9

Basic Function of Hearing

This chapter starts by discussing the most fundamental of questions regarding an auditory object: under what conditions does it exist? Two physical attributes limit the audibility of a frequency component of sound: the sound pressure level (SPL) and frequency. If the SPL is too low, nothing is heard. If it is too high, the sensation of sound is accompanied by the sensation of pain, and beyond some limit in the SPL the ears are destroyed. The threshold of pain is commonly taken as the upper limit of the SPL to generate auditory events in sound and voice techniques. On the other hand, sound components with too low frequencies are not perceived as sound, but more as a sensation of vibration, and sound components with too high frequency content are not perceived at all. The attributes also interact with tonal signals; the SPL threshold of audibility depends in a complicated manner on frequency. The first section in this chapter will discuss these issues.

When multiple sounds are presented to the subject, they influence each other's audibility. In spectral masking, sounds in different frequency regions make each other inaudible, and in temporal masking, the same happens in the temporal dimension. The basics of masking are discussed in the second and third sections of this chapter.

The last section in this chapter discusses the first steps of spectral analysis conducted in hearing; that is, the characteristics of the frequency bands in hearing. The processing of sound begins in the cochlea, where the sound is divided into narrowband time-domain signals, which are then processed in the brain more or less individually. When a spectrally broad sound is presented to a subject, a relevant question is what is the frequency resolution of the neural presentation? Two methods to measure the widths of the bands are described, and different estimates of human 'critical bandwidths' are reviewed.

9.1 Effective Hearing Area

The auditory system is able to receive and process a very wide range of different sounds. The first characterization of the abilities of hearing is to describe the working range of possible signal frequencies and amplitudes. Figure 9.2 defines the upper and lower limits of normal working hearing as well as equal loudness curves as functions of frequency. The useful range

---

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

of frequency is from about 20 Hz to 20 kHz, although higher and lower frequencies can be perceived if they are intense enough. Sounds below the lower frequency limit are called *infrasound* and those above the upper limit are called *ultrasound*.

The smallest amplitude of a tone causing an auditory event is called the *hearing threshold*. This threshold curve is measured as the *minimum audible field* (MAF), the sound pressure level of the weakest binaurally perceived sound from the front. The MAF level is measured in the free field after the subject is removed, in the position where the centre of the subject's head was.

The decibel scale of sound pressure (see Section 2.2.3) is defined so that 0 dB is close to the hearing threshold for a pure tone at 1 kHz, which corresponds to the reference pressure p_0 in Equation (2.14), $2 \cdot 10^{-5}$ Pa. The sensitivity of hearing is best at about 3–4 kHz, and the threshold increases at low (below 250 Hz) and high frequencies (above 5 kHz).

The upper amplitude limit is about 130 dB, beyond which a hearing percept changes to pain. Just above this level the hearing system is in danger of being injured, even for a short impulse of sound. Even at considerably lower levels, at about 85 dB, exposure for eight hours daily over a long period results in hearing loss (see Section 19.3.2).

Figure 9.1 illustrates sound pressure levels in dB for typical sound sources that are useful for ear-based level estimation. A good reference in the middle frequency range is to remember that typical speech measured from a distance of one metre is about 60–70 dB. This is the optimal level for communication, since it is not too loud to be harmful for the functioning of hearing but is loud enough to produce a good signal-to-noise ratio in most environments, and it also provides good conditions for temporal and spectral analysis to the auditory system.

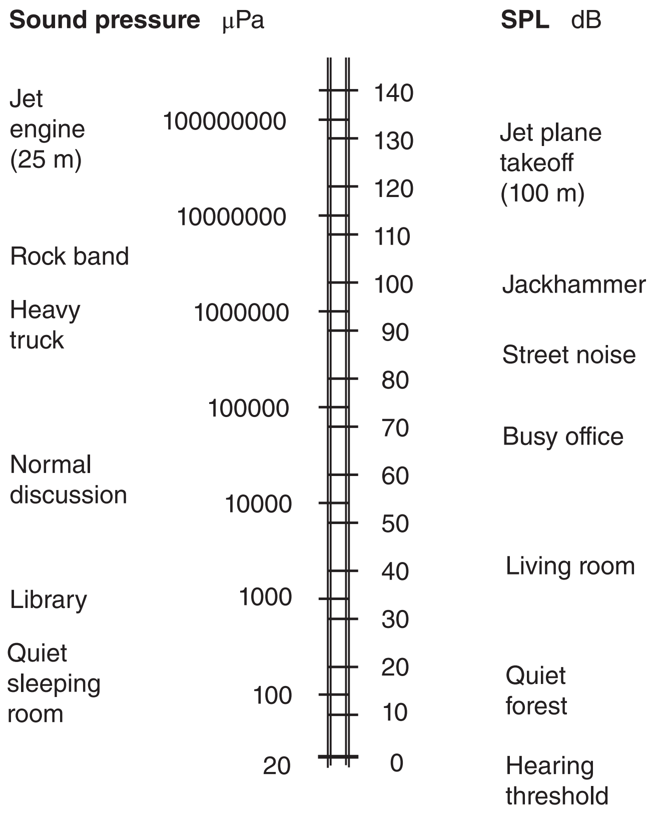

Figure 9.1 Sound pressure levels of typical sound sources and environments.

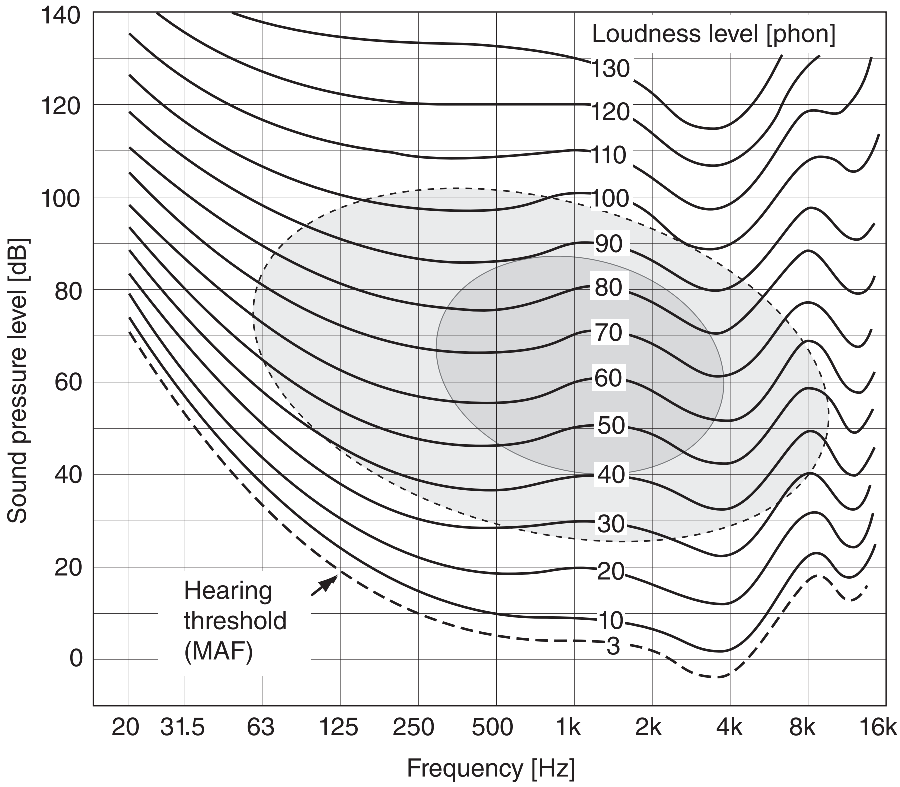

Figure 9.2 The working range and equal loudness contours of human hearing on the frequency–SPL plane. The lowest curve specifies the minimum audible field (MAF), the field producing a just audible sound to the subject, as measured in free-field conditions for a sound source in front of the subject. The uppermost curve specifies the approximate threshold of pain. Each separate curve specifies a constant of the perceived loudness for a pure tone as a function of frequency.

Although the full functional range of hearing is vast, about 130 dB for the level and 20 Hz to 20 kHz for frequency, a much smaller area of the total range is utilized in practical communication situations. For acoustic music the effective range is about that shaded light grey in Figure 9.2. In speech communication the most basic range is still more limited. For example, the frequency range for understandable speech communication over a telephone connection is from about 300 Hz to about 3.4 kHz and the amplitude range is typically 40–70 dB (the darker shade in Figure 9.2) (Fletcher, 1995).

9.1.1 Equal Loudness Curves

Sounds between the threshold of hearing and threshold of pain are perceived with increasing ‘strength’ or ‘volume’. This subjective feature of sound is called *loudness*. The *loudness level* has been defined such that the sound pressure level of a 1-kHz pure tone in dB has the same loudness level in phon units (SPL, Equation (2.14)).

Based on this definition, it is possible to measure the equal loudness curves using subjects who compare a 1-kHz reference tone at a given level and a test tone of another frequency by adjusting the latter to have the same perceived loudness. Curves in Figure 9.2 are averaged for a large number of subjects under conditions standardized in ISO-226 (2003). Loudness perception will be discussed in more detail in Section 10.2.

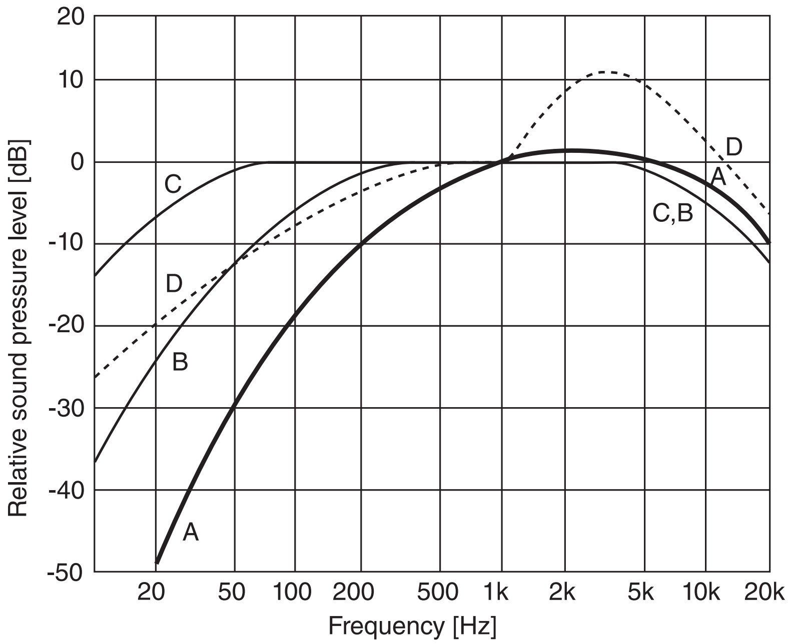

Figure 9.3 Weighting curves A, B, C, and D for sound level measurement.

9.1.2 Sound Level and its Measurement

Sound pressure was introduced in Section 2.2.2 as one primary physical measure of sound waves. The sound pressure level (SPL) in decibels, defined in Equation (2.14), is a logarithmic measure of amplitude, which is more convenient because of its large range and better corresponds to subjective perception. Even after this mapping to the new scale, sound pressure does not work well as a perceptual measure since the subjective level is strongly frequency-dependent, as can be seen easily from Figure 9.2.

In order to have a better, yet relatively simple, measure of the perceived level of sound, the concept of *sound level* is defined as a measure that is weighted with frequency so that the level roughly approximates the frequency sensitivity of hearing. Four different weighting curves are defined: A, B, C, and D, as shown in Figure 9.3. The *A-weighted* sound level is commonly used in noise measurements to characterize the perceived level and the risk of hearing loss. The A weighting slightly emphasizes the levels at mid frequencies and attenuates them at low and high frequencies. When the weighting curve is compared with the inverted equal-loudness curves, it is easy to see that the A weighting is just a very rough estimate. Technical simplicity and extensive use in practice are its advantages.

The other curves (B, C, and D) are rarely used. The unit of all sound levels is the decibel [dB], the same as for SPL, but it is common to also denote the weighting curve, for example, dB(A) for the A-weighted curve.

9.2 Spectral Masking

The inaudibility of soft sounds in the presence of louder sounds is very common in everyday listening scenarios. The phenomenon itself is present in all technical devices aiming to detect a weak signal in the presence of an interfering strong signal. The masking effect is an important characteristic of human hearing, which is commonly exploited in sound and voice technologies, as in the design of lossy audio codecs (see Section 15.3.1). The effect of spectral masking occurs when a sound with certain spectral content makes the detection of another sound with different spectral content harder, even though the spectra do not necessarily overlap. Spectral masking, that is, how the *masker* sound affects the detection threshold of the test sound, can be best described by plotting the masking threshold as a function of frequency.

9.2.1 Masking by Noise

Let us first investigate the effect of broadband noise on the masking threshold of a tone. The threshold is presented in Figure 9.4 when the masker sound is white noise with different power spectral densities. The test sound is a tone whose detection threshold has been measured in psychoacoustic experiments (Fastl and Zwicker, 2007). The dashed line shows the hearing threshold of the tone without the presence of the masker.

The masking threshold plots have a constant value up to about 500 Hz, after which they increase with a slope of 3 dB/octave (10 dB/decade). If spectral masking with a constant effect at all frequencies is desired, a *uniform masking noise* must be used, where the spectral density decreases correspondingly by 3 dB/octave at frequencies above 500 Hz.

The nature of spectral masking can be revealed by using narrowband noise as the masker. The masking thresholds then have the form shown in Figure 9.5. For example, a 160-Hz-wide noise masker with centre frequency 1 kHz and sound pressure level 60 dB generates a slightly asymmetric curve with its peak at 1 kHz, a few decibels below the level of the masker. The other curves shown in the figure for maskers with different centre frequencies f_c produce similar masking threshold curves.

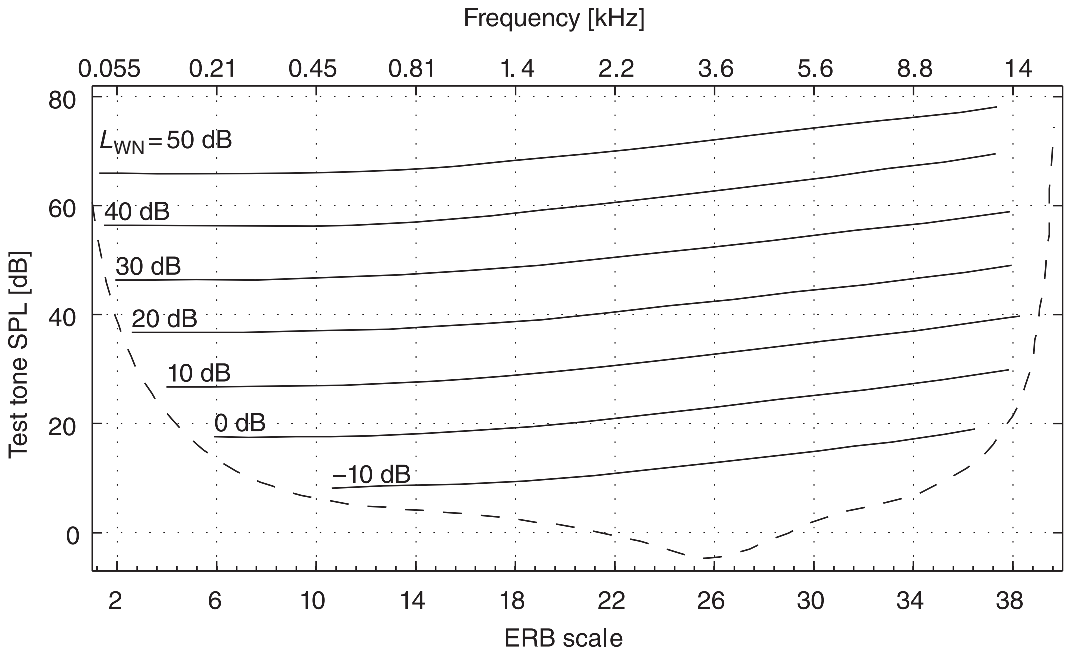

Figure 9.4 Masking thresholds (solid lines) in the presence of white noise with different power spectral density values L_{WN}. The ERB scale is an auditory frequency scale, and it will be described in Section 9.4.3 on page 167. The dashed line represents the threshold of audibility. Adapted from Fastl and Zwicker (2007).

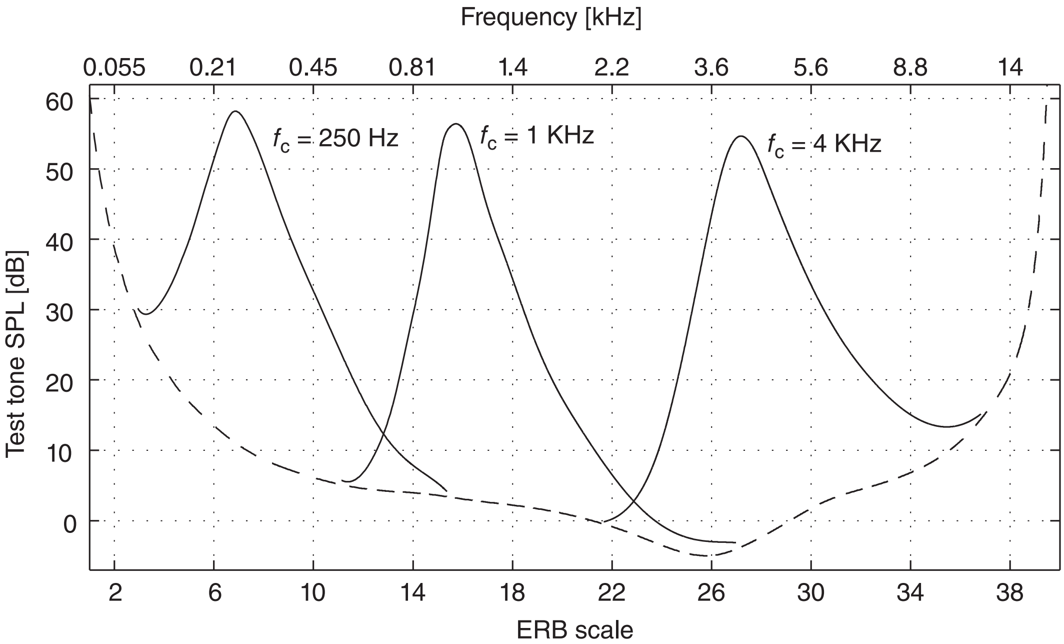

Figure 9.5 Masking threshold curves caused by narrowband noise with centre frequencies 250 Hz, 1 kHz, and 4 kHz with a sound pressure level of 60 dB. Adapted from Fastl and Zwicker (2007).

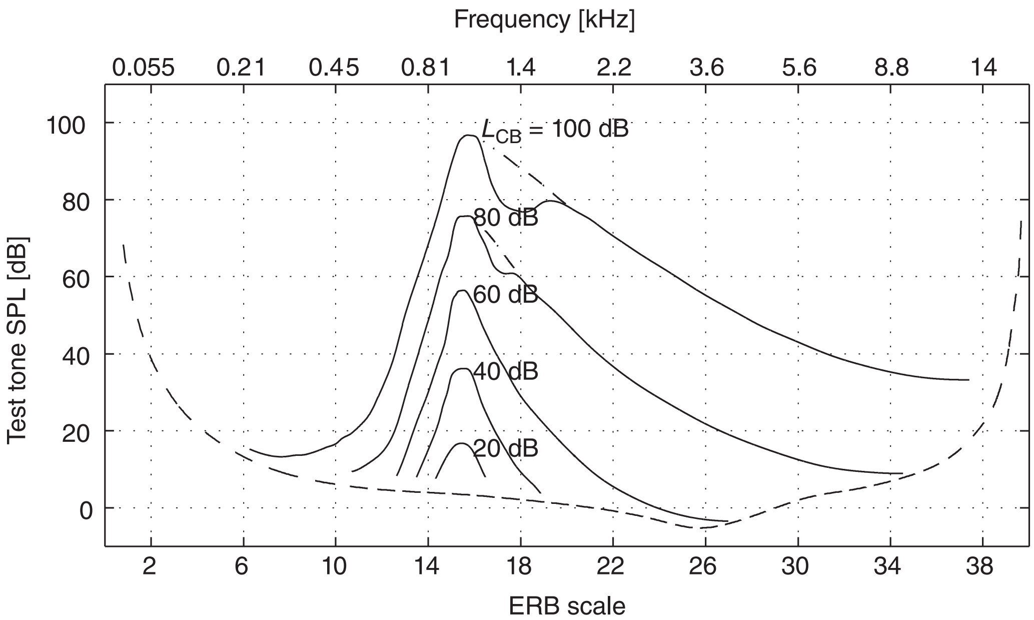

Figure 9.6 Masking threshold curves caused by narrowband noise with centre frequency 1 kHz at different sound pressure levels L_{CB}. Adapted from Fastl and Zwicker (2007).

When narrowband noise with a centre frequency of 1 kHz at different sound pressure levels is used as a masker, the masking curves shown in Figure 9.6 are obtained. When the level of the masker is increased, the curvature of the masking thresholds changes. With an increase in level, the drop in the threshold due to masking is shallower at frequencies above the centre frequency of the masker. However, the slope of the masking threshold below the centre frequency of

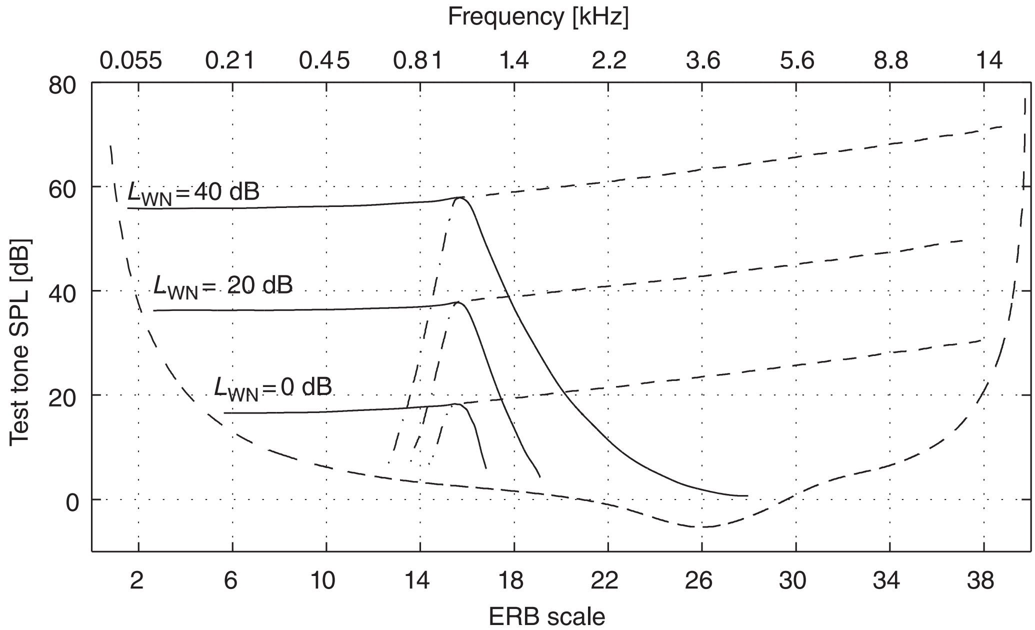

Figure 9.7 Masking threshold curves obtained with low-pass (solid line) or high-pass (dashed line) filtered white noise by a filter with a cutoff frequency of 1 kHz. Adapted from Fastl and Zwicker (2007).

the masker is not affected. The dips in the masking threshold curves with $L = 80$ and $L = 100$ dB at frequencies above the centre frequency of the masker are caused by the audibility of combination tones caused by non-linearities in hearing. The test sound itself becomes audible only when the level shown with a dashed line is exceeded.

If low-pass or high-pass filtered noise filtered at a cutoff frequency of 1 kHz is used as the masker, the masking threshold curves shown in Figure 9.7 are obtained, where the curves follow the results shown in Figures 9.5 and 9.6, but only on the half of the spectrum that has no signal.

9.2.2 Masking by Pure Tones

A pure tone – a sinusoid with fixed frequency – causes a masking effect similar to that of narrowband noise. However, due to the non-linear processing in hearing, in the presence of a second test tone, called the probe tone, close to the first tone (the masker tone) some beating and combination tones are also perceived. Beats are the periodic changes in the amplitude of the signal, which may be audible as roughness or fluctuations, as will be explained in Chapter 11. Thus, the presence of a probe tone with a tone masker can be detected with lower amplitudes than with a narrow bandnoise masker, as shown in Figure 9.8. In many cases, the probe tone itself is not audible, the listeners perceive only the beating or the combination tones (Fastl and Zwicker, 2007; Wegel and Lane, 1924).

9.2.3 Masking by Complex Tones

The masking effect caused by other sounds follows the same principle as with the special cases described above. The masking threshold caused by a harmonic tone complex consisting of the ten lowest partials is shown in Figure 9.9. Each harmonic can be thought to result in a partial masking pattern, and the total masking pattern is the sum of the individual patterns

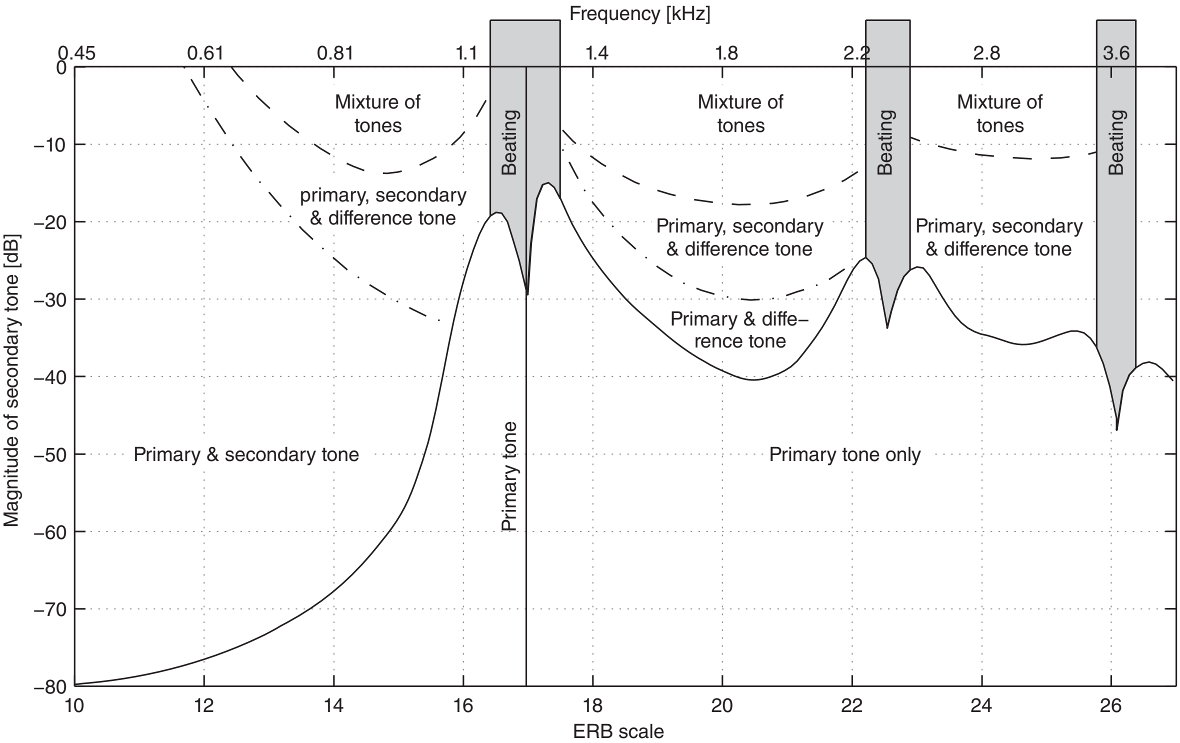

Figure 9.8 Masking threshold patterns of a probe tone in the presence of a tone masker (1.2 kHz). The probe and the masker tones interfere with hearing mechanisms at the frequencies shaded grey. The audibility regions of the masker tone, the test tone, and the difference tones are marked. Adapted from Wegel and Lane (1924).

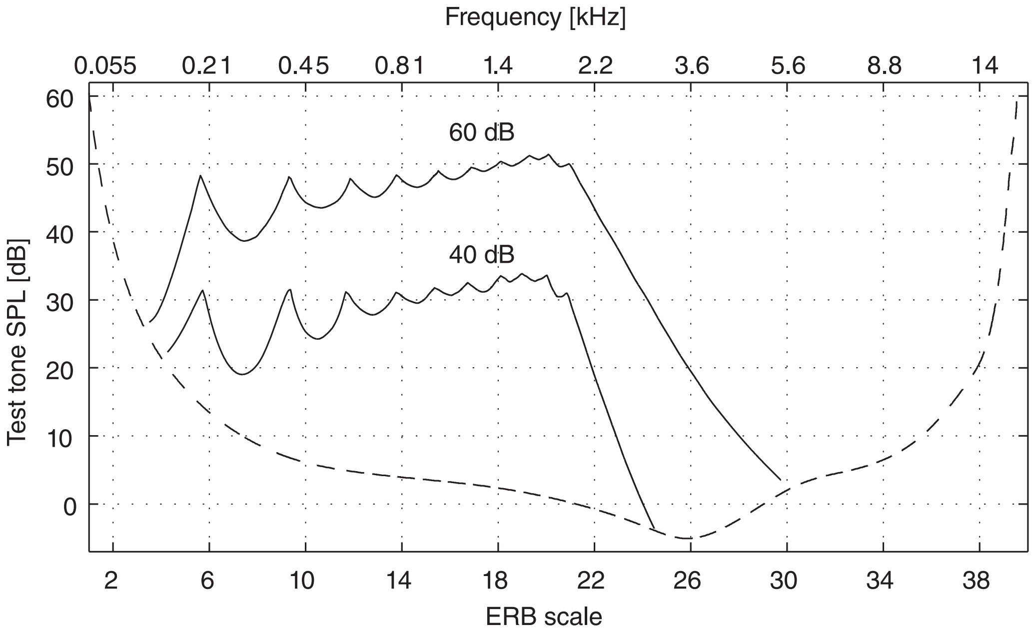

Figure 9.9 The masking patterns caused by a harmonic complex tone consisting of the ten first partials of equal level. The fundamental frequency of the masker is 200 Hz. The effect of the masker is shown for two different sound pressure levels. Adapted from Fastl and Zwicker (2007).

(Fastl and Zwicker, 2007). Similar results are obtained with instrument sounds similar to this spectrum. Overall, spectral masking is of great importance in music and audio reproduction. For example, the audibility of one instrument sound is limited by the masking effects caused by other, simultaneously playing instrument sounds.

9.2.4 Other Masking Phenomena

Co-modulation masking release is the decrease in the masking effect when the masker is amplitude-modulated with the same modulating function at all frequencies. Let us consider the case where the signal is a single sinusoid and the masker is a noise band around the frequency of the sinusoid. If the noise band is not modulated, the masking effect grows stronger when the noise bandwidth is extended. Interestingly, if the masker is amplitude-modulated coherently in all frequencies of the masker at the rate of 10–20 Hz, the masking effect deviates from the non-modulated case when the bandwidth of the masker exceeds 100 Hz (Hall *et al.*, 1984). This can be seen to imply that the hearing system groups different frequency bands together based on the similarity of the variation of the level in the bands.

In some cases the masking effect is reduced and cannot be explained with the masking effects discussed in the previous sections, which have been measured using simple sinusoids, noise, or transient sounds – types of masking called *energetic masking* (Watson, 2005). When the maskers and signals are more complex, reduced masking effects may be obtained, such as in speech-in-speech masking (Brungart, 2001). Such masking effects are called *informational masking* and can be thought to be caused by the ability of the human listener to segregate the signal and masker at a higher processing level.

Dau *et al.* (1997) suggest that the outputs of the auditory filters are processed by a bank of overlapping 'modulation filters' (analogous to auditory filters), each tuned to a different modulation frequency – *a modulation filter bank*. This helps the system to group signals sharing the same source, since different frequency bands originating from the same source are often modulated coherently. Although the model explains many aspects of modulation perception, the existence of the bank has not been proven, and is controversial (Plack, 2013).

9.3 Temporal Masking

The masking effect has been discussed so far only for continuous sounds. Sounds mask each other in time as well; that is, a sound affects the audibility of a preceding or following sound. A conceptual illustration of temporal masking is shown in Figure 9.10, both for a sound occurring before the masker, called *backward masking*, or *pre-masking*, and after the masker, called *forward masking*, or *post-masking*. Such thresholds of audibility are measured using psychoacoustic tests, where the listener hears a masker sound long enough (>200 ms), which is preceded or followed by a short burst of sound. Forward masking is, in general, more significant and more consistent a phenomenon than backward masking.

Backward masking has an effect only 5–10 ms before the onset of the masker, and relatively low-level sounds are masked. The effect seems to appear only in inexperienced listeners. Thus, backward masking is not interesting in the context of acoustic communication, and it is not discussed further in this book.

The forward masking effect, in turn, has an effect over a much longer period of time, about 150–200 ms after the offset of the masker. Relatively high-level sounds are also masked, thus it is relevant in the context of this book. The forward masking effect of a noise masker on a

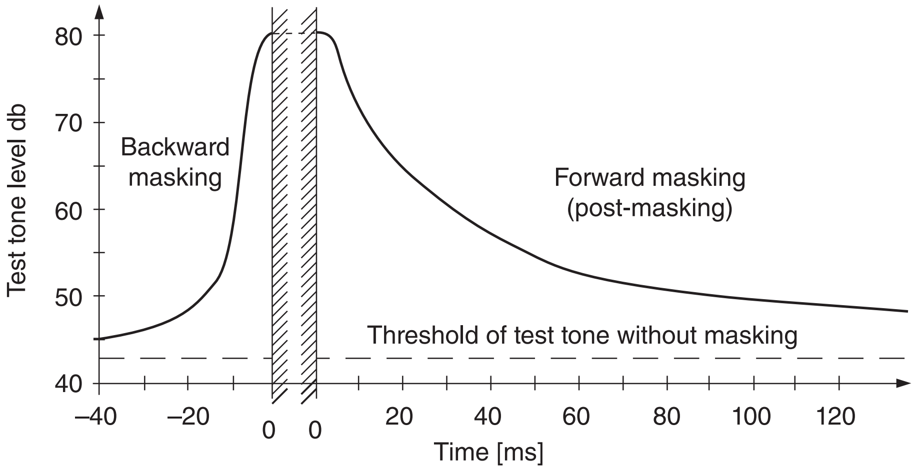

Figure 9.10 The temporal masking effects in hearing. A probe sound at a sufficiently low level arriving after a masker sound is not audible due to forward masking, and a probe sound at a low enough level arriving just before the masker sound is not audible due to backward masking. The vertical hatched lines represent the start and end times of the masker sound lasting at least 200 ms.

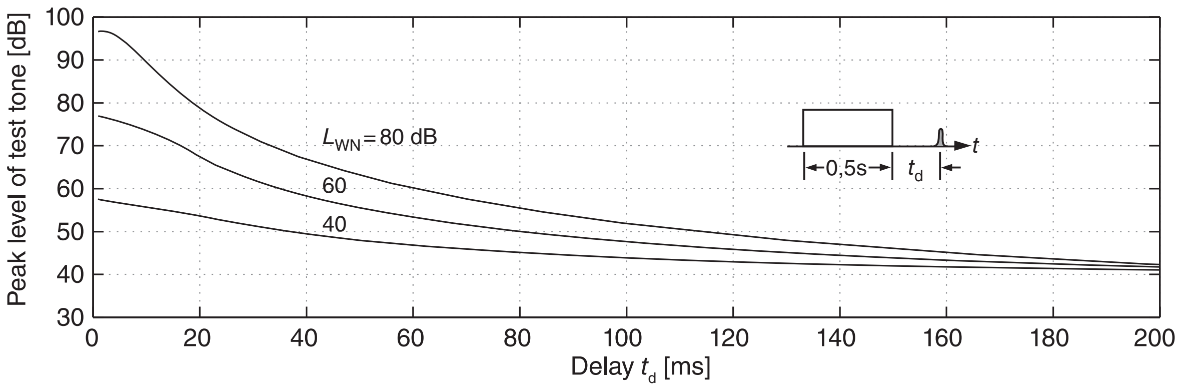

Figure 9.11 The forward masking effect at three different masker levels. Adapted from Fastl and Zwicker (2007).

probe tone is shown for three different levels of the masker on a linear time scale in Figure 9.11. The masking level decreases linearly after 5–10 ms from the offset, and after about 200 ms the threshold of hearing in silence is reached (Fastl and Zwicker, 2007).

The forward masking effect is a complex phenomenon, which also depends on the length of the masker sound. Figure 9.12 shows the effect for maskers lasting 5 ms and 200 ms. The effect is significantly milder with the shorter sound.

The masking effect has also been studied with periodically repeating bursts, modulated signals, pulses, and impulses (Duifhuis, 2005). For example, impulses or pulses can be used to measure the resolution of hearing of the temporal fine structure of sound, as shown in Figure 9.13. Using a masker composed of an impulse, the masking threshold has been measured for a probe impulse as a function of the temporal position with respect to the position of the masker impulse. A distinct pattern is seen, where the maximum value of the masking threshold pattern of about −10 to −5 dB is within a distance of 1 ms from the position

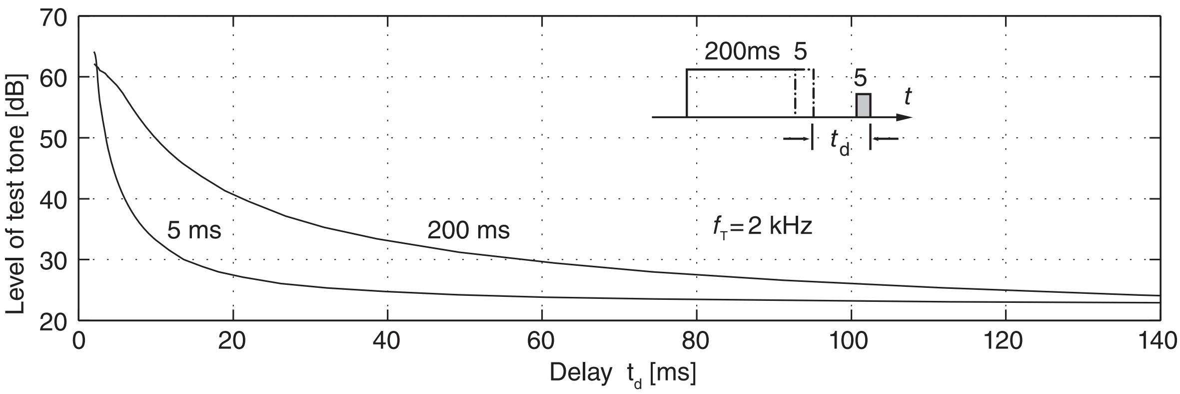

Figure 9.12 The forward masking effect for two different temporal lengths of the masker sound. Adapted from Fastl and Zwicker (2007)

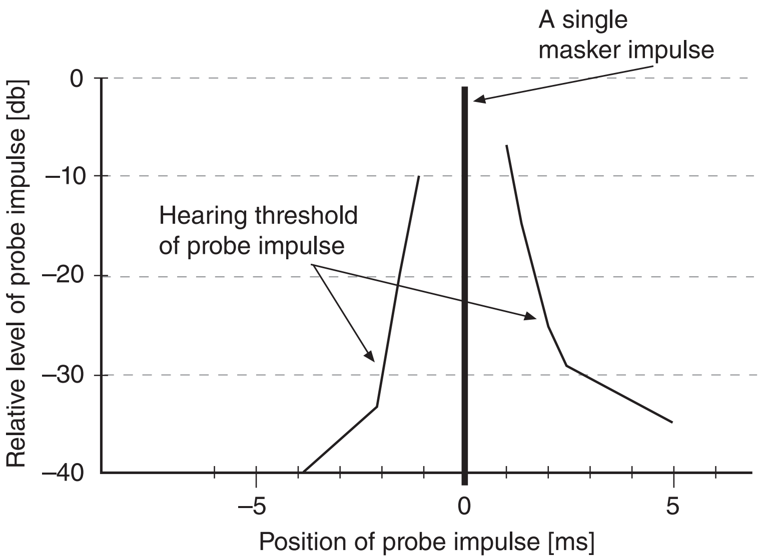

Figure 9.13 The temporal masking threshold pattern caused by an impulse. The probe signal is a single impulse at different positions in time with respect to the masker impulse. Adapted from Feth and O'Malley (1977).

of the masker impulse. The threshold decreases quickly at longer temporal distances from the masking impulse, reaching −40 dB after 5 ms. This shows that a relatively weak impulse is still perceived as an individual auditory event if it is not preceded or followed immediately by other impulses. If the impulses arrive within a time window of about 1–2 ms, they are merged into one auditory event, especially if one of them has a higher level. The value of 1–2 ms at which the impulses are just perceived as separate sounds can be treated as the best time resolution of hearing.

9.4 Frequency Selectivity of Hearing

The frequency masking patterns shown in Figure 9.5 imply that a narrowband sound presented to the listener affects the perception of other sounds in nearby frequencies. If two narrowband sounds at similar levels have different enough frequencies, our hearing resolves them into two separate auditory events. If they are located sufficiently close to each other in frequency, they are perceived as a single auditory event. The ability of our hearing to segregate sounds separated in frequency is called *frequency resolution*.

Frequency resolution and selectivity are important properties in terms of understanding the functioning of human hearing. Frequency resolution has been studied a lot, and slightly different results have been obtained with different approaches. Frequency selectivity is commonly thought to stem from cochlear processing; a broadband stimulus arriving at the cochlea is converted using mechanical and neural processing into a neural output. Each of the inner hair cells has a different best frequency to which it responds, but they also respond strongly to frequencies near the best frequency. This frequency region is called the *critical band* or the *auditory filter*. In practice, the width of this band, which depends on frequency, is of interest in sound and voice technologies.

9.4.1 Psychoacoustic Tuning Curves

The frequency selectivity of hearing measured from a single point on the basilar membrane was shown in Figure 7.11. A related measurement can be conducted applying psychoacoustic methods. In the measurement, the masking sound is usually narrowband noise, whose centre frequency is a parameter, and the test signal is often a tone, whose frequency and level (say, 10 dB) are kept constant when measuring a single curve. The task of the listener is to adjust the level of noise so that the test tone is just audible. The resulting functions obtained with this set-up are shown in Figure 9.14. The cochlear gains for the lowest signal levels in Figure 7.11 and psychoacoustic tuning curves in Figure 9.14 are clearly approximately reciprocal to each other, which verifies the validity of the approaches.

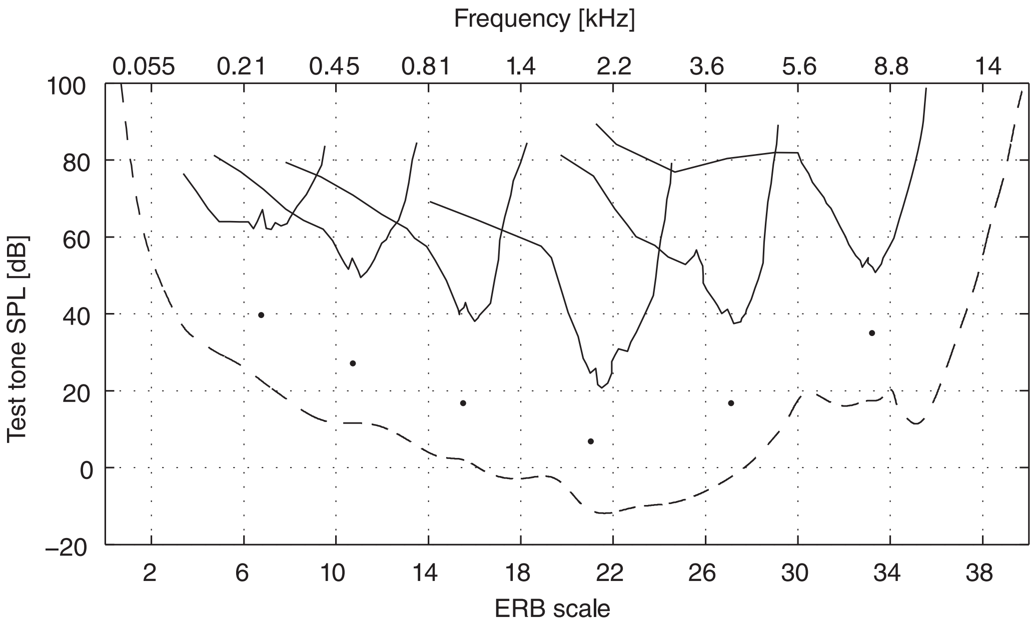

Figure 9.14 Psychoacoustic tuning curves measured using a low-level sinusoid as the signal and narrowband noise as the masker. The level of the sinusoid was 10 dB, and the frequencies of the tones were 0.25, 0.5, 1, 5, and 10kHz. Adapted from Vogten (1974).

Bark bandwidths

A classic method to measure the frequency resolution of human hearing is outlined in this section. A narrowband noise with a fixed centre frequency is used as the reference sound, to which the subject compares the test sound, band-limited noise whose sound pressure level and centre frequency are equal to the reference sound (see Figure 9.15). The change in perceived loudness is measured with some psychoacoustic test, and a schematic plot of such a test result is shown in Figure 9.16. Interestingly, the loudness is constant up to a certain value of bandwidth (Fastl and Zwicker, 2007). This bandwidth is 160 Hz for the center frequency 1 kHz, as shown in the figure, beyond which the loudness increases steadily. The knee point in the plot is thought to be the position where the test sound's spectrum spreads over more than a single critical band

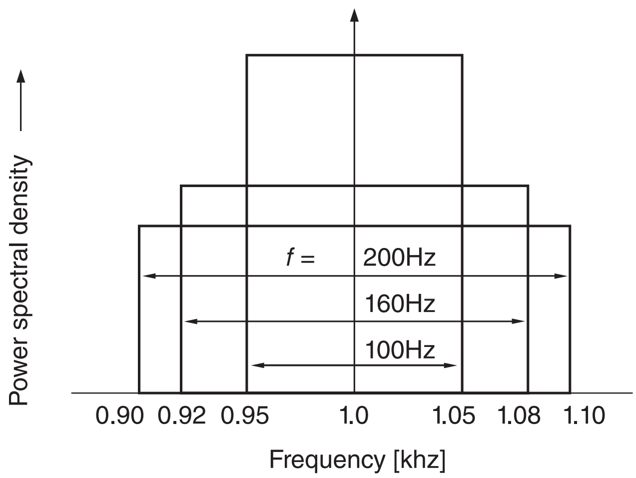

Figure 9.15 Measuring the critical bandwidth with band-limited noise having equal centre frequency and sound pressure level.

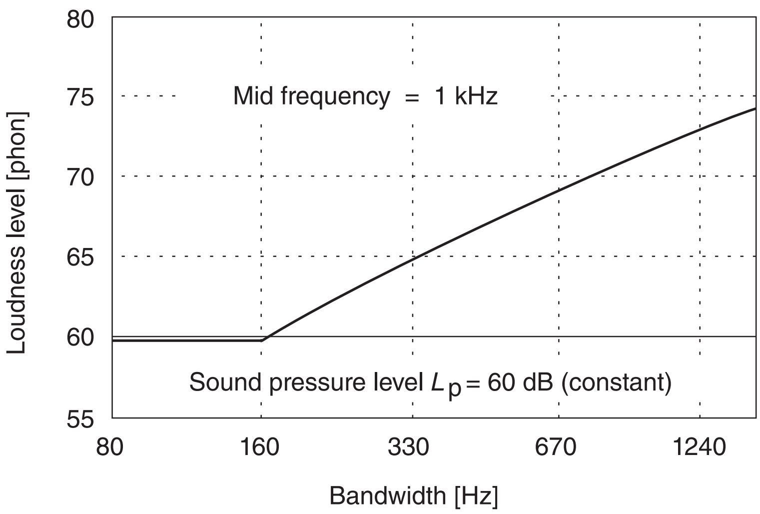

Figure 9.16 The perceived loudness as a function of the frequency bandwidth of the noise stimulus. The critical bandwidth is defined as the point on the curve above which the perceived loudness starts to increase. The centre frequency in this example is 1 kHz, and the width of the critical band is measured to be 160 Hz. Rossing *et al.* (2001).

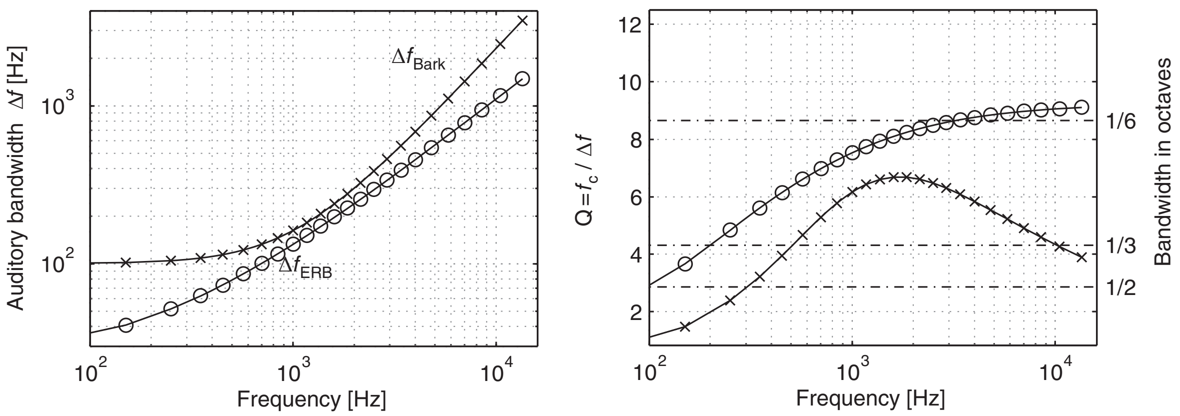

Figure 9.17 (a) Two estimates of the critical bandwidths Δf in Bark bandwidths (crosses) and ERB bandwidths (circles) as a function of the centre frequency f_c. (b) The same data plotted as Q values.

and so is processed inside more than one band. The knee point is regarded as the measure of the bandwidth of the critical band.

Zwicker named the critical bandwidths measured using this approach *Bark bandwidths* after Heinrich Barkhausen, who proposed the first subjective measure of loudness. The bandwidths Δf_{Bark} [Hz] measured via listening tests are estimated as

$$\Delta f_{Bark} = 25 + 75[1 + 1.4(f_c/1000)^2]^{0.69} \tag{9.1}$$

and are shown in Figure 9.17a as a function of the centre frequency f_c. The bandwidth is 100 Hz at low frequencies, and above 500 Hz it increases with frequency, being a bit less than 1/3 octave wide. Near the upper end of the audible frequency range the width is several kHz.

The fact that the perceived loudness increases (instead of, say, decreasing or staying constant) is an interesting phenomenon. The perceived loudness can be assumed to be related to the broadening of the excitation pattern on the basilar membrane with the level of the signal, as shown in Figure 7.12. The increase in loudness with signals having a broader spectrum can be explained by the mechanism whereby the auditory bands analysed to originate from the same source are integrated, and the total loudness is computed from the sum of the signals from each auditory band.

9.4.2 ERB Bandwidths

Another method to measure the bandwidth of the auditory filters in our hearing uses the concept of *ERB* (equivalent rectangular bandwidth) bands. ERB gives an approximation to the bandwidths of the filters in human hearing, using the unrealistic but convenient simplification of modelling the auditory filters as rectangular band-pass filters. The frequency widths of filters can be measured with different methods. In a commonly used listening test, bands of masking noise applied above and below the test signal, as shown in Figure 9.18, eliminate the possibility of hearing responding in the cochlea to the test tone outside the frequency region being tested, a phenomenon called *off-frequency listening*. The detection threshold of the tone is measured as a function of the width of the notch $2\Delta f$ to give an estimate of the frequency selectivity of

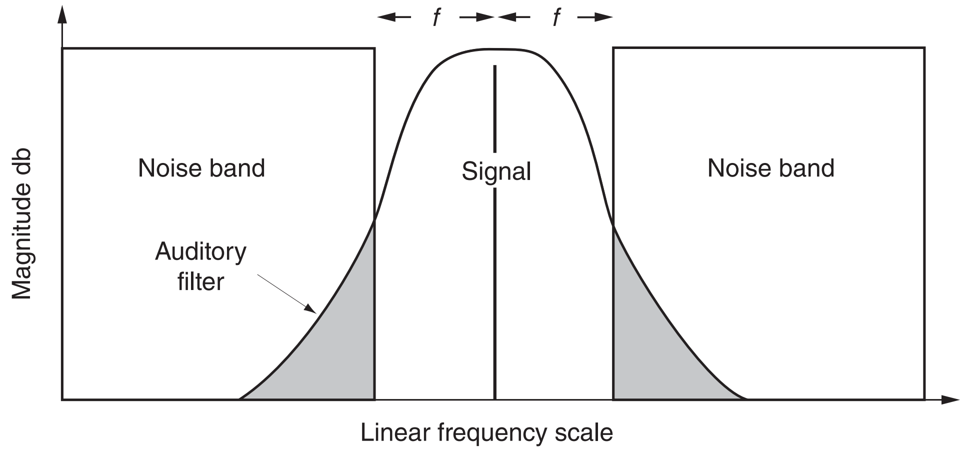

Figure 9.18 Estimation of the critical bandwidth by measuring the detection threshold of a tone masked by notched noise. The threshold is measured as a function of the width of the notch, which is further used to estimate the width of the critical band at different frequencies as ERB (equivalent rectangular bandwidth) bands.

human hearing. The obtained widths of critical bands (ERB) are typically 11–17% of the value of the centre frequency f_c, and the widths can be estimated as

$$\Delta f_{\mathrm{ERB}} = 24.7 + 0.108 f_c \tag{9.2}$$

(Glasberg and Moore, 1990). The bandwidth here also follows a logarithmic relationship with the centre frequency but over a larger range than the Bark bandwidth. This is seen from the fact that the relation between the ERB bandwidth and centre frequency in Figure 9.17b changes less with frequency.

The bandwidth of hearing also changes with level: the higher the level, the poorer the frequency resolution, as discussed previously. However, for acoustic communication technologies, it seems to be a fair assumption that the frequency window in human hearing used in the spectral analysis of complex sound signals is around 10–15% of the centre frequency. This is not to be confused with the other frequency scale in humans, the pitch scale, which will be discussed in the next chapter.

9.4.3 Bark, ERB, and Greenwood Scales

Both Bark and ERB bandwidths define a frequency scale. If the bandwidths are computed and stacked on top of each other starting from a very low frequency, a scale is obtained for both methods. With Bark bands, this scale is called the *Bark scale*, z_{Bark}, and it can be estimated from the frequency f as

$$z_{\mathrm{Bark}} = 13 \arctan(0.76 f/1000) + 3.5 \arctan(f/7500)^2. \tag{9.3}$$

Correspondingly, the frequency scale based on ERB bandwiths is derived by:

$$z_{\mathrm{ERB}} = 21.3 \log_{10}(1 + f/228.7). \tag{9.4}$$

Both scales are presented as logarithmic functions of frequency in Figure 9.19. The Bark scale is linear up to about 500 Hz, as was seen in Figure 9.15. Above 500 Hz it is roughly logarithmic.

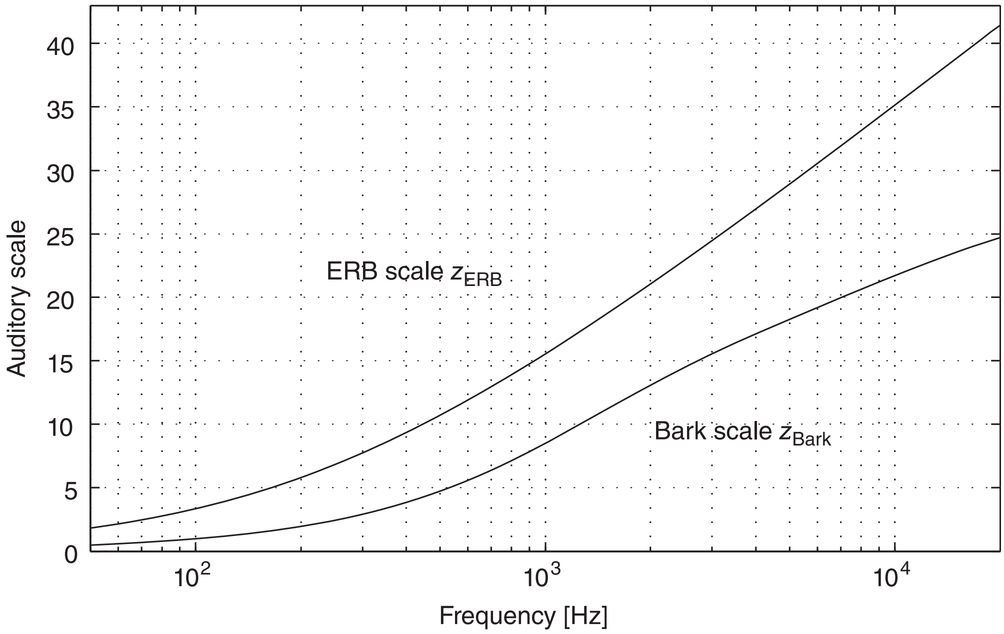

Figure 9.19 Bark and ERB frequency scales as functions of frequency.

The slope of the ERB scale is relatively straight at all frequencies with the logarithm of the frequency. Thus, it is logarithmic over a larger frequency region than the Bark scale.

The auditory scale also has an interesting relation to the anatomy of the inner ear. Each of the scales has been interpreted to have a linear relation to the position of resonance on the basilar membrane. This means that a constant change in the auditory frequency scale corresponds to a constant change in the resonance position on the basilar membrane at all frequencies. It seems that the scale based on ERB bands is closest to this relation, where a bandwidth of 1 ERB corresponds to 0.9 mm on the basilar membrane, which has about 90 inner hair cells.

To conclude, the ERB bandwidths and the ERB scale are commonly used in hearing sciences to approximate the auditory frequency resolution and scale, respectively. However, the Bark scale has not been abandoned. Many technical applications utilize it or related scales, for example the *mel* pitch scale is relatively commonly used in certain speech technologies, as will be discussed in Section 10.1.4.

A related auditory scale is the scale proposed by Greenwood (1990), where the resonance position on the basilar membrane x [mm] and the frequency f [Hz] have been found in mammals to have the relation

$$f = A(10^{ax} - k), \tag{9.5}$$

if x is known. If f is known, x is computed as

$$x = (1/a)\ \log_{10} \frac{(f + kA)}{A}, \tag{9.6}$$

where, for humans, $A = 165.4$, $a = 0.06$ and $k = 1$.

Summary

This chapter has drawn a picture of the first processing steps conducted in our hearing. The ear has a certain working range limited in frequency and in SPL. The range is actually very impressive, as the 120-dB range in SPL with the lowest levels very low and the frequency range starting from 20 Hz and ending at 20 kHz are hard to achieve with man-made devices.

Frequency masking effects and the measurement of critical bandwidths indicate that the frequency resolution of hearing corresponds to a bandwidth of about 1/6 octave at best. The best resolution is obtained at frequencies above 1 kHz with an SPL level below about 60 dB. Temporal masking effects are strong after the dying out of a loud sound. However, the temporal resolution is about 1–2 ms, which is quite remarkable in accuracy.

Further Reading

Due to a relatively long period of research and an active interest in auditory mechanisms, there exists a rich literature on this topic. For general introductions to the topic, some being more specific than in this book, the reader is referred to, for example, (Fastl and Zwicker, 2007; Moore, 2012; Plack, 2013; Schnupp *et al.*, 2011) and the appropriate chapters in Bregman (1990); Crocker (1997); Moore (1995); Yost (1994). A good source on the early investigations towards a quantitative formulation of auditory sensation and perception is Fletcher (1995).

References

Bregman, A. (1990) *Auditory Scene Analysis*. MIT Press.

Brungart, D.S. (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.*, **109**, 1101–1109.

Crocker, M.J. (ed.) (1997) *Encyclopedia of Acoustics*, Volume 3. John Wiley & Sons.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation. I. detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.*, **102**(5), 2892–2905.

Duifhuis, H. (2005) Consequences of peripheral frequency selectivity for nonsimultaneous masking. *J. Acoust. Soc. Am.*, **54**(6), 1471–1488.

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

Feth, L.L. and O'Malley, H. (1977) Influence of temporal masking on click-pair discriminability. *Percep. Psychophys.*, **22**(5), 497–505.

Fletcher, H. (ed.) (1995) *Speech and Hearing in Communication*. Acoustical Society of America.

Glasberg, B.R. and Moore, B.C.J. (1990) Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, **47**(1–2), 103–138.

Greenwood, D.D. (1990) A cochlear frequency-position function for several species – 29 years later. *J. Acoust. Soc. Am.*, **87**(6), 2592–2605.

Hall, J.W., Haggard, M.P., and Fernandes, M.A. (1984) Detection in noise by spectro-temporal pattern analysis. *J. Acoust. Soc. Am.*, **76**(1), 50–56.

ISO-226 (2003) Acoustics – normal equal-loudness-level contours. International Organization for Standardization.

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

Moore, B.C.J. (2012) *An Introduction to the Psychology of Hearing*, 6th edn. Brill.

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

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

Schnupp, J., Nelken, I., and King, A. (2011) *Auditory Neuroscience: Making Sense of Sound*. MIT Press.

Vogten, L. (1974) Pure-tone masking: A new result from a new method *Facts and Models In Hearing*, Springer. pp. 142–155.

Watson, C.S. (2005) Some comments on informational masking. *Acta Acustica United with Acustica*, **91**(3), 502–512.

Wegel, R. and Lane, C. (1924) The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.* **23**(2), 266–285.

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