10

Basic Psychoacoustic Quantities

As seen in the earlier chapters, the basic functioning of hearing can be characterized as a kind of time–frequency analysis of the ear canal pressure signals resembling a bank of band-pass filters. This chapter describes the psychoacoustic quantities at the lowest level of analysis: pitch, loudness, timbre, and duration, which are more or less related to the physical quantities frequency, level, magnitude spectrum, and time.

10.1 Pitch

Pitch is defined by the American National Standards Institute as 'that auditory attribute of sound according to which sounds can be ordered on a scale from low to high' (ANSI-S1.1, 2013). Pitch is perceived from many types of sounds, such as sinusoids, vocals, instrument sounds, and noisy sounds (Fastl and Zwicker, 2007; Hartmann, 1996). However, the definition is problematic in the sense that not all sounds have clear pitch, and some of them have more than one pitch. However, in the context of the basic properties of hearing, it is meaningful to restrict the discussion to relatively simple sounds producing a single, more or less salient pitch. The nearest counterpart of pitch in the physical world is the frequency of repetition of a signal portion, even though pitch depends on some other parameters as well.

Pitch is a relevant cue in speech communication. It is the primary cue that distinguishes between male, female, and child speakers. The prosody of speech is also perceived as pitch changes during spoken sentences. Of course, pitch plays a very central role in music, as melody is composed of the perception of successive changes in pitch. The pitch of sound produced by a physical object also conveys information about its geometry, physical properties, and size, for example, large barrels produce a different pitch from small cans.

10.1.1 Pitch Strength and Frequency Range

The clarity and salience of pitch perception, or the *pitch strength*, depends significantly on the nature of the sound. Pitch strength has been measured using psychoacoustic tests with different types of signal, as shown in Figure 10.1 (Fastl and Stoll, 1979; Fastl and Zwicker, 2007).

Communication Acoustics: An Introduction to Speech, Audio, and Psychoacoustics, First Edition.
Ville Pulkki and Matti Karjalainen.

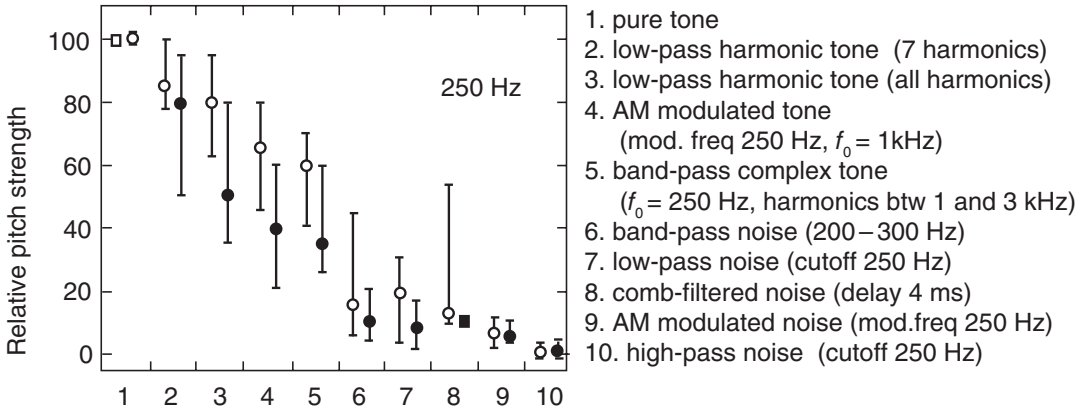

Figure 10.1 Pitch strength with different types of sounds producing a pitch of 250 Hz. The pitch strength is scaled so that 100% corresponds to a 250-Hz sinusoid. The unfilled markers show data measured using a pure tone as the reference, and filled markers denote results using comb-filtered noise (type 9) as the reference. The pre-assigned values for references (100 and 10, respectively) are marked with squares. Adapted from Fastl and Stoll (1979), and reprinted with permission from Elsevier.

The pitch strength is compared to the pitch of a 250-Hz sinusoid that serves as the reference on the scale. Periodic signals are clearly perceived to have higher pitch strength than aperiodic ones. Sinusoids produce the clearest pitch perception, closely followed by low-pass filtered harmonic complexes, both of which are perfectly periodic signals. Quite interestingly, an aperiodic sound, narrowband noise, is reported to have a pitch strength just below that of a low-pass filtered harmonic complex and a higher strength than an AM tone and band-pass filtered complex tones, both of which are periodic signals. The perception of some of the signals is analysed in detail below.

The frequency range of pitch perception has been studied by testing where musical melodies are still recognized correctly. The range covers repetition frequencies from about 30 Hz (Pressnitzer *et al.*, 2001) to 4 kHz (Plack, 2013). At lower frequencies, vibration is still perceived, but the musical tone is no longer recognized. At frequencies above 4 kHz the perception of pitch is also weak. A musical melody at these frequencies may just sound peculiar.

10.1.2 JND of Pitch

We begin our exploration of pitch perception by looking at the basic results of JND – just noticeable difference – measurements. As shown in the previous section, tones presented alone evoke the highest pitch salience, where the frequency of the tone defines the pitch in general. Thus, to get an idea of the best performance of hearing with pitch, measuring the smallest perceived change in the frequency of the tone $\Delta f = |f_1 - f_2|$, where f_1 and f_2 are the frequencies of two tones presented in succession to the subject, is of relevance. Δf then characterizes the JND of pitch. The result of such a measurement is shown in Figure 10.2. At the frequencies 250 Hz and 500 Hz, the JND is about 1 Hz, and it rises quickly to 200 Hz for 8-kHz tones (Sek and Moore, 1995). The same result plotted as a ratio with tone frequency is shown in the right panel. This shows that the ratio is at best about 0.3% and clearly rises at frequencies above 2 kHz; it also seems to rise at frequencies below 500 Hz. The poorer relative JND at higher frequencies can be explained simply by the neural loss of phase synchronization, while at low frequencies the decreasing sensitivity of hearing explains the degrading results.

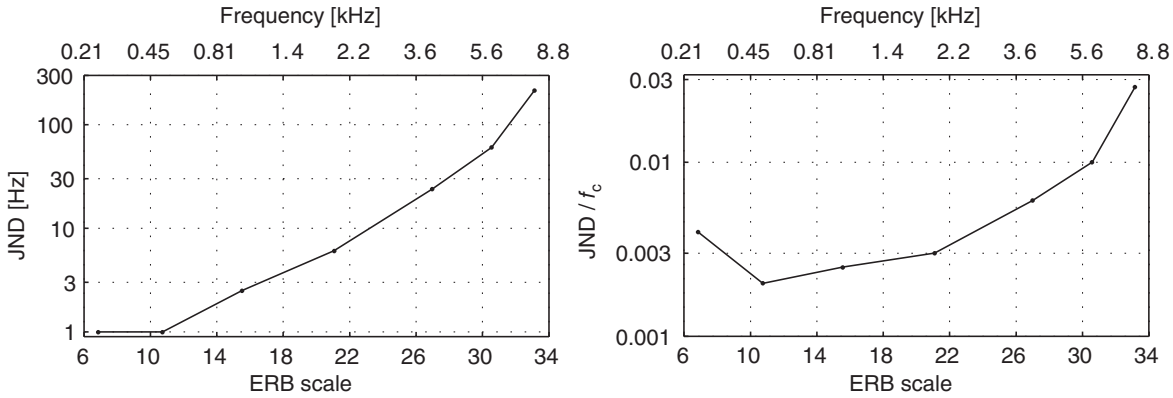

Figure 10.2 The left panel shows the JND of frequency for two successively presented tones, with the frequency difference on the *y*-axis, as a function of the mean frequency of the tones. The same data are plotted as a ratio between the JND and the mean frequency in the right panel. Adapted from Sek and Moore (1995).

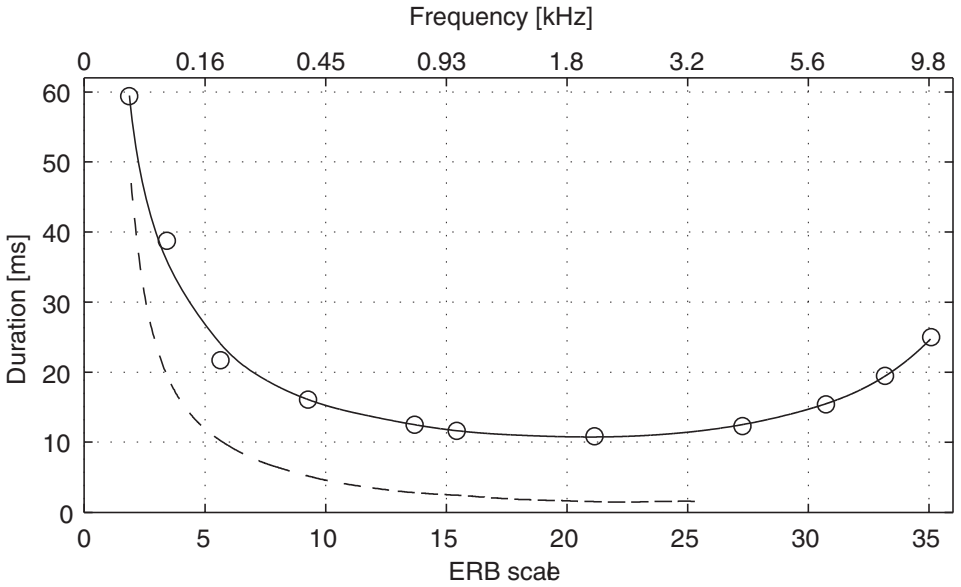

Figure 10.3 The minimum length of a tone burst required for pitch perception. The dashed line shows the length of two periods depending on frequency. Adapted from Burck *et al.* (1935).

10.1.3 Pitch Perception versus Duration of Sound

The formation of the percept of pitch and improving its precision requires a sample of a signal with non-zero length. In theory, at least two periods of a periodic signal are necessary before the pitch percept can be formed. The minimum length of a tone burst after which a pitch is perceived is shown in Figure 10.3. The dashed line shows the length of two periods of sine signals. The shortest time window required for pitch perception is seen to be between 400 Hz and 6 kHz, where the length is less than 20 ms. The performance of human hearing in pitch perception is remarkable at low frequencies; the required length for a tone burst is only slightly longer than two periods.

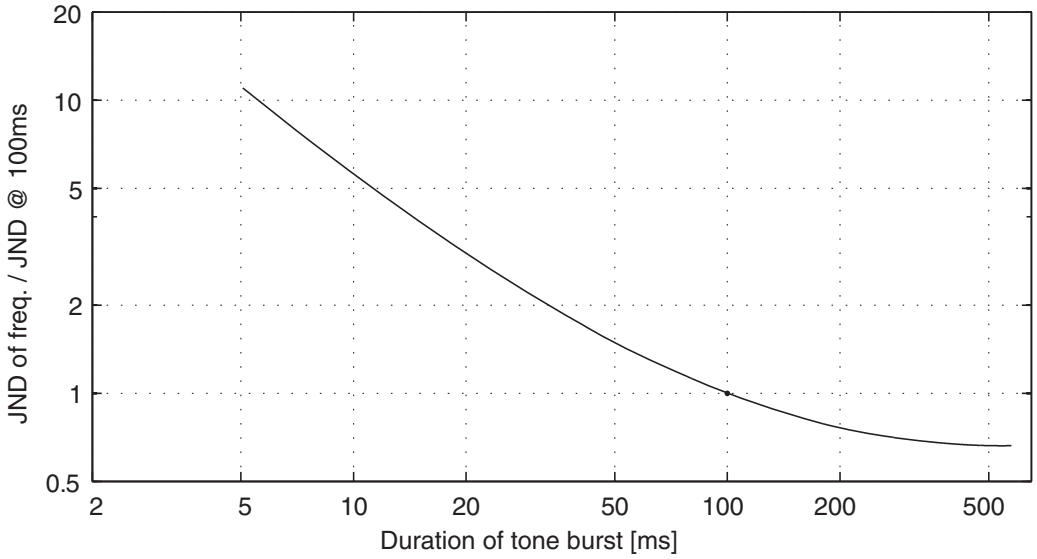

Figure 10.4 The normalized JND of pitch as a function of the duration of a 2-kHz tone. Adapted from Fastl and Zwicker (2007).

The best accuracy in pitch perception is achieved 100–200 ms after the onset of the sound. The normalized JND of pitch as a function of 2-kHz tone length is plotted in Figure 10.4. It can be seen that the JND decreases only slightly when increasing the duration to over 100 ms, and the JND is practically constant after 200 ms. The variation in the JND with duration differs somewhat across frequencies (Moore, 1973).

10.1.4 Mel Scale

A frequency scale was created using psychoacoustic tests to quantify pitch. In the tests, the task of the subject was to adjust a tone or a harmonic complex to a specified ratio of the reference sound for a range of ratios to obtain a subjective scale. An example task would be 'adjust the pitch of the test tone to be two times higher than the reference tone.' The phrase 'two times higher' may seem arbitrary, and different people may interpret the task differently. For example, raising the pitch of a musical note by one octave is commonly considered to 'double the height of the note'.

The scale thus obtained is called the *mel scale*, and the name of the unit is the *mel*, which comes from the English word *melody*. It turns out that when subjects are asked to double the pitch they double the frequency, on average, for frequencies below about 500–1000 Hz (Fastl and Zwicker, 2007). Above this frequency limit, subjects adjust the frequency to increase in considerably larger steps than the pitch (see Figure 10.5). It can be seen that when doubling the pitch of a 2-kHz tone, subjects have adjusted the frequency to about 15 kHz. This can be explained by the low sensitivity of hearing to pitch at high frequencies due to the inability of the cochlea to lock onto phases of tones there.

The mel scale can be approximated at low frequencies with relatively simple equations, such as

$$m = 2595 \log_{10}(1 + f/700), \tag{10.1}$$

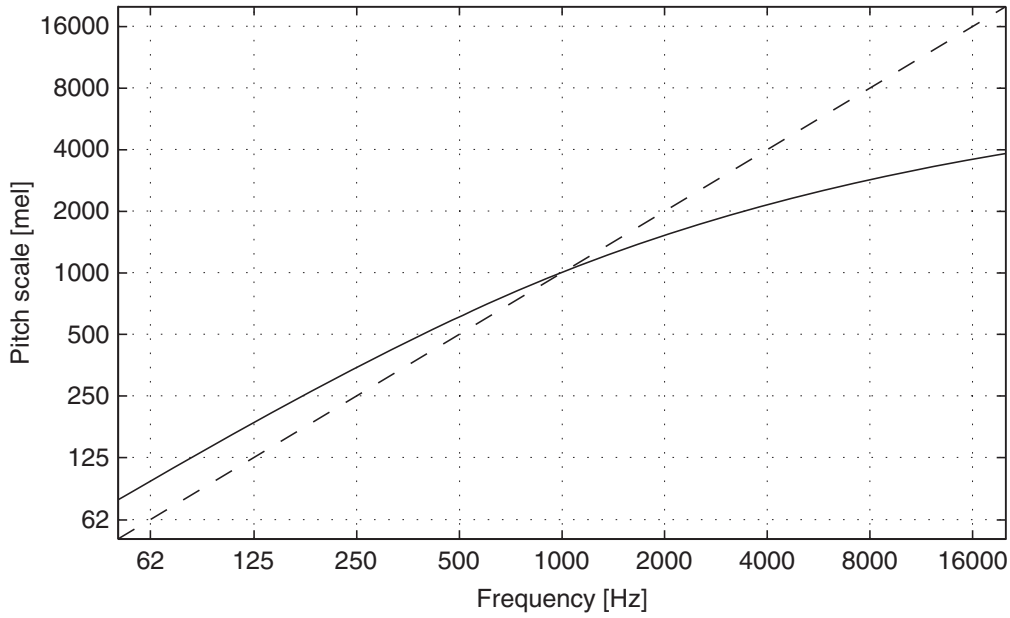

Figure 10.5 The mel scale defined with Equation (10.1) plotted as a function of frequency in Hz. The anchor point is 1000 mel ↔ 1000 Hz. The dashed line shows, for reference, the linear relationship with a slope of one.

where m is the mel value and f is frequency in Hz. The anchor point in the equation is 1000 Hz ↔ 1000 mel. 1 mel equals to there are also other equations proposed for this purpose.

10.1.5 Logarithmic Pitch Scale and Musical Scale

As will be discussed later in this book, musical notes can, in most cases, be interpreted simply as harmonic complexes. The pitch scales form sets of notes with different fundamental frequencies. The scales are logarithmic universally, since the distances between musical notes are defined as ratios between their fundamental frequencies. In simpler terms, when the fundamental frequencies of two musical notes have the ratio 2:1, their distance in frequency, or *interval*, is called an octave. Octaves are perceived to have similar width in pitch in different frequency areas, especially between about 100 and 1000 Hz. There are a number of methods to define the fundamental frequencies of notes within an octave, which will be discussed in Section 11.6.2. A simple method is equal temperament, where an octave is logarithmically divided into 12 semitones, and the intervals between successive notes are set as $\sqrt[12]{2}$.

The logarithmic frequency scale plotted against the piano keyboard illustrating the diatonic scale used in music is shown in Figure 10.6. The widths of intervals – octave, fifth, major third, and semitone – are also shown. The reference frequency commonly used in Western music is 440 Hz for the note A. The reference has changed with time, and the overall tendency has been upwards in pitch to obtain a brighter sound for classical orchestras (Rossing *et al.*, 2001). The figure also shows the relative frequency of occurrence of different tones in five randomly selected piano sonatas by Beethoven. It is clear that most of the notes reside in the frequency region between 50 Hz and 2000 Hz, although outliers also exist. Since the piano is an instrument with one of the largest ranges of notes, this frequency region can be taken in general as an example of the most commonly found range of pitches in tonal music.

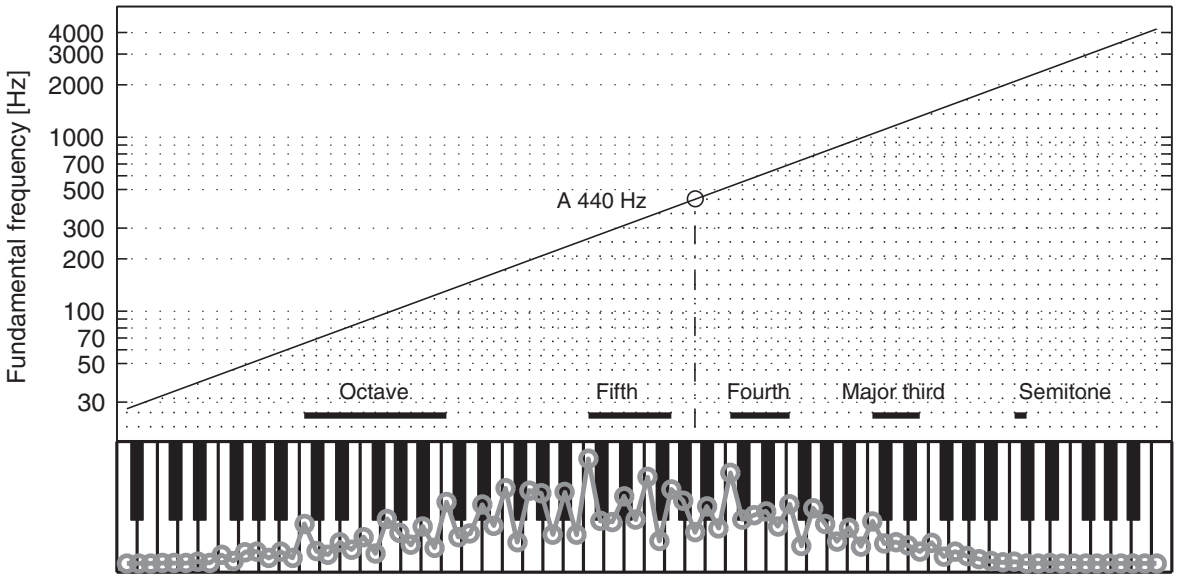

Figure 10.6 The logarithmic frequency scale and the diatonic scale illustrated as a piano keyboard. The grey bold line imposed on the keyboard illustrates the relative frequency of appearance of each note in five randomly selected Beethoven piano sonatas.

The mel scale defined in the previous section is essentially a linear scale up to about 1000–2000 Hz. It is thus in conflict with the logarithmic definition of musical scales. In addition, the JND of the frequency of a sinusoid is about 0.3% from 250 Hz to 2 kHz instead of being a constant value expressed in Hz, which bolsters use of the logarithmic pitch scale in this frequency region. Although the mel scale does not correspond to the musical pitch scale, it is still useful in some technical applications, such as speech recognition.

10.1.6 Detection Threshold of Pitch Change and Frequency Modulation

The JND of frequency depending on the frequency of the tone was shown earlier in Figure 10.2. Often, the frequency of continuous sound changes with time and varies around a centre frequency. The changing of frequency may or may not be audible as a change in pitch or some other perception. This change in frequency can be studied using frequency modulation. The JND of frequency modulation of carrier tones at different frequencies is shown as a function of modulation frequency in Figure 10.7.

A practical example of the perception of frequency modulation is present in mechanical audio devices such as LP players caused by the fluctuations in speed of the player. Figure 10.7 shows that humans are most sensitive to modulations at 4 Hz (Demany and Semal, 1989; Fastl and Zwicker, 2007).

10.1.7 Pitch of Coloured Noise

In addition to periodic signals, non-periodic signals may also give rise to a perception of pitch in many situations. A band-pass filtered noise signal with a single spectral peak is perceived to have pitch corresponding to the centre frequency of the peak. The salience of pitch perception is better the narrower the peak is.

A low-pass or high-pass filtered noise signal with steep roll-off causes a perception of pitch which matches the cutoff frequency. In the case of band-pass filtered noise, one or

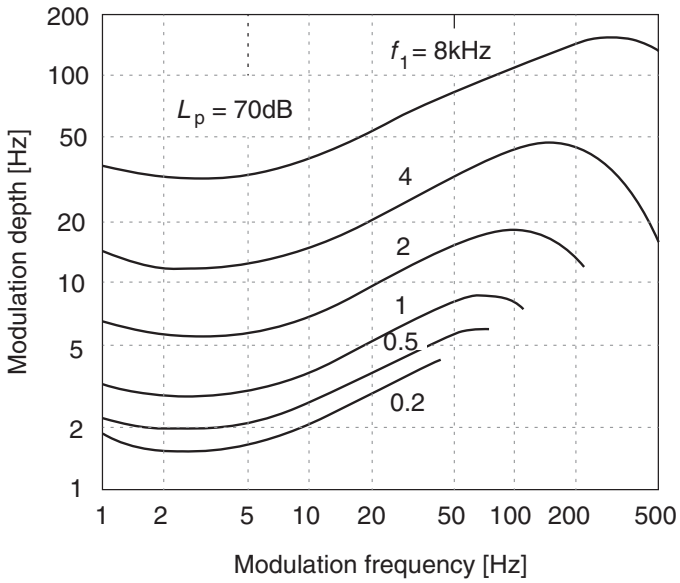

Figure 10.7 A schematic presentation of the detection threshold of frequency modulation as a function of the modulation frequency for different carrier frequencies. Adapted from Demany and Semal (1989).

two pitches may appear depending on the frequency characteristics. If the cutoff frequencies are far enough from each other and the roll-off curves are steep enough, two pitches may appear. Otherwise, for sufficiently shallow roll-off curves, a single pitch emerges, corresponding either to the centre of gravity of the sound energy in frequency or to one of the edge frequencies.

The preceding examples pertained to cases where the frequency spectrum of sound contained peaks or other variations. A weak pitch percept is also generated by modulating sinusoidally the amplitude of white noise (sinusoidally amplitude-modulated noise, or SAM noise). The pitch is perceived at the frequency of amplitude modulation, although the long-term spectrum of sound is flat.

10.1.8 Repetition Pitch

An example of pitch generated by coloured noise is *repetition pitch*. In repetition pitch, a broadband signal such as noise or a rapid transient is delayed and summed to the non-delayed signal. The perceived pitch corresponds to the frequency that is the reciprocal of the delay. In signal processing this is known as the *comb-filter effect*, where the spectrum of the signal has evenly spaced peaks and valleys. The frequency of the first peak in the comb-filter spectrum corresponds to the repetition pitch (Bilsen and Ritsma, 1969).

Repetition pitch often occurs outdoors, where a noisy sound is reflected from a wall or from hard ground and is summed to the direct sound in the ear canals. Relatively often it is simply perceived as colouring of sound and not as pitch. However, if the delay changes with time, as when the source or the receiver moves, the pitch is perceived more prominently.

The frequency of a single, sharp, moving spectral peak can be perceived as pitch, where the signal itself can also be a harmonic complex. An example of this phenomenon is the ‘wah wah’ effect pedal commonly used with guitars, which effectively performs peak filtering with a high-Q filter, the centre frequency of which is controlled. Another example is overtone singing, as described in Section 5.4.

10.1.9 Virtual Pitch

Pitch perception in all the cases presented so far, except for the SAM noise case, can be explained using the place theory of pitch perception, where the location of the maxima in the place-dependent response of the basilar membrane is responsible for the percept. However, as shown in Figure 7.12, the activation of the basilar membrane spreads to a larger range with an increase in level of the stimulus, and one could expect that the perceived pitch would change noticeably if only the maximum of the activation in frequency defined the perception of pitch. Such a noticeable change in pitch depending on the level does not occur, and another process is likely to be involved in pitch perception. There exist many other cases where the place theory does not at all explain the perception. A simple case is a harmonic sound, where the partials at higher frequencies are dominant. The perceived pitch follows the fundamental frequency in many cases, although the centre of gravity of the spectrum is elsewhere (Plack *et al.*, 2005).

An important special case of this is the phenomenon of perceiving the *missing fundamental*. If the lowest partial(s) of a harmonic tone complex are missing, the pitch of the sound still corresponds to the fundamental frequency of the tone complex. The sound is perceived somewhat differently, as if the colour of sound is a bit thinner, and the pitch strength may be weaker, but in many cases the pitch does not change. The perception of pitch as being the frequency of the missing fundamental is an example of *virtual pitch*, also known as *residue pitch*.

The mechanism leading to the perception of the missing fundamental can, at least partially, be understood by considering the half-wave rectification performed by the hair cells in the cochlea. If more than one partial is present in one critical band, non-linearity in the cochlea creates the presence of sum and difference signals in the output of the band. In practice, the fundamental frequency is present as amplitude modulation in the output. This notion calls for some sort of periodicity analysis that is applied to the auditory band outputs. The result of the analysis is then perceived as pitch.

10.1.10 Pitch of Non-Harmonic Complex Sounds

Pitch perception becomes an even more complex phenomenon if the sound consists of partials that are not necessarily whole number multiples of a common fundamental frequency. For example, church bells and other vibrating bars and plates generate tone complexes with partials in more or less inharmonic succession. There may be one or more pitches perceived from the bell sound, and the pitches may correspond to some high-level partials in the spectrum, or they may be virtual pitches generated by some partials.

An interesting experiment has been conducted to test the validity of place and timing theories of pitch perception. In the experiment, a set of harmonic partials is generated; for example, three to ten harmonics along with the fundamental frequency of, say, 100 Hz. Thus, the spectrum spans from 300 Hz to 1000 Hz in intervals of 100 Hz. If all partials are shifted upwards by 30 Hz, the output envelopes of hair cells should still have a strong modulation with 100 Hz frequency. However, the pitch perception rises a bit, even though the modulation in each auditory band remains the same (Plack *et al.*, 2005). It seems that the more complex the spectral relationships of the partials are, the harder it is to explain the perception using simple models.

10.1.11 Pitch Theories

Pitch is a remarkably stable percept. The SPL of sound and the direction of sound source do not have a noticeable effect on it. The only factors that seem to affect pitch perception are the magnitude spectrum and repetitive structures in the ear canal signals.

The mechanism leading to pitch perception has been explained by two theories: the *place theory* and the *timing theory* (Plack *et al.*, 2005).The place theory assumes that the frequency–place mapping occurring in the cochlea explains pitch perception. Unfortunately, this does not explain all pitch phenomena, and in some cases it is clear that a time-domain analysis of periodicity is also performed in hearing. The current assumption is that hearing analyses pitch using both mechanisms, but, unfortunately, they are not known well enough to be able to construct a functional model of the pitch perception mechanisms for all cases, although some quite advanced models already exist (Plack *et al.*, 2005).

10.1.12 Absolute Pitch

Human hearing is very accurate when comparing the characteristics of two sounds, but quite inaccurate in making absolute assessments of the characteristics of sounds in isolation. This is also true in the perception of pitch. An interesting exception is the ability of some people to evaluate pitch on an absolute scale without any reference, this ability is called *absolute pitch* (Rossing *et al.*, 2001).

Although musicians can improve their ability to assess pitch, the ability of a person with absolute pitch is about ten times more precise in assessing pitch than a person who is trained but does not naturally have this ability. It is as if people with this ability have an inner reference frequency. It has been known for this reference to change with increasing age. Some people also claim that the pitch of recorded music has changed with time. Absolute pitch is not necessarily connected to musical giftedness, though undoubtedly it may help an otherwise musical person in their musical training.

10.2 Loudness

Loudness is 'that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud' (ANSI-S1.1, 2013). Loudness perception is a relatively complex, yet consistent phenomenon. The theory describing loudness perception is one of the central theories of psychoacoustics. The theory is used below to explain different phenomena revealed by psychoacoustic experiments using different kinds of stimuli, starting with simple sinusoids and continuing gradually to more complex cases.

10.2.1 Loudness Determination Experiments

As with the concept of pitch, the concept of loudness is easiest to understand by first looking at how we perceive sinusoidal signals. Subjects are given the task of adjusting a test tone with the same frequency as the reference tone to two times (or half) the loudness of the reference tone. The adjusted tone then becomes the reference tone, and the process is continued. There may be individual differences in terms of how the phrase 'twice as loud' is interpreted, and the adjusted level may be different in each trial (Fastl and Zwicker, 2007). However, when the results from a large set of subjects and from many repetitions of the same task are averaged, and when an anchor point is selected, a psychophysical function is obtained which describes the relation between the level of a tone and perceived loudness. Such a function is presented in Figure 10.8 for a 1-kHz tone. The result is also presented for the same test made with white noise.

The unit of loudness, the *sone*, is defined by stating that a loudness of 1 sone is equivalent to the loudness of a 1-kHz tone at 40 dB SPL.

Figure 10.8 shows that the sound pressure level does not estimate perceived loudness directly, since different signal types at equal sound pressure levels do not generate an equal loudness

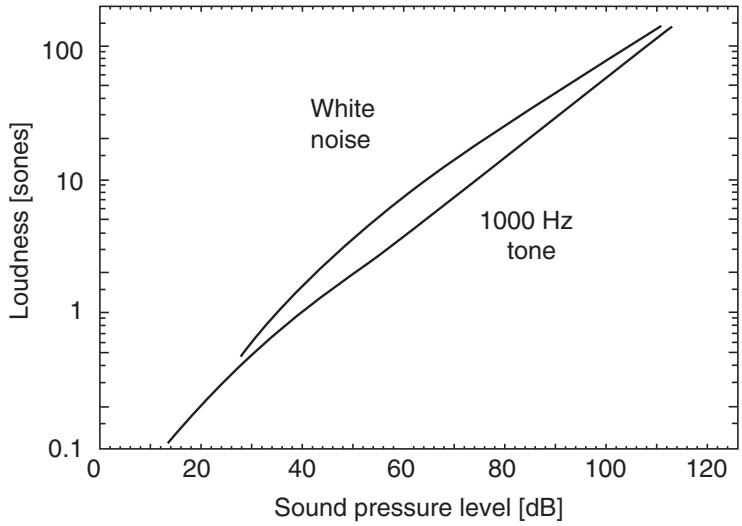

Figure 10.8 The loudness of a 1-kHz tone and white noise as a function of the sound pressure level. Reprinted from Canteretta and Friedman (1978) with permission from Academic Press.

perception. The figure also shows that the loudness of white noise is higher, although its sound pressure level is equal. This effect is actually a manifestation of the same phenomenon that was used to measure Bark bands in Section 9.4.1; the perceived loudness increases when the band-pass spectrum is made wider while keeping the sound pressure constant.

10.2.2 Loudness Level

The equal loudness curves presented in Figure 9.2 on page 155 show that the perception of loudness depends on the level and frequency of sinusoidal signals in a manner which cannot be easily expressed in mathematical terms. However, the curves can themselves be used to define the *loudness level*. The curves represent sound pressure levels of tones that are perceived to have equal loudness with reference values located at 1 kHz with 10 dB spacing in the sound pressure level. The unit of loudness level is the *phon*. It is defined such that at 1 kHz the sound pressure level in dB and the loudness level in phons have the same magnitudes.

The linear dependence between the SPL in dB and the loudness level in phons is different from the dependence between SPL and loudness in sones introduced in the previous section. The SPL in Figure 10.8 is expressed in dB and the loudness is expressed in sones on a logarithmic scale. A 1-dB change in SPL thus produces very different changes in the value of loudness in sones at low and high levels of the SPL. Technical applications often use the loudness level, since the JND of the SPL of sound events is constant at about 1 dB (or 1 phon) over a wide range of SPLs (Fastl and Zwicker, 2007). A change in a certain number of phons thus corresponds to a similar change in the perceived 'strength' of an auditory event, which is beneficial in auditory models and sound and voice techniques.

10.2.3 Loudness of a Pure Tone

The result of the 1-kHz loudness listening test discussed in Section 10.2.1 is shown in Figure 10.9, where one can see that at levels higher than 40 dB the loudness is doubled with an increase in level of 10 dB. This can be expressed as

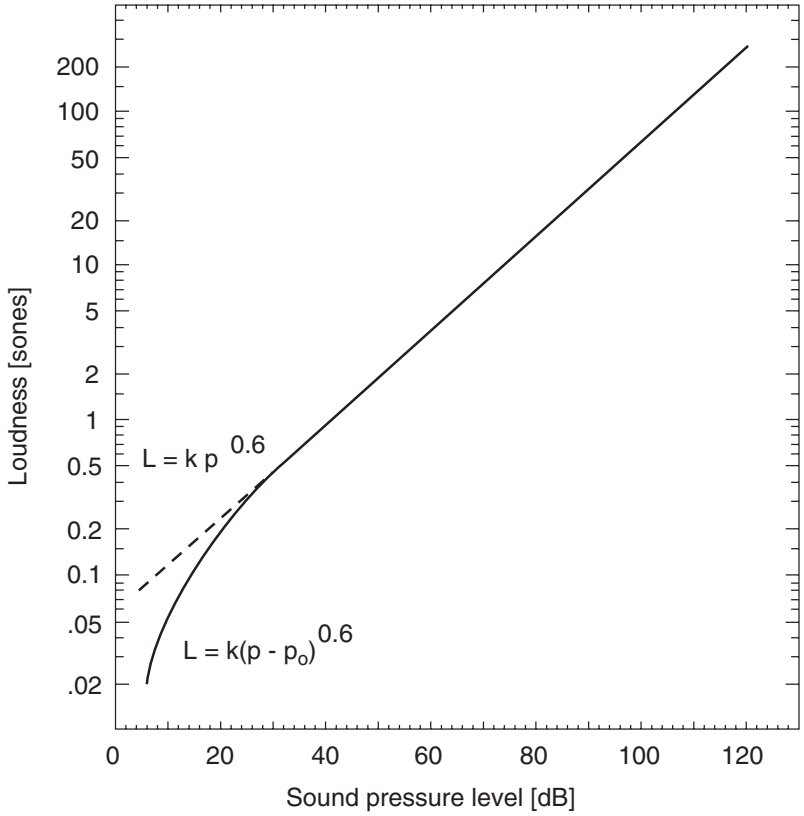

Figure 10.9 The loudness of a 1-kHz tone in sones as a function of the sound pressure level. Reprinted from Canteretta and Friedman (1978) with permission from Academic Press.

$$N = 2^{(L_L - 40)/10}, \tag{10.2}$$

where N is the loudness in sones and L_L is the loudness level of a tone at 1 kHz. Note that at 1 kHz, $L_L = L_P$ by definition, where L_P is the sound pressure level in dB. At levels below 40 phons, the curve is steeper. A theoretical relation between sound pressure and loudness can thus be derived as

$$N = \mathrm{k} \cdot p^{0.6}. \tag{10.3}$$

The curve can also be fitted to the results at low levels by introducing p_0, the pressure level near the threshold of hearing, into the equation:

$$N = \mathrm{k} \cdot (p - p_0)^{0.6}. \tag{10.4}$$

The constant k must be selected so that the loudness at a 40-dB sound pressure level matches 1 sone.

When the test sound is partially masked due to the presence of white noise, the perceived loudness of the signal is also affected, as shown in Figure 10.10. The threshold above which Equation (10.2) holds rises with the level of masking noise. When the level of the masker is increased, the loudness level of the tone decreases gradually before becoming inaudible.

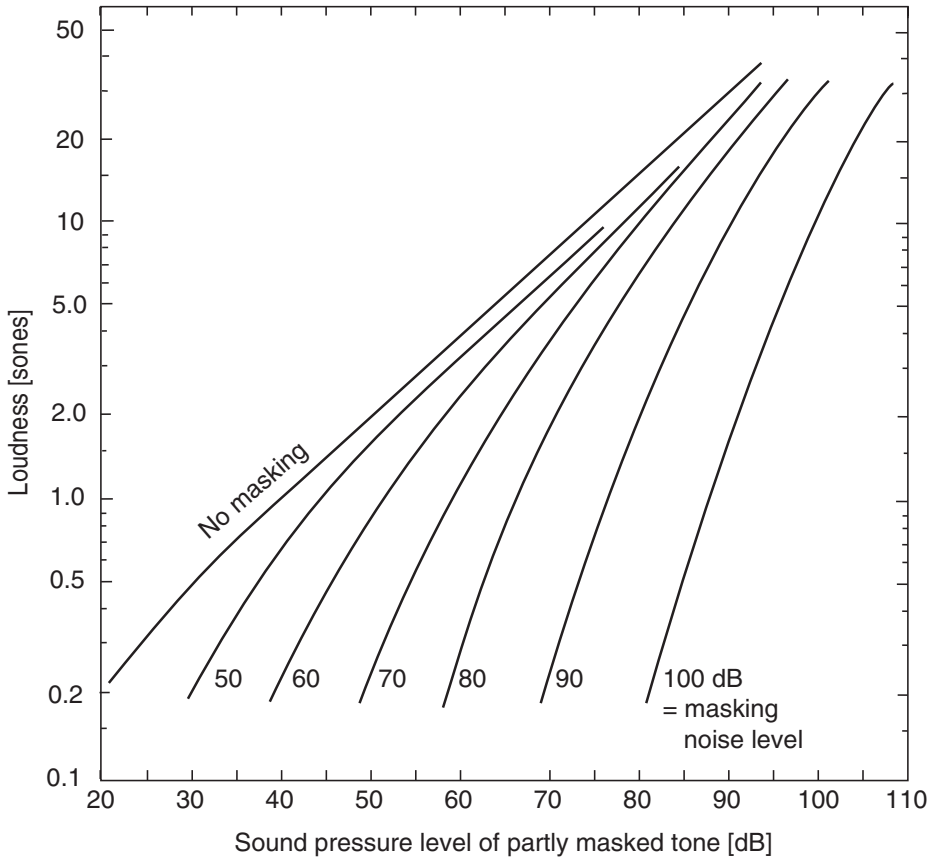

Figure 10.10 The loudness of a tone as a function of its SPL in the presence of masking white noise with tabulated SPL. Reprinted from Canteretta and Friedman (1978) with permission from Academic Press.

10.2.4 Loudness of Broadband Signals

A broadband sound is perceived to be louder than sound with a narrower band of equal sound pressure level, as shown by both the phenomenon used to define Bark bands in Section 9.4.1 and the higher loudness generated by white noise than a sinusoid shown in Figure 10.8. Our hearing thus analyses the loudness of different auditory bands separately and adds the obtained results to form a total loudness for the system, which turns out to be higher than it should be in principle. This implies that the mechanism detecting loudness accounts for the effect of the inherent broadening of the excitation of narrowband signals in the basilar membrane. It can be assumed that the mechanism uses the broadening as a loudness cue, and higher loudness is analysed with broader frequency content. This makes the system measuring the loudness dependent on the signal bandwidth itself, which could be thought to be an issue. However, very possibly this does not cause any deficiencies in the perception of the surrounding world, since no large evolutionary advantage can be seen to occur if the loudness of sound was not dependent on the spectral content of sound. On the other hand, an important task of hearing is to estimate if a sound source is approaching or if its distance is changing in general. In such cases, the loudness of the auditory event is compared to earlier similar events, and the relative loudness gives the cue of distance. Such a comparison does not require an accurate estimate of the sound pressure level, and thus a relative difference is already very usable.

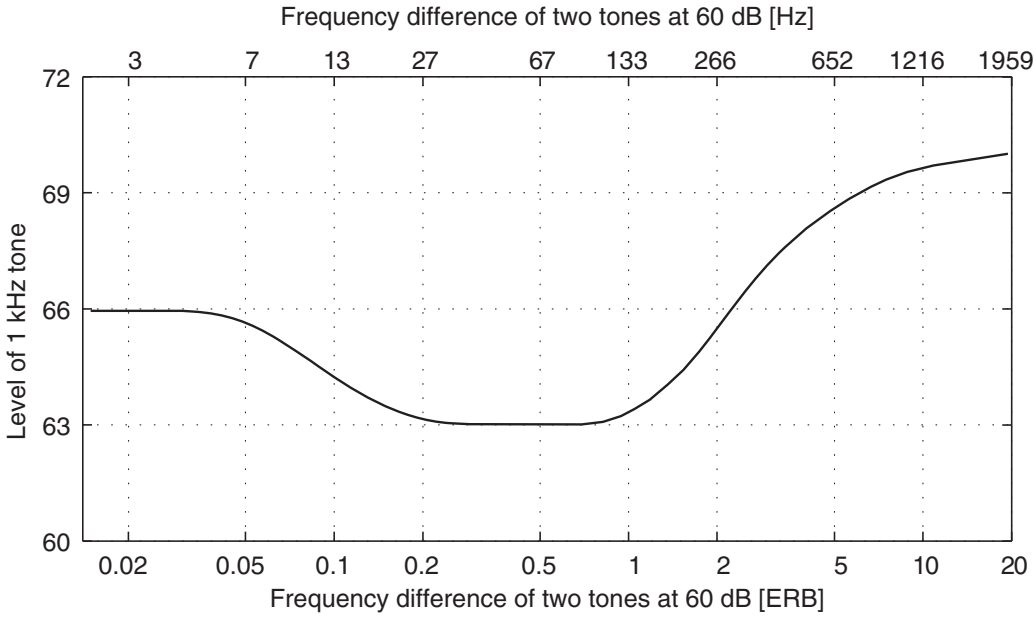

Figure 10.11 The level of a reference tone adjusted to match the loudness with a pair of tones having the frequency difference shown in the abscissa. The level of each of the tones in the pair is 60 dB. Adapted from Fastl and Zwicker (2007).

The formation of the loudness percept on the frequency axis can be investigated with different psychoacoustic tests. Figure 10.11 shows a result where the level of a 1-kHz tone was adjusted to match the perceived loudness with the loudness generated by a pair of concurrent tones. When the frequency of a pair of sinusoids differed less than a few hertz, the level of the reference tone was raised by 6 dB from the level of the non-summed tones. This effect can be explained by assuming that hearing reacts to the instantaneous maximum levels in the input and, due to beating, the maximum corresponds to a direct sum of the amplitudes (6 dB equals double the amplitude). When the frequency difference is from 20 Hz to 160 Hz, the adjusted level of the reference tone is only 3 dB higher than the non-summed test tones, which corresponds to a doubling of the power of the sound. The hearing can then no longer follow the maxima of beating, but only the average of the signal, corresponding to the sum of the powers. At frequency differences above 160 Hz, the tones appear in different critical bands, and the adjusted level of the reference tone approaches 10 dB, which corresponds to the sum of the loudnesses (Fastl and Zwicker, 2007). However, some aspects of this explanation are not supported by data shown by (Moore *et al.*, 1999).

10.2.5 Excitation Pattern, Specific Loudness, and Loudness

As already discussed, broadband sound with SPL equal to a narrowband sound is perceived to be louder and that sound in one frequency region affects the perception at other frequencies. Thus, we have to differentiate between the *excitation pattern*, *specific loudness*, and *loudness* (Fastl and Zwicker, 2007).

Let us assume that a sound stimulus entering the hearing mechanism has a power spectral density $S(f)$ on a linear scale. As the cochlea processes $S(f)$ on the auditory frequency scale,

a change of frequency axis is needed. The resulting spectrum is $S'(z)$, and the change can be computed on an arbitrary frequency scale z as

$$S'(z) = S[f(z)]\frac{\mathrm{d}f}{\mathrm{d}z}, \tag{10.5}$$

$f(z)$ is a function that transforms the linear frequency scale to an auditory one, for example, the inverse of Equation (9.4). The derivative $\mathrm{d}f/\mathrm{d}z$ represents the change in spectral density, as the density grows at higher frequencies.

The effect of each narrowband component of the input signal spreads in frequency in a manner similar to the effect of the narrowband masker in Figure 9.6. This spread can be characterized by a *spreading function* $B(z)$, which depends both on frequency and level, like the masking curves. For average levels of speech (about 60 dB SPL), a single-peak function has an approximate peak (at -3 dB) of about 1 ERB (or Bark), and the slopes are about 8 dB/ERB (10 dB/Bark) towards high frequencies, and about 20 dB/ERB (25 dB/Bark) towards low frequencies. The closest physiological correlate of the spreading function is the envelope of the vibration on the basilar membrane when excited by a single sinusoid.

The *excitation pattern* $E(z)$, that is, the effect of $S'(z)$ spread on the basilar membrane, can be thought to result from a convolution between the input signal power spectrum S' and the spreading function B,

$$E(z) = S'(z) \star B(z). \tag{10.6}$$

Note that the convolution, or the filtering operation, is computed here between two functions depending on auditory frequency.

The *specific loudness* N' can then be computed from the excitation pattern by scaling:

$$N'(z) = \mathrm{c}\, E(z)^{0.23}. \tag{10.7}$$

The constant c is chosen such that the SPL of 40 dB with a 1-kHz sinusoid results in 1 sone of loudness, and the exponent 0.23 is chosen so as to double the loudness when the SPL of a sinusoid increases by 10 dB.

The last concept, *loudness*, expressed in sones, is computed by integrating specific loudness over all the critical bands:

$$N = \int_0^M N'(z)\mathrm{d}z, \tag{10.8}$$

where M is the number of critical bands. If desired, the loudness level can be set to 0 sones at the hearing threshold with the following computation:

$$N = \mathrm{c}\int_0^M \max\{[E(z) - E_0(z)]^{0.23}; 0\}\mathrm{d}z, \tag{10.9}$$

where $E_0(z)$ is the excitation pattern corresponding to the hearing threshold, which can be interpreted as the background noise level of the hearing system.

Figure 10.12a shows the excitation patterns caused by a sinusoid (1 kHz, dashed line), and by uniform masking noise (solid line). Figure 10.12b, in turn, shows the corresponding specific loudness patterns in sones/ERB. The excitation pattern produced by the uniform masking noise follows the sensitivity function of hearing. The figures also show, at least in principle, how much the sounds mask each other, shown as the area with the cross hatching in Figure 10.12b.

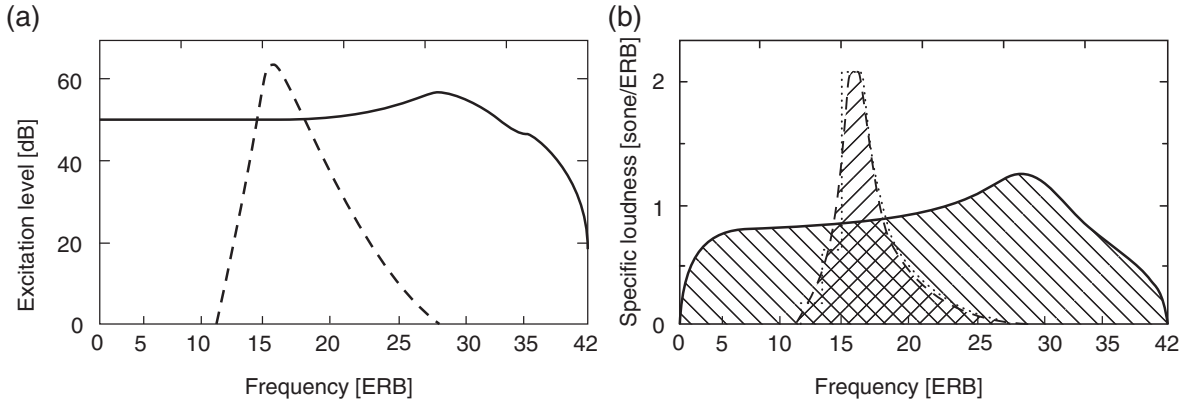

Figure 10.12 (a) Excitation patterns created by a sinusoid (dashed line) and by uniform masking noise (solid line) and (b) the corresponding specific loudness functions. Adapted from Fastl and Zwicker (2007).

These concepts are central to the understanding of the formation of the perception of loudness, and they are also widely used in computational models of loudness perception. Such models are discussed in more detail in Chapter 13.

10.2.6 Difference Threshold of Loudness

As discussed above, loudness depends on many acoustic parameters of a sound event, such as level, spectral content, and duration. It is also interesting to quantify the JND of loudness to see how small a change is detected at all. This has been researched as the JND of level, which has a very strong influence on loudness.

The detection threshold of amplitude modulation indicates some properties of human ability to perceive changes in loudness. As will be shown below, the greatest sensitivity is obtained when the modulation frequency is about 4 Hz. The detection threshold of the level of modulation of a 1-kHz tone and white noise is shown in Figure 10.13 as a function of the level of the signal. When the level of the carrier signal is increased from silence, the threshold decreases monotonically. With relatively low levels (20–50 dB) the detection threshold is about 1 dB, and at higher levels the threshold keeps decreasing, reaching about 0.2 dB at 100 dB. Interestingly, when the same test is conducted with white noise, the detection threshold is constant at all levels above 25 dB.

Interestingly, the detection threshold of an amplitude-modulated tone depends on its level, but such dependence is not found with noise. This can be understood when the excitation pattern of the tone on the basilar membrane is investigated in the cases of amplitude and frequency modulation. Schematic diagrams of these cases are shown in Figure 10.14. In the case of amplitude modulation, the shape of the excitation pattern depends on the level such that at maximum levels of the signal a much broader frequency range is covered than with the lowest levels. This implies that when the level is modulated, the frequency range is also modulated, which provides an additional cue to detect the modulation. The frequency coverage is, on the other hand, constant in the case of frequency-modulated tones, and such an additional cue is not available.

The dependence of the detection threshold of the modulation index as a function of the modulation frequency is shown in Figure 10.15. When the carrier signal is a tone, the result is similar to that obtained with frequency modulation in Figure 10.7. The threshold level

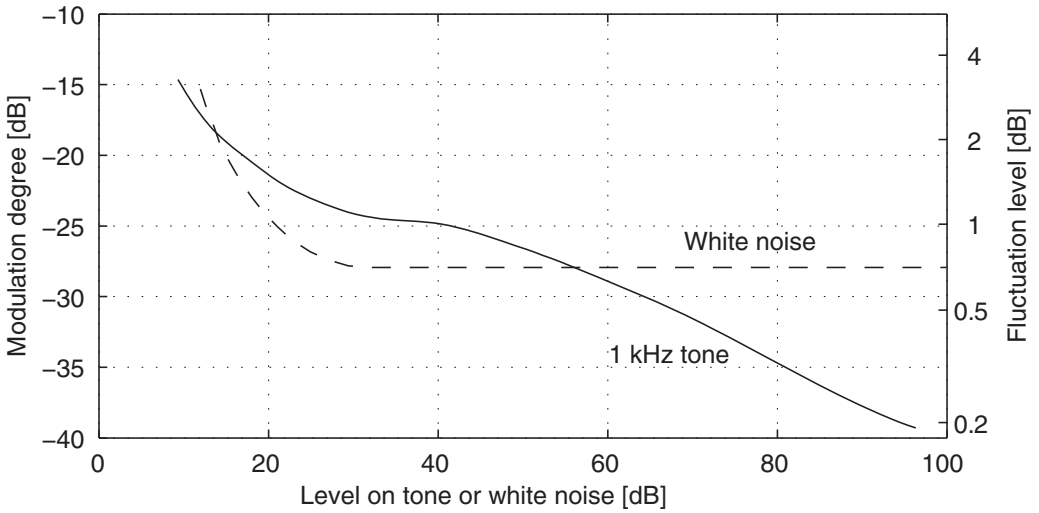

Figure 10.13 The just noticeable level of amplitude modulation of a 1-kHz tone (solid line) and white noise (dashed line). The modulation degree m (see Equation (3.4)) is expressed in dB on the left-hand y-axis, and the resulting fluctuation of level is shown on the right-hand y-axis. The modulation frequency is 4 Hz. Adapted from Fastl and Zwicker (2007).

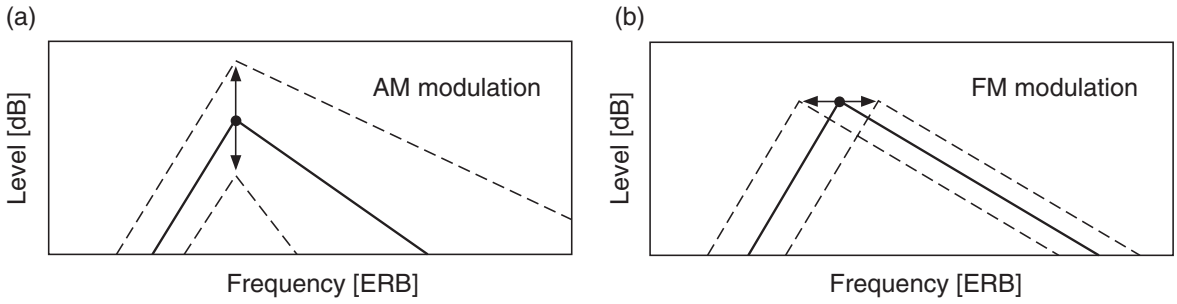

Figure 10.14 Excitation patterns in the cochlea for (a) an amplitude-modulated and (b) a frequency-modulated tone. Adapted from Fastl and Zwicker (2007).

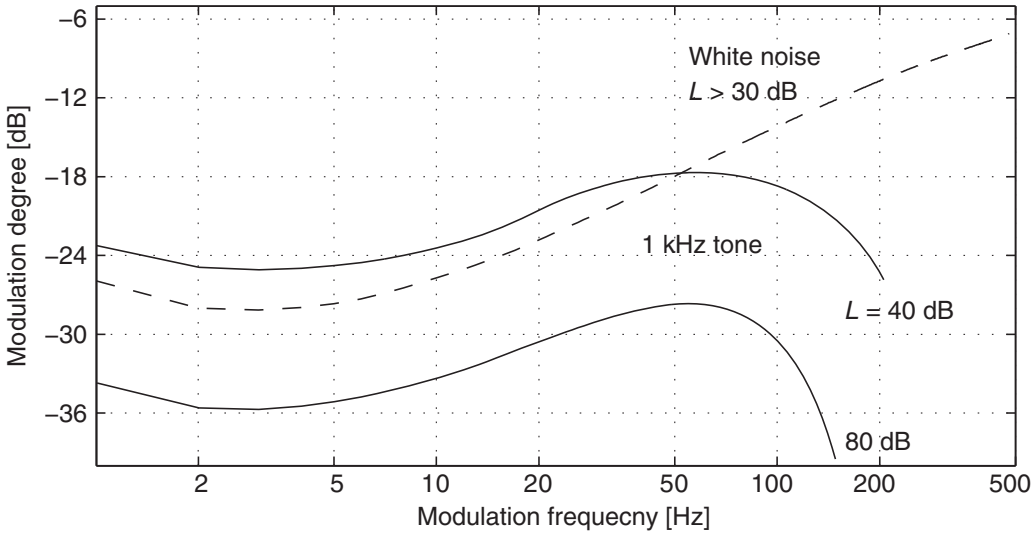

Figure 10.15 A schematic presentation of just noticeable amplitude modulation of a 1-kHz tone (solid line) at two different levels, and of white noise (dashed line) as a function of modulation frequency. Adapted from Fastl and Zwicker (2007).

is at its lowest with a modulation frequency of 4 Hz, and increases for other frequencies (Fastl and Zwicker, 2007). In the case of a sinusoidal carrier, when the modulation frequency reaches about half the width of the critical band, the threshold starts to decrease again. As the spectrum of the modulated signal spreads to an area wider than one critical band, hearing detects the change more easily. For a white noise carrier, the signal is broadband to begin with, and the spectral effect of modulation is not detectable.

10.2.7 Loudness versus Duration of Sound

The loudness percept requires some time to build up. The loudness level caused by a 2-kHz tone burst is shown in Figure 10.16. When the length of the burst decreases from 100 ms, the loudness decreases 10 phones each time the length is decreased to one tenth, indicating that the loudness percept is formed by integrating sound energy over time. This operation performed by our hearing is called *temporal integration*. For a sound burst lasting longer than 200 ms, loudness no longer increases (Fastl and Zwicker, 2007).

A similar temporal effect of a change in the level of sound is seen in measurements of the JND, as shown in Figure 10.17. When the burst length is decreased from 200 ms, the detection

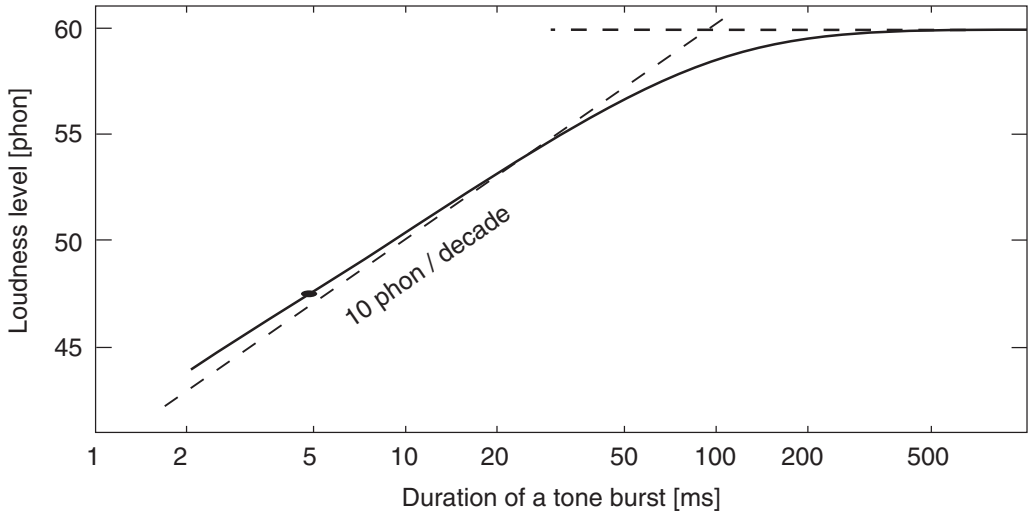

Figure 10.16 The dependence of loudness level on the duration of a tone burst with a frequency of 2 kHz and a sound pressure level of 57 dB. Fastl and Zwicke 2007.

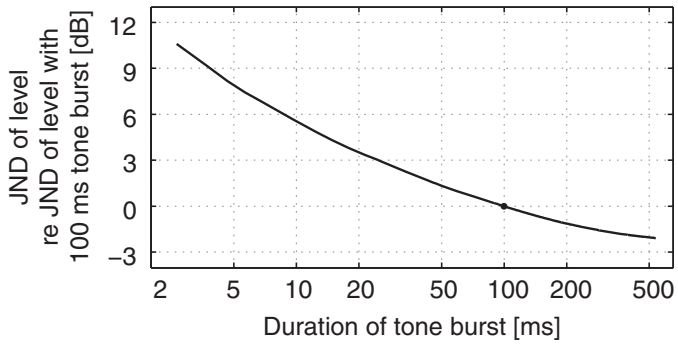

Figure 10.17 Just noticeable difference of the level of a 1-kHz tone as a function of the temporal length. The ordinate value is normalized with the JND of a 100-ms-long tone burst. Adapted from Fastl and Zwicker (2007).

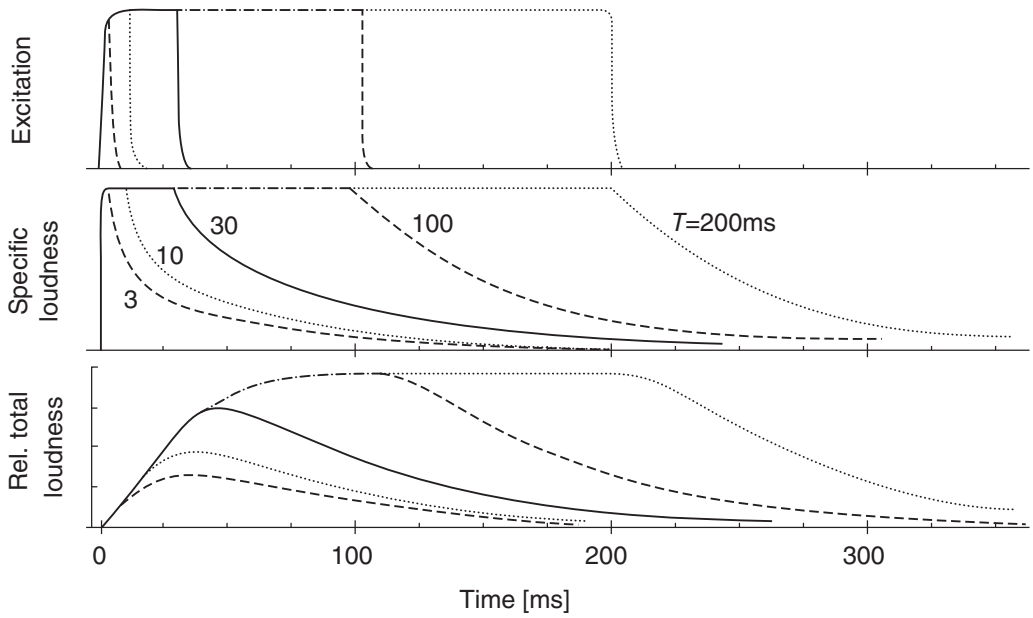

Figure 10.18 The change in loudness depending on time. Upper panel: sound pressure level. Middle panel: specific loudness. Lowest panel: relative total loudness. Tone bursts with length 3, 10, 30, 100, and 200 ms are used as stimuli in the simulations. Adapted from Zwicker (1984), and reprinted with permission from The Acoustical Society of America.

threshold increases monotonically (Fastl and Zwicker, 2007). The loudness of temporally shorter sounds is thus subject to added uncertainty in the hearing mechanism.

So far we have reviewed the perception of loudness caused by tone bursts. Speech and music are composed of sounds whose spectrum changes all the time. The perceived loudness can thus be expected to change all the time. However, it has been found that the loudness of such time-varying sounds seems to be the result of a short-time integration of sound power followed by a processing of the peak values by another mechanism. The perceived loudness is thus defined by loudness peak values. For example, a sound, in which 200-ms sound bursts alternate with 200-ms pauses is perceived to have loudness similar to a corresponding continuous sound.

Fastl and Zwicker (2007) present a model to estimate the loudness of sound with a dynamically changing level. The time-varying level of a tone-burst excitation with different lengths is shown in Figure 10.18 (topmost panel). The resulting loudness densities varying with time are shown in the middle panel, and the relative total loudness is shown in the lowest panel.

The specific loudness shown in the middle panel increases rapidly, matching the stimulus level, but decreases with a time constant that depends on the length of the corresponding stimulus. With short excitation signals, the specific loudness decreases rapidly, and with longer signals, the decrease is notably slower. The specific loudness represents the excitation in the inner ear, and the time integrated total loudness is assumed to be formed at higher stages of processing. The curves in the lowest panel represent the result after integration over time and frequency. The total loudness perception then corresponds to the highest peak in the curve.

10.3 Timbre

The tonal colour, or *timbre*, is a multidimensional psychoacoustic measure. When two sounds have the same pitch, loudness, and duration, timbre is what makes one particular musical sound different from another. For example, the same musical notes played by a piano and a trumpet

are easily distinguished by listeners as being different. The best physical explanation for this difference comes from the spectrum and its variation with time. Certain factors that affect the formation of timbre are discussed next.

10.3.1 Timbre of Steady-State Sounds

If the short-term amplitude spectrum of the sound is constant with time, the timbre is also constant in most cases. In some cases with periodic sounds, some alterations in the phase spectrum of sound but no changes in the amplitude spectrum may also change the perceived timbre. This effect is discussed more thoroughly in Section 11.5. However, in such steady-state conditions, the specific loudness (or the corresponding excitation pattern) represents the perceived timbre quite accurately if the sound is broadband noise.

The number of possible timbres is enormous. Theoretically, since the frequency resolution of hearing is 1 ERB and the entire auditory frequency range consists of 42 bands, and since the level resolution is 1 dB for a dynamic range of about 100 dB, there are about 100^{42} possible timbres. However, when the masking effect is taken into account, this number is smaller. Nevertheless, even for a dynamic range of only 6 dB, the number of timbres is $6^{42} > 4 \cdot 10^{32}$, which is still a colossal number.

10.3.2 Timbre of Sound Including Modulations

In many musical instruments, such as the piano or the human singing voice, the partials are modulated in amplitude, in frequency, or in both. These modulations show up as modulations in the auditory band signals, which give a new dimension to the perceived timbre. These fluctuations are visible in the *amplitude envelope* of the auditory bands. Hearing is especially sensitive to modulations at 4 Hz. Faster modulations (>10 Hz) lead to the sound being perceived as 'rough', and still faster modulations are not perceived, since the signal components spread over different critical bands (Fastl and Zwicker, 2007, Chapter 10). In general, the presence of modulation in the signal of an auditory channel makes that channel more detectable, and the rate of modulation can also be perceived as beating.

The *onset* of sound is also important in timbre perception. Sounds from many instruments have a distinctive onset, where the partials of the sound build up and possibly contain some transient-like, noisy components. The timbre during the onset defines the individual characteristics of the instrument in many cases. For example, some brass instrument sounds cannot be distinguished from each other if the onset is digitally removed.

10.4 Subjective Duration of Sound

Quantifying the duration of sound, the attribute used to order a set of sounds from 'short' to 'long', can be done in a similar manner to quantifying the loudness and pitch of sound. The subjective duration of a burst of tone is shown in Figure 10.19. The curves originate from a test conducted in a fashion similar to that for loudness tests; the task for the subjects was to find sounds with half or twice the duration of the reference sound. The unit of subjective duration is the *dura*, which corresponds to the perceived duration of 1 s of a 1-kHz tone (Fastl and Zwicker, 2007, Chapter 10). As shown in the figure, the subjective and physical duration of sound correspond well with each other for sounds longer than about 200 ms. For physical durations shorter than 200 ms, the subjective duration is perceived to be longer than the physical one.

The subjective duration of a relatively long sound surrounded by silence is thus quite accurately perceived. It is also interesting how accurately humans perceive the duration of silence.

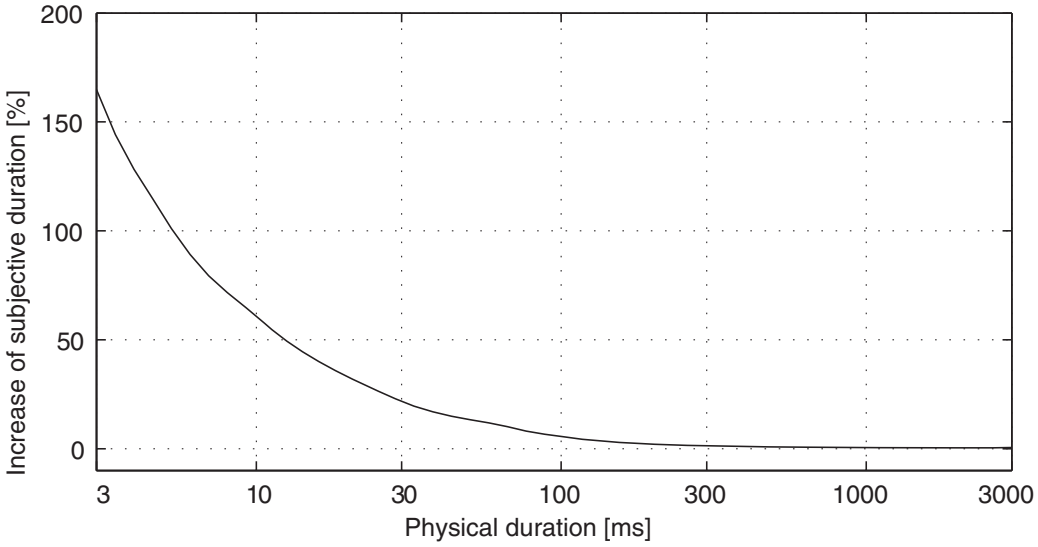

Figure 10.19 The relative subjective duration of sound as a function of the physical duration of a 1-kHz tone at an SPL of 60 dB. A similar result is obtained with noise. Adapted from Fastl and Zwicker (2007).

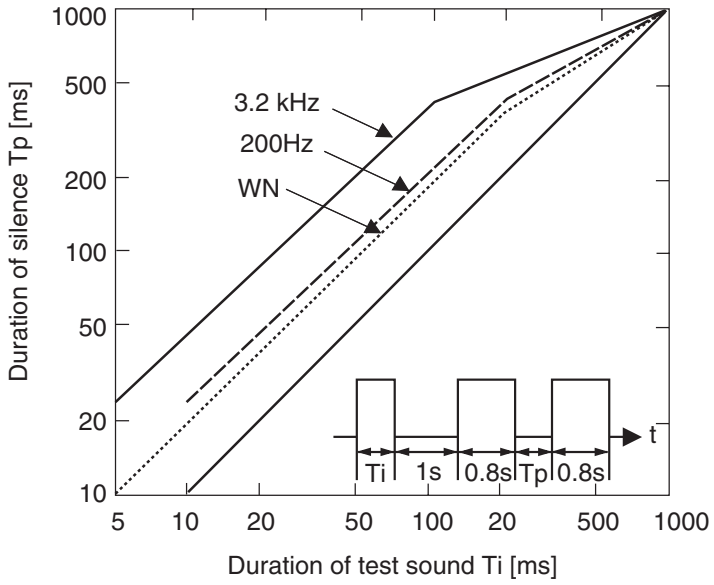

Figure 10.20 The physical duration of silence T_p between two bursts is adjusted to match the perceived length of a test sound. The physical duration of the test sound is T_i. Fastl and Zwicke 2007.

This has been studied by asking subjects to adjust the length of silence between two bursts so that it is equal to the length of a test sound. The result of this experiment is shown in Figure 10.20, which indicates the physical duration of the silence T_p when it has been adjusted to match the corresponding perceptual durations of the test sound. The physical duration of the test sound is T_i.

The physical duration of test sounds interpreted as durations of silence by subjects match well when the duration is near 1 s. The difference between the values, however, increases with a shortening test sound, with durations less than 100 ms being perceived as much longer than the length of the physical test signal (Fastl and Zwicker, 2007, Chapter 10). It is clear that this phenomenon has an effect on the perception of speech and music. For example, short silences

existing in speech are hard to perceive, at least partly due to this effect. It is also clear that the longer perceptual duration of the burst tone relative to silence is related to the time-masking phenomenon. The silence immediately after the burst is masked, as shown in Figure 10.18, which makes the silence appear shorter than it actually is.

Summary

The basic quantities of sound measured by the hearing system, namely pitch, loudness, timbre, and subjective duration, were introduced in this chapter. Based on how our hearing perceives these quantities, a general view of the functioning of hearing, which is coherent though rich in details, was drawn. In general, our hearing shows remarkable capabilities in analysing sounds arriving at the ear canals. Identifying these basic quantities is the first step taken to identify different sounds, such as speech, natural sounds, and sounds from musical instruments. The next chapter discusses some of the further analysis mechanisms used by our hearing, which then lead to our abilities to generate and enjoy music, and to organize complex sound scenarios into meaningful streams.

Further Reading

The basic quantities in psychoacoustics are also discussed in the corresponding chapters of the books by Fastl and Zwicker (2007), Moore (1995), and Yost (1994). Recent advances in theories of loudness and pitch perception are reviewed by Florentine *et al.* (2005) and Plack *et al.* (2005), respectively.

References

ANSI-S1.1 (2013) Acoustical terminology. Standards Secretariat, Acoustical Society of America.

Bilsen, F. and Ritsma, R. (1969) Repetition pitch and its implication for hearing theory. *Acta Acustica United with Acustica*, **22**(2), 63–73.

Burck, W., Kotowski, P., and Lichte, H. (1935) Die horbarkeit von laufzeitdifferenzen. *Elek. Nachr.-Techn.*, **12**, 355–362.

Canteretta, E.C. and Fridman, M.P. (eds)(1978) *Handbook of Perception*. Academic Press.

Demany, L. and Semal, C. (1989) Detection thresholds for sinusoidal frequency modulation. *J. Acoust. Soc. Am.*, **85**(3), 1295–1301.

Fastl, H. and Stoll, G. (1979) Scaling of pitch strength. *Hearing Res.*, **1**(4), 293–301.

Fastl, H. and Zwicker, E. (2007) *Psychoacoustics – Facts and Models*. Springer.

Florentine, M., Popper, A., and Fay, R.R. (eds) (2005) *Loudness*, volume 37. Springer.

Hartmann, W.M. (1996) Pitch, periodicity, and auditory organization. *J. Acoust. Soc. Am.*, **100**(6), 3491–3502.

Moore, B. (1973) Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.*, **54**(3), 610–619.

Moore, B.C.J. (ed.) (1995) *Hearing*. Academic Press.

Moore, B.C., Vickers, D.A., Baer, T., and Launer, S. (1999) Factors affecting the loudness of modulated sounds. *J. Acoust. Soc. Am.*, **105**(5), 2757–2772.

Plack, C.J. (2013) *The Sense of Hearing*. Psychology Press.

Plack, C.J., Oxenham, A.J., Fay, R.R., and Popper, A.N. (2005) *Pitch: Neural Coding and Perception*, volume 24. Springer.

Pressnitzer, D., Patterson, R.D., and Krumbholz, K. (2001) The lower limit of melodic pitch. *J. Acoust. Soc. Am.*, **109**, 2074–2084.

Rossing, T.D., Moore, F.R., and Wheeler, P.A. (2001) *The Science of Sound*, 3rd edn. Addison-Wesley.

Sek, A. and Moore, B.C. (1995) Frequency discrimination as a function of frequency, measured in several ways. *J. Acoust. Soc. Am.*, **97**, 2479–2486.

Yost, W.A. (ed.) (1994) *Fundamentals of Hearing – An Introduction*. Academic Press.

Zwicker, E. (1984) Dependence of post-masking on masker duration and its relation to temporal effects in loudness. *J. Acoust. Soc. Am.*, **75**, 219–223.