CHAPTER 10

Auditory Sensitivity

THRESHOLDS OF AUDIBILITY

A first step in understanding auditory perception is to understand the auditory system's sensitivity to frequency, amplitude, and starting phase. When we understand the sensitivity of the auditory system to these variables for sinusoidal stimuli, the results might form a complete picture of the absolute sensitivity of the auditory system to any stimulus because all acoustical stimuli can be defined in terms of sinusoids. However, as we will see in this and following chapters, the solutions are not that simple.

When a pure tone is presented to one ear, the auditory system is actually insensitive to starting phase. This does not mean, as Helmholtz suggested more than 100 years ago, that the auditory system is phase insensitive. For instance, if two sinusoids of different frequencies are added, the perceived quality of the sound may vary as the starting phases of the sinusoids are varied. Also, as mentioned in Chapter 2, if a sinusoid is presented to both ears and there is a change in the interaural (between-ears) phase difference, then observers report a change in the perceived location of the tone (see Chapter 12).

We can measure auditory sensitivity to frequency and level by determining the level required for a listener to detect the presence of a sinusoid at each of many frequencies. There will be very low and very high frequencies to which, no matter how intense the sinusoid, the auditory system is insensitive. These frequency limits define the bounds of the auditory system's sensitivity to frequency. The thresholds relating the smallest level required for detection to the frequency of the tone are called *thresholds of audibility*.

Thresholds of audibility are obtained in the laboratory in the following manner. For each frequency tested, a psychometric function is obtained either directly or indirectly using a psychophysical procedure (see Appendix D). Figure 10.1 demonstrates the results from this part of the experiment. The threshold is determined for each frequency by choosing a performance level, such as 75%, for $P(C)$. The thresholds of audibility are plotted as the threshold in decibels versus frequency. A listener who has a low *threshold* can also be described as being very *sensitive*. Thus, "high sensitivity" and "low threshold" mean the same thing.

Minimum audible field (MAF) thresholds are sound-pressure levels for pure tones at absolute threshold, measured in a free field. A listener listens in a room to tonal sounds presented over loudspeakers. In the absence of

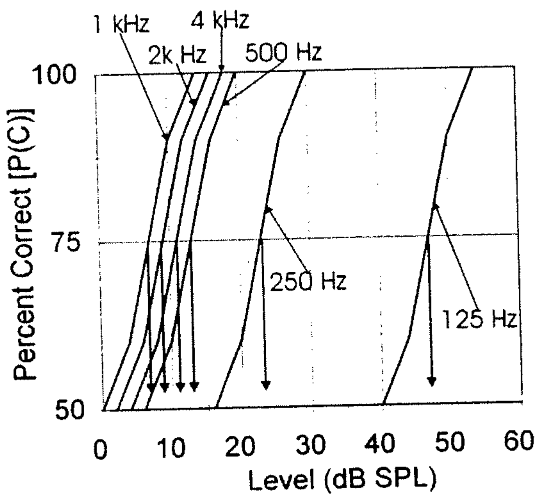

FIGURE 10.1 Psychometric functions obtained for six tones. The thresholds are obtained for these forced-choice psychophysical data at a *P*(C) of 75% (indicated by the arrows). Adapted with permission from Watson *et al.* (1972).

the listener, a microphone is placed at the position of the listener to calibrate the threshold levels. MAF thresholds are usually determined for listeners facing the source, listening with both ears (binaurally), at 1 meter from the sound source.

Minimal audible pressure (MAP) thresholds describe thresholds in terms of the sound pressure level at the observer's tympanic membrane. The listeners listen to sounds presented over earphones, and various procedures are used to determine the sound pressure that occurs at the tympanic membrane. Although the minimum audible pressure at the eardrum is independent of the type of sound source used to deliver the sound, most earphones are not calibrated in terms of the eardrum pressure they produce. To estimate the actual pressure at the tympanic membrane in a standardized manner, the sound-pressure level is estimated from the sound level in a test *coupler* attached to the earphone during calibration. Such couplers are designed to approximate the average acoustic properties of the ear of a listener with normal hearing and contain a microphone for estimating the sound pressure level that would exist at the tympanic membrane. Since there are different types of earphones and couplers, the threshold sound-pressure levels vary for each earphone–coupler combination. For a particular earphone–coupler combination, the threshold sound-pressure levels measured in the coupler are called the *Reference Equivalent Threshold Sound Pressure Levels* (RETSPLs), and these RETSPLs have been standardized for a variety of earphone–coupler combinations. Each earphone–coupler combination generates a different set of values for RETSPL.

Three classes of earphones, each with its own coupler type, are typically used for most hearing tests. *Supra-aural* phones fit over the pinna and are calibrated using a *"6-cc" coupler* (6 cc represents the average volume of the adult outer ear canal that lies between the earphone and the tympanic membrane). *Circumaural* phones fit completely over and around the pinna and are calibrated using an *artificial ear*. *Insert earphones* fit directly into the outer ear canal and are often calibrated in an *occluded-ear simulator*, sometimes called a *Zwislocki coupler*, after its founder Joseph Zwislocki, or in a 2-cc coupler (since 2 cc is the volume between the tip of the insert earphone and the tympanic membrane). Figure 10.2 and Table 10.1 show the RETSPLs for a few earphone–coupler combination examples.

As can be seen by comparing the MAF with the supra-aural or circumaural earphone MAP thresholds, they do not agree. MAF thresholds are always lower than MAP thresholds. Since the sound impinging upon the tympanic membrane should be the same

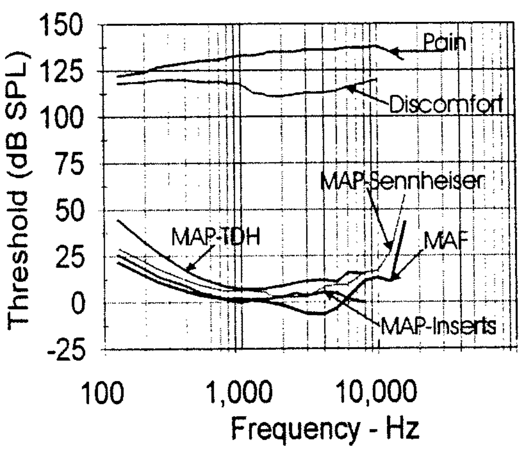

FIGURE 10.2 The thresholds of hearing in decibels of sound pressure level (dB SPL) are shown as a function of frequency for six conditions: RETSPL–Minimal Audible Field (MAF) thresholds, RETSPL–MAP thresholds for a supra-aural phone and 6-cc coupler, RETSPL–MAP thresholds for an circumaural phone and an artificial ear, RETSPL–MAP thresholds for an insert phone and a Zwislocki coupler, thresholds for pain, and thresholds for discomfort. The RESTPL measures are from ANSI 3.6-1996. The thresholds for pain and discomfort represent estimates of the upper limit of level that humans can tolerate.

TABLE 10.1 Thresholds of Audibility for a Number of Testing Conditions (from American National Standard ANSI 3.6-1996 Specifications for Audiometers)

Frequency (Hz)	Thresholds — RETSPL (dB SPL) MAF[a]	Supra aural[b]	Circum-aural[c]	Insert[d]
125	22	45	29.5	26
200	14.5	32.5	–	18
250	11	27	18	14
400	6	17	12	9
500	4	13.5	9.5	5.5
750	2	9	6.5	2
800	2	8.5	–	1.5
1,000	2	7.5	6.5	0
1,250	1.5	7.5	–	2
1,500	0.5	7.5	5.5	2
1,600	0	8	–	2
2,000	–1.5	9	3	3
2,500	–4	10.5	–	5
3,000	–6	11.5	3	3.5
4,000	–6.5	12	8.5	5.5
5,000	–3	11	9.5	5
6,000	2.5	16	9.5	2
8,000	11.5	15.5	16	0
10,000	13.5	–	21.5	–
12,500	11	–	27.5	–
16,000	43.5	–	58	–

[a]Binaural listening, free-field, 0° incidence.

[b]TDH type.

[c]Sennheiser HDA2000, EEC 318 with type 1 adaptor.

[d]HA-2 with rigid tube.

for a threshold response, and since great care is given to calibrating the actual sound-pressure levels for each procedure, all thresholds should be essentially the same. Thus, one might conclude from examining the difference between the MAF and MAP thresholds that a "missing 6 dB" (the average difference between MAF and MAP measures in the mid-frequency range) existed between sound field (MAF) and earphone (MAP). Experimental difficulties led to that erroneous conclusion. When estimates of MAP and MAF data are corrected for head diffraction, outer-ear resonances, and the type of calibration procedure used, the greatest difference between the two types of measurements is maximally 2.5 dB, indicating that the different reference sound-pressure levels required for calibrating different sound sources reflect differences in the calibration procedure, the fact that the head acts to diffract sound, and that the outer ear has a resonant frequency (see Chapter 6). Therefore, there is no "missing 6 dB."

The shapes of the MAP and MAF absolute threshold functions are approximately the same, showing a loss of sensitivity below approximately 1000 Hz and above approximately 4000 Hz. The loss of sensitivity declines at approximately 6 dB/octave at low frequencies below 1000 Hz and at approximately 24 dB/octave above 4000 Hz (the ability to estimate the slope of the high-frequency side of the thresholds of hearing function is much more difficult than at the low-frequency side due to the difficulty in measuring thresholds above 8000 Hz). Several factors governing the acoustic transfer function of the outer and middle ears (Chapter 6) have been suggested as explanations for the shape of the thresholds of hearing functions. For instance, the loss in middle-ear pressure gain at low and high frequencies (see Figure 6.8) helps explain the loss of sensitivity for these tonal frequencies.

The fact that we can detect tones over a range of frequencies from 20 to 20,000 Hz requires a very good loudspeaker or earphone system. Zero dB SPL corresponds to 20 μPa, which is an extremely small pressure. At this pressure and at a signal frequency of 1000 Hz, the tympanic membrane is vibrating through a total distance approximately equal to the diameter of a hydrogen atom. The horizontal axis in Figure 10.2 was constructed so that the intensity is in decibels of sound-pressure level, or dB SPL, that is, 0 dB SPL is equal to 20 μPa of pressure (see Chapter 3). Therefore, a number such as 50 dB SPL means that the threshold is 50 dB above a pressure of 20 μPa.

The curves in Figure 10.2 represent the thresholds in dB SPL from the ANSI standard on thresholds. These values have been standardized as the recognized thresholds and are the set of threshold values agreed upon when instruments are built and hearing tested. If hearing is tested using such a standard, the thresholds may be expressed in decibels relative to these standardized values. Here, the thresholds are expressed in dB HL (decibel of *hearing level*). For instance, a person with a 30-dB HL threshold at 1000 Hz measured with a supra-aural earphone has a threshold 30 dB above that shown in Figure 10.2 and Table 10.1, or the threshold is 37.5 dB SPL (i.e., 30 dB HL plus 7.5 dB SPL).

The thresholds of audibility define the smallest amount of pressure to which the auditory system is sensitive. What about an upper limit in terms of the most pressure the auditory system can tolerate? The upper curves in Figures 10.2 show an estimate of this upper limit. We can ask the listener to say when the sound is "felt," when *pain* is experienced, or when a tickling sensation is felt. The listener can also be asked when the sound level has become uncomfortable (*thresholds of discomfort*). These experiences indicate that the sound pressure is reaching a maximum. This maximum limit is approximately 120–130 dB SPL, and it remains relatively unchanged as a function of the frequency content of the stimulus. Thus, the *dynamic range* (difference between threshold in dB SPL and the maximum limit in dB SPL) of the auditory system is a function of frequency; it is approximately 125–135 dB at 1000 Hz, but 80–90 dB at 100 Hz.

DURATION

One variable that was not specified in the procedures used to obtain tonal thresholds is the duration of the sinusoid. From several points of view, we would expect that, the longer a sound lasts, the easier it is to hear. For the thresholds of audibility shown in Figure 10.2, the tones had a duration of more than 400 msec.

When time (*T*) is involved in measurement of intensity, we must be careful in noting whether the intensity is expressed in terms of units of power (P) or energy (E). Remember from Chapter 3 that

$$P = E/T \quad \text{and} \quad E = PT. \qquad (10.1)$$

To describe the auditory system's sensitivity to duration, the experimenter can keep either the signal constant in terms of units of power or energy, but not both, while changing its duration. Of course, if energy is maintained constant, then the power of the signal must change as the duration changes as expressed in equation (10.1).

Figure 10.3 shows the thresholds for different frequencies as a function of the duration of signals. In this figure, level is expressed in terms of units of power. Thus, for each duration a psychometric function was obtained in which observer performance was related to signal power. Then for each duration the threshold in units of power was determined from the psychometric functions and plotted in Figure 10.3.

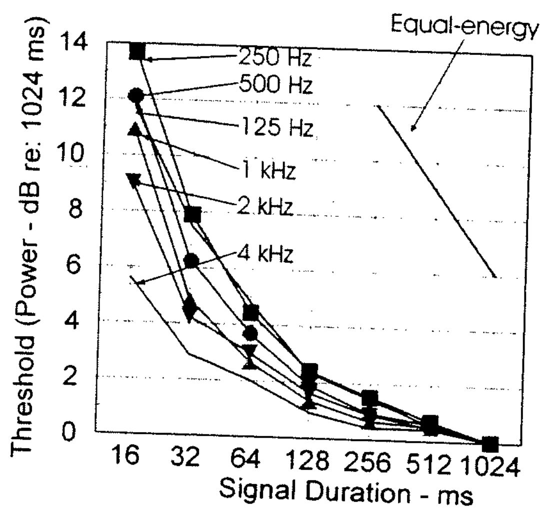

FIGURE 10.3 Thresholds in units of power for six tonal frequencies (ranging from 125 to 4000 Hz) displayed as a function of tonal duration. The thresholds are calculated and plotted in relation to the thresholds obtained at 1024 ms. For instance, the threshold for a 32-ms tone burst with a frequency of 250 Hz (circles) is approximately 8 dB higher than that of a 1024-ms (1-second) tone. The straight line with a slope of 3 dB/duration doubling indicates performance for equal energy detection. Adapted with permission from Watson and Gengel (1969).

Notice that at durations greater than approximately 250 to 500 msec the threshold in units of power for various tones does not change much; for durations greater then about 250–500 msec the tone does not become easier to detect. However, as the tone's duration is made shorter than 250 msec, the power of the tone must be increased for the observer to detect the tone. This increase is approximately equal to 8 to 10 dB of power increase for each 10-fold decrease in the duration of the tone, although this effect is slightly dependent on frequency. A 10-dB increase in power for each 10-fold decrease in duration means that the signal energy is remaining approximately constant for a constant level of the listener's performance. That is, equation (10.1) states that energy will remain constant if, as duration decreases, power increases. In other words, if $P = E/T$, then $\log P = \log (E/T)$, or $10 \log P = 10 \log E - 10 \log T$ (see Appendix B, Rule 2). Therefore, $10 \log E = 10 \log P + 10 \log T$ (if T is less than 1 sec). Thus, for a 10-fold change in duration, power must also change tenfold to keep energy constant (or for a doubling of duration signal power must change by 3 dB). For the data shown in Figure 10.3, it is important to realize that duration is expressed in relation to 1 second. Therefore, a change from 1 second to 1/10 of a second (100 msec) is a change of 1/10 or, in decibels, (10 log 1/10) or −10 dB. So as duration becomes shorter (less than 1 second), the power must become greater if energy is to remain constant. For example, if a 500-msec tone has a power of 80 dB, then the energy of this 500-msec tone is 77 dB; that is,

$$\begin{aligned} 10 \log E &= 10 \log P + 10 \log T, \\ E \text{ in dB} &= P \text{ in dB} + 10 \log T \\ &= 80 \text{ dB} + 10 \log (500 \text{ msec}/1000 \text{ msec}) \\ &\quad (1 \text{ sec} = 1000 \text{ msec}) \\ &= 80 \text{ dB} + 10 \log 0.5 \\ &= 80 - 3 \text{ dB} = 77 \text{ dB}. \end{aligned}$$

If the duration is changed from 500 to 50 msec and energy is to remain at 77 dB, then

$$\begin{aligned} 77 \text{ dB} &= P \text{ in dB} + 10 \log 50 \text{ msec}/500 \text{ msec} \\ &= P \text{ in dB} + 10 \log 0.10 \\ &= P \text{ in dB} - 10 \text{ dB} \end{aligned}$$

or

$$P \text{ in dB} = 87 \text{ dB } (77 + 10 \text{ dB}).$$

Thus, the power of the 50-msec tone must be 10 dB higher than that of the 500-msec tone, if energy is to remain constant.

Between approximately 10 and 300–500 msec, the energy of the signal must remain approximately constant for constant detection performance by the observer. Note, however, that the constant-energy property of the auditory system is only approximately true and depends on frequency. A complete understanding of auditory processing requires that we explain the exact form of the relation between signal detection and duration.

Once the duration of a sinusoidal signal decreases below 10 msec, much more power is required for detection than is needed to keep energy constant. It appears that the auditory system is not a constant-energy detector below 10 msec. In Chapter 4, we showed that a tone turned on and off spreads its energy over a large frequency region. As the duration of the tone becomes shorter and shorter, this frequency region over which the energy is spread becomes larger and larger. Thus, for very short-duration tones (less than 10 msec) the frequency region over which the energy of the tone is spread becomes so large that not all of the energy is contributing to its detectability. Because there is some energy at frequencies to which the ear is insensitive, there will be less energy in the region where the ear is sensitive. This in turn means that the total power or energy of the tone must be increased so that enough energy is in the auditory system's frequency region of sensitivity for the tone to be detected.

In other cases (e.g., for a low-frequency tone) the spread of energy associated with the short duration of the tone might produce energy in a frequency region (higher in frequency) where the auditory system is more sensitive (see Figure 10.2). The listener might detect the short-duration tone because of the energy available at these other more detectable frequencies. Therefore, the duration of the signal used to establish tonal threshold is important. If it is longer than approximately 300 msec, the thresholds represent intensity in units of power; if the signal is between 10 and 300 msec, the thresholds reflect approximately constant energy; for signals of durations that are less than 10 msec, the spread of energy makes determination of thresholds difficult.

TEMPORAL INTEGRATION

Notice in Figure 10.3 that further increases in duration beyond 300 msec do not change the detectability of a tone. It is as if the auditory system requires about 300 msec for maximal performance. If the signal is shorter than 300 msec, then its level expressed in units of power must be increased for maximal performance. This property of the detection of signals of different durations is often called *temporal integration*. Figure 10.4 diagrams the basic idea of temporal integration. A signal must have some critical amount of energy (depicted by the hashed area in Figure 10.4a) to

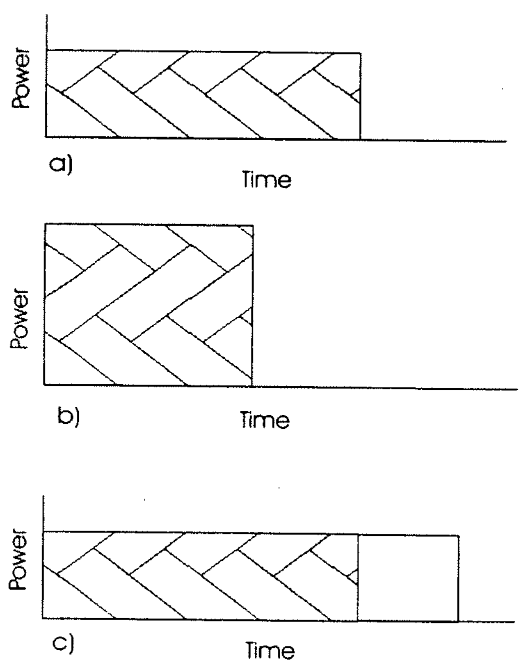

FIGURE 10.4 A schematic diagram depicting the concept of temporal integration. The hashed area at the top Figure 10.4a represents the energy of a signal required to detect the presence of a tone. In Figure 10.4b, the signal is shorter than that in Figure 10.4a; thus, the power of this shorter signal must be increased over that shown in Figure 10.4a to yield a just barely detectable signal. In Figure 10.4c, the signal duration is longer than the time required to achieve the necessary energy for detection; thus, there is no need to change the power of the signal in order for the signal to remain just detectable.

be detected, and once the signal contains that amount of energy (that area), it is detectable. In addition, the process of summing the power (integration) to generate the required energy is completed by 300 msec (the *integration time*). This means that, if the duration is less than 300 msec (the estimate of the integration time from Figure 10.4), then the power of the signal must be increased for the signal to be detected (the height of the rectangle must be increased to achieve the required area, as shown in Figure 10.4b). For durations greater than 300 msec, the threshold expressed in units of power remains constant for a constant level of performance, since the required integration time has been met or exceeded (Figure 10.4c).

The estimate of the integration time for auditory processing can vary a great deal from stimulus condition to stimulus condition. For instance, detection of a tone requires an integration time of about 300 msec; however, detection of a click (impulse) stimulus requires an integration time of only a few milliseconds. Thus, the integration time of auditory processing depends on the type of signal being processed. Estimates range from 1–2 msec to more than 500 msec.

DIFFERENTIAL SENSITIVITY

Although the thresholds of audibility define the frequency and intensity range of the auditory system, it does not describe our sensitivity to *changes* in intensity and frequency. In the early 1800s, Weber observed that it was easy to distinguish between a 1- and a 2-pound weight but not so easy to differentiate a 100-pound weight from a 101-pound weight, although both pairs of weights differ by 1 pound. It was found that the difference between two weights that could just be detected was proportional to the value of the smaller weight. That is, if the *just-noticeable difference* (jnd) for a 1-pound weight was 0.1 pound, then the jnd for the 100-pound weight was 10 pounds. Weber and Fechner stated this relation in equation form:

$$\Delta S/S = \text{constant},$$

where ΔS is the just noticeable physical difference in some stimulus value, and S is the smaller of the two values being discriminated. The ratio $\Delta S/S$ is called the *Weber fraction*. Psychoacousticians have attempted to determine whether the Weber fraction for the level, frequency, and duration of sinusoids is a constant and, if so, over what range of levels, frequencies, and durations.

Extreme care must be taken in studying differential sensitivity. As mentioned previously, not all of the energy of a tone is at the tonal frequency if the tone is very short or if it is turned on and off abruptly. This spread of energy also arises if the frequency, or level, or phase of an ongoing tone is suddenly changed. Thus, we cannot simply study differential sensitivity by changing the level or frequency of an ongoing tone and asking a listener if a change is detected. The listener will probably hear the change because of detecting a click resulting from the spread of energy to other frequencies rather than an intensity, frequency, or duration change per se. In most studies, two tones are presented: one tone is the standard tone and the second is called the comparison tone, which is presented with a slight change in intensity, frequency, or in some cases duration (although, as we will discuss later, there are other methods for measuring duration discrimination).

FREQUENCY DISCRIMINATION

Figure 10.5 shows the value of threshold Δf (frequency difference) required to just discriminate a change in frequency from a given frequency, f. The data are plotted as threshold Δf versus f, with the various curves representing different levels. These data represent the values of the difference threshold (Δf) for frequency. That is, psychometric functions in which the listener's performance was related to various frequency differences could be obtained for many different base frequencies, and then the difference threshold for frequency calculated from the psychometric functions.

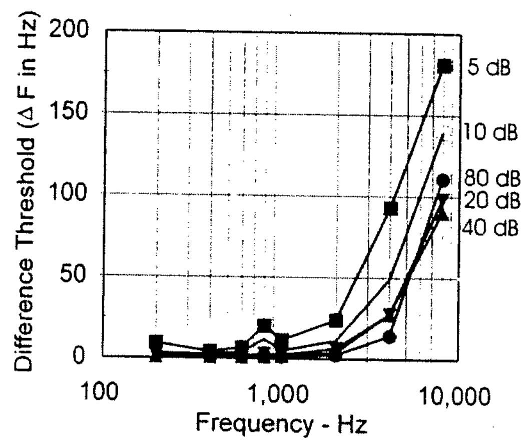

FIGURE 10.5 The value of threshold Δf (Hz) required to just discriminate between two different frequencies is shown as a function of the base frequency for five stimulus levels, expressed in dB SL. Adapted with permission from Weir *et al.* (1977).

As can be seen, the value of threshold Δf increases as f increases above 1000 Hz. In the mid-frequency region, this increase is enough to maintain the Weber fraction for frequency, $\Delta f/f$, approximately constant. This can be seen in Figure 10.6, in which the Weber fraction ($\Delta f/f$) is plotted as a function of f. Notice that over an intermediate range of frequencies this fraction is nearly constant at approximately $\Delta f/f = 0.002$ (or 0.2%). This means that at low frequencies threshold Δf can be as small as 1 Hz. (For instance, if f is 800 Hz and $\Delta f/800 = 0.002$, then $\Delta f = 800 \times 0.002 = 1.6$ Hz.)

INTENSITY DISCRIMINATION

Figure 10.7 demonstrates the value of threshold ΔI in dB required for the observers to detect a difference in level from the initial

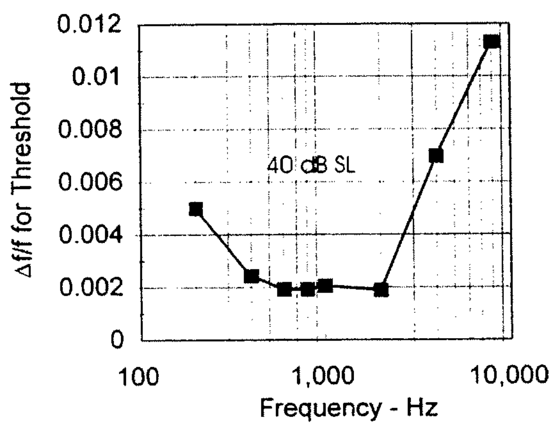

FIGURE 10.6 The value of $\Delta F/F$ (the Weber fraction for frequency) is shown as a function of frequency for the 40 dB SL data plotted in Figure 10.5. The results show that over an intermediate range of frequencies the Weber fraction is approximately constant.

value I. The value of threshold ΔI in dB is plotted as a function of I, and the different curves represent the results using different tonal frequencies and a wideband noise. Threshold ΔI in dB is the difference threshold for level.

Notice that the auditory system is sensitive to approximately a 0.5- to 1.0-dB change in level across a broad range of frequencies and levels. The fact that threshold ΔI in dB is nearly a constant as a function of I implies that $\Delta I/I$ (not in decibels) for pressure, energy, or power (or the Weber fraction) is a constant. This results from the fact that the logarithm (decibels) of a ratio (such as the Weber fraction) is equal to a difference between the logarithm of the divisor and that of the dividend (see Appendix B).

Let $\Delta I/I = c$, a constant. This is the Weber fraction. By slightly changing this equation, we get: $(\Delta I/I) + 1 = c + 1 = K$, another constant. We can write this last result as: $(\Delta I/I) + 1 = (\Delta I + I)/I = K$, which is another form of the Weber fraction. Now by expressing this ratio in decibels, we have: $10 \log [(\Delta I + I)/I] = 10 \log (\Delta I + I) - 10 \log (I) = 10 \log K = C$, another constant. This last equation is the same as the decibel difference between the more intense stimulus, $I + \Delta I$, and the less intense stimulus, I. This is the same as ΔI in dB or ΔI "in decibels." Thus, if ΔI in dB is a constant, then $\Delta I/I$, not in dB, is also a constant. ΔI "in decibels" is plotted along the left-hand axis in Figure 10.7. From the relations shown above, ΔI in decibels is a simple transform of $(\Delta I + I)/I$ or $\Delta I/I$, as shown on the right-hand axis of Figure 10.7. These relationships also show the various forms that the Weber fraction can take when intensity discrimination is studied.

Intensity discrimination data for a variety of tones of different frequencies and a wide-

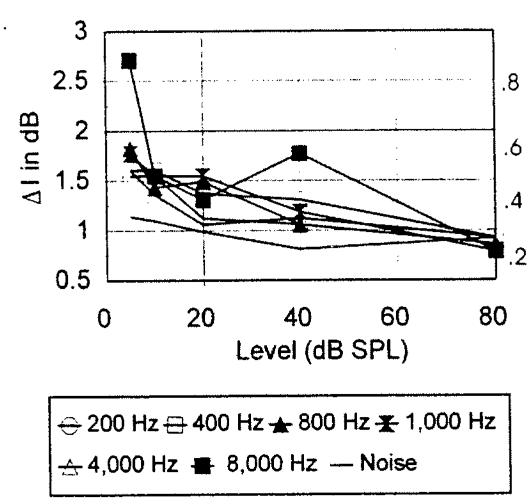

FIGURE 10.7 The value of threshold ΔI in decibels (the difference in decibels between the more and less intense tones) required for threshold discrimination is shown as a function of overall tonal level in dB SL. Data for six tonal frequencies and a wideband noise are shown. On the right-hand axis is $\Delta I/I$ not in decibels. Adapted with permission from Jesteadt *et al.* (1977) and Viemeister (1974).

band noise are shown in Figure 10.7. Notice that the thresholds for threshold ΔI in dB remain very near to 0.5–1 dB for a wide range of levels for the wideband noise, but that the threshold ΔI in dB decreases slightly, but steadily, as the level of the tones increases. The fact that the thresholds for tones are not constant when plotted in terms of ΔI in dB means that they do not strictly follow the Weber fraction. This fact is often referred to as the *near miss to the Weber fraction*. The "near miss" applies to tonal but not to wideband noise stimuli.

TEMPORAL DISCRIMINATION

If we asked a listener to detect the difference between a 50-msec sinusoid and a 60-msec sinusoid, additional variables aside from the tone's duration would be present that might aid the listener in detecting a difference in duration. Recall from Figure 10.3 that, because tones of different durations sometimes have different thresholds, the 50- and 60-msec tones might appear different in detectability. Also, the spectrum of pulsed tones depends on the duration of the tone (Chapter 4), so that the two tones differ in spectra. These two additional variables (spectra and detectability) make the study of temporal discrimination difficult.

To measure sensitivity to duration, some investigators have attempted to avoid these confounding problems by presenting the listener with what they call acoustic markers. To measure duration discrimination the stimuli may consist of the following. A standard stimulus in which a tone of 170-msec duration is followed 10 msec later by an identical 170-msec tone. And, the comparison stimulus in which the two 170-msec tones are separated by 20 msec instead of by 10 msec. The listener's task is to decide whether these two stimuli are different in time separation between the two 170-msec tonal markers. If the durations are correctly judged to be different, then one assumes that the auditory system can detect this 10-msec difference in duration between the markers.

The data in Figure 10.8 show the results from such an experiment. At various standard separations of T milliseconds between the two tonal markers, the value of the additional amount of time (ΔT) required to detect the increase in duration is shown for the condition when the markers were 1000 Hz and 85 dB SPL. Thus, as the temporal separation (duration) increases, a greater and greater change in the temporal separation is required to make a temporal discrimination. Although threshold ΔT increases as a function of T, it does not increase at a rate such that $\Delta T/T$ (Weber fraction) is equal to a constant. Notice that at T equal to approximately 1 msec threshold ΔT is approximately 2 msec, so the Weber fraction is

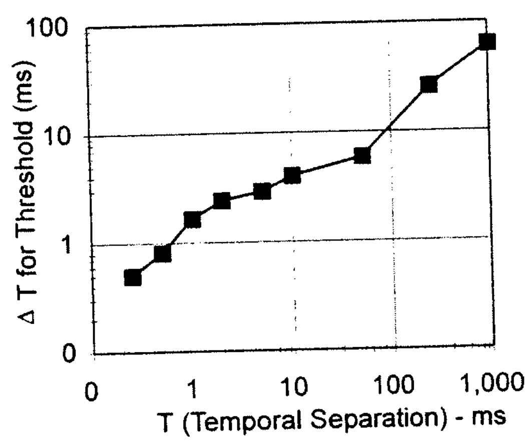

FIGURE 10.8 Value of threshold ΔT (change in temporal separation between two tonal markers in ms) is shown as a function of T (standard separation between the two tonal markers in ms). Adapted with permission from Abel (1971).

2.0. At 300 msec of T, the value of threshold ΔT is 30 msec, so the Weber fraction is 0.1.

Although these results, plus those of other investigations, show that the difference in time required for temporal discrimination increases as the standard time increases, the exact nature of the temporal relationships depends on many different stimulus conditions.

TEMPORAL MODULATION TRANSFER FUNCTIONS

The conditions used to obtain the data of Figure 10.8 are based on a single change in the temporal property of sound. Most real-world sounds have fluctuating temporal changes. Detection of sinusoidally amplitude-modulated (SAM) wideband noise can be used to measure auditory sensitivity to fluctuating temporal changes. Recall from Chapter 4 (see Figure 4.16 and equation (4.6)) that sinusoidal amplitude modulation for a noise can be written as $[1 + m \sin(2\pi F_m t)]n(t)$, where m is modulation depth, F_m is modulation rate, and $n(t)$ is the noise carrier stimulus.

The basic task for the listener, as shown in Figure 10.9, is to detect which wideband noise stimulus (panel **a** or **b**) is sinusoidally amplitude modulated. A psychometric function is obtained for each rate of amplitude modulation (F_m) relating performance, such as $P(C)$, to depth of modulation (m). Threshold modulation depth is determined from these psychometric functions and is plotted as a function of modulation rate such as that shown in Figure 10.10. The value of threshold m (threshold modulation depth, where m ranges from 0 to 1) is often expressed in decibel units as 20 log m, where 0 dB means 100% modulation (m = 1) and the more negative threshold depth becomes in decibels, the smaller the depth (for m = 0 the noise is unmodulated and 20 log m = $-\infty$; if m = 0.5, then 20 log m = -6 dB).

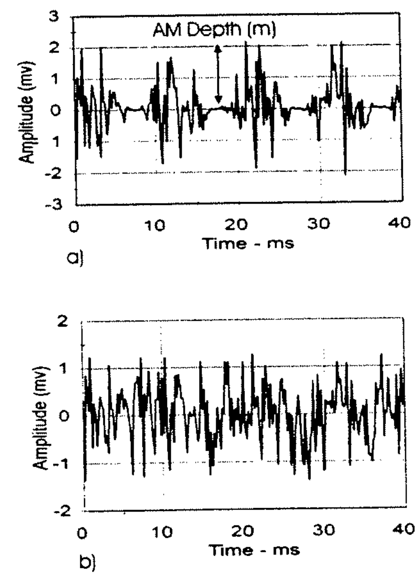

FIGURE 10.9 An unmodulated (**b**) and a sinusoidally amplitude-modulated noise (**a**) used to determine a temporal modulation transfer function. The listener is to determine which sound is amplitude modulated, and the depth of modulation (where AM depth = m) is adjusted to determine threshold.

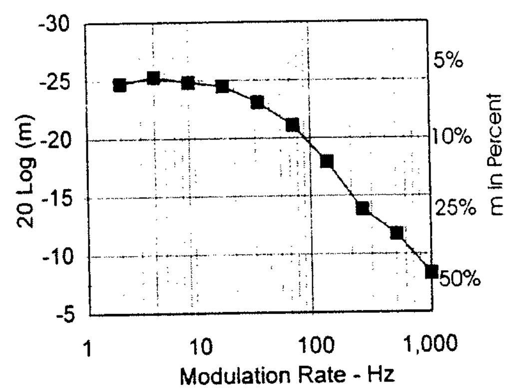

FIGURE 10.10 The temporal modulation transfer function (TMTF) for a wideband noise stimulus showing threshold depth of amplitude modulation (expressed in decibels as 20 log m, where m is the threshold depth of modulation) as a function of the rate of sinusoidal modulation. On the right-hand axis is shown modulation depth (m) in percent. Adapted with permission from Bacon and Viemeister (1985) and Viemeister (1979).

As can be seen, the ability to detect sinusoidal amplitude modulation of noise remains fairly constant up to modulation rates of approximately 50 Hz, and as modulation rate increases beyond 50 Hz the thresholds decline, indicating that fluctuations in the amplitude of the noise are more difficult to detect. That is, as modulation rate increases (the time between amplitude peaks in the modulated noise becomes shorter), the depth of modulation must be increased in order for the listener to detect the presence of modulation.

The shape of the data curve in Figure 10.10 is like that of a lowpass filter (see Chapter 5 and Figure 5.2). Thus, imagine that the auditory system acts like a lowpass filter, attenuating the depth of modulation for modulation rates above its cutoff frequency (which according to Figure 10.10 would be approximately 50 Hz), making it difficult to detect amplitude modulation for high modulation rates. Such a lowpass filter describes the way in which amplitude-modulated stimuli might be transformed by the auditory system. These data are, therefore, often described as determining the *temporal modulation transfer function* (TMTF) of the auditory system.

SUMMARY

The threshold of audibility describes the auditory system's sensitivity to level and frequency. The thresholds are measured in either an MAF or MAP paradigm, both of which involve a calibration procedure. The duration of the tone used to determine absolute threshold is crucial. If the duration is between about 10 and 300 msec, its energy must remain approximately constant for a constant level of detection by the observer. For durations of 300 msec or longer, the power must be held constant, whereas for durations shorter than 10 msec much more energy is required for tonal detection because short-duration tones are affected by spread of energy. Temporal integration is often used to describe these duration effects. The data from listeners detecting differences in frequency obey, to a first approximation, the Weber–Fechner law, with $\Delta f/f$ equaling approximately 0.002. The Weber fraction also applies approximately to discriminations of differences in tonal level. In this case, only a 0.5 to 1.0 dB difference is necessary for reliable discrimination performance. Temporal discrimination is difficult to study due to detectability and spectral changes associated with changing duration. The temporal modulation transfer function is used to describe detection of amplitude modulation for different rates of modulation.

SUPPLEMENT

Measures of auditory thresholds have existed since the early part of the twentieth century, and Sivan and White (1933) conducted the first thorough study of them. Parts of Chapters 1, 2, 3, and 4 in the textbook by Moore (1997) cover topics related to auditory sensitivity. Chapter 2 by Green in the book by Yost, Popper, and Fay (1993) also contains information relevant to this chapter. Laming (1988) provides an interesting discussion of differential sensitivity in audition and vision that pertains to some observations made in this chapter.

Because thresholds of hearing are used to define hearing loss for the hearing impaired, they have been standardized both in the United States and internationally. The ANSI standards establish the thresholds of audibility (sometimes called an *audiogram*) for young adults with normal hearing, as outlined in

Table 10.1. Copies of these and other standards may be obtained either from ANSI, 1430 Broadway, New York, New York 10018, or from The Acoustical Society of America, Standards Manager, Suite 1NO1, 2 Huntington Quadrangle, Melville, New York 11747-4502.

Sometimes an individual may have a slightly different threshold in the frequency region near 4000 Hz. Because this is the frequency region near the resonant frequency of the outer ear, it is usually assumed that alterations in normal thresholds in this frequency region are due to some consequences of these resonances.

The difference between the MAF and the MAP thresholds has been of concern to auditory scientists for years. The differences are almost entirely explained if the impedance properties of the middle and inner ear, along with the resonances of the outer ear, are carefully measured and properly combined. Killion (1978) and Yost and Killion (1997) describe these calculations.

The ability to measure thresholds at very high frequencies (above 8 kHz) is difficult because of the resonance and standing-wave acoustics of the outer ear. There are large differences in the acoustics at or near the tympanic membrane depending on the type of sound source (external speaker, supra-aural, or insert) used to present the sound. A major difference between measuring thresholds with a supra-aural and an insert earphone is the effect of internal noise due to the *occlusion effect*. When the outer ear is closed with a supra-aural earphone cushion, the effects of internal noise such as that caused by breathing and the pulsing of blood through the arteries near the ear canal interfere with the ability to detect sounds since the ear is occluded by the earphone. Deeply seated insert phones produce less of an occlusion effect and, therefore, produce lower thresholds than supra-aural phones at low frequencies (see Yost and Killion, 1997).

As mentioned in Chapters 6 and 7, the inner ear can be vibrated directly via bone conduction. A common hearing test is to vibrate the mastoid (bone of the skull behind the ear) or the forehead with a bone vibrator. One can then measure the bone vibration force required to produce a just-detectable sound for a variety of frequencies in much the same manner as the air-conducted thresholds of Figures 10.1 and 10.2 were obtained. The difference between the air-conducted (through earphones) and bone-conducted thresholds may be used to estimate the site (in the middle ear versus the inner ear or nervous system) of a hearing abnormality. The values given in Table 10.2 show the force (relative to 1 dyne) required for normal bone-vibrated thresholds, when a vibrator, meeting the specifications of the standard, is applied to the mastoid (column 1) and the forehead (column 2). These values can then be used to determine if a person has a bone-conducted hearing loss in the same way the values in Table 10.1 are used to determine if someone has an air-conducted hearing loss.

The concepts of temporal integration may be expressed in the following formula:

$$T(I - I_\infty) = C,$$

where T is the tonal duration, I is the tonal level at threshold and I_∞ is the threshold level for a very long (greater than 1 second) tone, and C is a constant. Thus, in order to maintain the value of C constant, I must decrease as T increases, as is consistent with the data shown in Figure 10.3. Recent discussion of temporal integration and alternative methods to account for changes in auditory perception as a function of temporal variables can be found in Viemeister and Plack (1993).

TABLE 10.2 RETFLs for Audiometric Bone Vibrators, from ANSI 3.6-1996 (see Table 10. 1)

Frequency (Hz)	Mastoid locations (decibels) relative to 1 dyne	Forehead location (decibels) relative to 1 dyne
250	67	79
400	61	74.5
500	58	72
750	48.5	61.5
800	47	59
1000	42.5	51
1250	39	49
1500	36.5	47.5
1600	35.5	46
2000	31	42.5
2500	29.5	41.4
3000	30	42
4000	35.5	43.5
5000	40	51
6000	40	51
8000	40	50

To be consistent with the definitions given in Chapters 2 and 3, we should refer to intensity discrimination as level discrimination. However, since the term "intensity discrimination" is used so widely, we have used it to describe sensitivity to a change in sound level. Riesz (1928) made the first accurate estimates of intensity discrimination thresholds by using a beating stimulus. Riesz reasoned that the ability to hear a slowly beating sinusoid must relate to the ability to detect the change in level that is occurring over time. As Jesteadt *et al.* (1977) argue, this method produces somewhat lower intensity discrimination thresholds than the method they used (Figure 10.7). The beating may also have some additional spectral information that aids the listener in discrimination. Thus, a more accurate estimate of intensity discrimination is probably obtained in a forced-choice procedure, using two tones of slightly different levels.

Table 10.3 outlines the different ways one might calculate these levels in estimating intensity discrimination thresholds. As Grantham and Yost (1982) showed, there are a variety of methods used to calculate the intensity discrimination thresholds. The more intense tone may be generated by adding two tones, S (the signal) and M (the masker). The less intense tone is M. (See Chapter 11 for a discussion of signals and maskers.) When adding two tones, the phase relationship (α) between the tones (see Appendix A) must be used to compute the summed level. Table 10.3 indicates the formulas used to compute a variety of measures of intensity discrimination. The entries marked a and b in Table 10.3 are the two used for the data shown in Figures 10.6–10.8. The work of Viemeister (1974) should be consulted in regard to the near-miss to Weber's law, the differences between tonal and noise intensity discrimination, and issues pertaining to coding of level by the auditory system (see also Green, 1993, and Viemeister and Plack, 1993).

Shower and Biddulph (1931) used a frequency modulation (see Chapter 4) technique to measure frequency discrimination thresholds, for approximately the same reason that Riesz used beats to measure intensity discrimination thresholds. The results of Weir *et al.* (1977), as displayed in Figures 10.4 and 10.5, are in fair agreement with these earlier data. Jesteadt and Bilger (1974) discuss problems in measuring frequency discrimination thresholds.

The potential power of the TMTF is based on treating the obtained TMTF as a lowpass filter transfer function for predicting auditory sensitivity to other forms of amplitude modu-

TABLE 10.3 Expressions for Four Common Measures of Intensity Discrimination in Terms of Power, Amplitude, and the Weber Fraction (the more intense stimulus is the sum of the signal plus masker and the less intense stimulus is the masker)

Measure of intensity discrimination	Power	Amplitude	Weber fraction
Increment power (ΔI)	$P_{m+s} - P_m$	$\frac{1}{2}V_s^2 + V_m V_s \cos\alpha$	$(\Delta I/I) = \Delta I$
Relative increment power (Weber fraction, $\Delta I/I$)	$(P_{m+s} - P_m)/P_m$	$V_s^2/V_m^2 + 2(V_s/V_m)\cos\alpha$	$\Delta I/I$[b]
Difference limen ("ΔI in dB")	$10 \log P_{m+s}$ $-10 \log P_m$	$10 \log [V_s^2/V_m^2 +$ $2(V_s/V_m)\cos\alpha + 1]$	$10 \log [(\Delta I/I + 1)]$[a] $= 10 \log [(\Delta I + I/i]$
Signal-to-masker	P_s/P_m	V_s^2/V_m^2	$[(\Delta I/I + \cos^2\alpha)^{1/2} -$ $\cos\alpha]^2$

[a]Refers to form used on left axis in Figure 10.8.

[b]Refers to form used on right axis in Figure 10.8.

P = stimulus power, V = stimulus amplitude; s = signal stimulus, m = masker stimulus; α = phase angle of addition between signal and masker. All entries not in decibels. All computations assume measurements made into a nominal 1-ohm load, such that $P = 1/T \int_0^T x^2(t)\,dt$, where T (in units of seconds) is the signal duration and $x(t)$ is the time-domain waveform. See Chapter 11 for a discussion of the signal-to-masker ratio.

lation that might be imparted to a sound. The TMTF predicts how we lose our sensitivity to modulation as the rate of modulation increases. One such form of modulation is square-wave (see Chapter 4) modulation, in which the noise is turned on and off abruptly in a repeating fashion that happens if the noise is multiplied by a square wave rather than by a sine wave. The TMTF does a good job of describing a listener's ability to detect square-wave modulation. One extreme version of square-wave modulation is a noise with a temporal gap in the middle. A sound with a temporal gap is also similar to stimuli like those discussed for Figure 10.8 in which two identical sounds mark a temporal interval. A great deal of work has been done on listener ability to detect temporal gaps of different widths for assessing temporal auditory sensitivity (see Viemeister and Plack, 1993).

One can also measure modulation detection for sinusoidal carriers (see Dau *et al.*, 1997). However, recall that sinusoidally amplitude modulating a sinusoidal carrier produces a stimulus with sideband components (see Chapter 4) that are separated from the carrier frequency by a frequency difference equal to the modulation frequency. Thus, at fast rates of modulation (high modulation frequencies) these sideband components are at frequencies that differ significantly from the carrier frequency. As a result, listeners might detect these sideband components as their cue for discriminating an unmodulated tonal carrier from a modulated one. Such sideband detection would confound the ability to measure modulation processing per se.

The measure of the temporal processing abilities of the auditory system is sometimes referred to as temporal acuity. Green (1971) and Viemeister and Plack (1993) provide excellent reviews of some of these concepts and experiments. Viemeister and Plack (1993) describe the procedures used to estimate temporal processing time from TMTF experiments. Such TMTF estimates of temporal processing time are usually much shorter than most estimates of temporal integration times. Viemeister and Plack (1993) suggest a possible way to reconcile the different estimates of temporal processing time.

CHAPTER 11

Masking

Sounds in our environment rarely occur in isolation; often many stimuli occur either simultaneously or close together in time. The study of masking is concerned with the interaction of sounds. The experimenter is interested in the amount of interference one stimulus can cause in the perception of another stimulus. A change in stimulus threshold is a typical measure of the amount of interference produced. Tonal masking, for instance, deals with the change in tonal threshold of one tone associated with the interference, or masking, produced by another tone.

TONAL MASKING

We might guess that there is probably a great deal of interaction or masking between two stimuli with frequencies that do not differ by much. To investigate this interaction, the following experiment is performed. The listener is asked to detect the presence of a weak-intensity sinusoid (perhaps 5 dB SL) at one frequency (perhaps 1000 Hz)—the *signal tone*. Another tone (the *masking tone*), usually with a frequency different from that of the signal, is presented simultaneously with the signal. The level of the masker that leads to threshold detection of the signal is used as the indicator of the amount of masking the masker has provided for the signal. If the masker is very intense and the listener can still detect the signal, the masker is not very effective in interfering with detection of the signal. On the other hand, if a weak-intensity masker causes the signal to be undetected, then the masker is effective in interfering with signal processing. Figure 11.1 shows the results from just such an experiment, in which the listener was presented with either the signal-plus-masker followed by the masker or the masker followed by the signal-plus-masker. In this two-alternative, forced-choice procedure, the listener had to decide whether the signal appeared the first time (first observation interval) or the second time (second observation interval). At masker frequencies either lower or higher than the signal frequency, the masker level required for threshold detection was higher than when the masker frequency was near the signal frequency. This indicates that frequencies different from the signal frequency are not as effective at masking as those near the signal frequency. Figure 11.1 shows a family of masking

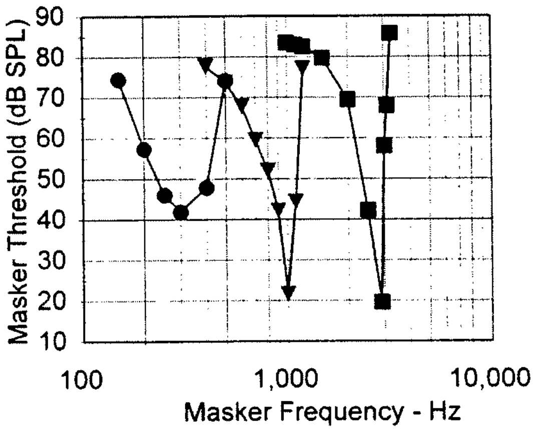

FIGURE 11.1 Three psychophysical tuning curves for simultaneous masking are shown. The different curves are for conditions in which the signal frequency was 300, 1000, and 3000 Hz. Based on data from and adapted with permission from Wightman *et al.*, (1977).

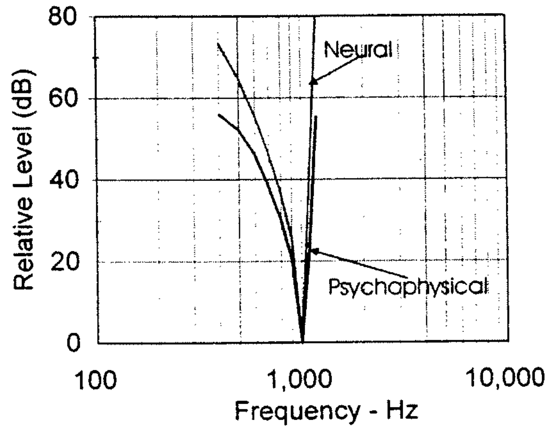

FIGURE 11.2 A comparison of an auditory nerve-tuning curve (from Figure 9.5) and a psychophysical simultaneous tuning curve (from Figure 11.1). The auditory nerve-tuning curve is narrower than the psychophysical tuning curve.

curves obtained with different signal frequencies. In general, their shapes are similar.

Both intuition and knowledge about the neural activity of the auditory periphery should enable you to understand the shape of the curves in Figure 11.1. Consider the "tuning curves" for auditory nerve fibers described in Chapter 8. These curves show that a particular neuron is most sensitive to one frequency of stimulation and that, as the frequency of stimulation differs from the nerve's best frequency, a higher level of stimulation is required to drive the neuron at its threshold value. The similarity between the procedure used to obtain the neural tuning curves and that described above to obtain the masking data shown in Figure 11.1 has led to the name *psychophysical tuning curves* for these masking functions. Not only are the procedures similar, but so are the derived functions. Both the psychophysical and neural tuning curves have a sharp tip at the center or test frequency, and the high-frequency sides of the curves are steeper than the low-frequency sides. However, the psychophysical tuning curves are not as narrow as the neural tuning curves. These comparisons are shown in Figure 11.2.

A variety of interactions can take place when two tones are presented together. These interactions are important, both for a greater understanding of how the auditory system functions and in terms of recognizing the precautions that must be taken when measuring masking. In Chapter 4, the phenomenon of beats was described. When two tones are close together in frequency, the time-domain waveform has a modulated pattern that is called beating, and we often experience a waxing and waning in loudness or beats. In this situation (i.e., when two tones are close together in frequency, one with frequency $f1$ and the other with frequency $f2$), we hear a tone at a frequency equal to the average of the two frequencies $(f1 + f2)/2$) wax and wane in loudness at a rate equal to the difference between the two frequencies ($f1 - f2$). The "best-beating" sensation for a continuously presented sound is usually heard at a rate of 3–5 Hz (the difference between $f1$ and $f2$ is equal to 3–5 Hz). The beats are strongest when the amplitudes of the two tones are equal. Thus, it is possible in masking experiments that, when the masker frequency is close to the signal frequency, the listener will hear beats. This is especially true because under these conditions the signal and masker are likely to be close together in level (see Figure 11.1). One way to avoid, or at least reduce, the extra sensation of beats during a masking experiment is to present a very short-duration signal. Since the best beats occur with a rate of 3–5 Hz, the signal must be 200 to 333 msec long for just one period of the beat to occur (i.e., the period of 3–5 Hz is 333 to 200 msec). Thus, if the signal is 30 msec long, as it was for the data shown in Figure 11.1, such a short portion of the beat period is presented that the listener usually does not hear any loudness change or beating.

Another independent property of auditory systems that could influence detection of the signal in a masking experiment is the nonlinearity of the ear (see Chapter 5). This nonlinearity can produce audible tones (aural harmonics and combination tones) in addition to the signal and masker, especially at high stimulus levels. Although aural harmonics and combination tones can occur in a masking experiment, they are usually not detected in a psychophysical tuning experiment because of the low signal level and short signal duration.

To indicate how beats, aural harmonics, and combination tones can influence results from a study of masking, consider the following experiment conducted by Wegel and Lane (1924). They asked listeners to detect the presence of signals of different frequencies while an 80 dB SL, 1200-Hz tone served as the masker. For the fixed 1200-Hz masker, the level of the signal was adjusted until the subject could just barely discriminate a difference between the masker and the signal plus masker. The data are shown in Figure 11.3 as the threshold of the signal (expressed in dB SL) versus the frequency of the signal. These data indicate that there is an *upward spread of masking*, in that a masker of a given frequency masks higher-frequency signals more than lower-frequency signals. Because signal frequency was varied in this experiment, the masking data are sometimes referred to as *masking patterns* to differentiate this pattern from the psychophysical tuning curve patterns shown in Figures 11.1 and 11.2, where the signal frequency was kept constant.

The shaded areas in Figure 11.4 indicate that, when the signal was close in frequency to the masker, the observer heard beats that indicated the presence of the signal. Notice that beats were also reported at harmonics (2400 and 3600 Hz) of the 1200-Hz masker. If a nonlinear system is excited with a single frequency, higher harmonics of the input frequency can be detected. When a pure tone of one frequency is presented to the auditory system at a high level, listeners report hearing tones at that frequency and tones at frequencies equal to harmonics (usually the first two or three harmonics) of the frequency presented. These audible higher harmonics are called *aural harmonics*, and their presence indicates that the auditory system is nonlinear. Figure 11.3 shows that at a signal frequency of 2400 Hz (the second aural harmonic of the masker frequency, 1200 Hz) and at 3600 Hz (the third aural harmonic of the masker) the signal frequency is beating with the masker harmonics. In other words, the intense 80-dB SPL, 1200-Hz masker has produced aural har-

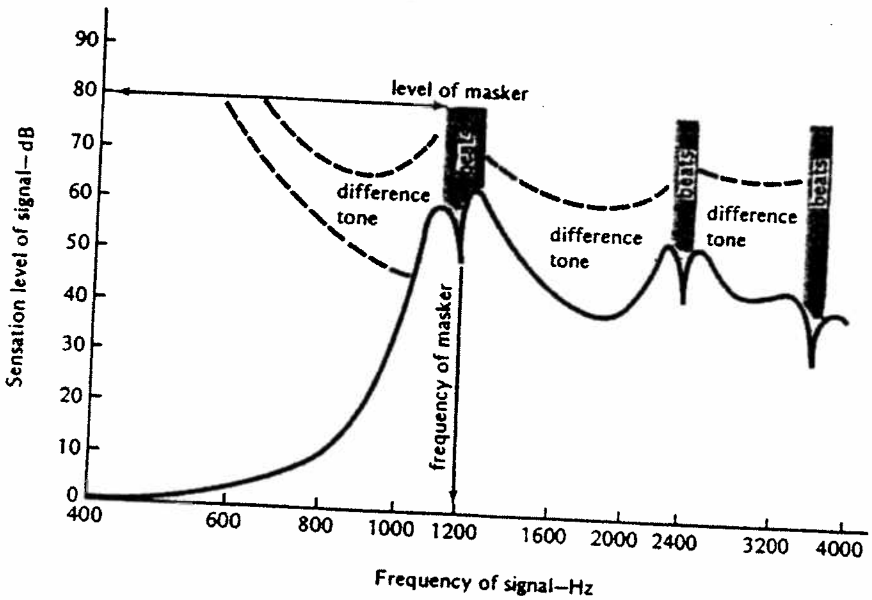

FIGURE 11.3 Masked threshold of signal in dB SL as a function of signal frequency in Hz with a 1200-Hz, 80-dB SL masker. The figure indicates that, in addition to detecting the masker and the signal, the observer can also detect beats and nonlinear difference tones. Adapted with permission from Wegel and Lane (1924).

monics at 2400 and 3600 Hz. When the signal frequency is close to those of these aural harmonics (e.g., 2403 or 3603 Hz), the beating occurs between the signal frequency and that frequency produced by the nonlinear properties of the auditory system (e.g., a 3-Hz beat is generated by a 2403-Hz signal and the 2400-Hz second aural harmonic produced by the 1200-Hz masker).

In addition to the aural harmonics produced by the nonlinearity of the auditory system, combination tones are present when the signal and masker are presented simultaneously. The two types of combination tones (produced by the signal and masker) heard in this experiment were the *primary and secondary difference tones*. The frequency of the primary difference tone is equal to the difference in frequency between the masker and signal. The

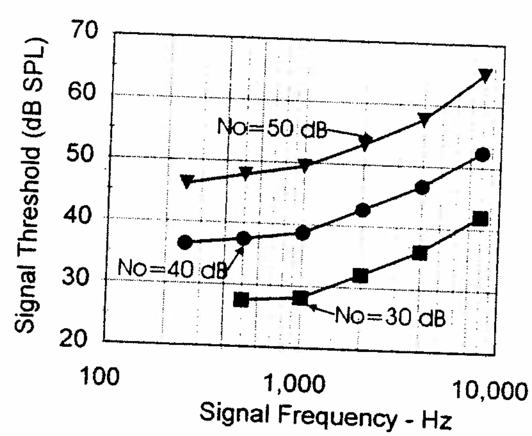

FIGURE 11.4 The signal level (dB SPL) required for noise-masked threshold is shown as a function of signal frequency for three levels (expressed in spectrum level, N_o) of the wideband masking noise. Based on data from and adapted with permission from Reed and Bilger (1973).

frequency of the secondary difference tone (also called the *cubic difference tone*; see also Chapters 8, 9, and 13) is produced by the difference between twice the masker frequency minus the signal frequency, or twice the signal frequency minus the masker frequency (see Appendix A). Wegel and Lane's pure tone masking experiment shows that listeners do detect these difference tones. Although both beats and combination tones can be heard when two tones are presented together, the two phenomena (beats and combination tones) represent very different aspects of hearing.

NOISE MASKING

Since white noise (see Chapter 4) contains a wide range of frequencies, we would expect it to mask tones of many different frequencies. In a noise-masking experiment, a broadband white Gaussian noise, whose spectrum level (N_o, see Chapter 4) was varied is used to mask a tonal signal with different frequencies. The masked threshold of the signal was measured. The data from this type of experiment are shown in Figure 11.4. With no noise present, the thresholds of hearing show that the threshold for a tone depends to a large extent on the tone's frequency. However, as the noise background level is increased, the masked threshold for the pure tone is less dependent on the frequency of the tone. Another important aspect of these data is that, above an N_0 of approximately 20 dB, an increase in the spectrum level of the noise means that the signal level must be increased by approximately the same amount for the signal to be detected (i.e., for each decibel increase in the level of the masker, the signal must be increased by the same amount to maintain constant detection). The data indicate that over a wide range of levels and frequencies the signal energy must be 5–15 dB more intense than the spectrum level of the noise for the signal to be detected, as shown in Figure 11.5.

In these masking experiments, the ratio of the signal energy to the spectrum level of the noise is used to describe the masked threshold. The signal-to-noise ratio (E/N_0) is expressed as the energy of the signal (E) divided by noise power per unit bandwidth (N_0). In decibels, the signal-to-noise ratio is equal to the signal energy (in dB) minus the spectrum level (in dB). The data in Figure 11.5 from the noise-masking experiments suggest that E/N_0 must be approximately 5–15 dB for the signal to be detected, with E/N_0 being lower for low signal frequencies and increasing as the signal frequency increases.

CRITICAL BAND AND THE INTERNAL FILTER

The data of Figures 11.1–11.3 suggest that those masker frequencies near that of the sig-

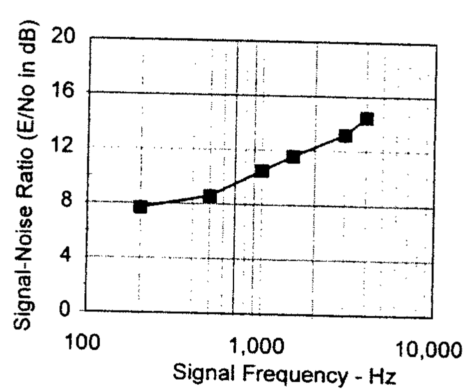

FIGURE 11.5 The signal- (energy) to-noise (spectrum level) ratios required for detection are shown as a function of signal frequency. Adapted with permission from Reed and Bilger (1973).

nal are important in determining masking. Thus, we might expect that, as the noise was passed through a narrower and narrower bandpass filter (i.e., a narrower and narrower band of noise is providing the masking), detection of a tone whose frequency was in the center of the noise's passband would become easier. Conversely, one expects that a signal with a frequency that was not in the passband of the filter would be very easy to detect because there would be no energy in the masker at the same frequency as the signal frequency.

Fletcher performed a band-narrowing experiment in 1940 and made some assumptions about the frequency region of the noise that would be effective in masking the tone. He assumed that some sort of "internal filter" was centered around the frequency of the signal and that the total noise power coming through that internal filter determined the amount of masking for the signal. That is, detection of a tonal signal is determined by the amount of total power present in a narrow range of frequencies. This narrow range of frequencies is determined by the internal filter. Figure 11.6 shows this idealized internal filter schematically for two noise spectra. In Figure 11.6a, the noise spectrum is much broader than the passband of the internal filter, and hence the maximum amount of masking occurs because the maximum amount of total power is coming through the filter. In Figure 11.6b, the spectrum of the masking noise is narrower than the passband of the internal filter, and the signal is easier to detect than that indicated in Figure 11.6a because less than a maximum amount of noise power is coming through the filter. Fletcher called the internal filter the *critical band*, because the frequencies within the passband of the internal filter were *critical* for masking.

A band-reject filtered noise such as that shown in Figure 11.7 offers an excellent masker for estimating the shape of the critical band. Thresholds for detecting a tonal signal whose frequency is centered in the spectral notch of the band-reject noise are determined as a function of the spectral width of the spectral notch, as indicated in Figure 11.7a. The resulting masking data (Figure 11.7b) can be used to estimate the shape of the critical band. It is assumed that the total power coming through the filter determines the amount of masking. As the spectral notch is widened, there will be less total power coming through the filter and therefore less masking, as indicated in Figure 11.7b.

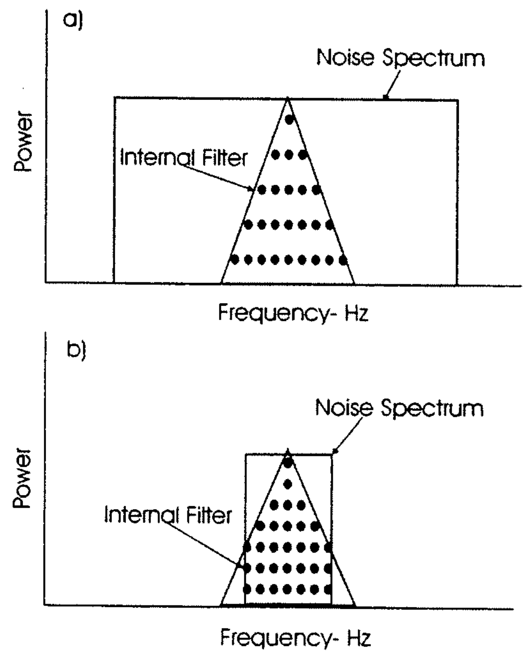

FIGURE 11.6 Schematic diagram of the "internal filter" (the triangle). (**a**) Broadband noise "produces" maximum power (maximum area) at the output of the internal filter. (**b**) The bandwidth of the noise is less than the bandwidth of the internal filter (less area under the filter). There is more masking for a signal whose frequency is at the center of the internal filter for the broadband noise (**a**) than for the narrow-band noise (**b**).

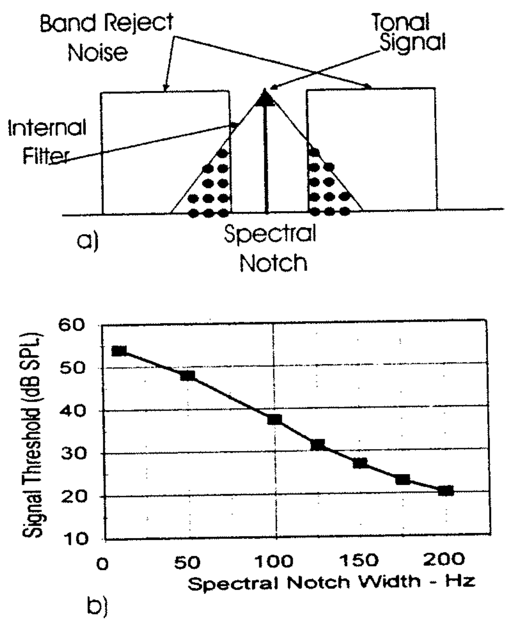

FIGURE 11.7 (a) A noise band with a spectral notch or gap is used to mask a signal whose frequency is in the center of the spectral gap. (b) The masked thresholds for detecting a 1000-Hz signal are shown as a function of increasing the spectral notch of the band-reject noise. Adapted with permission from Patterson and Moore (1989).

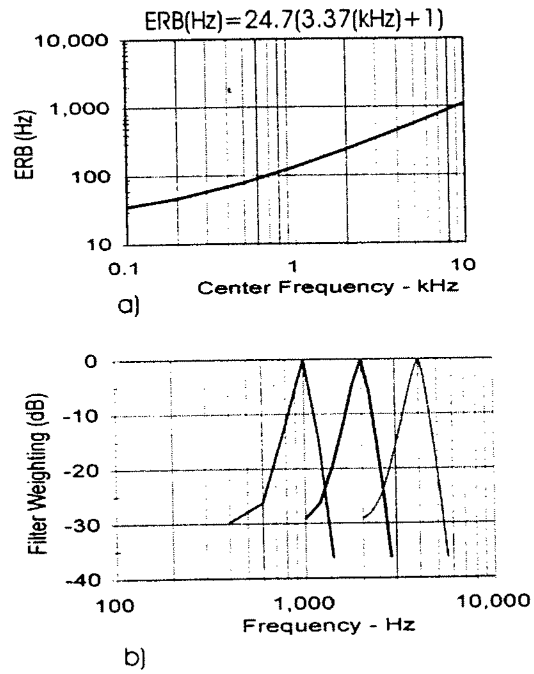

FIGURE 11.8 (a) The width (in Hz) of the estimated critical band (ERB or Equivalent Rectangular Bandwidth) is shown as a function of signal frequency. Data were obtained from experiments in which noises with spectral gaps masked tonal signals. The equation in the Figure 11.8a shows a fit to the data that can be used to estimate ERB critical bandwidths for critical bands centered at any frequency. In the equation, "f" is the center frequency of the critical band (i.e., the signal frequency) and is expressed in terms of kHz. Adapted with permission from Moore (1997). (b) Estimates of three critical-band internal filters based on the ERB experiments described in Figure 11.7 are shown. (The filters are based on the Rounded Exponential filter model, *roex*, from Patterson and Moore, 1989).

The shape of these masking data can be used to estimate the shape of the internal, critical band filter. The bandwidth of the derived critical band filter (called the *equivalent rectangular bandwidth,* ERB) can be obtained for different signal frequencies. Figure 11.8 shows the ERB as a function of frequency, and the results indicate that the width of the critical band is proportional to its center frequency (i.e., the signal frequency). The bottom part of Figure 11.8 displays estimates of auditory filters centered at different center frequencies (signal frequency). The estimated shape of the filters does not change much with signal frequency, but bandwidth does. The width of the critical band (or ERB) also increases with increasing signal level as indicated in Figure 11.9. This increase in critical bandwidth with level may represent nonlinear properties of cochlear transduction as discussed in Chapters 8 and 16.

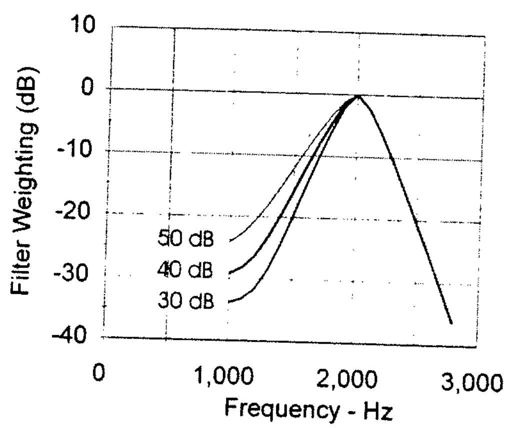

FIGURE 11.9 Estimates of the critical-band internal filters from experiments in which the level of tonal signal was increased in the notched-noise–masking experiment. The width of the filters increase with increasing level. (The filters are based on the Rounded Exponential filter model, *roex*, from Rosen and Baker, 1994).

RELATIONSHIP BETWEEN EXCITATION PATTERNS AND CRITICAL BANDS

We have used the concept of an internal filter (or critical band) to explain masking data when the masker contains frequencies different from the signal. The data of Figures 11.1 and 11.2 involving the psychophysical tuning curve and those of Figures 11.7, 11.8, and 11.9 involving noise maskers with a spectral gap can be used to derive estimates of the shape and bandwidth of the critical band. In these experiments, the signal is kept fixed in frequency, and we assume that the listener detects the signal by monitoring the critical band centered on the signal frequency.

For the masking pattern data shown in Figure 11.3, the frequency of the masker was kept constant and the signal frequency changed. Thus, the listener is assumed to monitor a different critical band for each signal frequency (each critical band with a center frequency at the signal frequency). In order to explain the masking pattern data, it is assumed that the masking tone stimulates a number of different neurons, one neuron with its best frequency at the masker's frequency and other neurons with best frequencies near that of the masker's frequency. The neuron with its best frequency equal to that of the masker would be stimulated the most, and the other neurons would be stimulated less depending on how close their best frequencies were to that of the masker and on the overall level of the masker. That is, the masking tone sets up a pattern of excitation in the auditory nerve. If we imagine that the detection of a signal tone masked by a masking tone is mediated by the *excitation pattern*, we can explain the results shown in Figure 11.3. That is, the excitation caused by the masker spreads to critical bands located above and below the masker in frequency. When the listener monitors the critical band centered on the signal frequency, the critical band will contain energy due to the spread of excitation from the masker (assuming that the masker and signal are not too far apart in frequency and the level of the masker is sufficiently high for the excitation to spread to that critical band), and this energy will mask the signal whose frequency is at the center of the critical band. Figure 11.10 describes how an excitation pattern can be obtained from critical-band internal filters. Notice the similarity in the shape of the excitation pattern of Figure 11.10b and the masking pattern data of Figure 11.3.

TEMPORAL MASKING

In the masking experiments just described, the masker and signal occurred simultaneously in time. There are many acoustical events in which two stimuli follow one an-

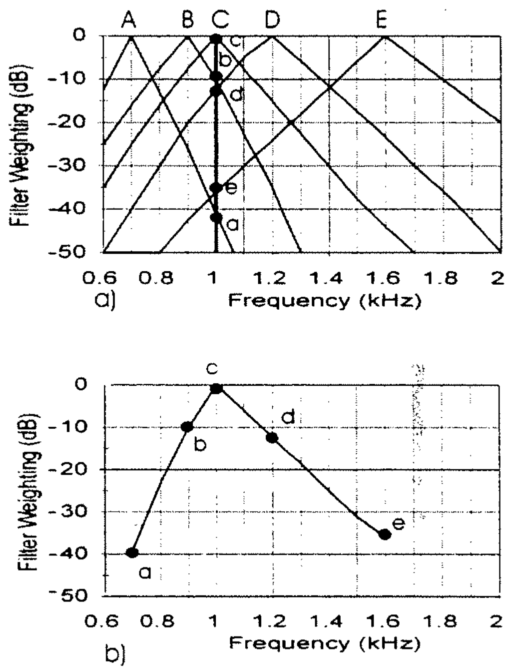

FIGURE 11.10 (a) Five critical-band internal filters are shown. (b) If a masker tone is placed at 1000 Hz (1 kHz), the masking pattern is obtained by assuming that the excitation produced in any critical band by the masker at 1000 Hz is equal to the amount of energy coming through that critical band at 1000 Hz. Thus, for the filter labeled "A" the amount of excitation for this filter at 1000 Hz is "a" or -40 dB. Thus, in the excitation pattern (Figure 11.10b) at the frequency equal to the center frequency of filter "A" (700 Hz), the amount of excitation is -40 dB (point a). Similar calculations can be made for filters B, C, D, and E yielding the other points b, c, d, and e on the excitation pattern (Figure 10.10b).

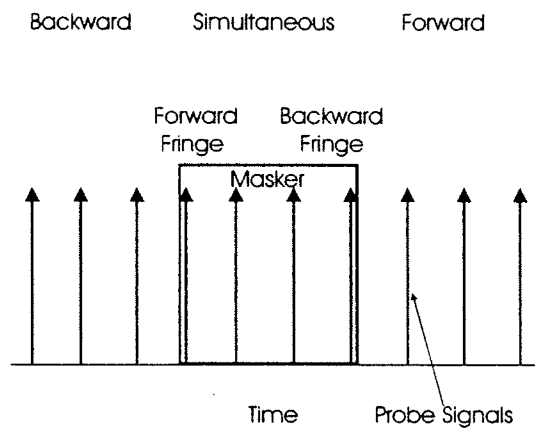

FIGURE 11.11 Schematic diagram of temporal positions of probe signals in relation to a pulsed masker. Signals can occur before the masker (backward masking), during the time of the masker presentation (simultaneous masking), or after the masker (forward masking). Forward and backward fringes also produce masking.

other. For instance, in music the notes usually appear sequentially in time, and in speech words appear in sequence. Psychoacousticians have therefore studied the amount of masking provided for a signal that occurs before or after the masker.

Figure 11.11 is a schematic diagram of the stimulus conditions used in studies of *temporal masking*. Signals or probe tones can occur at different times relative to the masker (rectangle in Figure 11.11). The signal probes in the middle of the masker represent the simultaneous-masking conditions we have already studied. When the signal is presented near the beginning or end of the masker, *backward fringe masking* or *forward fringe masking* occurs. When the signal precedes the masker in time, the condition is called *backward masking*; when the signal follows the masker in time, the condition is *forward masking*.

Various stimuli have been used in temporal masking studies (tones, noises, speech, clicks), and the results shown in Figure 11.12 demonstrate the salient data from these experiments. More masking occurs in the fringe conditions than in the simultaneous situation, such that a signal placed in the forward fringe is masked more than one placed in the backward fringe (the masking that occurs when the signal is in the forward fringe is sometimes referred to as

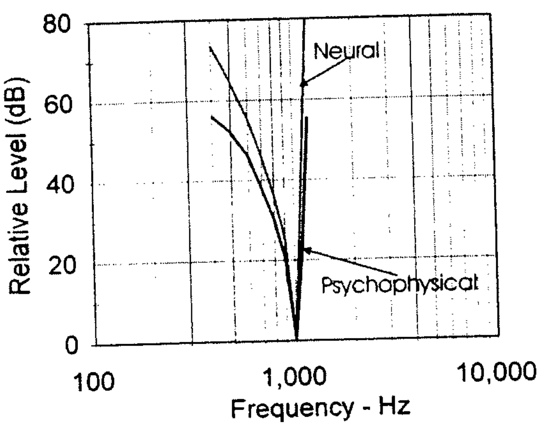

FIGURE 11.12 Schematic diagram of the relative change in signal thresholds as a function of the temporal position of the probe signal in relation to the masker.

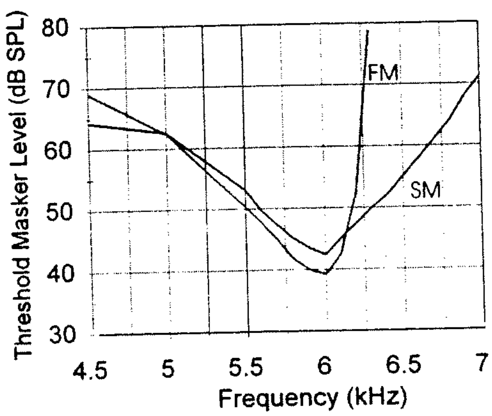

FIGURE 11.13 A comparison of a forward (FM) and a simultaneous masked (SM) psychophysical tuning curve obtained for similar stimulus conditions. The signal was 6000 Hz. Forward-masked tuning curves are very similar to neural tuning curves obtained from the eighth nerve (see Figures 7.17 and 9.5). Based on data from and adapted with permission from Moore (1978).

overshoot, as if the extra masking caused by the masker onset has overshot the masking caused by the steady-state portion of the masker). Forward masking of a stimulus can take place when a temporal difference between the two stimuli is between 75 and 100 msec, and backward masking occurs up to 50 msec. Thus, the amount of backward masking declines more quickly than does the amount of forward masking as a function of increasing the temporal separation between the signal and masker.

TONAL–TEMPORAL MASKING

At the beginning of this chapter, we described the psychophysical tuning curve. If there is forward masking, then we might expect that the effect of the forward masker would be frequency dependent, as it was for simultaneous masking (Figure 11.1). The data shown in Figure 11.13 marked as FM were obtained in the same manner as described for Figure 11.1, except in this case the signal appeared immediately after the masker was turned off (forward masking, FM). These forward masking results show that the masker does influence signal detection when it does not overlap the signal. The effects are about the same as when the signal and masker overlap in time (Figure 11.1). Figure 11.13 shows a direct comparison of psychophysical tuning curves obtained in simultaneous masking (SM) and in forward masking (FM). As can be seen, this comparison indicates that the psychophysical tuning curve is sharper in forward masking than in simultaneous masking. The sharper tuning curve means that the auditory system is better able to detect the presence of the signal in forward masking than in simultaneous masking for the same masker frequency.

These tuning curves are based on one pure tone masking another pure tone. We have already studied the use of noise maskers, but what happens for other complex maskers? Let us consider the case of using a two-tone complex as the masker. In this experiment, the actual masker (*M*) will be fixed at a particular level (40 dB SPL) and frequency (1000 Hz), and the level of the signal (the signal also has a frequency of 1000 Hz) is varied to determine a threshold. In a test or baseline condition, the signal threshold is determined when the masker is equal in frequency to the signal (i.e., both are at 1000 Hz). In the test conditions, a second tone is added to the masker, so that the masking stimulus consists of two tones: the initial masker *M* and the second masking tone *SU*. The level of the second tone will be 20 dB above the level of the masker, or *M*, tone; the second tone, *SU*, is therefore 60 dB SPL. We will then vary the frequency of the second tone (*SU*) and determine the threshold for detecting the signal for each value of *SU*. Finally, we will perform this experiment for both simultaneous masking and forward masking.

Figure 11.14 shows the two stimulus conditions and the results. The vertical axis in each figure is the change in signal threshold from the baseline condition. Recall that in the baseline condition only the masking tone (*M*) was presented and the signal and masker were equal in frequency (1000 Hz). The solid horizontal line at 0 decibels represents this masking condition. Thus, if the second tone, *SU*, provides masking in addition to that caused by *M* presented alone, the signal thresholds should increase above 0 dB. As can be seen, in simultaneous masking most masking (about 20 dB of masking) occurs when the second tone, *SU*, is equal in frequency to the masking tone, *M* (*SU*, *M* and the signal are all 1000 Hz). As the second tone (*SU*) becomes different in frequency from the masking tone (*M*), the amount of threshold changes above 0 dB decreases until the difference between the two masking tones is so large that only the original masking tone, *M*, continues to provide masking (i.e., masking is back at the baseline amount of masking, 0 dB). These data are similar to those shown in Figure 11.4.

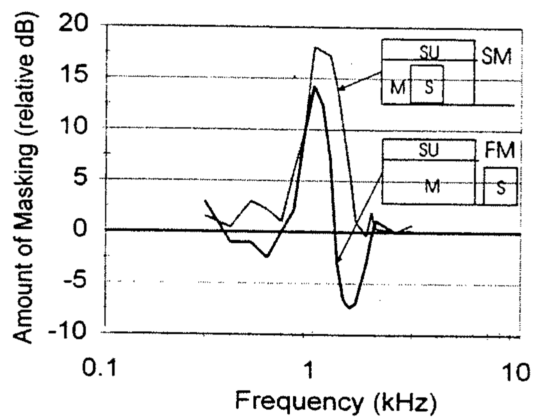

FIGURE 11.14 A comparison of two-tone tonal masking in simultaneous (the SM condition shown with the light curve) and forward masking (the FM condition shown with the dark curve). The frequency of the suppressor tone (*SU*) is shown on the horizontal axis and the level of the 1000-Hz signal (*S*) required for threshold detection is shown on the vertical axis. Signal threshold is shown relative to that required to detect the signal when only the masking tone (*M*) was present. The masking tone is presented at 40 dB SPL and with a frequency of 1000 Hz. The suppressor tone is presented at 60 dB SPL. In the simultaneous-masking condition, the signal is presented in the temporal middle of the masking stimulus (either *M* alone or *M* and *SU* presented together), while in the forward-masking condition the signal is present after the offset of the masking stimulus. Signal thresholds above 0 dB mean that the suppressor tone (*SU*) increased the amount of masking provided by the masker tone (*M*), thresholds below 0 dB mean that the suppressor tone (*SU*) caused less masking than that produced by the masker tone (*M*) when it was presented alone (e.g., *SU* suppressed the masking ability of *M*), and thresholds at 0 dB mean that only the masker tone (*M*) was providing masking. Adapted with permission from Shannon (1974).

Let us compare this simultaneous-masking effect to what happens in the forward-masking condition. Notice that in forward masking the values of signal threshold are negative when the second masking tone, *SU*, is slightly greater in frequency than the masking tone, *M*. That is, when *SU* is added to *M*, signal threshold can be lower than that obtained when just *M* is presented. In this example, the 1000-Hz signal is easier to detect when the 1000-Hz masker (*M*) is present along with a second tone (*SU*) of slightly higher frequency (e.g., 1250 Hz). It is as if the second tone (*SU*) has made the 1000-Hz masking tone (*M*) a less effective masker. The second tone (*SU*) is sometimes referred to as a *suppression tone* because the second tone (*SU*) is suppressing the masking ability of the masking tone (*M*).

Labeling the second tone the *suppression tone* allows us to describe the two-tone masking effect. Care should be used, however, in drawing too many conclusions about the nature of this suppression-like effect. Some investigators refer to the effect as "unmasking" in that the second tone has "unmasked" the effect of the masker (*M*). "Unmasking" is viewed as a more neutral term than suppression because it does not imply that the second tone (*SU*) actually interacts in a direct fashion with *M*. The unmasking or suppression phenomenon is another factor that must be considered in attempting to account for the way in which the auditory system operates when two or more stimuli exist in the environment.

SUMMARY

Masking of a tonal signal by a tonal masker has been used to study the interaction of sounds occurring simultaneously and to probe the frequency selectivity of the auditory system. Psychophysical tuning curves are often used to describe the masking effect of one tone on another tone. The results indicate that low-frequency tones mask high-frequency signals more than high frequencies mask low frequencies. Sometimes, due to the presence of beats and combination tones, the exact relationship between tonal signal threshold and tonal masker frequency is difficult to determine. Beats indicate that the auditory system has a limited frequency-resolving power but that it can follow the amplitude of the input stimulus. Combination tones indicate the extent to which the auditory system is nonlinear. When white Gaussian noise is used as a masker for tonal signals, the ratio of signal energy (*E*) to masker spectrum level (N_o) required for masked threshold is approximately 5 to 15 dB as the spectrum level of the noise or the frequency of the signal is varied over a considerable range. Critical bands and excitation patterns are used to estimate the bandwidth properties of the frequency-resolving capability of the auditory system. Fringe masking and forward and backward masking are used to determine the interaction of sounds not occurring simultaneously. Tuning curves obtained in forward masking are sharper than those obtained in simultaneous masking. Psychophysical suppression or unmasking can occur when the masker consists of two or more frequencies.

SUPPLEMENT

Throughout the discussion of masking we have stressed the concept of the "internal filter." The psychophysics of hearing suggests that the nervous system operates as if there is a bank of bandpass filters that process sound. We have already learned (in Chapters 7–9) that both the biomechanics and the neural aspects of auditory physiology demonstrate the existence of neural tuning that resembles bandpass filtering. A challenge for auditory scientists is to determine to what extent the neural tuning observed in the cochlea and auditory nerve accounts for the psychophysics of masking. Many of these issues and data are reviewed in the book edited by Moore (1986), *Frequency Selectivity in Hearing,* in Chapter 3 of a textbook by Moore (1997), by Moore and Patterson (1986), and by Moore in a chapter in the book edited by Yost, Popper, and Fay (1993).

Another variable that can affect a listener's ability to detect signals in masking experiments is "off-frequency listening." The basic idea behind off-frequency listening is that the listener may be able to perform the detection task by listening in a frequency region different from that where the signal occurs (i.e., off the signal frequency or off frequency). In many situations, the signal may excite a wide frequency region, and a larger signal-to-masker ratio may exist in a frequency region that is different from that of the signal. Using a noise with a spectral notch (see Figure 11.4) reduces the ability of the listener to "listen off-frequency" to detect the signal in the notch-noise–masking procedure, because this noise energy exists in all frequency regions, except in the narrow region near the signal. An excellent discussion of off-frequency listening can be found in Moore's (1986) book.

Patterson and Moore (1986) used a particular filter shape called the rounded exponential (*roex filter*) to fit masking data like those shown in Figure 11.7. The roex filter function can be written as

$$W(g) = (1 + pg) \exp(-pg),$$

where $W(g)$ is the linear (non-decibel) value of the filter output ($0 \le W(g) \le 1$); $g = |f - fo|/f$, where fo is the signal frequency and f is a frequency on the filter function; p is determined by the bandwidth and slope of the filter such that, the higher the value of p, the more sharply tuned the filter, and p is obtained by finding the best fitting function $W(g)$ to the data; "exp" is the exponential argument. Once p is determined from the data, the ERB or the equivalent rectangular bandwidth (see Supplement to Chapter 5) can be obtained from the roex filter as:

$$\text{ERB} = 4fo/p.$$

The *gammatone filter* (see Patterson *et al.*, 1995 and Rosen and Baker, 1994) is another filter function used to describe the shape of the critical band filter. The time domain (the *impulse function*) description of the gammatone function is:

$$t^{n-1} \exp(-2\pi bt) \cos(2\pi fot),$$

where n and b *are constants and* fo is the CF of the filter. In the frequency domain the filter function is approximately equal to

$$(1 - r)\,[1 + \{(f - fo)^2\}/b^2]^4,$$

where fo is the filter's CF, f is a frequency on the filter, b controls the filter's sharpness (like p for the roex filter), and r is proportional to the slope of the filter. In these measurements, filter bandwidth is sometimes referred to in terms of ERBs. One can estimate the frequency associated with an ERB via this equation:

$$F = (\text{ERB} - 24.7)/107.94,$$

where F is expressed in kHz (Glasberg and Moore, 1990).

Weber (1983) has suggested that there are at least three explanations for the narrower forward-masked tuning curves (see also Lutfi, 1988). Psychophysical suppression appears to be much more effective for high-frequency suppression tones on a lower-frequency masker tone than for low-frequency suppression tones on a high-frequency masking tone. Suppression by low-frequency suppression tones on the masker tone has been observed, but usually only when the suppression tone is fairly high in level (see Shannon, 1974). Although comparisons between psychophysical suppression and two-tone neural suppression (see Chapter 7) are tempting, there are some arguments that the two may not be the same (see Moore and Glasberg, 1982).

The two stimuli in the intensity-discrimination experiment described in Chapter 10 differ only in that one is more intense than the other. The more intense stimulus could have been produced by adding two tones of the same frequency: *Tone A* with a level of *I* in decibels and tone *B* with a level such that when it was added to tone *A* the summed level in decibels would equal $I + \Delta I$ in dB (recall that when two stimuli of equal frequency are added the sum depends on the phase difference between two tones). Thus, data from the intensity experiment could be plotted as the level of tone *B* required for the subject to determine that it was added to tone *A* versus the level of tone *A*. The fact that the level of tone *B* must be raised above absolute threshold due to the presence of tone *A* is defined as *tone A masking tone B* (see Table 10.3).

Fletcher observed that, when the spectrum of the noise was broad (i.e., the internal filter contained the maximum total noise power required for masking), the power of the just detectable masked signal (the signal at masked threshold) was equal to the total power contained within the critical band. Fletcher's observation makes it possible to predict the width of the critical band without performing a band-narrowing or band-reject experiment. According to Fletcher's assumptions, masked signal power (*Ps*) is equal to the power of the noise in the critical band (*Pncb*): *Ps* = *Pncb*. Since *Pncb* = $N_o \times$ CBW (see Chapter 4), where *No* is noise spectrum level and CBW is an estimate of the critical bandwidth, CBW = Ps/N_0. Expressed in decibels, 10 log CBW = *Ps* in dB – N_o in dB, and 10 log CBW is referred to as the *critical ratio*. Thus, CBW can be estimated directly from a single masking threshold. However, the signal level (*Ps*) required for detection depends on how efficient the auditory system is at detecting the signal, as well as on the width of the critical band. Thus, if the total power within the critical band does not *equal* the power of the masked signal at threshold (i.e., *Ps* does not equal *Pncb*), then the basic assumption of the critical ratio calculation is violated and the critical ratio will not yield a valid estimate of the width of the critical band.