12 Pitch

In this chapter, we will deal with several attributes of sounds grossly classified as **pitch**, along with several associated topics. Like "loudness," the word "pitch" denotes a perception with which we are all familiar. Pitch is generally described as the psychological correlate of frequency, such that high-frequency tones are heard as being "high" in pitch and low frequencies are associated with "low" pitches (ANSI, 2004). However, we saw in Chapter 9 that not all changes in frequency are perceptible. Instead, a certain amount of frequency change is needed before the difference limen (DL) is reached. In other words, the frequency difference between two tones must be at least equal to the DL before they are heard as being different in pitch. Moreover, we shall see that pitch does not follow frequency in a simple, one-to-one manner along a monotonic scale from low to high. Instead, the perception of pitch appears to be multifaceted, and it may be that there are various kinds of pitch. In addition, although we know that pitch involves both the place and temporal mechanisms of frequency coding discussed in earlier chapters, the precise interaction of frequency and temporal coding is not fully resolved.

Pitch can be expressed in a variety of ways. Perhaps the most common approach is to express pitch and pitch relationships in terms of musical notes and intervals. Another very useful method is to ask listeners to find the best match between the pitch of a certain sound and the pitches of various pure tones, and then to express the pitch of that sound in terms of the frequency of matched pure tone. For example, if a complex periodic sound is matched to a 500-Hz pure tone, then that sound has a pitch of 500 Hz. Similarly, if the pitch of a certain noise is matched to a 1700-Hz tone, then the noise is said to have a pitch of 1700 Hz. Another way to approach pitch is to construct a pitch scale analogous to the sone scale of loudness (Chap. 11), which will be our first topic.

MEL SCALES OF PITCH

Stevens, Volkmann, and Newman (1937) asked listeners to adjust the frequency of a tone until its pitch was one-half that of another (standard) tone. The result was a scale in which pitch is expressed as a function of frequency. This scale was revised by Stevens and Volkmann (1940), whose subjects had to adjust the frequencies of five tones within a certain frequency range until they were separated by equal pitch intervals. In other words, their task was to make the distance in pitch between tones A and B equal to that between tones B and C, C and D, and so on. Stevens and Volkmann also repeated the earlier fractionalization experiment except that a 40-Hz tone that was arbitrarily assigned a pitch of zero. Based on direct estimates of the pitch remaining below 40 Hz and extrapolations, 20 Hz was identified as being the lowest perceptible pitch. This estimate agreed with Bekesy's (1960) observation that the lowest frequency yielding a sensation of pitch is approximately 20 Hz.

Stevens and Volkmann's (1940) revised **mel scale** is shown by the solid curve in Fig. 12.1a. On this graph, frequency in hertz is shown along the abscissa and pitch in units called **mels** is shown along the ordinate. The reference point on this scale is 1000 mels, which is defined as the pitch of a 1000-Hz tone presented at 40 phons. Doubling the pitch of a tone doubles the number of mels, and halving the pitch halves the number of mels. Thus, a tone that sounds twice as high as the 1000-mel reference tone would have a pitch of 2000 mels, while a tone that is half as high as the reference would have the pitch of 500 mels. The dotted curve shows what the relationship would look like *if* frequency and pitch were the same (if mels = hertz). Notice that frequency and pitch correspond only for low frequencies. However, the solid (mel scale) curve becomes considerably shallower than the dotted curve as frequency continues to rise. This reveals that pitch increases more slowly than frequency, so that the frequency range up to 16,000 Hz is focused down to a pitch range of only about 3300 mels. For example, notice that tripling the frequency from 1000 to 3000 Hz only doubles the pitch from 1000 to 2000 mels.

Other pitch scales have also been developed (e.g., Stevens et al., 1937; Beck and Shaw, 1962, 1963; Zwicker and Fastl, 1999), and it should not be surprising that the various scales are somewhat different from one another due to methodological and other variations. For example, the solid curve in Fig. 12.1b shows the mel scale formulated by Zwicker and Fastl (1999). This scale is based on ratio productions in which listeners adjusted the frequency of one tone to sound half as high as another tone. The reference point on the Zwicker and Fastl scale is 125 mels, which is the pitch of a 125-Hz tone. As in the upper frame, the dotted curve shows what the relationship would look like *if* pitch in mels was the same as frequency in hertz. Here, too, we see that pitch increases much slower than frequency, with the frequency range up to 16,000 Hz is compressed into a pitch range of just 2400 mels. For example, a 2:1 change in pitch from 1050 to 2100 mels involves increasing the frequency by a factor of more than 6:1, from 1300 to 8000 Hz.

There is reasonably good correspondence between pitch in mels, critical band intervals in barks (see Chap. 10), and distance along the basilar membrane (e.g., Stevens and Volkmann, 1940; Scharf, 1970; Zwicker and Fastl, 1999; Goldstein, 2000). Zwicker and Fastl (1999) suggested that 100 mels corresponds to one bark and a distance of approximately 1.3 mm along the cochlear partition. These relationships are illustrated in Fig. 12.2. Goldstein (2000) reported that the Stevens–Volkmann (1940) mel scale is a power function of the frequency–place map for the human cochlea (Greenwood, 1990).

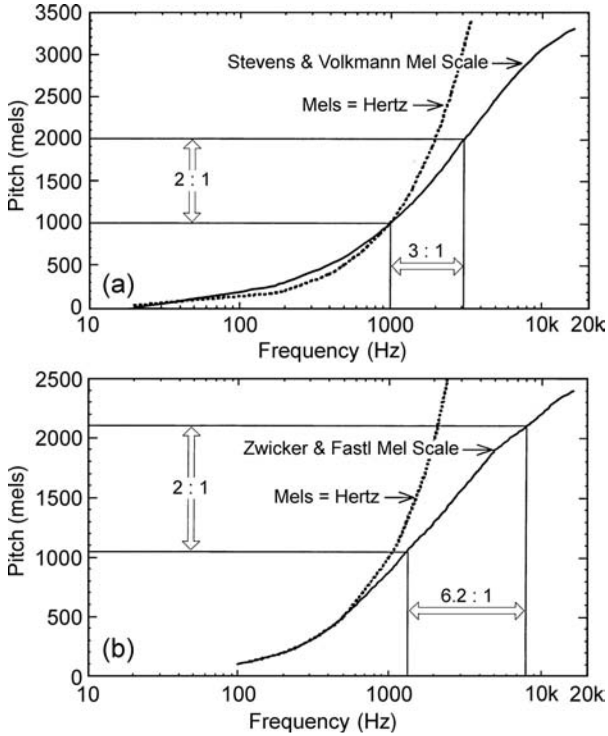

Figure 12.1 The relationship between frequency in hertz and pitch in mels based on the findings of (a) Stevens and Volkmann (1940) and (b) Zwicker and Fastl (1999). The dotted curves labeled Mels = Hertz illustrate what the relationship would look like if pitch in mels was the same as frequency in hertz.

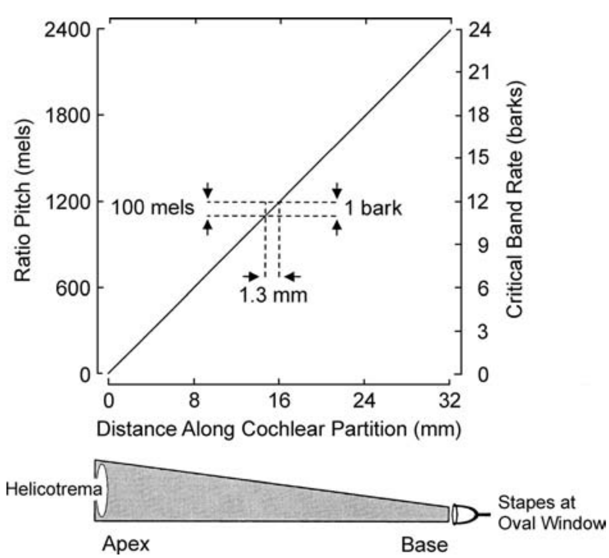

Figure 12.2 Illustration of the relationships between distance along the cochlear partition, pitch in mels, and critical band rate in barks. Notice the orientation of the basilar membrane. Based on scales by Zwicker and Fastl (1999).

BEATS, HARMONICS, AND COMBINATION TONES

The topics of audible beats, harmonics, and combination tones are often discussed in the context of how we perceive frequency. We will divert our attention to these phenomena early in this chapter not only because they are topics of interest in their own right, but also because they will be encountered as tools used to study various aspects of pitch perception in the discussion that follows.

We have seen in several contexts that a pure tone stimulus will result in a region of maximal displacement along the basilar membrane according to the place principle. Now, suppose that a second tone is added whose frequency (f_2) is slightly higher than that of the first sinusoid (f_1), a in Fig. 12.3. If the frequency difference between the two tones ($f_2 - f_1$) is small (say 3 Hz), then the two resulting excitation patterns along the cochlear partition will overlap considerably so that the two stimuli will be indistinguishable. However, the small frequency difference between the two tones will cause them to be in phase and out of phase cyclically in a manner that repeats itself at a rate equal to the frequency difference $f_2 - f_1$. Thus, a combination of a 1000-Hz tone and a 1003-Hz tone will be heard as a 1000-Hz tone that **beats**, or waxes and wanes in level, at a rate of three times per second. This perception of aural beats therefore reflects the limited frequency-resolving ability of the ear.

If the two tones are equal in level, then the resulting beats will alternate between maxima that are twice the level of the original tones and minima that are inaudible due to complete out-of-phase cancellation. Such beats are aptly called **best beats**. Tones that differ in level result in smaller maxima and incomplete cancellation. As one would expect, the closer the levels of the two tones, the louder the beats will sound.

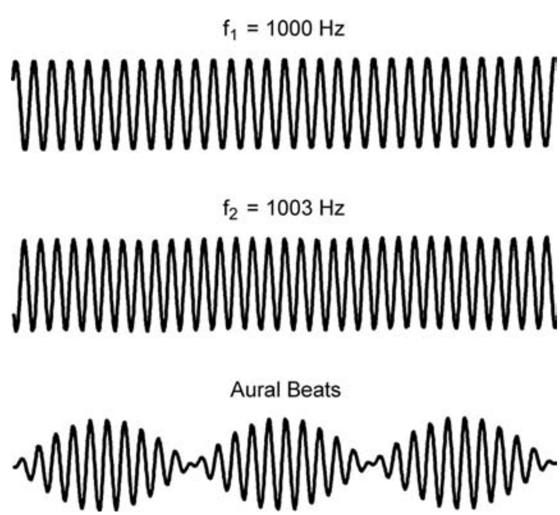

Figure 12.3 Tones of slightly different frequency, f_1 and f_2, result in beats that fluctuate (wax and wane) at a rate equal to the difference between them ($f_2 - f_1$).

As the frequency difference between the two tones widens, the beats become faster. These rapid amplitude fluctuations are perceived as **roughness** rather than as discernible beats, as discussed earlier in the chapter. Further widening of the frequency separation results in the perception of the two original tones, in addition to which a variety of combination tones may be heard. Combination tones, as well as aural harmonics, are the result of **nonlinear distortion** in the ear.

Distortion products are those components at the output of the system that were not present at the input. A simple example demonstrates how nonlinear distortions produce outputs that differ from the inputs. Consider two levers, one rigid and the other springy. The rigid lever represents a **linear system**. If one moves one arm of this lever up and down sinusoidally (the input), then the opposite arm will also move sinusoidally (the output). On the other hand, the springy lever illustrates the response of a **nonlinear system**. A sinusoidal input to one arm will cause the other arm to move up and down, but there will also be superimposed overshoots and undershoots in the motion, due to the "bounce" of the springy lever arms. Thus, the responding arm of the lever will move with a variety of superimposed frequencies (distortion products) even though the stimulus is being applied sinusoidally.

The simplest auditory distortion products are **aural harmonics**. As their name implies, these are distortion products, which have frequencies that are multiples of the stimulus frequency. For example, a stimulus frequency f_1, when presented at a high enough level, will result in aural harmonics whose frequencies correspond to $2f_1$, $3f_1$, etc. Therefore, a 500-Hz primary tone (f_1) will result in aural harmonics that are multiples of 500 Hz ($2f_1 = 1000$ Hz, $3f_1 = 1500$ Hz, etc.). Two primary tones are illustrated in Fig. 12.4. The 800-Hz primary tone (f_1) is associated with an aural harmonic at 1600 Hz ($2f_1$) and the 1000-Hz primary (f_2) is associated with the aural harmonic at 2000 Hz ($2f_2$).

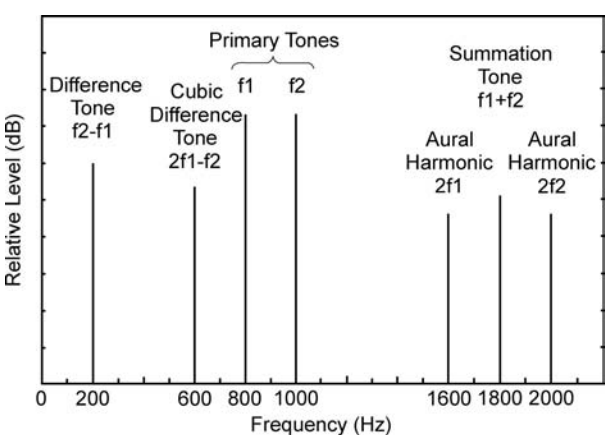

Figure 12.4 Examples of various auditory distortion products associated with stimulus (primary) tones of 800 Hz (f_1) and 1000 Hz (f_2).

If two primary tones f_1 and f_2 are presented together, nonlinear distortion will result in the production of various **combination tones** due to the interactions among the primaries and the harmonics of these tones. For convenience, we will call the lower-frequency primary tone f_1 and the higher one f_2. There are several frequently encountered combination tones that we shall touch upon. [See Boring (1942) and Plomp (1965) for interesting historical perspectives, and Goldstein et al. (1978) for an classic review of the compatibility among physiological and psychoacoustic findings on combination tones.]

It is necessary to devise methods that enable us to quantify the aspects of combination tones. A rather classical method takes advantage of the phenomenon of aural beats discussed above. Recall that best beats occur when the beating tones are of equal amplitudes. Generally stated, this technique involves the presentation of a probe tone at a frequency close enough to the combination tone of interest so that beats will occur between the combination and probe tones. Characteristics of the combination tone are inferred by varying the amplitude of the probe until the subject reports hearing best beats (i.e., maximal amplitude variations). The **best beats method**, however, has been the subject of serious controversy (Lawrence and Yantis, 1956, 1957; Meyer, 1957; Chocolle and Legouix, 1957), and has been largely replaced by a cancellation technique (Zwicker, 1955; Goldstein, 1967; Hall, 1975). The **cancellation method** also employs a probe tone, but in this method, instead of asking the subject to detect best beats, the probe tone is presented at the frequency of the combination tone, and its phase and amplitude are adjusted until the combination tone is canceled. Cancellation occurs when the probe tone is equal in amplitude and opposite in phase to the combination tone. The characteristics of the combination tone may then be inferred from those of the probe tone that cancels it. A lucid description of and comparison among all of the major methods has been provided by Zwicker (1981).

Techniques such as these have resulted in various observations about the nature of combination tones. The simplest combination tones result from adding or subtracting the two primary tones. The former is the **summation tone** ($f_1 + f_2$). As illustrated in Fig. 12.4, primary tones of 800 Hz (f_1) and 1000 Hz (f_2) will result in the production of the summation tone $1000 + 800 = 1800$ Hz. We will say little about the summation tone except to point out that it is quite weak and not always audible. On the other hand, the **difference tone** ($f_2 - f_1$) is a significant combination tone that is frequently encountered. For the 800- and 1000-Hz primaries in the figure, the difference tone would be $1000 - 800 = 200$ Hz. The difference tone is heard only when the primary tones are presented well above threshold. Plomp (1965) found, despite wide differences among his subjects, that the primaries had to exceed approximately 50 dB sensation level in order for the difference tone to be detected.

The **cubic difference tone** ($2f_1 - f_2$) is another significant and frequently encountered combination tone. These distortion products appear to be generated by the active processes in the cochlea and have already been encountered when distortion-product otoacoustic emissions were discussed in Chapter 4. For the 800- and 1000-Hz primary tones in Fig. 12.4, the resulting cubic difference tone is $2(800) - 1000 = 600$ Hz. A particularly interesting aspect of the cubic difference tone is that it is audible even when the primaries are presented at low sensation levels. For example, Smoorenburg (1972) demonstrated that $2f_1 - f_2$ is detectable when the primaries are only 15 to 20 dB above threshold, although he did find variations among subjects.

When the primary tones exceed 1000 Hz, the level of the difference tone $f_2 - f_1$ tends to be rather low (approximately 50 dB below the level of the primaries); in contrast, the difference tone may be as little as 10 dB below the primaries when they are presented below 1000 Hz (Zwicker, 1955; Goldstein, 1967; Hall, 1972a, 1972b). On the other hand, the cubic difference tone $2f_1 - f_2$ appears to be limited to frequencies below the lower primary f_1, and its level increases as the ratio f_2/f_1 becomes smaller (Goldstein, 1967; Hall, 1972a, 1972b, 1975; Smoorenburg, 1972). Furthermore, the cubic difference tone has been shown to be within approximately 20 dB of the primaries when the frequency ratio of f_2 and f_1 is on the order of 1.2:1 (Hall, 1972a; Smoorenburg, 1972). The student with an interest in this topic should also refer to the work of Zwicker (1981) and Humes (1985a, 1985b) for insightful reviews and analyses of the nature of combination tones.

An interesting attribute of combination tones is their stimulus-like nature. In other words, the combination tones themselves interact with primary (stimulus) tones as well as with other combination tones to generate beats and higher-order (secondary) combination tones, such as $3f_1–2f_2$ and $4f_1–2f_2$. Goldstein et al. (1978) have shown that such **secondary combination tones** have properties similar to those of combination tones generated by the primaries.

MUSICAL PITCH

As already mentioned, pitch is usually considered in musical terms. Here, tones are separated by perceptually relevant intervals. The intervals themselves are based on *ratios* between the frequencies of the tones (f_2/f_1) rather than on the differences between them ($f_2 - f_1$). For example, the principal interval is the **octave**, which is a 2:1 frequency ratio. It is helpful to discuss musical pitch scales with reference to Fig. 12.5, which depicts a standard piano keyboard with 88 keys.[1] The white keys are grouped in sets of 7, which are labeled in the order

C, *D*, *E*, *F*, *G*, *A*, *B*, next *C*.

The frequency of a particular C is exactly twice the frequency of the prior one so that each octave is divided into seven intervals. This order is called the *major scale* and is also associated

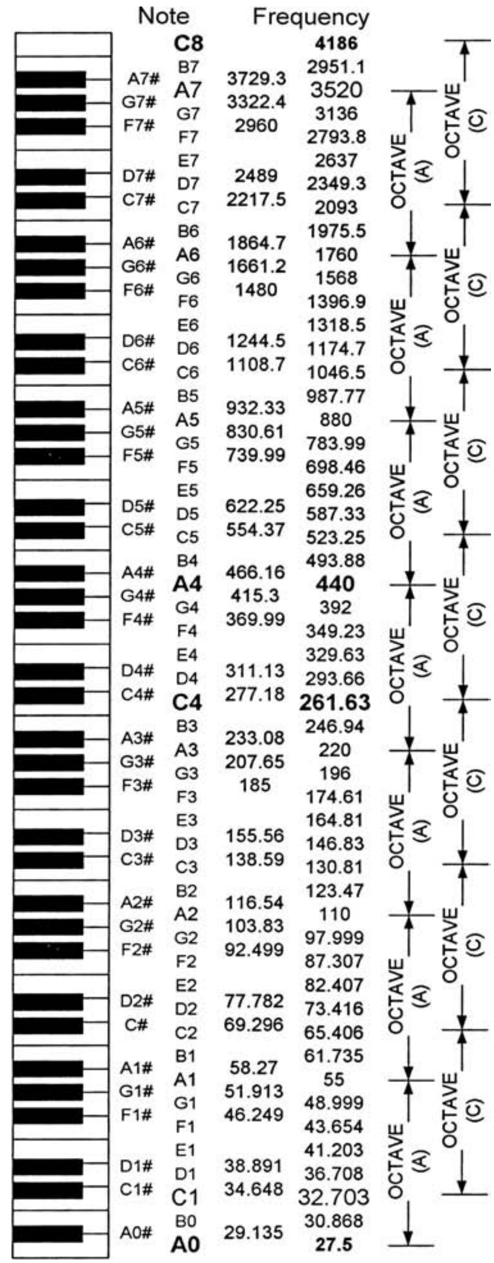

Figure 12.5 The musical scaling of pitch intervals (equal temperament) with reference to the keys on the piano keyboard.

with the familiar singing scale (*do*, *re*, *mi*, *fa*, *so*, *la*, *ti*, *do*). Octaves can also be divided into 12 intervals, which are labeled in the order

C, *C#*[2], *D*, *D#*, *E*, *F*, *F#*, *G*, *G#*, *A*, *A#*, *B*, next *C*.

[1] For an online application that provides the musical note corresponding to a given frequency, see Botros (2001).

[2] The symbol # is read as "sharp"; hence, *C#* is C-sharp, *D#* is D-sharp, etc.

These 12-note groupings form the *chromatic scale* and correspond to the sets of 12 (7 white and 5 black) keys on the piano. Intervals within octaves are discussed below.

The musical scaling of pitch may be viewed in terms of two attributes, height and chroma (Révész, 1913; Bachem, 1948, 1950; Shepard, 1964). **Tone height** pertains to the monotonic increase from low to high pitch as frequency rises, and it is represented by the vertical arrow in Fig. 12.6. It has been suggested that mel scales of pitch are related to the dimension of tone height (Ward and Burns, 1982; Goldstein, 2000). In contrast, chroma is related to a perceptual similarity or sameness among tones that are separated by 2:1 frequency intervals, called **octave equivalence**. As a result, these tones have the same pitch class or chroma and share the same name. For example, every doubling of frequency beginning with 27.5 Hz in Fig. 12.5 produces a note called *A*, labeled *A0* at 27.5 Hz through *A7* at 3520 Hz. (The note called *A4*, which has a frequency of 440 Hz, is most commonly used standard or reference pitch for musical tuning.) Similarly, every doubling of frequency starting with 32.703 Hz in Fig. 12.5 is a note called *C*, from *C1* at (32.703 Hz) to *C8* (4186 Hz). It should now be apparent that chroma changes from C through B within each octave, and that this patterns recycles at a higher frequency beginning at the next C. For example, *C1*, *D1#*, and *F1* differ in terms of chroma, whereas *A3*, *A5*, and *A6* have the same chroma but different tone heights. Thus, chroma is conceptualized in the form of the helix in Fig. 12.6.

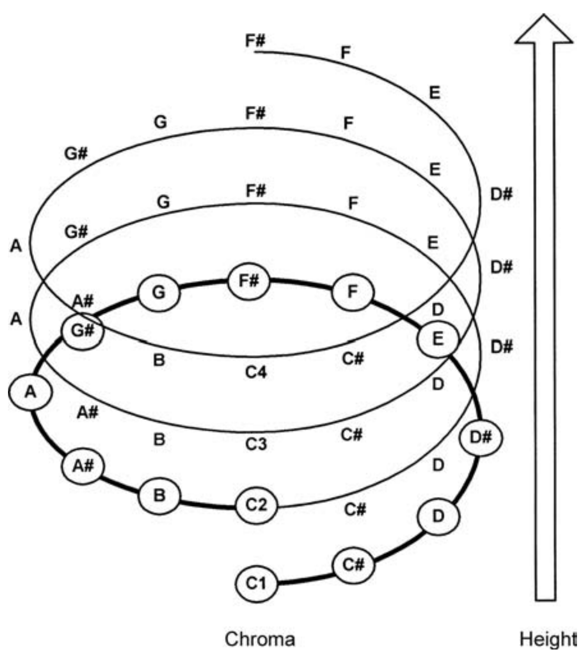

Figure 12.6 Chroma changes going around the helix, and this pattern of tone classes is repeated every octave. Tone height increases monotonically with frequency, as suggested by vertical arrow. Notice that tones with the same chroma and different tone heights appear above each other. Based on Shepard's (1964) conceptualization.

It is interesting to note in this context that Demany and Armand (1984) found that 3-month-old infants were less sensitive to frequency shifts of exactly one octave than to smaller or larger shifts. These findings provide evidence of octave equivalence, supporting the existence of chroma perception in babies. Moreover, a typographically organized representation of musical keys in the rostromedial prefrontal cortex was identified by functional magnetic resonance imaging (fMRI) in musically experienced listeners performing musical perception tasks (Janata et al., 2002).

Just as there is a 2:1 frequency ratio between one C and the next one, the intervals within an octave also take the form of frequency ratios; however, a variety of different scaling schemes are available. For example, the **just (diatonic) scale** is based on three-note groupings (triads) having frequency ratios of 4:5:6, in order to maximize the consonance of thirds, fourths, and fifths. The **Pythagorean scale** concentrates on fourths and fifths. In contrast, the **equal temperament scale** employs logarithmically equal intervals. These scales are compared in Table 12.1. Each scale has advantages and limitations in music, although we will concentrate on the equal temperament scale. Several informative discussions of musical scales are readily available (e.g., Campbell and Greated, 1987; Rossing, 1989).

The equal temperament scale divides the octave into logarithmically equal **semitone** intervals corresponding to the groupings of 12 (7 white and 5 black) keys on the piano (Fig. 12.5). In order to split the 2:1 frequency ratio of a whole octave into 12 logarithmically equal intervals from C to the next C, each interval must be a frequency ratio of $2^{1/12}{:}1$ or 1.0595:1 (an increment of roughly 6%). For example, the semitone interval from *C* to *C#* constitutes a frequency ratio of $2^{1/12}{:}1$, which is the equivalent to 1.0595:1; two semitones from *C* to *D* is $(2^{1/12} \times 2^{1/12}){:}1 = 2^{2/12}{:}1 = 1.12246{:}1$; and 12 semitones corresponding to one octave from one *C* to the next *C* is $2^{12/12}{:}1 = 2{:}1$.

Each equal temperament semitone is further divided into 100 logarithmically equal intervals called **cents**. The use of cents notation often facilitates pitch scaling. For example, cents notation makes it easier to compare the musical intervals involved in the equal temperament, just, and Pythagorean scales (Table 12.1). Since there are 100 cents per semitone and 12 semitones per octave, there are 1200 cents per octave. Thus, 1 cent corresponds to a frequency ratio[3] interval of $2^{1/1200}{:}1$ or 1.00058:1, which is an increment of less than 0.06%. This is considerably smaller than the relative difference limen for frequency ($\Delta f/f$) of about 0.002 encountered in Chapter 9, which corresponds to 0.2%. In this sense, it takes more than a few cents to achieve a perceptible pitch change. In fact, a study of perceptible tuning changes by Hill and Summers (2007) found that the difference

[3] A *frequency ratio* (f_2/f_1) can be converted into *cents* (c) with the formula $c = 1200 \times \log_2(f_2/f_1)$; or $c = 3986.31 \times \log_{10}(f_2/f_1)$ if one prefers to use common logarithms.

Table 12.1 Frequency Ratios and Cents for Musical Intervals with the Equal Temperament, Just, and Pythagorean Scales

Interval		Frequency ratio (f_2/f_1)			Cents		
From C to . . .	Name	Equal temperament	Just	Pythagorean	Equal temperament	Just	Pythagorean
C (itself)	Unison	$2^{0} = 1.0595$	$1/1 = 1.0000$	$1/1 = 1.0000$	0	0	0
C#	Minor second	$2^{1/12} = 1.0595$	$16/15 = 1.0667$	$256/243 = 1.0535$	100	112	90
D	Major second	$2^{2/12} = 1.1225$	$10/9 = 1.1111$	$9/8 = 1.1250$	200	182	204
D#	Minor third	$2^{3/12} = 1.1892$	$6/5 = 1.2000$	$32/27 = 1.1852$	300	316	294
E	Major third	$2^{4/12} = 1.2599$	$5/4 = 1.2500$	$81/64 = 1.2656$	400	386	408
F	Fourth	$2^{5/12} = 1.3348$	$4/3 = 1.3333$	$4/3 = 1.3333$	500	498	498
F#	Tritone	$2^{6/12} = 1.4142$	$45/32 = 1.4063$	$1024/729 = 1.4047$	600	590	588
G	Fifth	$2^{7/12} = 1.4983$	$3/2 = 1.5000$	$3/2 = 1.5000$	700	702	702
G#	Minor sixth	$2^{8/12} = 1.5874$	$8/5 = 1.6000$	$128/81 = 1.5802$	800	814	792
A	Major sixth	$2^{9/12} = 1.6818$	$5/3 = 1.6667$	$27/16 = 1.6875$	900	884	906
A#	Minor seventh	$2^{10/12} = 1.7818$	$7/4 = 1.7500$	$16/9 = 1.7778$	1000	969	996
B	Major seventh	$2^{11/12} = 1.8877$	$15/8 = 1.8750$	$243/128 = 1.8984$	1100	1088	1110
(Next) C	Octave	$2^{12/12} = 2.0000$	$2/1 = 2.0000$	$2/1 = 2.0000$	1200	1200	1200

limen is about 10 cents and that category widths approximated 70 cents.

Recall that frequency is coded by both temporal and place mechanisms, with the former taking on a more important role for lower frequencies and the latter predominating for higher frequencies. Since auditory nerve firing patterns reflect phase locking for stimulus frequencies as high as roughly 5000 Hz (Chap. 5), it would appear that the temporal mechanism is operative up to about this frequency. With these points in mind, it is interesting to be aware of several lines of evidence indicating that the perception of pitch chroma is limited to the frequencies below roughly 5000 Hz, although pitch height continues to increase for higher frequencies.

The restriction of chroma perception to 5000 Hz and below has been shown in various ways. Bachem (1948) and Ohgushi and Hatoh (1989) found that listeners with absolute pitch could identify the chroma of pure tones was limited to the frequencies up to about 4000 to 5000 Hz. **Absolute (perfect) pitch** is the very rare ability to accurately identify or produce pitches in isolation, that is, without having to rely on a reference tone (Ward, 1963a, 1963b). Ward (1954) found that musicians could adjust the frequency of a (higher) tone to be an octave above another (lower) tone, providing both tones were less than 5500 Hz. In addition, Atteave and Olson (1971) found that the ability of musicians to transpose a familiar three-tone sequence (in effect, a simple melody) deteriorated above about 5000 Hz. Semal and Demany (1990) asked musicians to transpose a sequence of two pure tones until the higher tone in the pair was perceived to be "just above the upper limit of musical pitch." On average, this limit occurred at approximately 4700 Hz.

Consonance and Dissonance

In addition to their perceptual similarity, two simultaneously presented tones differing by octave intervals are also perceived as being consonant. Such combinations of two or more simultaneous tones are often referred to as **chords**. **Consonance** simply means that when two sounds are presented together they result in a pleasant perception; in contrast, **dissonance** refers to sounds that appear unpleasant when presented together. Consonance versus dissonance for pairs of tones depends upon the difference between the two frequencies, and how well these frequencies can be resolved by the ear (Plomp and Levelt, 1965; Plomp and Steeneken, 1968; Kameoka and Kuriyagawa, 1969; Schellenberg and Trehub, 1994). Dissonance occurs when two tones are close enough in frequency to be less than a critical band apart. In this case, they are not completely resolved because their vibration patterns interact along the basilar membrane, leading to a sensation of **roughness** due to rapid beats between the two tones. This roughness is perceived as being unpleasant or dissonant. In contrast, the two tones will not interact if they are separated by more than a critical band, in which case there is no roughness, and if they are experienced as consonant.

Complex tones, such as musical notes, contain other frequencies in addition to their fundamentals. In music, the terms **partials** and **overtones** are used to refer these frequencies and are usually used as synonyms for harmonics. (However, one should be aware that many instruments, such as chimes and pianos, also produce "inharmonic" partials at frequencies that are not exact multiples of the fundamental). Thus, dissonance can be introduced by interactions among these partials, as well as between the fundamentals. With these points in mind, consonance is associated with notes composed of frequencies that are related by simple (small) integer ratios, such as 2:1 (octave, e.g., *C4* and *C5*), 3:2 (perfect fifth, e.g., *C4* and *G4*), and 4:3 (perfect fourth, *C4* and *F4*). In contrast, dissonance occurs when their frequencies are related by complex (large) ratios such as 45:32 (tritone, e.g., *F3* and *B3*). In terms of the musical intervals, two tones will be consonant when they are separated by 0 and

12 semitones (i.e., unison and octave), and dissonance occurs when the two notes are separated by intervals of 1–2, 6, and 10–11 semitones.

In contrast to two-tone combinations, the perception of chords becomes more involved when three or more tones are involved. A consonant three-tone chord (triad) is perceived as being stable when the spacing between the low and middle tones is different than the interval between the middle and high tones. However, **tension** is experienced when these two intervals are equal, in which case the chord is perceived as being unstable or ambiguity even though there is no dissonance (Meyer, 1956). Cook and colleagues have shown that the harmonious perception of triads depends on the effects of both dissonance and tension involving the fundamentals and partials of the notes that make up the chord (Cook, 2001; Cook and Fujisawa, 2006; Cook and Hayashi, 2008).

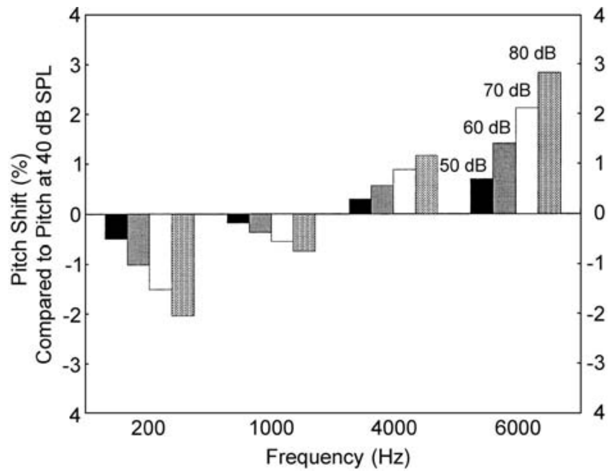

Figure 12.7 Change in pitch (pitch shift) compared to the pitch at 40 dB SPL for tones presented at sound pressure levels of 50, 60, 70, and 80 dB, at selected frequencies. Pitch shifts are upward for high frequencies and downward for low frequencies by about 2% to 3% or less. Based on data by Zwicker and Fastl (1999).

PITCH AND INTENSITY

In a classic study by Stevens (1935), subjects were asked to adjust the intensity of a tone until it had the same pitch as a standard tone of slightly different frequency. Results for one subject who was a "good responder" showed that increasing the intensity of the tone increased its pitch for frequencies 3000 Hz and above, and lowered its pitch for frequencies 1000 Hz and below. The pitch stayed essentially constant as intensity was varied for tones between 1000 and 3000 Hz.

Although Stevens' study has frequently been cited to illustrate how intensity affects pitch, subsequent studies did not find large pitch changes associated with intensity increases (Morgan et al., 1951; Ward, 1953; Cohen, 1961; Terhardt, 1979; Zwicker and Fastl, 1999). Figure 12.7 shows how the pitches of pure tones change as the sound pressure level is increased above a reference level of 40 dB (Terhardt, 1979; Zwicker and Fastl, 1999). Increasing level does cause pitch to fall for low-frequency tones and to rise for high-frequency tones. However, the sizes of these pitch shifts are quite small, amounting to less than 2% to 3% as the level increases from 40 to 80 dB.

PITCH OF COMPLEX SOUNDS

Pitch perception is not limited to pure tones. On the contrary, real-world sounds are complex, and common experience reveals that pitch perceptions are associated with them. When dealing with *aperiodic* sounds, the perceived pitch is related to the spectrum of the noise (Small and Daniloff, 1967; Fastl, 1971). In general, the pitches of low- and high-pass noises are related to their cut-off frequencies, and the pitches of band-pass noises are associated with their center frequencies. Moreover, the distinctiveness of the pitch sensation (pitch strength) elicited by narrow bands of noise can be quite strong, but it lessens as the bandwidth gets wider and becomes quite weak when the critical band is exceeded.

For complex *periodic* sounds, such as those produced by the human voice and many musical instruments, consider, for example, a woman's voice that is composed of many harmonics of 220 Hz (220, 440, 660, 880, 1100, 1320 Hz, etc.). Comparing the pitch of her voice to various pure tones, a match would be made with a 220-Hz tone. Here, the pitch is that of the fundamental frequency and is dominated by its lower harmonics. Classical experiments emphasized the importance of the first five harmonics in determining the pitch of complex tones (e.g., Plomp, 1964, 1967; Ritsma, 1967; Bilsen and Ritsma, 1967). However, the dominant range does not appear to be a fixed, and it is affected by factors such as the frequency of the fundamental, the levels of the harmonics, and duration (e.g., Patterson and Wightman, 1976; Moore et al., 1984, 1985; Gockel et al., 2005, 2007).

What is odd, however, is that woman just described would still have a vocal pitch of 220 Hz even if the 220-Hz fundamental frequency is missing (in which case the lowest harmonic actually present would be 440 Hz). Demonstrating this phenomenon does not even require a trip to the laboratory—it is actually an every day occurrence because telephones do not provide any frequencies below about 300 Hz. This perception of the **missing fundamental** was first described by Seebeck in 1841, and was reintroduced a century later by Schouten (1940).

The missing fundamental is perhaps the best-known example of what has descriptively been called **periodicity pitch**, **virtual pitch**, **residue pitch**, **low pitch**, and **repetition pitch**. Another case is provided by signals that are periodically interrupted or amplitude modulated (Thurlow and Small, 1955; Burns and Viemeister, 1976, 1981). For example, Thurlow and Small (1955) found that if a high-frequency tone is interrupted periodically, then the subject will perceive a pitch corresponding to the frequency whose period is equal to the interruption rate. Thus, if the high-frequency tone is interrupted every 10 ms (the period of a 100-Hz tone), then subjects will match the pitch of the interrupted high-frequency tone to that of a 100-Hz tone.

Studies by Ritsma (1962, 1963) indicate that the existence region of virtual pitches extends up to about 800 Hz, but others have shown that periodicity pitches can be perceived as high as roughly 1400 Hz with a large enough number of harmonics (Plomp, 1967; Moore, 1973).

The classical *resonance-place theory* would suggest that the missing fundamental is due to energy present at the fundamental frequency as a result of distortions. In other words, the difference tone $f_2 - f_1$ would be the same as the missing fundamental since, for example, $1100 - 1000 = 100$ Hz. However, this supposition is not true because the missing fundamental differs from combination tones in several dramatic ways. For example, the missing fundamental is heard at sound pressure levels as low as about 20 dB (Thurlow and Small, 1955; Small and Campbell, 1961), whereas difference tones are not heard until the primary tones are presented at sound pressure levels of 60 dB or more (Bekesy, 1960; Plomp, 1965). Also, if a probe tone is presented to a subject at a frequency close to that of a difference tone (which is actually represented at a place along the basilar membrane), then aural beats are heard. However, beats do not occur when the probe is added to the missing fundamental (Schouten, 1940).

Further evidence against the supposition that the missing fundamental is the result of energy at the apex of the cochlea due to distortions (or other means) comes from masking studies (Licklider, 1954; Small and Campbell, 1961; Patterson, 1969). These experiments demonstrated that masking of the frequency range containing the missing fundamental does not obliterate its audibility. In other words, real low-frequency tones and difference tones can be masked, but the missing fundamental cannot be.

The concept that the missing fundamental results from distortions due to interactions among the harmonics within the cochlea is further weakened by studies which preclude this by using dichotic stimulation (Houtsma and Goldstein, 1972) or by presenting the harmonics sequentially rather than simultaneously (Hall and Peters, 1981). Dichotic stimulation refers to the presentation of different stimuli to the two ears (see Chap. 13). Houtsma and Goldstein (1972) presented one harmonic to each ear and asked their subjects to identify melodies based upon the perception of missing fundamentals. If the missing fundamental were really the result of interacting harmonics in the cochlea, then the subjects in this dichotic experiment would not hear it because only one tone was available to each cochlea. They found that the missing fundamental was perceived when the harmonics were presented separately to the two ears. This finding indicates that the phenomenon occurred within the central auditory nervous system, since this is the only region where the harmonics were simultaneously represented.

Hall and Peters (1981) showed that subjects could hear the missing fundamental when presented with three harmonics one after the other rather than simultaneously. Their stimuli were sequences of three harmonics (e.g., 600, 800, and 1000 Hz) each lasting 40 ms and separated by 10 ms pauses. An interaction among the harmonics was precluded because they were present at different times. These stimuli were presented alone (in quiet) and also in the presence of a noise. Pitch discrimination and matching tests revealed that their subjects heard the pitches of the harmonics in quiet, but that they heard the missing fundamental in noise.

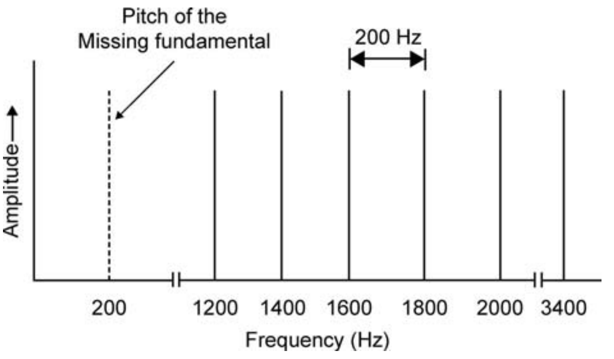

Figure 12.8 Spectrum illustrating the perception of the missing fundamental. Based on Patterson (1973), with permission of *J. Acoust. Soc. Am.*

Schouten's (1940, 1970) *residue theory* proposed that the perception of this missing fundamental is based upon the *temporal pattern* of the complex periodic sound's waveform. Consider a complex periodic tone containing energy only above 1200 Hz, spaced as shown in Fig. 12.8. This spectrum shows energy only for higher harmonics of 200 Hz (1200, 1400, 1600 Hz, etc.), but no energy at the 200-Hz fundamental frequency. Nevertheless, subjects presented with this complex tone will match its pitch to that of a 200-Hz tone. This would occur because the auditory system is responding to the period of the complex periodic tone (5 ms or 0.005 s), which corresponds to the period of 200 Hz ($1/0.005$ s $= 200$ Hz). If the components were separated by 100 Hz (e.g., 2000, 2100, 2200, 2300 Hz), then the waveform would have a period of 10 ms or 0.01 s, corresponding to 100 Hz. This can occur because all of the harmonics are separated by the same frequency difference (200 Hz in our example), but auditory filters (critical bandwidths) get wider as frequency increases. As a result, lower harmonics would fall into separate low-frequency auditory filters and would thus be perceived separately from one another. However, two or more higher harmonics will fall into the same (wider) high-frequency auditory filter. Interference between the harmonics within an auditory filter constitutes a complex periodic wave that repeats itself over time at the rate of the fundamental.

Raising each of the frequencies making up the complex by the same increment causes a slight increase in the pitch of the missing fundamental, as illustrated in Fig. 12.9. Here, the components of the complex sound have been increased by 60 Hz compared to the frequencies in the preceding figure so that

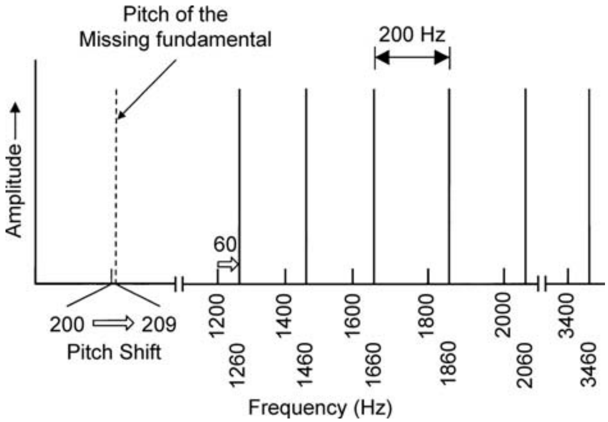

Figure 12.9 Spectrum illustrating the pitch shift of the missing fundamental. Based on Patterson (1973), with permission of *J. Acoust. Soc. Am.*

they are now 1260, 1460, 1660 Hz, etc. Notice that the missing fundamental is now matched to 209 Hz even though the components frequencies are still 200 Hz apart. This **pitch shift** of the missing fundamental was also originally described by (Schouten, 1940) and has been confirmed by many others (e.g., deBoer, 1956; Schouten et al., 1962; Smoorenburg, 1971; Patterson, 1973; Wightman, 1973a; Buunen et al., 1974). It has been suggested that the pitch shift may be based on the fine structure of the repeated waveform (e.g., Schouten 1940; Thurlow and Small, 1955; deBoer, 1956), as illustrated in Fig. 12.10. When the harmonics are exact multiples of 200 Hz (1200, 1400 Hz, etc.), there is exactly 5 ms between equivalent peaks in the repeated waveform. The upward shift of the harmonics by 60 Hz results in a slightly shorter interval between the equivalent peaks as the waveform is repeated, so that the pitch shifts upwards a bit. Ambiguities of pitch would result when the nearly but not exactly equivalent peaks are compared, as in Fig. 12.10.

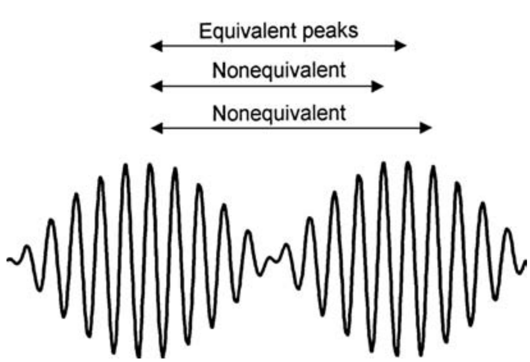

Figure 12.10 Comparison of the intervals between equivalent and nonequivalent peaks on repetitions of a waveform. Pitch ambiguity results when the pitch-extracting mechanism compares the peak of one repetition to peaks not equivalent to it on the next repetition of the waveform. Based on Schouten et al. (1962), with permission of *J. Acoust. Soc. Am.*

More recent explanations of virtual pitch involve *pattern perception models*. These explanations begin with a frequency analysis and go on to evaluate the pitch using some kind of higher-level processing. Pattern recognition may be thought of as the extraction of those similar attributes of a stimulus that allow it to be recognized as a member of a class in spite of variations in details. For example, the letter "A" is recognized whether it is printed upper- or lowercase, in italics, or even in this author's illegible penmanship.

Terhardt's (1974, 1979; Terhardt et al., 1982a, 1982b) *virtual pitch model* involves the perception of an auditory gestalt calling upon previously learned cues. These cues are subharmonics associated with various frequencies in complex sounds and are learned in the course of one's ongoing experience with the complex periodic sounds in speech. When we hear a sound, dominant resolved components are provided by the initial spectral analysis, and then the pitch processor compares the subharmonics of these components to those of the learned cues. The pitch is assigned based on the greatest degree of coincidence. The *optimum processor model* proposed by Goldstein (1973; Gerson and Goldstein, 1978) looks for the best match between a template of harmonics and the components of the complex periodic sound resolved by the initial frequency analysis. The virtual pitch is based on the fundamental frequency of the best-fitting template. In the *pattern transformation model* (Wightman, 1973b) the neural pattern resulting from a Fourier analysis at the peripheral level is examined by a pitch extractor that looks for important features shared by stimuli having the same pitch. Interested students will find detailed discussions of these and other models in several reviews (e.g., Plomp, 1976; Houtsma, 1995), in addition to the original sources.

Several studies provide physiological support for Shouten's explanation of virtual pitch (e.g., Brugge et al., 1969; Javel, 1980; Horst et al., 1986). For example, Javel (1980) found that both the envelopes and fine structures of amplitude-modulated tones were represented in auditory nerve firing patterns. However, perceptual evidence has revealed that his residue theory fails to provide a complete explanation for the perception of complex tones. The main problem is that it depends on interactions among the higher harmonics of a complex sound that are unresolved because they fall inside the same critical band filter(s). Yet, recall that the lower harmonics are dominant in determining the pitch of complex tones. The previously described results of Hall and Peters (1981) and Houtsma and Goldstein (1972) are also problematic for the Shouten's residue model. Recall that virtual pitches were heard even when the harmonics could not interact because they were presented sequentially or in different ears, implicating the involvement of more central mechanisms.

These issues are strengths for the pattern perception models, which rely upon the resolved lower harmonics and central pitch processors. However, these models also fall short of providing a complete explanation of virtual pitch. A key limitation here is that in spite of fact that the lower harmonics are the most important ones for hearing a virtual pitch, it is possible for a virtual pitch to be perceived (although generally not as well) based on just the higher frequency harmonics (e.g., Moore, 1973; Houtsma and Smurzynski, 1990). Thus, neither approach by itself provides a comprehensive account of virtual pitch perception, and it is not surprising that models involving aspects of both approaches have been suggested (e.g., Moore, 1977, 1997; van Noorden, 1982).

TIMBRE

Timbre, often referred to as **sound quality**, is typically defined as the sensory attribute of sound that enables one to judge differences between sounds having the same pitch, loudness, and duration (e.g., Plomp, 1970; ANSI, 2004). For example, it is timbre that distinguishes between the same musical note played on a violin, piano, guitar, or French horn; different vowels spoken by the same person; or a normal voice compared to one that is hoarse, harsh, or breathy. These examples show that timbre is multidimensional, being affected by both the ongoing or steady-state features of a sound as well as by its dynamic characteristics. The term **tone color** is sometimes used to refer to the timbre of steady-state sounds.

The interplay between the steady-state and dynamic attributes of timbre is illustrated in Fig. 12.11, which shows the results of a study in which listeners were asked to make similarity judgments among pairs of notes played on different instruments (Iverson and Krumhansl, 1993). An analysis of these judgments revealed that they involved two dimensions corresponding to (1) steady-state spectral features (a dull-bright continuum), shown vertically in the figure, and (2) dynamic characteristics (separating percussive from blown instruments), shown horizontally. Notice that the trumpet, saxophone, and tuba differed considerably vertically but were quite close in terms of the horizontal dimension. On the other hand, the piano and saxophone were similar along the vertical dimension but far apart horizontally. The piano and tuba were perceptually far apart along both continua.

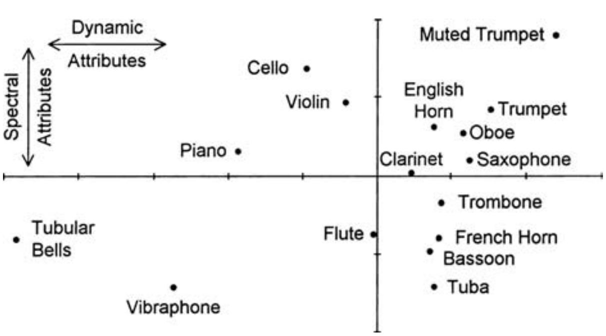

Figure 12.11 Two-dimensional similarity judgments among various musical instruments. Vertical dimension represents steady-state (spectral) characteristics. Horizontal dimension represents dynamic (amplitude envelope) characteristics. *Source*: Modified from Iverson and Krumhansl (1993), with permission of *J. Acoust. Soc. Am.*

The steady-state and dynamic features of timbre have been studied in some detail (e.g., Berger, 1964; Strong and Clark, 1967; Wedin and Goude, 1972; von Bismark, 1974a, 1974b; Grey, 1977; Grey and Gordon, 1978; Wezel, 1979; Krumhansl, 1989; Iverson and Krumhansl, 1993; McAdams et al., 1999). Among the *steady-state* features of a sound that influence its timbre are the shape of the spectral envelope, its centroid, or center of gravity,[4] whether the spectral envelope is smooth or irregular, and whether there are noise-like (aperiodic) components present. (The vertical dimension in Fig. 12.11 is mainly associated with the centroid frequency of the spectrum.) For example, sounds with spectra emphasizing the higher frequencies have a bright character, whereas sounds emphasizing the lows are heard as dull. The locations of resonance peaks or formants distinguish between different vowels (see Chap. 14) and also between various musical instruments. Timbre is also affected when certain harmonics are absent. For example, a hollow quality is heard when complex tones are composed of just odd harmonics (i.e., even harmonics are missing).

The *dynamic* features affecting timbre include the amplitude envelope of the waveform (or how amplitude changes over time) and changes in the spectral envelope over time. (The horizontal dimension in Fig. 12.11 is associated with the amplitude envelope.) It is important to stress that these dynamic aspects include the onset (attack) and offset (release or decay) characteristics of a sound, as well as those occurring during its overall duration. For example, notes played on one musical instrument can sound like they are being played on another instrument when onsets are removed from the remainder of the sound, or when a recording of a piano section is played backward instead of forward.

REFERENCES

American National Standards Institute, ANSI S1–1994 (R2004). 2004. *American National Standard Acoustical Terminology.* New York, NY: ANSI.

Attneave, F, Olson, RK. 1971. Pitch as a medium: A new approach to psychophysical scaling. *Am J Psychol* 84, 147–166.

Bachem, A. 1948. Chroma fixation at the ends of the musical frequency scale. *J Acoust Soc Am* 20, 704–705.

Bachem, A. 1950. Tone height and chroma as two different pitch qualities. *Acta Psychol* 7, 80–88.

[4] The centroid of the spectrum is its average frequency, weighted according to amplitude. The usual formula for it is

$$c = \frac{\sum f_i a_i}{\sum a_i}$$

where c is the centroid, f_i is the frequency of a component in the spectrum, and a_i is the amplitude of that component.

Beck, J, Shaw, WA. 1962. Magnitude estimations of pitch. *J Acoust Soc Am* 34, 92–98.

Beck, J, Shaw, WA. 1963. Single estimates of pitch magnitude. *J Acoust Soc Am* 35, 1722–1724.

Bekesy, G. 1960. *Experiments in Hearing*. New York, NY: McGraw-Hill.

Berger, KW. 1964. Some factors in the recognition of timbre. *J Acoust Soc Am* 36, 1888–1891.

Bilsen, FA, Ritsma, RJ. 1967. Repetition pitch mediated by temporal fine structure at dominant spectral regions. *Acustica* 19, 114–116.

Boring, EG. 1942. *Sensation and Perception in the History of Experimental Psychology*. New York, NY: Appleton-Century.

Botros, A. 2001. *Frequency to Musical Note Converter*. Available at http://www.phys.unsw.edu.au/music/note/.

Brugge, JF, Anderson, DJ, Hind, JE, Rose, JE. 1969. Time structure of discharges in singled auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J Neurophysiol* 32, 386–401.

Burns, EM, Viemeister, NF. 1976. Nonspectral pitch. *J Acoust Soc Am* 60, 863–869.

Burns, EM, Viemeister, NF. 1981. Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *J Acoust Soc Am* 70, 1655–1660.

Buunen, TJF, Festen, JM, Bilsen, FA, Van Den Brink, G. 1974. Phase effects in a three-component signal. *J Acoust Soc Am* 55, 297–303.

Campbell, M, Greated, C. 1987. *The Musician's Guide to Acoustics*. New York, NY: Schirmer.

Chocolle, R, Legouix, JP. 1957. On the inadequacy of the method of beats as a measure of aural harmonics. *J Acoust Soc Am* 29, 749–750.

Cohen, A. 1961. Further investigations of the effects of intensity upon the pitch of pure tones. *J Acoust Soc Am* 33, 1363–1376.

Cook, ND. 2001. Explaining harmony. The roles of interval dissonance and chordal tension. *Ann N Y Acad Sci* 930(1), 382–385.

Cook, ND, Fujisawa, TX. 2006. The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Emper Mus Rev* 1, 106–126.

Cook, ND, Hayashi, T. 2008. The psychoacoustics of harmony perception. *Am Sci* 96, 311–319.

deBoer, E. 1956. *On the residue in hearing*. Ph.D. Dissertation. Amsterdam, The Netherlands: University of Amsterdam.

Demany, L, Armand, F. 1984. The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76, 57–66.

Fastl, H. 1971. Über Tonhöhenempfindungen bei Rauschen. *Acustica* 25, 350–354.

Gerson, A, Goldstein, JL. 1978. Evidence for a general template in central optimal processing for pitch of complex tones. *J Acoust Soc Am* 63, 498–510.

Gockel, HE, Carlyon, RP, Plack, CJ. 2005. Dominance region for pitch: Effects of duration and dichotic presentation. *J Acoust Soc Am* 117, 1326–1336.

Gockel, HE, Moore, BCJ, Carlyon, RP, Plack, CJ. 2007. Effects of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *J Acoust Soc Am* 121, 373–382.

Goldstein, JL. 1967. Auditory nonlinearity. *J Acoust Soc Am* 41, 676–689.

Goldstein, JL. 1973. An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54, 1496–1516.

Goldstein, JL. 2000. Pitch perception. In: AE Kazdin (ed.), *Encyclopedia of Psychology*, Vol. 6. Oxford, UK: Oxford University Press, 201–210.

Goldstein, JL, Buchsbaum, G, Furst, M. 1978. Compatibility between psychophysical and physiological measurements of aural combination tones. *J Acoust Soc Am* 63, 474–485.

Greenwood, DD. 1990. A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87, 2592–2605.

Grey, JM. 1977. Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am* 61, 1270–1277.

Grey, JM, Gordon, JW. 1978. Perceptual effects of spectral modifications of musical timbres. *J Acoust Soc Am* 63, 1493–1500.

Hall, JL. 1972a. Auditory distortion products f_2–f_l and $2f_1$–f_2. *J Acoust Soc Am* 51, 1863–1871.

Hall, JL. 1972b. Monaural phase effect: Cancellation and reinforcement of distortion products f_2–f_l and $2f_1$–f_2. *J Acoust Soc Am* 51, 1872–1881.

Hall, JL. 1975. Nonmonotonic behavior of distortion product $2f_1$–f_2; Psychophysical observations. *J Acoust Soc Am* 58, 1046–1050.

Hall, JW, Peters, RW. 1981. Pitch for nonsimultaneous successive harmonics in quiet and noise. *J Acoust Soc Am* 69, 509–513.

Hill, TJW, Summers, IR. 2007. Discrimination of interval size in short tone sequences. *J Acoust Soc Am* 121, 2376–2383.

Horst, JW, Javel, E, Farley, GR. 1986. Coding of fine structure in the auditory nerve. I. Fourier analysis of period and interspike histograms. *J Acoust Soc Am* 79, 398–416.

Houtsma, AJM. 1995. Pitch perception. In: BCJ Moore (ed.), *Hearing*. 2nd ed. San Diego, CA: Academic Press, 267–297.

Houtsma, AJM, Goldstein, JL. 1972. The central origin of the pitch of complex tones: Evidence from musical interval recognition. *J Acoust Soc Am* 51, 520–529.

Houtsma, AJM, Smurzynski, J. 1990. Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87, 304–310.

Humes, LR. 1985a. Cancellation level and phase of the (f_2–f_l) distortion product. *J Acoust Soc Am* 78, 1245–1251.

Humes, LR. 1985b. An excitation-pattern algorithm for the estimation of ($2f_l$–f_2) and (f_2–f_l) cancellation level and phase. *J Acoust Soc Am* 78, 1252–1260.

Iverson, P, Krumhansl, C. 1993. Isolating the dynamic attributes of musical timbre. *J Acoust Soc Am* 94, 2595–2563.

Janata, P, Birk, JL, van Horn, JD, Leman, M, Tillmann, B, Bharucha, JJ. 2002. The cortical topography of tonal structures underlying western music. *Science* 298, 2167–2170.

Javel, E. 1980. Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones. *J Acoust Soc Am* 68, 133–146.

Kameoka, A, Kuriyagawa, M. 1969. Consonance theory. Part I: Consonance of dyads. *J Acoust Soc Am* 45, 1451–1459.

Krumhansl, CL. 1989. Why is musical timbre so hard to understand? In: J Nielzen, O Olsson (eds.), *Structure and Electroacoustic Sound and Music*. Amsterdam, The Netherlands: Elsevier (Excerpta Medica), 43–53.

Lawrence, M, Yantis, PJ. 1956. Onset and growth of aural harmonics in the overloaded ear. *J Acoust Soc Am* 28, 852–858.

Lawrence, M, Yantis, PJ. 1957. In support of an "inadequate" method for detecting "fictitious" aural harmonics. *J Acoust Soc Am* 29, 750–751.

Licklider, JCR. 1954. Periodicity pitch and place pitch. *J Acoust Soc Am* 26, 945.

McAdams, S, Beauchamp, JW, Meneguzzi, S. 1999. Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. J Acoust Soc Am 105, 882–897.

Meyer, LB. 1956. *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Meyer, MF. 1957. Aural harmonics are fictitious. *J Acoust Soc Am* 29, 749.

Moore, BCJ. 1973. Some experiments relating to the perception of complex tones. *Q J Exp Psychol* 25, 451–475.

Moore, BCJ. 1977. *An Introduction to the Psychology of Hearing.* London, UK: Academic.

Moore, BCJ. 1997. *An Introduction to the Psychology of Hearing*, 4th ed. London, UK: Academic Press.

Moore, BCJ, Glasberg, BR, Peters, RW. 1985. Relative dominance of individual partials in determining the pitch of complex tones. *J Acoust Soc Am* 77, 1853–1860.

Moore, BCJ, Glasberg, BR, Shailer, MJ. 1984. Frequency and intensity limens for harmonics within complex tones. *J Acoust Soc Am* 75, 550–561.

Morgan, CT, Garner, WR, Galambos, R. 1951. Pitch and intensity. *J Acoust Soc Am* 23, 658–663.

Ohgushi, K, Hatoh, T. 1989. On the perception of the musical pitch of high frequency tones. *Proceedings of the 13th International Congress on Acoustics* (Belgrade), 3, 27–30

Patterson, RD. 1969. Noise masking of a change in residue pitch. *J Acoust Soc Am* 45, 1520–1524.

Patterson, RD. 1973. The effects of relative phase and the number of components on residue pitch. *J Acoust Soc Am* 53, 1565–1572.

Patterson, RD, Wightman, FL. 1976. Residue pitch as a function of component spacing. *J Acoust Soc Am* 59, 1450–1459.

Plomp, R. 1964. The ear as a frequency analyzer. *J Acoust Soc Am* 36, 1628–1636.

Plomp, R. 1965. Detect ability threshold for combination tones. *J Acoust Soc Am* 37, 1110–1123.

Plomp, R. 1967. Pitch of complex tones. *J Acoust Soc Am* 41, 1526–1533.

Plomp, R. 1970. Timbre as a multidimensional attribute of complex tones. In: R Plomp, GF Smoorenburg (eds.), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden, The Netherlands: Sijthoff, 397–414.

Plomp, R. 1976. *Aspects of Tone Sensation.* London, UK: Academic.

Plomp, R, Levelt, WJM. 1965. Tonal consonance and critical bandwidth. *J Acoust Soc Am* 38, 548–560.

Plomp, R, Steeneken, HJM. 1968. Interference between two simple tones. *J Acoust Soc Am* 43, 883–884.

Révész, G. 1913. Zur Grundlegung der Tonspchologie. Leipzig: Veit. [English translation: de Courcy, G. 1954. *Introduction to the Psychology of Music*. Norman, France: University of Oklahoma Press.]

Ritsma, R. 1962. Existence region of the tonal residue. I. *J Acoust Soc Am* 34, 1224–1229.

Ritsma, R. 1963. Existence region of the tonal residue. II. *J Acoust Soc Am* 35, 1241–1245.

Ritsma, RJ. 1967. Frequencies dominant in the perception of the pitch of complex tones. *J Acoust Soc Am* 42,191–198.

Rossing, TD. 1989. *The Science of Sound*. Reading, MA: Addison-Wesley.

Scharf, B. 1970. Critical bands. In: JV Tobias (ed.), *Foundations of Modern Auditory Theory*, Vol. I. New York, NY: Academic Press, 157–202.

Schellenberg, EG, Trehub, SE. 1994. Frequency ratios and the perception of tone patterns. *Psychon Bull Rev* 1, 191–201.

Schouten, JF. 1940. The residue, a new concept in subjective sound analysis. *Proc K Ned Akad* 43, 356–365.

Schouten, JF. 1970. The residue revisited. In: R Plomp, GF Smoorenburg (eds.), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden, The Netherlands: Sijthoff, 41–58

Schouten, JF, Ritsma, R, Cardozo, B. 1962. Pitch of the residue. *J Acoust Soc Am* 34, 1418–1424.

Seebeck, A. 1841. Beohachtungen uber einige Bedingungen der Entstehung von Tonen. *Ann Phys Chem* 53, 417–436.

Semal, C, Demany, L. 1990. The upper limit of "musical" pitch. *Music Perception* 8, 165–176.

Shepard, RN. 1964. Circularity of judgments of relative pitch. *J Acoust Soc Am* 36, 2346–2353.

Small, A, Campbell, RA. 1961. Masking of pulsed tones by bands of noise. *J Acoust Soc Am* 33, 1570–1576.

Small, A, Daniloff, TG. 1967. Pitch of noise bands. *J Acoust Soc Am* 41, 506–512.

Smoorenburg, G. 1971. Pitch perception of two-frequency stimuli. *J Acoust Soc Am* 48, 924–942.

Smoorenburg, GF. 1972. Audibility region of combination tones. *J Acoust Soc Am* 52, 603–614.

Stevens, SS. 1935. The relation of pitch to intensity. *J Acoust Soc Am* 6, 150–154.

Stevens, SS, Volkmann, J, Newman, EB. 1937. A scale for the measurement of the psychological magnitude pitch. *J Acoust Soc Am* 8, 185–190.

Stevens, SS, Volkmann, J. 1940. The relation of pitch to frequency: A revised scale. *Am J Psych* 53, 329–353.

Strong, W, Clark, M. 1967. Perturbations of synthetic orchestral wind-instrument tones. J Acoust Soc Am 41, 277–285.

Terhardt, E. 1974. Pitch, consonance, and harmony. *J Acoust Soc Am* 55, 1061–1069.

Terhardt, E. 1979. Calculating virtual pitch. *Hear Res* 1, 155–182.

Terhardt, E, Stoll, G, Seewann, M. 1982a. Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *J Acoust Soc Am* 71, 671–678.

Terhardt, E, Stoll, G, Seewann, M. 1982b. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J Acoust Soc Am* 71, 679–688.

Thurlow, WR, Small, AM. 1955. Pitch perception of certain periodic auditory stimuli. *J Acoust Soc Am* 27, 132–137.

von Bismark, G. 1974a. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica* 30, 146–159.

von Bismark, G. 1974b. Sharpness as an attribute of timbre of steady sounds. *Acustica* 30, 159–172.

van Noorden, L. 1982. Two-channel pitch perception. In: M Clynes (ed.), *Music, Mind and Brain*. New York, NY: Plenum, 217–225.

Ward, WD. 1953. The subjective octave and the pitch of pure tones. Ph.D. Dissertation. Cambridge, MA: Harvard University.

Ward, WD. 1954. Subjective musical pitch. *J Acoust Soc Am* 26, 369–380.

Ward, WD. 1963a. Absolute pitch. Part I. *Sound* 2(3), 14–21.

Ward, WD. 1963b. Absolute pitch. Part II. *Sound* 2(4), 33–41.

Ward, WD, Burns, EM. 1982. Absolute pitch. In: D Deutch (ed.), *The Psychology of Music*. London, UK: Academic Press, 431–451.

Wedin, L, Goude, G. 1972. Dimensional analysis of the perception of instrumental timbre. *Scand J Psychol* 13, 228–240.

Wezel, DL. 1979. Timbre space as a musical control structure. *Comput Music J* 3, 45–52.

Wightman, FL. 1973a. Pitch and stimulus fine structure. *J Acoust Soc Am* 54, 397–406.

Wightman, FL. 1973b. The pattern-transformation model of pitch. *J Acoust Soc Am* 54, 407–416.

Zwicker, E. 1955. Der Ungewohnliche Amplitudengang der Nichtilinearen Vertzerrungen des Ohres. *Acustica* 5, 67–74.

Zwicker, E. 1981. Dependence of level and phase of the $(2f_l{-}f_2)$-cancellation tone on frequency range, frequency difference, and level of primaries, and subject. *J Acoust Soc Am* 70, 1277–1288.

Zwicker, E, Fastl, H. 1999. *Psychoacoustics: Facts and Models*, 2nd ed. Berlin, Germany: Springer-Verlag.