of melodic structure for melody recognition [Dowling and Fujitani 1971; Dowling 1978; Massaro et al. 1980]. Contour is more important than conformance to a diatonic scale for the memory of melodies [Davies 1979]: scalar conformance, i.e., the exact sizes of pitch intervals, are "processed separately" from contour [Eiting 1984] and only become important in longer melodies, in which a tonal framework is well established [Edworthy 1985]. Infants, like adults, recognize melodies by their contours, in spite of variations of pitch register and interval size [Trehub 1987], and children learn melodic contour before they learn chromatic interval categories [Dowling 1982; Davidson 1985].

In general, emotion in speech is transmitted by "phonetic content, gross changes in fundamental frequency, the fine structure of the fundamental frequency, and the speech envelope amplitude, in that order" [Lieberman and Michaels 1962, p. 248]. With the possible exception of phonetic content, each of these has a musical correlate. The musical correlate of "gross changes in fundamental frequency" is contour. The musical correlate of the "fine structure of the fundamental frequency" is the amplitude envelope of single musical tones; the fine structure of a single tone can carry a range of emotional meanings [Clynes and Walker 1982]. Musical correlates of speech envelope amplitude are dynamic and accent (i.e., loudness); these, too, are important carriers of emotional meaning in music.

Speech is regularly interrupted by *breathing*. Similarly, musical passages (especially melodies) are organized into *phrases* [Révész 1953], whose durations are comparable with those of breathing (roughly, from 1 to 10 s, including both inhaling and exhaling). This is true not only for music in which phrases correspond to the breathing of the performer (singing, wind instruments) but also in music on instruments which are independent of the performer's breath (strings, keyboards, etc.).

According to Terhardt [1974a, 1976], a prerequisite for the perception of harmony in music is the ability to perceive complex tone sources as single entities, even though the spectrum of a complex tone contains several separately audible components. This ability is supposed to be acquired early in life, as the pitch patterns of the audible harmonics of complex tones become familiar to the auditory system. Experimental evidence on the perception of "missing fundamentals" by infants [Clarkson and Clifton 1985] suggests that this ability is present at the age of seven months. It is probably present at two months: infants of this age perceive speech intonation [Tonkova-Yampol'skaya 1973]. Further experimental evidence suggests that it is present as early as one month after birth: Eimas et al. [1971] found that one-month-old infants categorize speech sounds in much the same way as adults. Eimas et al. concluded that "the means by which the categorical perception of speech . . . is accomplished may well be part of the biological makeup of the organism" [1971, p. 306]. The prenatal conditioning hypothesis provides an alternative (or additional) explanation: these infants had been picking up information from speech sounds not only for the four weeks following birth but also for twenty weeks before birth.

If the ability to perceive complex tones as single entities is acquired prenatally, then there must be a period between the physiological development of the foetal auditory system and birth when the mother's voice is "heard" not as a single sensation but as a simultaneity of five to ten separate sensations corresponding to the lower harmonics (Sect. 6.1.1). This idea adds new meaning to Schenker's [1906] description of the harmonic series as the "chord of nature". Musical chords (or tone simultaneities generally) might be special kinds of "auditory images" [McAdams 1984] associated not with specific environmental sources of sound but with voices heard before birth, especially the mother's voice – the original *corps sonore* (cf. [Rameau 1721], Sect. 1.4.1). This may explain why chord progressions in Western music tend to be preferred to unaccompanied melodies, even though single complex tones are more consonant than chords.

Infants are more sensitive to semitone changes in melodies based on the major triad than to semitone changes in other melodies [Cohen et al. 1987]. According to Trehub [1987, p. 638], "This implies that the diatonic context has greater coherence or stability for infants . . . . It is nevertheless unclear whether diatonic structure in general or major triad tones in particular, with their simple harmonic relations (4 : 5 : 6), are responsible for the observed effects. Moreover, one cannot entirely rule out possible contributions of prior exposure to music, however limited." According to the prenatal conditioning hypothesis, the results of Cohen et al. are explicable in terms of prenatal familiarity with the pitch pattern of the 4th, 5th and 6th harmonics of complex tones.

Sounds heard regularly before birth have a calming effect when heard again after birth. This is true for sheep [Vince et al. 1982a] and for humans [Salk 1962]. As we have seen, music includes sounds similar in many ways to sounds heard before birth. This may explain music's wide range of therapeutic applications: it may be used to improve mood and attitude, reduce tension, reduce heart rate and blood pressure, and alleviate psychosomatic illness [Hanser 1985 and references therein]. The prenatal conditioning hypothesis may also explain why music helps learning, especially in learning-disabled students [Gfeller 1983]: subliminal prenatal associations may alleviate anxiety, and promote confidence.

"Music is inherently referential . . . in a very special way it 'means itself'" [Thomson 1983, pp. 4,5]. With the exception of programme music (e.g. Beethoven's *Pastoral Symphony*), music has little explicit meaning in relation to the human environment, at least not as much as the other arts; its meaning is syntactic rather than semantic [Booth 1981]. Despite this lack of specific associations, music is vivid and powerful. Musical rhythm, melody and harmony express deep emotions and abstract ideas in specific ways [L. B. Meyer 1956; Cooke 1959]. For an individual, this may be due to cultural conditioning, for prenatal conditioning, or both. The nature of cultural conditioning may in turn depend on the role of prenatal conditioning in the historical development of musical styles.

Roederer [1984, abstract] considered that "A most basic issue in the study of music perception is the question of why humans are motivated to pay attention to, or create, musical messages, and why they respond emotionally to them, when such messages seem to convey no real-time relevant biological information as do speech, animal utterances and environmental sounds." He provided three answers to this question, concerning speech acquisition, speech communication and social integration. In a similar vein, Sloboda [1985, p. 266] observed that "it is not at once clear how musical behaviour makes for better adapted individuals that are more likely to survive". He concluded that it is society as a whole (rather than the individual) that benefits from music: "Music, perhaps, provides a unique mnemonic framework within which humans can express, by the temporal organization of sound and gesture, the structures of their knowledge and of social relations" (p. 267).

The Darwinistic approaches of Roederer and Sloboda are similar to the prenatal conditioning hypothesis in that they help to explain why we have music at all. Prenatal conditioning additionally explains the central importance of certain kinds of structure – rhythm, melody and harmony – in music. These structures are cross-cultural, but they are exploited in different ways in different cultures.

How could the prenatal conditioning hypothesis be tested? If prenatal conditioning by heartbeat, footstep and voice sounds has a *direct* influence on musical ability, then the children of bedridden mothers should be less sensitive to rhythm (especially, the relationship between strict rhythms and dance), and children who are born deaf but regain their hearing as a result of an operation (and consequently acquire average hearing ability) should be altogether insensitive to rhythm, melody and harmony. However, such people could also become sensitive to musical structures from *postnatal* experience. In general, the influence of cultural conditioning on an individual's perception of music may be greater than the influence of prenatal conditioning; the effects of prenatal conditioning on music may only reveal themselves in the long term, as musical style develops over generations and centuries. If this is the case, the theory of prenatal conditioning (like Terhardt's theory of pitch, Gibson's theory of direct perception, Darwin's theory of evolution – indeed, *all* theories, according to Popper [1972]) may well be unprovable.

3.2 Consonance

3.2.1 Introduction

Musical sounds are said to be consonant if they are perceived to "sound well" with each other (*con sonare*). An understanding of consonance is essential for an understanding of tonality: the contrast between consonance and dissonance contributes to "tension" and "resolution" [Nielsen 1983] and thereby to a sense of "forward motion" [Forte 1962, p. 15] in tonal music.

The same musical intervals have been described as consonant or dissonant at different stages of Western musical history. For example, the perfect fourth (P4) was regarded as a consonance in its harmonic function as inversion of the P5, but as a dissonance in its melodic function as a suspension to the third. A simple modern definition of the (harmonic) consonance of musical intervals states that intervals contained in the major and minor triads (m3, M3, P4, P5, m6, M6, P8) are consonant, while others (m2, M2, TT, m7, M7) are dissonant; chords are consonant if they contain no dissonant intervals [Apel 1970]. These definitions apply primarily to simultaneities. In the case of sequential tones, the additional (melodic) role of pitch distance needs to be taken into account (steps are more "consonant" than leaps).

Musical consonance is enhanced by *familiarity* [Cazden 1972]. The more often a relatively dissonant tonal sound or sequence is heard, the more consonant it is judged to be [Heyduk 1975], and discords are objected to less, the more familiar they are [Valentine 1914]. The ability to appreciate the consonance of musical chords is learned by exposure to ordinary complex tones, musical chord progressions, etc. [Terhardt 1974a, 1976]. It seems possible that the ability to appreciate consonance in scales with radically "stretched" intonation could similarly be learned [Pierce 1966; E.A. Cohen 1984].

Consonance also depends on *context*. Chords sound more consonant in familiar contexts (e.g. contexts in which they are diatonic [Gardner and Pickford 1943, 1944; Cazden 1972]). Familiar contexts contain redundancies and therefore carry less information than unfamiliar contexts [Hiller and Isaacson 1959; A.J. Cohen 1962]. Roberts [1986] reported that "Chords presented in a 'traditional' context were judged as being more consonant than the same chords presented in a 'untraditional' context" (p. 169) and "Chords were judged [by untrained listeners] as being more consonant when they are resolved by the following chord than when they are left unresolved at the end of a 'traditional' context" (p. 170). This suggests an expansion of the usual music-theoretical concept of consonance to include sequential (as well as simultaneous) pitch relationships.

In a general approach, sounds at all levels in hierarchies of musical perception may be said to exhibit consonance and dissonance. At the most analytic level, the simultaneous pure tone sensations evoked by a complex tone produced by a musical instrument (string, pipe, etc.) are more consonant than those evoked by a nonharmonic sound such as that of a bell. The simultaneous complex tone sensations evoked by a musical chord are more consonant if the chord has a strong, unambiguous root. The sequential tones of an unaccompanied melody or the sequential chords of a harmonic progression are more consonant if they cover small pitch distances and remain within a standard pentatonic or heptatonic scale. As a rule, consonant sounds are relatively easy to group perceptually; dissonant sounds are not so easily grouped (Sect. 2.3.3).

Consonance should not be confused with *preference*, although in certain cases these terms turn out to be equivalent [e.g. Roberts 1986]. Relatively consonant sounds are not necessarily preferred to relatively dissonant sounds. If this were the case, single tones would always be preferred to chords. Instead, the relationship between dissonance (or complexity, or information flow) and preference (pleasantness, likeability, affective judgment) normally takes the form of an inverted U curve [Smith and Cuddy 1986 and references therein]. For relatively low degrees of dissonance, preference increases with increasing dissonance, while for relatively high degrees, preference decreases with increasing dissonance.

The peak of the inverted U curve is the *optimal dissonance* (or complexity, or information flow) of a piece of music. Degrees of dissonance or complexity which are either lower or higher than the optimum are liked or enjoyed less than degrees near the optimum: music which is too consonant or simple may be boring or irritating, and music which is too dissonant or complex is hard to listen to. Optimal musical complexity depends on the listener, and musical experience [Vitz 1964]. Typically, exposure increases tolerance to (or liking of) dissonance and complexity, causing the peak of the inverted U to move to the right.

Optimal dissonance gradually increased during the history of Western music, as more dissonant sound structures developed [Lundin 1947]. In two-part writing, for example, the generally preferred amount of dissonance changed from that of the fifth in the Middle Ages to that of the third in the Renaissance and later periods. (A notable exception to this was the use of thirds in Medieval England.) The preferred dissonance of chords, with occasional exceptions, increased from that of the major and minor triads in the Renaissance, Baroque and Classical periods to that of more complex chords and tone simultaneities in various romantic, impressionist, jazz, non-diatonic and atonal styles.

In recent decades, this process seems to have slowed or even stopped. Modern audiences prefer different kinds of dissonance in different amounts. Some prefer music with low dissonance (e.g. Baroque and Classical styles), some prefer high dissonance (e.g. atonal styles) and some prefer intermediate degrees of dissonance (late romantics, impressionists, jazz).

Consonance may be divided into *sensory* aspects [Terhardt 1974a, 1976], otherwise known as tonal consonance [Plomp and Levelt 1965], and *cultural* aspects [Lundin 1947; Cazden 1972].

Contributors to the sensory consonance of tone simultaneities include *roughness* and *tonalness* (Sect. 3.2.2). The main culturally conditioned aspect of the consonance of tone simultaneities is the root of a chord (Sect. 3.4.2). Contributors to the sensory consonance of sequential sounds are *pitch commonality* and *pitch proximity* (Sect. 3.2.3). These may be regarded as sensory bases of the culturally conditioned aspects known as harmonic relationship and voice leading (melodic relationship). A further aspect of cultural consonance, involving all lower levels in some way, is the *tonality* of melodies and harmonic progressions (Sects. 3.4.3, 6.2.3–5).

Culturally conditioned aspects of consonance are determined by historical developments. Over the centuries, these, in turn, may be influenced by sensory properties of musical sounds. Many cultural aspects of musical consonance may thus be described as *indirectly sensory.* Indirectly sensory aspects of consonance are conditioned by exposure to aspects of a musical style which appear to have developed as a direct consequence of specific sensory constraints. Although they are sensory in origin, they are only perceived by culturally conditioned listeners.

3.2.2 Roughness and Tonalness

The apparent *roughness* of a sound depends in a complicated way on all its physical properties [Terhardt 1968b, 1974b; Aures 1984]. The simplest case of a rough sound is an amplitude-modulated (beating) pure tone, produced by superposition of two equal-amplitude pure tones of slightly different frequency. The pitch of the tone depends mainly on its *carrier* frequency (the mean of the two original frequencies); the roughness of the tone, on its *modulation* (or beat) frequency (the difference between the two original frequencies). Roughness is evoked by tones with modulation frequencies in the range 20–300 Hz, reaching a maximum at around 70 Hz. Roughness also depends on carrier frequency, but to a lesser extent: it is most pronounced for carrier frequencies near 1 kHz. Contributions to the roughness of a complex sound may be summed over groups of carrier frequencies falling in different critical bands.

An example of a very rough musical sound is a dyad of the complex tones C_4 (middle C) and $D\flat_4$. This dyad contains several amplitude-modulated pure tone components, each formed by the superposition of pure tone components a semitone apart. The fourth harmonics of the two tones combine to produce a waveform with a modulation frequency of 60 Hz and a carrier frequency of 1080 Hz. This waveform, according to the data presented in the previous paragraph, sounds very rough.

In relatively small quantities, roughness can add richness to a sound. For example, full complex tones below about middle C are slightly rough [Terhardt 1974b]: their harmonics are less than 300 Hz apart, so superpositions of neighbouring harmonics have modulation frequencies of less than 300 Hz. This contributes to the rich tone quality of the 'cello. The clarinet, by contrast, has a dark, strangely hollow quality below middle C (in its *chalumeau* register [Kennan 1970]). This is partly due to *lack* of roughness: the even-numbered harmonics of clarinet tones are generally inaudible, so the spacing between odd-numbered harmonics is effectively twice that of the 'cello.

Aures [1984] defined *tonalness* (*Klanghaftigkeit*) to increase as the number and audibilities of pure tone components in a sound increase. In the present study, the term *pure tonalness* is used for this concept. Pure tonalness decreases as the amount of masking among pure tone components increases. It is higher in consonant than in dissonant musical chords, as harmonics of different tones coincide more often, and therefore mask each other less, in consonant chords.

The consonance of musical chords depends partly on the extent to which they evoke complex tone sensations such as "virtual bass notes" [Terhardt 1977]. This is quantified by the parameter *complex tonalness,* the degree to which a sound contains clearly audible complex tone components, i.e. the degree to which it contains clearly audible pure tone components in harmonic pitch patterns. Quantitatively, complex tonalness may be defined most simply to correspond to the audibility of the most audible complex tone component in a sound (Sect. 4.4.3): it is a measure of the ease with which the most salient complex tone sensation in the chord is perceived in the holistic mode of perception, i.e. when no other tone sensation is simultaneously noticed. Complex tones have fewer pure tone components than chords, but they have greater complex tonalness: the pure tone components of complex tones are masked less than those of chords, and are therefore more audible.

Complex tonalness may be enhanced in a musical chord by arranging the tones of the chord so that their fundamental frequencies describe part of a harmonic series. Examples are the major triad and the major-minor (dominant) seventh chord. In this way, complex tonalness influences the chord vocabulary of tonal music (Sect. 6.1.4).

The roughness and tonalness of musical chords are negatively correlated. Roughness is associated with the presence of more than one pure tone component in a single critical band [Plomp and Levelt 1965]. Such components mask each other [Fletcher 1940], reducing tonalness. Sounds in which pure tone components are widely spaced (e.g. octave-spaced tones) have low roughness and high tonalness.

The sensory consonance of a musical chord may be enhanced by playing one note (e.g. the melody) louder than the rest. This reduces roughness and enhances tonalness, as follows. The roughness of a musical chord depends mainly on contributions from pairs of pure tone components of roughly equal amplitude belonging to different complex tones (notes). The roughness contribution from a pair of components falls rapidly as their amplitudes become different, i.e. as the degree of amplitude modulation of the resultant beating tone becomes smaller [Terhardt 1974b]. In a musical chord, roughness may therefore be reduced by playing one note louder than the rest. The same technique enhances the complex tonalness of the chord, simply by increasing the audibility of the most audible tone in the chord. This may explain why it is necessary to play a melody considerably louder than its accompaniment (e.g. on the piano) in order to produce a "singing tone".

3.2.3 Pitch Commonality and Pitch Distance

The present study proposes a new terminology, and a newly structured theory, for the consonance of sequential sound pairs. Like Terhardt's theory of the consonance of tone simultaneities, the present theory has two aspects: pitch commonality, and pitch distance (or its opposite, pitch proximity).

The *pitch commonality* of a sequential pair of sounds is the degree to which the sounds have pitches in common. More precisely, it is the degree to which the sounds evoke tone sensations whose pitches coincide across the two sounds. In categorical perception, pitches up to half a semitone apart or even more may be perceived to coincide in this way (Sects. 2.5.3, 5.6.3).

The pitch of a complex tone is ambiguous (Sect. 2.4.4). Alternative possible pitches lie at important musical intervals (octave, fifth, etc.) above and below the pitch most often heard (the pitch corresponding to the fundamental frequency). So complex tones separated by important musical intervals have pitches in common; they are to some extent "confusable", and therefore possess a certain affinity (*Klangverwandtschaft*) for each other [Terhardt 1983]. This affinity may be regarded as the sensory basis for their consonance.

The idea of pitch commonality is by no means new. In the 19th century, Sechter observed that harmonic relationships between musical chords depended on the presence of common tones, either actual or implied, acting as "harmonic connectives" (*harmonische Bildungsmittel*) or "intermediate fundamentals" (*Zwischenfundamente*) [Watson 1982]. According to this theory, triads of C major and D minor have the tone A in common: the tone A is actually present in the D minor triad, and implied (as a Rameauian *basse fondamentale*) by the C major triad. Helmholtz [1863] thought that musical sounds were harmonically related if they had *frequencies* in common [Cohen 1982]. Western music theory texts recommend that chords in a progression should have *notes* in common, especially in the case of chromatic chords (Sect. 1.1.3). Chords with notes or frequencies in common generally have pitches in common.

The *cycle of fifths* in music theory (Fig. 1.1) may be regarded as a simplified or idealized version of the more general concept of pitch commonality. The pitch commonality of two chords depends on the degree to which they have notes in common, the harmonic relationship between their roots, and their conformance to a particular diatonic scale (Sect. 5.7.3). The pitch commonality of two keys depends on the number of notes the two scales have in common (i.e. the difference between their key signatures, taking into account common additional accidentals such as raised leading notes). Important pitches in musical chords generally coincide with actual notes, or lie at strong harmonic intervals such as octaves and fifths away from actual notes (see Fig. 6.3, Sect. 6.1.5); and diatonic scales are made up of chains of fifths. Consequently, distance around the cycle of fifths reflects harmonic relationships between both musical chords and musical keys.

Small intervals (such as seconds and thirds) are more common than larger intervals in melodies [Jeffries 1974], and chords in a contrapuntal texture "progress" or "lead" better if the voices of the chords traverse small intervals. In a general approach, these effects may be understood in terms of *pitch distance.* In general, the consonance of sequential tones is higher if they are closer to each other on a continuous pitch scale.

The importance of pitch distance is reflected by the importance of the interval of a semitone in the resolution of dissonant or chromatic chords. In the words of Forte [1962, p. 11]:

> "The leading note is only one instance of an important melodic law, the *law of the half step.* According to this law the strongest, most binding progression is the half-step progression."

In a similarly vein, Goldman [1965, p. 84] emphasized

> "the fact that the semitone gravitates more strongly to its upper or lower neighbor than does the whole tone. The force of the dominant seventh chord is largely the result of the leading-tone's drive to the tonic and the fourth's drive to the third."

Pitch distance is difficult to define quantitatively in the case of pairs of complex sounds such as musical chords. In analytical perception, several tone sensations (mostly corresponding to actual notes) are noticed at the same time in each chord. The overall apparent pitch distance between two chords may be assumed to depend in some way on the pitch distances between all sequential pairs of tone sensations (Sect. 4.6.2).

3.3 Musical Pitch

3.3.1 Octave Equivalence

Octave equivalence is a musical universal [Hardwood 1976]. Musics from all over the world, including many non-Western musics (see e.g. Nettl 1956], exhibit octave equivalence: notes an octave apart are interpreted by musicians as similar or identical in harmonic function, and are given similar or identical names. Octave equivalence is "one of the most fundamental axioms of tonal music" [Forte 1962, p. 7], allowing the same scale degree (e.g. the tonic) to be played in all pitch registers, and by all pitched instruments and voice types.

Octave equivalence aids in composition, performance and improvisation. It reduces the number of different compositional possibilities, and so makes music easier to organize and remember. Pitch terminology is simpler when notes an octave apart are called by the same name. Notation would also be simplified if notes an octave apart looked similar on paper: surprisingly, this possibility is not used to its full advantage in Western notation [cf. Read 1987].

The sensory basis for octave equivalence has been described by Terhardt [1972, 1974a] and further examined in the experiments of the present study (Chap. 5). *Simultaneous* complex tones an octave apart blend so well that the result sounds almost like a single tone, because all harmonics of the upper tone coincide with even-numbered harmonics of the lower tone. Single complex tones contain prominent pitches an octave apart, so *sequential* tones an octave apart have many pitches in common and are therefore more or less "confusable" with each other [Terhardt 1983] (Sects. 5.4–6).

Of the above, the simultaneous effect implies a greater degree of "equivalence" [Deutsch 1982a]. Transposition of one of the voices of a chord through an octave can make very little difference to the sound of the chord. By contrast, transposing a note in a melody through an octave changes the character of the melody altogether, especially if the melody's contour is affected. So melodies in which both pitch class and contour are maintained but pitch register is not (i.e. tones appear in the wrong octave) are difficult to recognize and recall [Deutsch 1972b; Dowling and Hollombe 1977; Deutsch and Boulanger 1984].

In a psychoacoustical approach, the difference in consonance between the octave and the fifth is only one of degree [Terhardt 1976]. The singular and categorical importance of the octave in music is not immediately obvious from the pitch properties of complex tones and intervals. Perhaps the main reason why the octave, and not some other interval, became *the* equivalence interval in music is that *combinations* of octaves are consonant. The fifth could not be used as an "equivalence" interval in the way that the octave is, as combinations of fifths (ninths, thirteenths, etc.) are much less consonant than combinations of octaves.

3.3.2 The Chromatic Scale

A musical scale consists of discrete pitches or pitch categories separated by specific intervals. It is a framework within which tonal music may be composed, performed and improvised [Pressing 1978; Terhardt 1979b]. The historical development of musical scales is influenced, but not predetermined, by the physical properties of tones [Cazden 1954; Shirlaw 1957].

The chromatic scale has provided a basis for the composition of tonal music in the West since all twelve scale degrees came into common usage in the late Middle Ages. The heptatonic modes and diatonic scales may all be regarded as subsets of the chromatic scale.

The chromatic scale need not be equally tempered [Rasch 1983]: any standard heptatonic scale with provision for five extra notes per octave, so that the resultant twelve notes are roughly equally spaced and every note has a reasonably tuned upper and lower perfect fifth, may be regarded as chromatic. In modern practice, the most common or average tuning of the chromatic scale is that of *equal temperament*, in which each semitone covers the same distance on a frequency level (or log frequency) scale. The principle of twelve-tone equal temperament was first clearly articulated in the Europe by Mersenne in 1635 [Apel 1970]; its invention over a thousand years earlier by the Chinese had been hidden from the West by the language barrier [McClain 1979]. Bach promoted equal temperament in keyboard instruments with the publication of the two parts of his *Wohltemperiertes Klavier* in 1722 and 1744.

The chromatic scale is cross-cultural, but not universal. It was in use in ancient China [McClain 1979] and later evolved independently in India and Persia [Burns and Ward 1982], before emerging in the West. More recently, it has spread from the West to other cultures [Nettl 1986], although largely for political and social (rather than intrinsically musical) reasons. These observations suggest that the chromatic scale has a sensory basis – that its development both in the West and elsewhere was influenced by universal sensory properties of complex tones. This conclusion is not supported by simplistic treatments based on the harmonic series (Sect. 1.2.3): more sophisticated musical and psychoacoustical explanations are required.

The chromatic scale may be generated by successively traversing rising fifth intervals in one direction and octaves in the other (the cycle of fifths). After a total of twelve fifths and seven octaves, a coincidence occurs. If these intervals are exactly tuned to frequency ratios of 3 : 2 and 2 : 1 respectively, the coincidence is not exact: the start and end points are 0.24 semitone (one "Pythagorean comma") apart. However, this distance is negligible by comparison to the cumulative uncertainties in the sizes of the twelve fifths and seven octaves, given that the sizes of fifths and octaves are generally uncertain by 0.05–0.1 semitone, even in excellent musical performances [Sundberg 1982]. In musical practice, then, the cycle of fifths is closed and complete.

The major-third interval has played an important harmonic role in Western music since the sixteenth century (earlier in England). The lower note of the major third, like that of the perfect fifth, often corresponds to the root of a chord. According to Terhardt, this is due to familiarity with the pitch pattern of complex tones: the perfect fifth occurs between the 2nd and 3rd harmonics, the major third occurs between the 4th and 5th harmonics, and the 2nd and 4th harmonics are octave-equivalent to the fundamental. Taking the harmonic role of the major third into account, the chromatic scale in diatonic music may also be generated by three intervals: the octave, the perfect fifth and the major
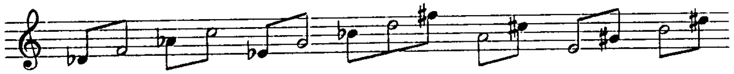
Fig. 3.1. Generation of the chromatic scale by combining octaves, fifths and major thirds. *Open noteheads*: standard heptatonic scale (generated by fifths and octaves). *Full noteheads:* non-heptatonic (chromatic) notes

third [Schoenberg 1911; Shirlaw 1957; Longuet-Higgins 1979]. In this approach, the standard heptatonic scale (the basis of the medieval modes, later specified by conventional key signatures) may be generated by octaves and fifths, while other, non-heptatonic notes lie at major-third intervals away from heptatonic notes (Fig. 3.1).

This derivation of the chromatic scale explains only the *harmonic* function of each non-heptatonic note, in a way which corresponds to its enharmonic spelling. *Melodic* functions and enharmonic spelling depend primarily on semitone relationships, i.e. pitch distance (e.g. C# → D, D♭ → C). The enharmonic spellings G♭ and A# are absent from the figure, as they have no harmonic function in C major or A minor. These notes may nevertheless have melodic functions, as chromatic auxiliaries to F and B respectively.

Twelve-note equal temperament approximates the small-number frequency ratios of musical intervals (especially the fifth) remarkably closely. Other equal temperaments with this property include (amongst others) 19-, 31- and 53-fold subdivisions of the octave [Yasser 1929; Fokker 1966; Pikler 1966; Hoerner 1976; De Klerk 1979; Yunik and Swift 1980; see also Balzano 1982). None of these has succeeded in materializing into new harmonic styles. The following psychoacoustical analysis suggests that this failure is due to more than just musical conservatism.

Harmonics of a complex tone which are mistuned by up to half a semitone are still perceived to belong to the tone; harmonics mistuned by a whole semitone or more are not [Moore et al. 1985]. This may be understood in terms of familiarity with the spectral pitch patterns of complex tones. The exact sizes of the intervals between spectral pitches (octave, fifth, fourth, ...) vary due to pitch shifts [Terhardt 1979a] and, sometimes, due to physical deviations from harmonicity (e.g. in pianos [Schuck and Young 1943]). These variations must be accommodated by the auditory system if complex tone sources are to be consistently recognized. Given that the root of a chord is a virtual pitch associated with a harmonic pattern of spectral pitches from different musical notes [Terhardt 1974a] (Sect. 3.4.2), division of the musical octave into 19 or more parts would allow adjacent scale steps to contribute to the same root. Thus, adjacent scale steps would no longer have the distinct harmonic functions that they have in the twelve-tone system.

The above argument ignores the role of roughness, and hence of exact frequency ratios, in harmony (Sects. 1.2.4, 3.2.2). In Indian *sitar* music, roughness appears to have a strong influence on intonation, as melody tones and their harmonics beat with the harmonics of the sustained drone. This may be why interval gradations (intonations?) of less than a semitone have structural significance in Indian music theory (although this is not necessarily the case in practice, see Jhairazboy and Stone [1963]).

Musical scales normally have less than twelve notes per octave, perhaps due to limitations of perceptual information capacity [Miller 1956]. Commonly used subsets of the chromatic scale are the standard pentatonic scale (e.g. in China) and standard heptatonic scale (e.g. in the West and in India). They are common because of their harmonic coherence (they describe unbroken sections of the cycle of fifths); because they each contain only two interval sizes and no adjacent half steps [Pressing 1978]; and because of their asymmetry, which prevents them from being mapped onto themselves by transposition within the octave [Balzano 1980; Cross et al. 1985].

World musics use a wide variety of other musical scales which may be regarded as subsets of the chromatic scale (when defined as a set of pitch *categories*, not exact pitches). In addition, there are scales which do not easily fit into the chromatic mould, e.g. in Indian, Thai and Indonesian music. Such scales may be based on quite different sensory and cultural premises from those which underlie the standard pentatonic and heptatonic scales.

3.3.3 Intonation

The three major theoretical intonations (tunings) in Western music are *Pythagorean* (based on frequency ratios containing only powers of two and three), *just* or *pure* (in which powers of five also appear) and *equal temperament* (in which semitones correspond to equal frequency ratios). In Pythagorean temperament, major intervals are wider, and minor intervals narrower, than in equal temperament; the reverse is the case for just intonation. For example, an equally tempered major third spans exactly 4 semitones, a Pythagorean major third (with a frequency ratio of 81:64) spans 4.08 semitones, and a just major third (5:4) spans 3.86 semitones.

The differences between these tunings are generally imperceptible to non-musicians [Roberts 1986] and are smaller than typical variations in intonation even in excellent performances [Seashore 1938; Burns and Ward 1982]. There is nevertheless an extensive literature on the subject of intonation, perhaps because many researchers believe (at least at the outset of their research) that the musical meaning of an interval is somehow embodied by its frequency ratio (Sect. 1.2.2), and that the "actual" frequency ratios of musical intervals revealed by intonation research might lead to a better understanding of harmony and tonality.

The optimal intonation for sustained harmony (simultaneous tones) is often found in experimental studies to lie on the "just" side of equal temperament [Norden 1936; O'Keefe 1975; Roberts and Matthews 1985], although sometimes equal temperament is preferred to just [e.g. Roberts 1986]. Just temperament causes harmonics to coincide, minimizing beats [Helmholtz 1863] and thereby roughness [Terhardt 1968b].

The musical instrument which is perhaps most susceptible to the dissonant effect of beats produced by tempered tuning is the pipe organ. The introduction of equal temperament in pipe organs was resisted until long after it had become the norm in other keyboard instruments such as the piano; in England, for example, it was not widely accepted until the nineteenth century [Williams 1968].

Just intonation only applies when individual complex tones are exactly harmonic. Piano tones, for example, are not exactly harmonic: the frequencies of their pure tone components are slightly stretched relative to a harmonic series [Ward and Martin 1961]. Further, the "exact frequency ratio" principle only applies when there is no vibrato. Even in barbershop singing, where vibrato is not normally noticeable, variations in intonation of 0.1 – 0.2 semitones may fail to give rise to beats and roughness: the periods of time over which individual voices are periodic, and hence exactly harmonic, are limited [Hagerman and Sundberg 1980].

Intonation of *melodies* in Western music tends to the Pythagorean side of equal temperament, especially in violin playing [Greene 1937; Nickerson 1949; Cazden 1954; Fransson et al. 1974). Pythagorean intonation emphasizes the difference between major and minor intervals, and tends to "anticipate" resolution of dissonance through semitone steps. Since beats play no role in the perception of sequential tones (Sect. 1.2.4), an explanation of quasi-Pythagorean intonation of melodies in terms of frequency ratios (such as 81 : 64 for the major third) is far-fetched [cf. Ward 1962].

The distinction between melodic and harmonic intonation is by no means clear cut. In one study [Shackford 1961, 1962] measurements were made of frequency ratios between tones in string trio performances. No systematic difference was found between the intonation of simultaneous and sequential tone pairs. Similarly, in a test involving intonation of both melodic and harmonic intervals by 48 wind players, Duke [1985] found that, overall, there was no consistent tendency to play either sharper or flatter than equal temperament.

The intonation of some scales and melodies, e.g. in Swedish folk music [Tjernlund et al. 1972], shows the opposite tendency to that normally found in the melodies of classical Western music: major intervals are played smaller than equal and just temperament, and minor bigger. In this way, the spacing of the seven tones of the standard heptatonic scale becomes more uniform relative to their chromatic (equally tempered) spacing. Roughly equally spaced scales (including 12-note equal temperament) have the advantage that expressive and random deviations from the centre of a scale step (pitch category) may more easily be accommodated without risking confusion with neighbouring steps.

Experiments on intonation are complicated by the effect of *pitch shift* [Terhardt 1972, 1974a]. The pitch of a pure tone or a pure tone component depends on its intensity [Stevens 1935] and on the simultaneous presence of other sounds which mask the tone [Egan and Meyer 1950; Webster et al. 1952; Terhardt and Fastl 1971]. Both these effects are related to masking patterns, i.e. to the shape of the masked threshold [Webster and Schubert 1954; Hesse 1987]. Pitch shifts of pure tones can amount to as much as a semitone, or even more than a semitone. The pitch of a complex tone or complex tone component (e.g. in a musical chord) is more stable with respect to changes in level and masking [Lewis and Cowan 1936; Schouten 1940; Stoll 1985].

The intonation of sequential complex tones in music is affected not so much by the pitch shift of the main tone sensation evoked by each tone as by shifts in the sizes of the *intervals* between the main tone sensation and subsidiary tone sensations, especially those at the upper and lower octaves. These intervals are slightly "stretched" due to pitch shifts [Terhardt 1970, 1972, 1974a]. Given that sequential tones an octave apart are tuned by lining up near-coincident pitches (i.e. maximizing pitch commonality), this explains why the octave interval between sequential complex tones in music tends to be tuned to a frequency ratio slightly larger than 2 : 1 [Corso, 1954; Terhardt 1971; Sundberg and Linqvist 1973; Terhardt and Zick 1975]. Octave stretching is also found between sequential pure tones [Ward 1954; Walliser 1969a; Terhardt 1970].

According to the above authors, perceptually optimal octave intervals between sequential pure and complex tones correspond to frequency level differences of 12.1 – 12.4 semitones, depending on pitch register. A stretch of about 0.15 semitones is typical between sequential tones in the middle pitch register [Fransson et al. 1974]. Intonation of simultaneous complex tones normally involves a compromise between the requirements of *pitch*, by which octaves tend to be stretched, and *roughness*, by which octaves are normally "pure", i.e. not stretched [Terhardt 1974a, 1976]. On average, it may be assumed that musical intervals are stretched by about 0.1 semitone per octave.

Because of octave stretch, octave-periodic scales and intonations (including equal temperament) tend to be slightly stretched throughout. This is especially true in the case of the piano, whose tones have stretched harmonics [Ward and Martin 1961]. Hence the results of experiments in which performed sequential intervals are found to be tuned consistently larger than equal temperament are explicable either in terms of familiarity with piano tuning (a cultural explanation) or in terms of optimization of pitch commonality (a sensory explanation). On instruments without octave stretching (e.g. electronic and pipe organs), very high tones can sound flat by comparison to very low tones, so much so that large intervals can be miscategorized: for a listener used to a stretch of 0.1 semitone per octave, an interval of five octaves could be heard as four octaves and a major seventh. The extent of this effect depends on the individual pitch shifts of the tones, which in turn depend on their level and spectral composition.

Small intervals are sometimes found to be tuned consistently smaller than equal temperament. This effect has been reported in the case of minor seconds [Ward 1970] and seconds and thirds [Rakowski 1976]. It presumably enhaces pitch *proximity*, without affecting chromatic pitch and interval categories.

3.4 Tonality

3.4.1 Introduction

In a general definition, the concept of tonality "embrances the main structural components of the tonal composition; within it are expressed the highly diversified events and multiple relationships which form a . . . musical unity" [Forte 1962, p. 79]. In this view, tonality includes all aspects of consonance described in the previous sections.

More specifically, tonality may be defined as the perceptual organization of a passage of music around a tonic or key. Music theory is not very specific about what "tonic" and "key" mean, however. Berry [1976, p. 22] regards the tonic as "a specific pitch-class-complex of resolution", i.e. as a pitch class, dyad class, chord class or scale. Similarly, Wilding-White [1961] regards tonality as a function of a pitch set, which could be a tone, a chord or a scale.

The kind of musical element around which tonality revolves depends on the kinds of musical element making up the passage of music in question. The tonic of a *melody* may be regarded as a specific pitch or note [Erickson 1984]. More often, octave equivalence is assumed, and the tonic of a melody is regarded as a pitch class (e.g. "C" instead of "middle C"). The tonic of a *chord progression* may be regarded as a specific (root position) chord, often appearing at the start and/or the end of the progression. Given octave equivalence, however, it is more common to specify the tonic as a *chord class*: as a set of pitch classes (e.g. CEG) rather than a set of actual pitches (e.g. $C_3G_3C_4E_4$). Similarly, the root of a chord is normally specified not as an actual pitch but as a pitch class.

Diatonic tonality arose in Western music soon after composers began to treat the major and minor triads as consonant and harmonically functional (around the sixteenth century). This suggests a link between the tonic of a diatonic scale and the root of a chord.

3.4.2 The Root of a Chord

The root of a chord may be defined most simply as the note or pitch after which the chord is named in Western music terminology and theory. According to this definition, the root of C–E–G is C simply because this chord is called "C major". In music theory, the root may be regarded as that single bass note which most satisfactorily represents the function of a chord in a harmonic progression. If the chord C–E–G in a chord sequence were to be replaced by a single bass note, the note which would least disturb the harmonic progression would be C. Roots are normally understood to be independent of the voicing (including inversion) and relative accentuation of the notes of a chord.

A chord's root determines its diatonic function [cf. Riemann 1893]. The concept of the root is therefore essential for the understanding of Western harmony. Yet despite centuries of music-theoretical development, there still exists no widely accepted theory of the root's nature and origins.

Rameau [1721] compared the notes of a musical chord with the harmonics of a single complex tone (such as a typical musical tone) and hypothesized a correspondence between the root of a chord and the fundamental of a harmonic series of frequencies. During the nineteenth century, Rameau's approach was regarded as the basis for a "vertical" view of harmony, complementing the "horizontal" view based on the diatonic scale and figured bass [Watson 1982]. Theorists such as Bruckner tended to place more emphasis on the "vertical" aspect of harmony, asserting that not only triads and sevenths but also ninths were explicable in terms of Rameau's theory. Others [e.g. Schenker 1906] emphasized the "horizontal" aspect, claiming that only triads were derived from the harmonic series ("Nature") while sevenths and ninths were produced by melodic or horizontal elaboration of triads (by the "Artist").

Music theories based on frequency ratios and the harmonic series have never been able to explain satisfactorily the nature and root of the second most common chord in mainstream Western music, the minor triad (Sect. 1.2.3). The most commonly advocated frequency ratio representation of the minor triad, 10 : 12 : 15, suggests that the root of the chord lies a major third below the conventional root, i.e. that the triad is part of a major seventh chord (8 : 10 : 12 : 15). Furthermore, the theory suggests that the minor triad is considerably more dissonant than the major; in musical practice, however, these two triads are about equally consonant.

A commonly proposed solution to this theoretical dilemma was the theory of *harmonic dualism*. Theorists such as Zarlino, Rameau, Tartini, Hauptmann and Riemann advocated that "since major and minor triads were supposed to produce opposite psychological effects on the listener they must therefore be based on opposite principles" [Jorgenson 1963, p. 31]. Accordingly, they proposed that the major triad was based on harmonics 1, 3 and 5 (or 4, 5 and 6), while the minor triad was based on the corresponding *subharmonics* (or *harmonic undertones*). Thus the single note C generated both the C major triad (through its overtones) and the F minor triad (through is undertones).

The theory of undertones has some serious faults. It still fails to predict the root of the minor triad – the root of F minor isn't C! And the underlying theory itself is implausible: subharmonics are not found in naturally occurring sounds.

According to the *phonic theory of reinforced resonance* of Oettingen and Helmholtz [Jorgenson 1963; Christensen 1987], the tones of the minor triad each have one harmonic in common. For example, G_6 is the 6th harmonic of C_4, the 5th of $E\flat_4$, and the 4th of G_4. Again, this does not explain why the root of C minor is C. Nor, for that matter, do combination tones (Sect. 1.2.5).

The music-theoretical root of the minor triad may be correctly predicted by treating it as a distorted form of the major – a major triad with a lowered third, or with a different mediant dividing the fifth. Advocates of this idea included Rameau, Hindemith, Tovey and Schenker [Jorgenson 1963]. The trouble with this approach is (again) the implication that the minor triad is categorically inferior to or less fundamental than the major.

In a psychoacoustical approach, all chords, including the major and minor triads, are initially treated simply as tone simultaneities, i.e. as sounds composed of tones which happen to occur at the same time. The ability to perceive a chord in this way is assumed to be acquired by familiarization with complex tones in the auditory environment.

According to Terhardt [1974a, 1982], the root of a chord is a virtual pitch, i.e. a complex tone sensation. This observation alone is not very useful, as there are virtual pitches at *all* the notes of the chord. The root is different in that the spectral pitches in its harmonic pitch pattern arise from more than one complex tone (note). In other words, the root is the implied fundamental of a group of pure tone components belonging to different complex tones. For example, the root A of the chord A–C–E may be produced by pure tone components at A (from the note A), at E (from the notes A, C or E), at C# (from the note A), at G (from A or C) and at B (from A or E). In a given chord, there are generally several complex tone sensations of this kind: the root is *ambiguous*.

The perception of a chord as a tone simultaneity is appropriately modelled by psychoacoustical methods based on universal aspects of hearing (Chap. 4). This enables specific predictions to be made concerning the roots of musical chords. The music-theoretical root of any chord class in mainstream harmony theory (including the minor triad) turns out to be the pitch class with the highest calculated salience (Sect. 6.1.6); and the root of a chord in a specific voicing generally corresponds to the most prominent pitch class in the bass region (Sect. 6.1.5).

The pitch pattern of a complex tone may be recognized even if parts of the pattern are missing, or extra elements are added (Sect. 2.3.3). Similarly, in a psychoacoustical approach, the root of a chord may still be perceived if notes corresponding to harmonics of the root are missing, or if notes not corresponding to harmonics are added. For example, the root of the C major triad is weakened, but not changed, if the note E (which corresponds to the fifth harmonic of C) is replaced by E♭ (which doesn't correspond to any normally audible harmonic of C); the root (C) is maintained by the strong root implication of the fifth C-G.

A mere tone simultaneity has limited "meaning" in a musical context. To be musically meaningful, a chord must be perceived not only as a tone simultaneity but also as a *musical element*. The ability to perceive a chord in this way is acquired by cultural conditioning (Sect. 3.1.1). It appears to involve an awareness of the harmonic nature of pitch intervals between the noticed tone sensations of the chord, and of the relationship between noticed tone sensations and the root of the chord.

3.4.3 The Tonic of a Scale

Melodies from all over the world exhibit tonality in that they centre on a particular "tonic" [Erickson 1982]. The most universal way of "tonicizing" a pitch is by direct emphasis: repetition, duration and accentuation. Another apparently universal effect is that the tonal centre tends to be near the centre of the pitch range or tessitura [Erickson 1984], so that the average pitch distance between the tonic and other scale notes is as low as possible. Both these sensory affects influence the tonics of world scales, including the "finals" of church modes in Western music of the Middle Ages.

By contrast, the tonic of a *diatonic* scale is analogous to the root of an elaborated broken chord. There is no sensory basis for the root of a broken chord: pure tones must be simultaneous, or very close to simultaneous, to form complex tone sensations [Hall and Peters 1981]. Yet both children and adults with experience of Western harmony extract the tonal hierarchy of a major scale from the sound of a broken major triad [Cuddy and Badertscher 1987]. This suggests that the perception of the root of a broken chord can develop by exposure to unbroken chords with clear roots, such as major triads, minor triads and major-minor ("dominant") sevenths. Such chords were first regularly heard in Western music in the late Middle Ages and Renaissance; diatonic tonality came into being soon after. Another musical development just preceding the emergence of diatonic tonality was the use of perfect or V-I cadences. As these began to influence tonicization, hierarchical nature of diatonic tonality began to depend on root relationships between notes a fifth apart [Schenker 1906].

The emergence of diatonic tonality reduced the number of theoretically available scales from seven (modes starting on different degrees of the standard heptatonic scale) to two (the major and minor scales). In practice, the change was not as great as implied by these figures, as some of the modes of Medieval theory were rarely used, and the melodic minor scale had different ascending and descending forms. But it certainly was significant: it paved the way for the drama of the classical sonata form, and the expressive possibilities of romantic harmony.

The tonic of a diatonic scale turns out to be the *root of the chord formed by scale notes not involved in tritone relationships* [Parncutt 1987a]. Tritone relationships eliminate B and F from C major, and B, F, D and A*b* from C harmonic minor. The remaining notes (C, D, E, G and A in C major, and C, E*b* and G in C harmonic minor) are tonally more stable than the eliminated notes, which in turn are more stable than non-scale (chromatic) notes. The root of the chord C–D–E–G–A is ambiguous: its most important root is C (hence the chord symbol C^{6add9}), its subsidiary root, A (A^{m7add4}). This explains why the tonic of the C major scale is usually C, and occasionally A (in the case of the descending melodic minor scale on A). The tonic of the chord C–E*b*–G is clearly C. The tonic is also the lowest note in a chain of fifths formed by the more stable scale degrees: the tonic of C major is the lowest note in the chain C–G–D–A–E, and the tonic of C minor is the lowest note of C–G (the only fifth among the notes C, E*b* and G).

The above procedure is based on sensory aspects of musical tones; but the application of the procedure to the determination of tonics is clearly specific to the Western diatonicism. In the words (and italics) of Cazden [1954, p. 290], "The natural potentials of tone act as a *limiting condition* of the art of music, but not as a *determining cause* of musical relations." The tritone is the least tonal interval class (not counting the major seventh, which is the inversion of the minor second): it combines low harmonic relationship with low pitch proximity. The avoidance of tritones dates from Medieval times; note, however, that it was still possible in the Middle Ages for a tritone to be included above the final (e.g. in the Lydian mode FGABCDEF). The importance of chord roots for diatonic tonality dates to the Renaissance, when major and minor triads began to be used as musical elements (rather than as more or less coincidental tone simultaneities in Medieval counterpoint); and the importance of the perfect fifth interval dates to the emergence of the V-I cadence at roughly the same time. As explained in the last section, the root of a chord has a sensory basis, and the fifth is the interval which has the strongest influence on the root of a chord.

Musically useful subsets of the chromatic scale seldom have two semitones in a row, or "adjacent half-steps" [Pressing 1978]. The only single chromatic alterations to the scale CDEFGAB which do not produce adjacent half-steps are C#, Eb, F#, G#, Ab and Bb. According to the above procedure, the tonic of the scale C# DEFGAB is the root of the chord D–E–A. The most likely root of this chord is D, as both E and A correspond to harmonics of D; A is a subsidiary root, as D does not correspond to a harmonic of A. Applying this procedure to all the above chromatic alterations produces the following predicted tonics: are D, C, G, A, C and F (respectively). Thus, only C#, F#, G# and Bb suggest modulations to new keys, while Eb and Ab suggest instead a shift toward the tonic minor.

In this light, the harmonic minor scale may be regarded as a chromatic alteration of the major in which only those notes which are not crucial for the determination of the tonic have been flattened. Conversely, the major scale may be regarded as a chromatic alteration of the harmonic minor, in which only those notes which are not crucial for the determination of the tonic have been sharpened. Major and minor scales on the same tonic may thus be regarded as belonging to one and the same "major/minor" tonality [Goldman 1965].

In the above examples of chromatic alterations to the standard heptatonic scale, the notes C#, F# and G# function as *leading notes*. Leading notes have an important *melodic* function: they lie at semitone intervals away from tonally strong pitches (here, D, G and A respectively). In the light of the proposed procedure for the determination of the tonic of a scale, rising leading notes (such as these three) also have a *harmonic* function: their tritone relationship with the subdominant (scale degree IV, the lowest note in a chain of fifths formed by *all* scale notes) prevents the subdominant from becoming the tonic. Falling leading notes have no such harmonic function; this may explain why they are much less common in diatonic music than rising leading notes.

The fifth C–G in the key of C major/minor is "authentic" in the sense that its root corresponds to the tonic. The fifth F–C is "plagal": its upper note (not its root) is the tonic [Shirlaw 1957]. The status of the F–C fifth may be changed to "authentic" by replacing the scale note B by Bb. This eliminates the tritone relationship against the F (making it a tonic candidate) and creates a new tritone between Bb and E (preventing Bb, the lowest element in the new chain of fifths, from being the tonic).

Sharps relative to a prevailing key signature lie at major-third intervals above prevailing scale notes; flats, at major-third intervals below scale notes (Fig. 3.1, Sect. 3.3.2). The root-implying property of the major-third interval means that flats relative to a prevailing key signature are more likely to act as roots of chords than sharps. In other words, flats are more *salient* than sharps. This explains the finding of Thompson and Cuddy [1986] that modulation to flat keys is more noticeable than modulation to sharp keys, i.e. adding *n* flats to (or removing *n* sharps from) a key signature produces a more distant modulation than adding *n* sharps (or removing *n* flats).

The diatonic scales may be regarded as the main musically conditioned aspect of the perception of Western melodies (just as streaming is its main sensory basis, Sect. 2.4.6). The diatonic scales are so familiar that it is useful to model melodic perception by matching the pitches of a melody to a rigid *diatonic template* [Jordan and Shepard 1987] (Sect. 2.4.3). This approach is consistent with the following experimental data: listeners quickly recognize fragments of diatonic scales in melodies, and expect subsequent tones to belong to the same scale [Francès 1972; Deutsch and Feroe 1981]; well-structured melodies (i.e. tonal melodies) are remembered more easily than atonal melodies [Dewar et al. 1977]; and nondiatonic tones in diatonic melodies are recognized less often than diatonic tones [Krumhansl 1979].

The rarest interval in the major scale is the tritone: it occurs only between scale degrees VII and IV. This allows the major scale (and therefore its perceptual "template") to be specified fully and unambiguously by two notes a tritone apart and one other note [Browne 1981]. This set of three pitches is therefore sufficient for the recognition of the scale and its tonic [Butler and Brown 1984]. For example, the notes B, F and any other "white note" (C, D, E, G, A) are only diatonic in the key of C major. The notes B, F and a "black note" could only belong to F#/Gb major. These considerations do not apply for the harmonic minor scale, as it includes two tritones (II–VI as well as VII–IV).

3.4.4 Major/Minor and Emotion

The major/minor distinction has been central to the emotional meaning of diatonic music at least since the Renaissance [Wienpahl 1959]. Other things being equal (tempo, timbre, texture, etc.), music in major keys is supposedly appropriate for the expression of positive emotions (happiness, brightness, confidence, victory, ...) and music in minor keys expresses negative emotions (sadness, darkness, defeat, tragedy). There is no doubt that Western listeners are sensitive to the emotional connotations of major and minor: recent research has gone so far as to demonstrate a connection between preference for minor tonality and oral dependency, i.e. a longing to be supported and nurtured [Juni et al. 1987]. However, the *origins* of the emotional connotations of major/minor are by no means clear.

Almost by definition, the "majorness" or "minorness" of a piece of music is determined by the "majorness" of "minorness" of intervals above the tonic. The most important interval in this regard is the third; the second most important, the sixth. It follows from this that the tonal structure of a piece needs to be firmly established before the piece can take on an unambiguously "major" or "minor" flavour. As yet, experimental studies on the emotional meaning of intervals [Maher 1980] and on the perceived major/minorness of triads [Crowder 1985] have not involved clear, unambiguous tonal contexts, and so have thrown little light on the emotional meaning of specific tonal progressions according to music theorists [e.g. Cooke 1959].

Musical styles develop under the continuous influence of sensory (universal) aspects of musical sounds (Sect. 3.1.1). The following sensory aspects are relevant for their emotional meaning: (i) the emotional meaning of pitch patterns in speech, and hence in melodies [Lieberman and Michaels 1962; Nilsonne and Sundberg 1984], and the emotional meaning of subtle pitch changes in individual musical notes [cf. Clynes 1977; Senju and Ohgushi 1987]; (ii) the presence of the pitch pattern of the major triad in the pure tone sensations of a single complex tone; and (iii) the dissonance and root ambiguity (multiplicity) of the minor triad, which is slightly greater than that of the major (Sects. 6.1.3, 6.1.4). One or more of the above three factors could contribute indirectly to the "happy/sad" (or "equilibrium/tension", or "pleasure/unpleasure") aspects of major/minor in Western music.

(i) According to Forte [1962, p. 11], "If both whole-step and half-step progressions are available from a given note the half-step progression will be preferred." This explains the melodic function of the leading notes of diatonic scales. It also explains the tendency of the third step of the major scale to rise to the fourth, and of the third and sixth steps of the minor scale to fall to the second and fifth respectively. This suggests that the emotional meaning of major/minor may be associated with rising and falling of pitch.

A problem with this approach to the understanding of major/minor is that the relationship between the intonation and emotional meaning of speech may be too complex to permit such simple generalizations. Intonation of statements, question, imperatives and exclamations are similar in different languages [Tonkova-Yampol'skaya 1973] but there are also important differences, even within languages (dialects and regional variations).

(ii) The "major" pitch pattern of the pure tone sensations of a complex tone could contribute to the emotional meaning of major/minor via prenatal conditioning (Sect. 3.1.2). If isolated speech vowels heard by the foetus sound like musical chords at some stage of prenatal development, major triads heard after birth may be associated more strongly with the security of prenatal experience than minor triads or other chords.

(iii) The dissonance and root ambiguity of the minor triad may have gradually contributed to the emotional meaning of major/minor during the historical emergence and development of diatonic tonality. Major triads more clearly and positively define the tonic of a passage than minor triads [Cuddy and Lyons 1981]; the "sadness" of minorness may be associated with a feeling of uncertainty or indecisiveness about the position of the tonic in minor keys.

The emotional characteristics of the Medieval church modes and ancient Greek modes were rather different from those of the diatonic scales [Révész 1953]. This is not surprising: these modes were used only for melody and polyphony, in a weak tonal context. They were not used for triadic harmony. The above three hypothetical sensory contributions to the emotional meanings of intervals apply only to chord progressions, or to melodies and contrapuntal passages which imply chord progressions.

3.4.5 Chord Progressions

The consonance of a chord progression depends on: the consonances of individual chords in the progression (Sect. 3.2.2); the consonances of pairs of chords, i.e. how well they are perceived to go with each other (Sect. 3.2.3); the unifying effect of melodic streaming (Sect. 2.4.6); and the strength of the tonal structure (tonality) of the progression (Sect. 3.4).

These contributions to the consonance of a progression are normally positively correlated with each other: progressions with strong tonal structure often contain consonant chords (such as the tonic), consonant chord relationships (such as those between the tonic and other chords), and coherent melodic lines. Thus, Baroque chord progressions are usually quite consonant, and twelve-tone chord progressions quite dissonant, on all four counts. Other combinations of the above factors are common: some impressionist styles have consonant chords but dissonant chord relationships and weak melodic and tonal structure; and some jazz styles have dissonant chords, but consonant chord relationships and strong melodic and tonal structure.

Chord progressions without passing notes are *homorhythmic*: all parts move in synchrony. In *polyphonic* (or contrapuntal) music, by contrast, the parts move in independent rhythms. The model to be developed in Chap. 4 concerns the perception of simultaneities and of pitch relationship between simultaneities, i.e. of homorhythmic music.

Ideally, perceptual grouping (Sect. 2.3.3) in homorhythmic music is entirely vertical – tone sensations are grouped only because they are close in time. In polyphonic music, perceptual grouping is ideally only horizontal, due to grouping in pitch and timbre (streaming). In reality, perceptual grouping of both kinds occurs in both styles, but vertical grouping is more pronounced in the case of homorhythmic music, and horizontal grouping is more pronounced in polyphonic music. The present study is mainly concerned with homorhythmic music, in which vertical grouping is assumed to predominate.

The term "chord progression" in this study refers to a limited number of chords, perceived as a group – in a similar way that the tones of a melodic phrase are grouped. All parts of a chord progression are assumed to be simultaneously accessible to the attention of the listener just after the progression has been completed. The duration of a chord progression is therefore limited to the maximum duration of auditory sensory memory, i.e. 2–10 s (Sect. 2.2.3). If the chords of the progression follow each other at a maximum rate of about $4\,\mathrm{s}^{-1}$ (the rate at which consonants are typically produced in speech), and the duration of short-term memory is 5 s, then the maximum number of chords a progression may contain is $4 \times 5 = 20$. The progressions analyzed in this study (Sect. 6.3.1) comprise no more than ten chords and may therefore be assumed to be perceptible as wholes.