5.6 READING

The relevant sections of Pickles and Møller provide good introductions to the physiology of frequency selectivity:

Pickles, J. O. (1988). *An introduction to the physiology of hearing* (2nd ed.). London: Academic Press.
Møller, A. R. (2000). *Hearing: Its physiology and pathophysiology*. New York: Academic Press. Chapters 3 and 6.

Details of cochlear frequency selectivity can be found in:

Robles, L., & Ruggero M. A. (2001). Mechanics of the mammalian cochlea. *Psychol. Rev.*, 81, 1305–1352.

For an introduction to the psychophysics of frequency selectivity:

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). London: Academic Press. Chapter 3.

For an excellent overview of auditory compression, from a physiological and psychophysical perspective:

Bacon, S. P., Fay, R. R., & Popper, A. N. (Eds.). (2004). *Compression: From cochlea to cochlear implants*. New York: Springer-Verlag.

6

Loudness and Intensity Coding

Sounds are largely characterized by variations in intensity across frequency and across time. To use speech as an example, vowel sounds are characterized by variations in intensity across frequency (e.g., formant peaks), and consonants are characterized (in part) by variations in intensity across time (e.g., the sudden drop in intensity that may signify a stop consonant, such as /p/). To identify these sounds, the auditory system must have a way of *representing* sound intensity in terms of the activity of nerve fibers, and a way of making *comparisons* of intensity across frequency and across time. The purpose of this chapter is to show how information regarding sound intensity is analyzed by the auditory system. This chapter examines how we perceive sound intensity, and discusses the ways in which sound intensity may be represented by neurons in the auditory system.

6.1 THE DYNAMIC RANGE OF HEARING

At the start, let us consider the range of sound levels with which the auditory system can cope. The *dynamic range* of a system is the range of levels over which the system operates to a certain standard of performance. To determine the dynamic range of human hearing, we need to know the lower and upper level limits to our ability to process sounds effectively.

6.1.1 Absolute Threshold

Absolute threshold refers to the lowest sound level a listener can perceive in the absence of other sounds. This is usually measured for pure tones at specific frequencies—to find, for example, a listener's absolute threshold at 1000 Hz. If you have had a hearing test, the audiologist may have measured the *audiogram* for one or both of your ears. The audiogram is a plot of absolute threshold as a function of the frequency of a pure tone. In a clinical setting, audiograms are often plotted in terms of hearing loss (the lower the point on the graph, the greater the hearing loss) relative to the normal for young listeners.

The lower curve in Fig. 6.1 shows a typical plot of absolute threshold as a function of frequency for young listeners with normal hearing, measured by presenting pure tones from a loudspeaker in front of the head. Note that sensitivity is

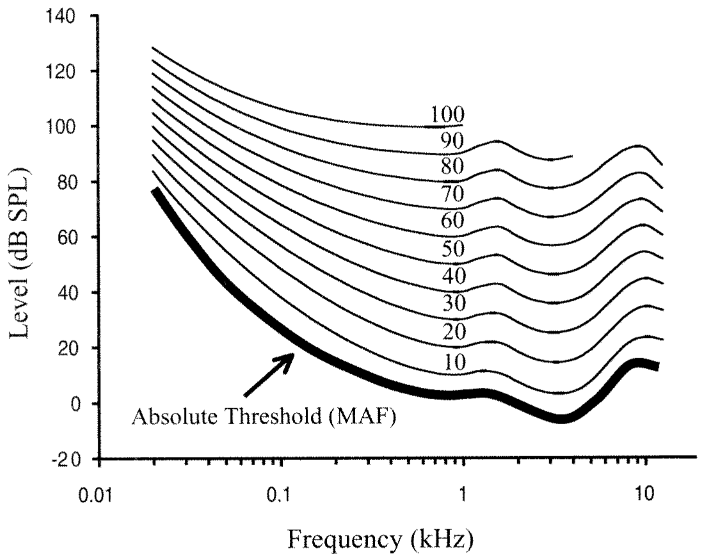

FIG. 6.1. Equal loudness contours. Each curve represents the levels and frequencies of pure tones of equal loudness, measured relative to a 1000-Hz pure tone. A pure tone corresponding to a point on a given curve will sound as loud as a pure tone at any other point on the same curve. The level of the 1000-Hz tone for each curve (in dB SPL) is shown on the figure above each loudness contour. Also shown is the lowest detectable level (absolute threshold) at each frequency. Stimuli were presented in the "free field," with a sound source directly in front of the listener. The data are re-plotted from the latest draft ISO standard (ISO/DIS 226).

greatest (threshold is lowest) for sound frequencies between about 1000 and 6000 Hz, and declines for frequencies above and below this region. The region of high sensitivity corresponds to resonances in the outer and middle ear, and sounds in this frequency range are transmitted to the cochlea more efficiently than sounds of other frequencies (see Section 4.1.1). The frequency range of normal hearing is about 20 Hz to 20 kHz in humans (the range often extends to higher frequencies in other mammals). Near the extremes of this range, absolute threshold is very high (80 dB SPL or more): Although not shown in Fig. 6.1, thresholds increase rapidly for sound frequencies above 15 kHz or so.

6.1.2 The Upper Limit

How loud is too loud? The sound level at which we start to feel uncomfortable varies considerably between individuals. I am used to quite high sound levels. Because I have been in rock bands for a large proportion of my adult life 110 dB SPL would not cause me extreme discomfort. However, 110 dB SPL would be quite uncomfortable for many people, and if the level were raised much above 120 dB SPL, most of us would experience something akin to physical pain, almost irrespective of the frequency of the sound. Exposure to these high sound levels for only a short time may cause permanent damage to the ear.

Another way of approaching this problem is to ask over what range of sound levels can we make use of differences in level to distinguish sounds. Although we may be able to hear a sound at 140 dB SPL, we may not be able to distinguish it from a sound at 130 dB SPL by using the auditory system. Not many experiments have been conducted at very high levels on humans, for obvious ethical reasons, but it appears that our ability to detect differences between the levels of two sounds begins to deteriorate for sound levels above around 100 dB SPL, although discrimination is still possible for levels as high as 120 dB SPL (Viemeister & Bacon, 1988). These results are discussed in more detail in Section 6.3.2.

In summary: The lower and upper limits of hearing suggests that the dynamic range of hearing (the range of levels over which the ear operates effectively) is about 120 dB in the mid-frequency region (1000–6000 Hz) and decreases at low and high frequencies. 120 dB corresponds to an increase in pressure by a factor of one million and an increase in intensity by a factor of one million million. In other words, the quietest sounds we can hear are about one million million times less intense than sounds near pain threshold.

6.2 LOUDNESS

6.2.1 What Is Loudness?

Loudness can be defined as the perceptual quantity most related to sound intensity. We use words like "quiet" and "loud" to refer to sounds that we hear in our daily lives ("turn down the TV, it's too loud"). Strictly speaking, however, loudness refers to the *subjective* magnitude of a sound, as opposed to pressure, intensity, power, or level, which refer to the *physical* magnitude of a sound. If I turn up the amplifier so that the intensity of a sound is increased, you *perceive* this as a change in loudness. It is *not* accurate to say "this sound has a loudness of 50 dB SPL." Decibels are units of *physical* magnitude, not subjective magnitude.

6.2.2 Loudness Matching: Effects of Frequency, Bandwidth, and Duration

Because loudness is a subjective variable does this mean that it cannot be measured? Fortunately, the answer is no. One of the obvious things we can do is to ask listeners to vary the level of one sound until it seems as loud as another sound. In this way, we can compare the loudness of sounds with different spectral and/or temporal characteristics. An example is shown in Fig. 6.1. The figure shows *equal loudness contours*, which are curves connecting pure tones of equal loudness, all matched to a 1000-Hz pure tone. Note that, like the absolute threshold curve, the equal loudness curves dip in the middle, indicating a region of high sensitivity around 1000–6000 Hz. The loudness curves flatten off slightly at high levels. Loudness does not vary as much with frequency at high levels. It follows that the growth of loudness with level is greater at low frequencies than at high frequencies.

The *loudness level* (in units called *phons*) of a tone at any frequency is taken as the level (in dB SPL) of the 1000-Hz tone to which it is equal in loudness. From Fig. 6.1 we can see that a 100-Hz pure tone at 40 dB SPL is about as loud as a 1000-Hz pure tone at 10 dB SPL. So a 100-Hz tone at 40 dB SPL has a loudness level of about 10 phons. Similarly, any sound that is as loud as a 1000-Hz tone at 60 dB SPL has a loudness level of 60 phons.

Loudness matching can also be used to measure the effects of bandwidth on loudness, by varying the level of a sound of fixed bandwidth until it matches the loudness of a sound with a variable bandwidth. Experiments like these have shown that increasing the bandwidth of a noise with a *fixed overall level* (so that the spectrum level decreases as bandwidth is increased) results in an *increase* in loudness, *if* the noise is wider than the auditory filter bandwidth and presented at moderate levels (see Fig. 6.2). If the noise is narrower than the filter bandwidth, then variations in bandwidth have little effect on the loudness of a sound with a fixed overall level. To look at this in another way, if the power of a sound is distributed over a wider region of the cochlea, then the loudness may increase. Loudness is not determined simply by the level of a sound: It also depends on the spectral distribution.

Finally, loudness matching can be used to measure the effects of duration on loudness by varying the level of a sound of fixed duration until it matches the loudness of a sound with a different duration. Up to a few hundred milliseconds, the longer the sound, the louder it appears. At mid levels, a short-duration pure tone has to be much higher in level than a long-duration tone to be judged equally loud

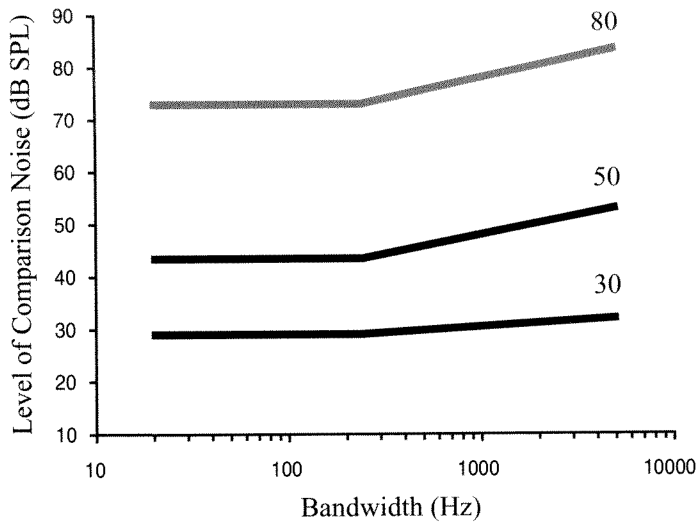

FIG. 6.2. Variations in loudness with bandwidth for a noise band geometrically centered on 1420 Hz. The three curves show the level of a comparison stimulus judged to be as loud as the noise band, for different overall levels of the noise band (30, 50, or 80 dB SPL). For a given overall level, the overall level does not change as bandwidth is increased (hence the *spectrum* level decreases with increases in bandwidth). The comparison stimulus was a noise band geometrically centered on 1420 Hz with a bandwidth of 2300 Hz. Based on Zwicker, Flottorp, and Stevens (1957).

(Buus, Florentine, & Poulsen, 1997). The difference is less at low and high levels. The mid-level effect is probably related to mid-level compression on the basilar membrane (see Section 5.2.3): At mid levels a greater change in the *physical* level of the tone is necessary to produce the change in basilar membrane vibration (or change in excitation level) required to compensate for the increase in duration. For long duration sounds, it is hard to judge the "overall" loudness as opposed to the loudness at a particular instant, or over a short time period, and loudness matches may become very variable.

6.2.3 Loudness Scales

Loudness matching can provide important information about how loudness is influenced by some stimulus characteristics (e.g., frequency, bandwidth, and duration), but it cannot tell us directly how loudness changes with sound *level*. Loudness matching cannot provide a number that corresponds directly to the magnitude of our sensation. If we had such a set of numbers for different sound levels, we would be able to construct a *loudness scale* describing the variation in subjective magnitude with physical magnitude.

One of the ways we can measure the variation in loudness with level is to simply present two sounds and ask a listener to give a number corresponding to how much louder one sound seems than the other. Alternatively, we can ask a listener to adjust the level of one sound until it seems, say, twice as loud as another sound. These two techniques are called *magnitude estimation* and *magnitude production* respectively, and have been used with some success to quantify the subjective experience of loudness. Steven's power law (Stevens, 1957, 1972) is based on the observation that, for many sensory quantities, the subjective magnitude of a quantity scales with the *power* of the physical magnitude of that quantity. For loudness:

$$L = kI^{\alpha} \tag{6.1}$$

where L is loudness, I is sound intensity, and k is a constant. Loudness quantified in this way is expressed in units called *sones*, where one sone is defined as the loudness of a 1000-Hz pure tone with a level of 40 dB SPL. A sound that appears to be four times as loud as this reference tone has a loudness of 4 sones, and so on. The exponent, α, appears to be somewhere between 0.2 and 0.3 for sound levels above about 40 dB SPL, and for frequencies above about 200 Hz. For sound levels below 40 dB SPL, and for frequencies less than 200 Hz, loudness grows more rapidly with intensity (the exponent is greater). Figure 6.3 illustrates the growth of loudness with sound level for a 1000-Hz pure tone.

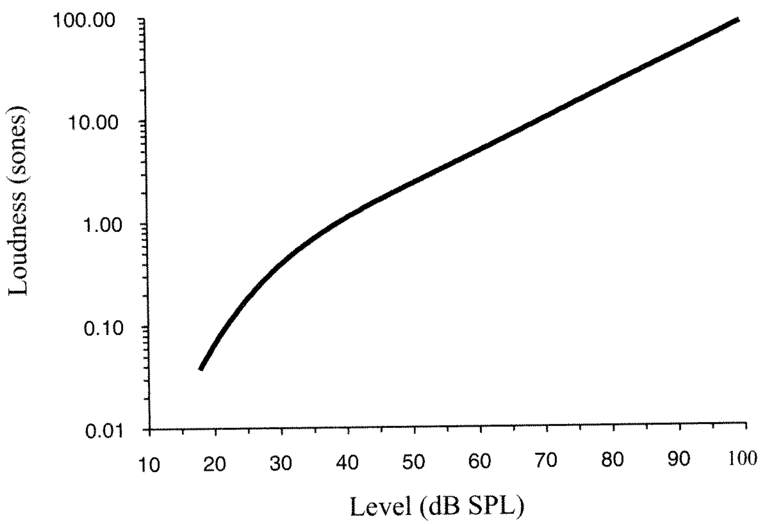

FIG. 6.3. The relation between loudness (plotted in sones on a *logarithmic* scale) and sound level, for a 1000-Hz pure tone. The curve is based on magnitude estimation and magnitude production data collected by Hellman (1976).

To give a better idea of the loudness scale, imagine that you want to increase the level of a sound so that it seems twice as loud as it did before. What increase in level (in dB) would you need? If we assume, for the sake of argument, that the exponent is 0.3, then the sound level would have to increase by 10 dB. This corresponds to an increase in intensity by a factor of 10. Loudness is a *compressed* version of sound intensity: A ten-fold increase in intensity produces a doubling in loudness. If you want your new guitar amplifier to sound four times as loud as your old one you have to increase the power rating (in Watts) by a factor of 100!

It is interesting to note that the function relating sound level to loudness is similar to the function relating sound level to the velocity of basilar membrane motion. Both the basilar membrane response function and the loudness growth function are steeper at low levels than they are at higher levels, when plotted on logarithmic axes as in Fig. 5.4 and Fig. 6.3. Indeed, loudness may be roughly proportional to the *square* of basilar membrane velocity (Schlauch, DiGiovanni, & Reis, 1998). Of course, sounds are detected by virtue of the vibrations that they elicit on the basilar membrane. It not surprising that basilar-membrane compression determines the growth of loudness.

6.2.4 Models of Loudness

The effects of sound level and bandwidth can be explained by models of loudness based on the excitation pattern (Moore, Glasberg, & Baer, 1997; Zwicker & Scharf, 1965). These models calculate the "specific" loudness at the output of each auditory filter, which is a compressed version of stimulus intensity in each frequency region, reflecting the compression of the basilar membrane. You can think of the specific loudness at a particular center frequency as being the loudness at the corresponding place on the basilar membrane (see Moore, 2003, p. 134). The final loudness of the sound is taken as the *sum* of the specific loudness across center frequency, with the frequencies spaced at intervals proportional to the ERB at each frequency (higher center frequencies are spaced further apart than lower center frequencies). Because this spacing of frequencies corresponds to a roughly constant spacing in terms of distance along the basilar membrane (see Section 4.2.2), the process is broadly equivalent to a summation of specific loudness along the length of the basilar membrane. In other words, loudness may be a measure of the total activity of the basilar membrane.

Consider the effect of bandwidth on loudness described in Section 6.2.2. If the bandwidth is doubled, then the excitation spreads to cover a wider region of the cochlea. If the overall intensity is held constant, then the intensity per unit frequency will be halved. However, the reduction in the specific loudness at each center frequency will be much less than one half because of the compression (Fig. 6.4): The input/output function has a shallow slope, so that any change in physical intensity will result in a much smaller change in the intensity of basilar-membrane vibration. The modest reduction in specific loudness at each center

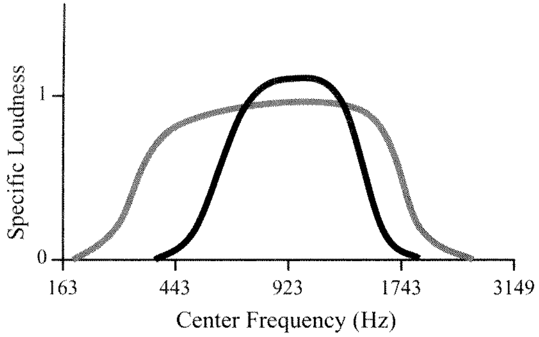

FIG. 6.4. Specific loudness patterns for two bands of noise with different bandwidths but with the *same overall level*. In models of loudness, the loudness of a sound is assumed to be equal to the area under the specific loudness pattern, so that the narrow band of noise (black line) has a much lower loudness than the wide band of noise (grey line). The bandwidth of the noise has doubled, so that the spectral density of the noise has been halved. However, the specific loudness at each center frequency reduces by much less because of cochlear compression. Note that the center frequencies are spaced according to an ERB scale (equal distance corresponds to equal number of ERBs). Based on Moore (2003, Fig. 4.5).

frequency is more than compensated for by the increase in bandwidth, because doubling bandwidth with a constant spectral density can produce a large increase in the overall loudness (the output from the different center frequencies are added linearly). It follows that spreading the stimulus energy across a wider frequency range, or across a wider region of the cochlea, can increase the loudness.

Loudness models can be used to estimate the loudness of an arbitrary sound. Their success suggests that our sensation of loudness is derived from a combination of neural activity across the whole range of characteristic frequencies in the auditory system.

6.3 HOW IS INTENSITY REPRESENTED IN THE AUDITORY SYSTEM?

One of the basic properties of auditory nerve fibers is that increases in sound level are associated with increases in *firing rate*, the number of spikes the fiber produces every second (see Section 4.4.2). The answer to the question in the heading seems obvious, therefore: Information about the intensity of sounds is represented in terms of the *firing rates of auditory neurons*. If only it were that simple. . . .

6.3.1 Measuring Intensity Discrimination

Intensity discrimination refers to our ability to detect a difference between the intensities of two sounds. In a typical experiment (see Fig. 6.5), the two sounds

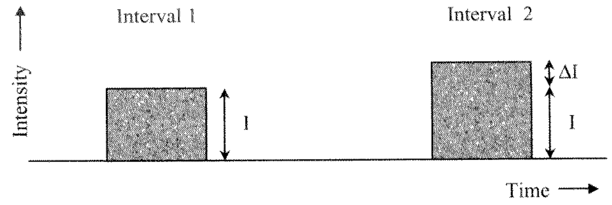

FIG. 6.5. The stimuli for a typical intensity discrimination experiment. The listener's task is to pick the observation interval that contains the most intense sound (interval 2 in this case). The interval containing the most intense sound would be randomized from trial to trial. The sound in interval 1 is called the *pedestal*, because it can be regarded as a baseline sound to which the increment (of intensity ΔI) is added.

to be compared are presented one after the other, with a delay between the two *observation intervals* (the times during which the stimuli are presented) of half a second or so. The listener is then required to pick the most intense sound. Over the course of several such trials, the intensity difference between the two sounds is reduced until the listener can no longer perform the task above a certain criterion (say, 71% correct).

The *just-noticeable difference* (jnd) in intensity can be expressed in many ways. Two of the most popular are the *Weber fraction* (expressed in dB), and ΔL. The Weber fraction is generally used in perception research to refer to the ratio of the smallest detectable change in some quantity to the magnitude of that quantity. For auditory intensity discrimination:

$$\text{Weber fraction} = \Delta I / I, \tag{6.2}$$

where I is the baseline (or *pedestal*) sound intensity, and ΔI is the smallest detectable change in that intensity (the difference between the intensity of the higher-level sound and the intensity of the lower-level sound, or pedestal, when the listener can just detect a difference between them). Expressed in dB, the equation becomes:

$$\text{Weber fraction (in dB)} = 10 \times \log_{10}(\Delta I / I). \tag{6.3}$$

It follows from Equation 6.3 that, if you need to double the intensity of one sound to detect that its intensity has changed (i.e., if ΔI is equal to I), then the Weber fraction is 0 dB [$10 \times \log_{10}(1) = 0$]. If the jnd corresponds to an increase in intensity that is less than a doubling (i.e., if ΔI is less than I), then the Weber fraction in dB is *negative* (the logarithm of a number between zero and one is negative).

ΔL is the jnd expressed as the ratio of the intensity of the higher-level sound to the intensity of the lower-level sound when the listener can just detect a difference between them:

$$\Delta L = 10 \times \log_{10}\{(I + \Delta I)/I\} \tag{6.4}$$

Note the distinction between ΔL and the Weber fraction. If you need to double the intensity of one sound to detect that its intensity has changed, then ΔL is about 3 dB $[10 \times \log_{10}(2) \approx 3]$. Unlike the Weber fraction in dB, ΔL can never be negative, because $(I + \Delta I)/I$ is always greater than 1. If the jnd is very large (ΔI is much greater than I), then the Weber fraction in dB and ΔL are almost the same.

6.3.2 The Dynamic Range Problem

Just how good are we at intensity discrimination? The Weber fraction for wideband white noise (noise with a flat spectrum), is about −10 dB (corresponding to a ΔL of about 0.4 dB) and is roughly *constant* as a function of level, for levels from 30 dB SPL up to at least 110 dB SPL (Miller, 1947, Fig. 6.6). For levels below 30 dB SPL, the Weber fraction is higher (performance is worse). In Miller's experiment, these levels refer to the *overall* level of a noise that has a flat spectrum between 150 and 7000 Hz. A constant Weber fraction implies that the smallest detectable increment in intensity is proportional to the pedestal intensity, a property called *Weber's law* that is common across sensory systems. A Weber fraction of −10 dB means that we can just detect a 10% difference between the intensities of two wideband noises.

It has been found that the Weber fraction for pure tones *decreases* with level (performance improves), for levels up to about 100 dB SPL. This has been called the "near miss" to Weber's law (McGill & Goldberg, 1968). Figure 6.6 shows data from Viemeister and Bacon (1988). Although the Weber fraction increases again for levels above 100 dB SPL, performance is still good for these very high levels:

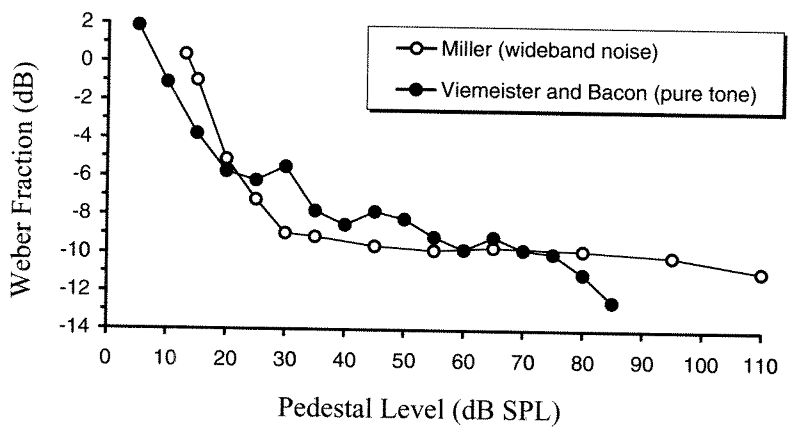

FIG. 6.6. Intensity discrimination for a wideband noise and for a 1000-Hz pure tone, as a function of level. In both experiments listeners were required to detect a brief increment in a continuous sound. Data are from Miller (1947) and Viemeister and Bacon (1988).

In the same study, Viemeister and Bacon measured a Weber fraction of −6 dB for a pure-tone pedestal with a level of 120 dB SPL.

This is where we have a problem. Human listeners can discriminate intensities at levels as high as 120 dB SPL for pure tones. However, most auditory nerve fibers—the high spontaneous rate fibers—are saturated by 60 dB SPL (see Section 4.4.2). That is, if you increase the level of a 60-dB SPL pure tone with a frequency equal to the characteristic frequency of a high spontaneous rate fiber, the neuron will not increase its firing rate significantly. Most neurons cannot *represent* sound levels above 60 dB SPL in terms of their firing rates alone, and cannot provide any information about changes in level above 60 dB SPL. Furthermore, the minority of fibers that do have wide dynamic ranges, the low spontaneous rate fibers, have shallow rate-level functions compared to the high spontaneous rate fibers. A shallow rate-level function means that changes in level have only a small effect on firing rate, and so these fibers should not be very sensitive to differences in intensity. How can we possibly be so good at intensity discrimination at high levels?

6.3.3 Coding by Spread of Excitation

One explanation for the small Weber fraction at high levels for pure tones (and other stimuli with restricted bandwidths), is that listeners are able to use information from the whole excitation pattern. At low levels, only a small region of the basilar membrane is stimulated (the region surrounding the place tuned to the pure tone's frequency), but as the level is increased, a wider area is stimulated. The extra information traveling up the auditory nerve may benefit intensity discrimination at high levels for two reasons. First, although most nerve fibers with characteristic frequencies close to the frequency of the pure tone may be saturated at high levels, neurons with characteristic frequencies remote from the frequency of the tone (representing the skirts of the excitation pattern, or regions of the basilar membrane remote from the place of maximum vibration) will receive less stimulation and may not be saturated. These neurons will be able to represent the change in level with a change in firing rate.

Figure 6.7 shows a simulation of the activity of high spontaneous rate and low spontaneous rate fibers as a function of characteristic frequency, in response to a 1000-Hz pure tone presented at different levels. These plots can be regarded as neural excitation patterns (see Section 5.4.4). The peak of the excitation pattern is saturated for the high spontaneous rate fibers at high levels because those neurons with characteristic frequencies close to the frequency of the tone are saturated. For example, a high spontaneous rate fiber with a characteristic frequency of 1000 Hz (equal to the pure-tone frequency) does not change its firing rate as the level is increased from 80 to 100 dB SPL. However, fibers tuned lower and higher than the stimulation frequency are *not* saturated. These fibers can represent the change in level (e.g., observe the change in firing rate of the 2000-Hz fiber as level is increased from 80 to 100 dB SPL).

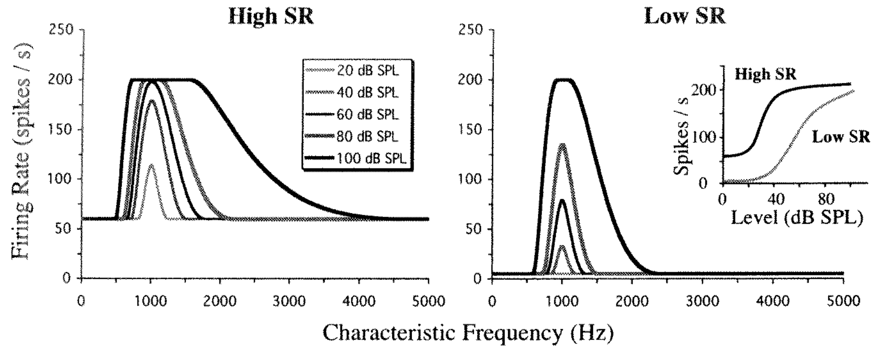

FIG. 6.7. A simulation of firing rate as a function of characteristic frequency, for representative high spontaneous rate (left panel), and low spontaneous rate (right panel) auditory nerve fibers, in response to a 1000-Hz pure tone presented at different levels. The rate-level functions of the two fibers in response to a tone at characteristic frequency are illustrated to the right of the figure.

A second possible reason for the benefit of spread of excitation is that listeners may combine information from across the excitation pattern to improve performance (Florentine & Buus, 1981). The more neurons that are utilized, the more accurate is the representation of intensity. Information from the high-frequency side of the excitation pattern may be particularly useful, as this region grows more rapidly with stimulus level than the center of the excitation pattern. This arises from the compressive growth of the central region of the excitation pattern compared to the linear growth of the high-frequency side (see Section 5.4.4). Note in Fig. 6.7 that, at moderate levels, the change in firing rate with level is greatest on the high-frequency side for the high spontaneous rate fibers.

Some researchers have tested the hypothesis that information from the skirts of the excitation pattern is used to detect intensity differences, by masking these regions with noise. Figure 6.8 shows the results of such an experiment (Moore & Raab, 1974), in which intensity discrimination for a 1000-Hz pure tone was measured with and without the presence of a band-stop masking noise (noise that has a notch in the spectrum). The noise would have masked information from the skirts of the excitation pattern, yet addition of the noise resulted in only a small deterioration of performance at high levels, removing the near miss and resulting in Weber's law behavior (as also observed for wideband noise).

The pattern of firing rates across the auditory nerve may provide more information about the level of pure tones than is present in a single nerve fiber or in a group of nerve fibers with similar characteristic frequencies. The use of this extra information may explain the reduction in the Weber fraction with level, referred to as the near miss to Weber's law. However, even if the use of the excitation pattern is restricted by the addition of masking noise, performance at high levels is still

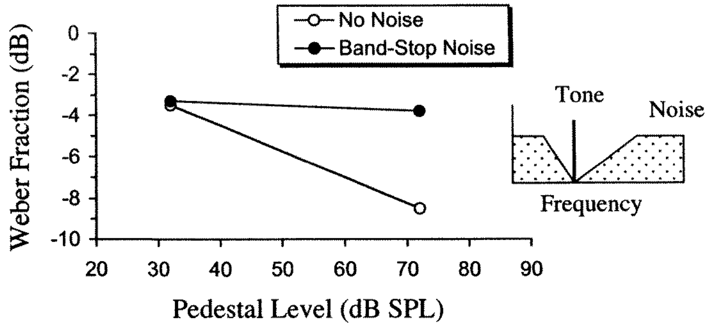

FIG. 6.8. Intensity discrimination for a 1000-Hz pure tone, with and without a band-stop noise. The noise had cutoff frequencies of 500 Hz and 2000 Hz (see schematic spectrum). The data are those of one listener (BM) from Moore and Raab (1974).

very good. It seems that the auditory system can represent high sound levels using neurons with only a narrow range of characteristic frequencies, many of which must be saturated.

6.3.4 Coding by Phase Locking

In some circumstances, intensity may be coded by the pattern of phase locking in auditory neurons. Recall that nerve fibers will tend to fire at a particular phase of the fine structure of the sound waveform (see Section 4.4.4). For a pure tone in the presence of noise, for example, the auditory system may be able to detect the tone by detecting regularity in the otherwise irregular firing pattern (noise produces irregular patterns of phase locking). Furthermore, increasing the intensity of a pure tone in the presence of a noise can increase the degree of regular, synchronized firing to the tone, even when the nerve fiber is saturated and cannot change its firing rate. The *pattern* of activity in auditory neurons can change with level, even if the spikes per second do not. This may help the auditory system to represent complex sounds at high levels, since each nerve fiber will tend to phase lock to the dominant spectral feature close to its characteristic frequency, for instance a formant peak (Sachs & Young, 1980). The spectral profile can be represented by the changing pattern of phase locking with characteristic frequency, rather than by changes in firing rate with characteristic frequency.

In situations in which intensity discrimination for a pure tone is measured in a noise, such as the band-stop noise experiment illustrated in Fig. 6.8, nerve fibers may represent changes in level by changes in the degree of *synchronization* of spikes to the pure tone. The band-stop noise that was used to prevent the use of the skirts of the excitation pattern may incidentally have increased the effective dynamic range of the fibers! However, intensity discrimination in band-stop noise is still possible at high levels and at high frequencies (Carlyon & Moore, 1984) above the frequency at which phase locking to fine structure is thought to break down. Although phase locking may contribute to intensity discrimination in some situations, it does not look like the whole story. Despite this, the possibility that intensity is represented, and stimuli are detected, by virtue of changes in the pattern of firing rather than in the overall firing rate, is a very interesting one, and is currently a subject of investigation (Carney, Heinz, Evilsizer, Gilkey, & Colburn, 2002).

6.3.5 The Dynamic Range Solution?

The psychophysical data described in Section 6.3.3 suggest that Weber's law is the characteristic of intensity discrimination when a relatively small number of auditory nerve fibers responding to a restricted region of the cochlea are used, even for stimulation at high levels. Given that only a small number of auditory nerve fibers have large dynamic ranges, and that these fibers have shallow rate-level functions, we would expect intensity discrimination to be much better at low levels than at high levels (in contrast to the psychophysical findings). An analysis of auditory nerve activity by Viemeister (1988) confirms this interpretation. Viemeister took a representative range of rate-level functions based on physiological data. He also took into account the *variability* in firing rate for each neuron (the amount the firing rate varies over the course of several identical stimulus presentations). The shallower the rate-level function or the greater the variability, the less is the sensitivity of the fiber to changes in level. The overall variability in the representation can be reduced by combining information across a number of fibers (assumed to have similar characteristic frequencies). In this way, Viemeister was able to predict Weber fractions as a function of level.

The results of Viemeister's analysis are shown in Fig. 6.9. The figure shows the best possible intensity-discrimination performances that could be achieved based on the firing rates of a group of 10 neurons and a group of 50 neurons. The curves show that there is much more firing-rate information at low levels (around 30 dB SPL) than at high levels, as expected from the characteristics and relative numbers of the high and low spontaneous rate fibers. The *distribution* of firing-rate information in the auditory nerve as a function of level does not correspond to human performance. However, a comparison of these curves with the human data in Fig. 6.8 shows that only 50 neurons are needed to account for discrimination performance for pure tones in band-stop noise over a fairly wide range of levels. Even if the number of usable fibers is restricted by the band-stop noise, we might still expect several hundred, perhaps even several thousand, to provide useful information about the vibration of a small region of the basilar membrane. According to Viemeister's analysis, this would decrease the predicted Weber fractions even further.

It is probable, therefore, that there is enough firing-rate information in the auditory nerve to account for human performance across the entire dynamic range

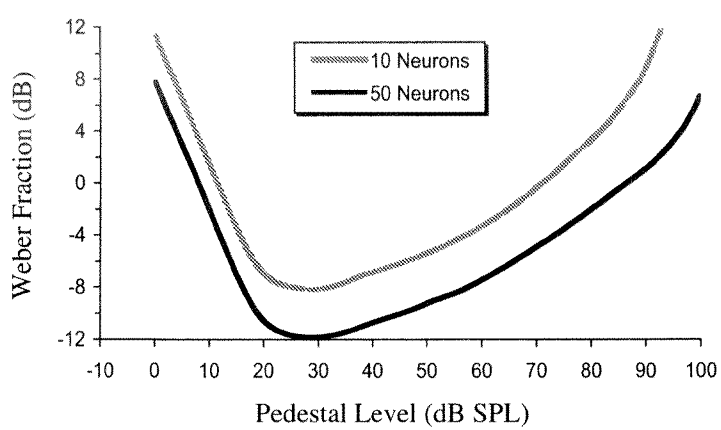

FIG. 6.9. Weber fractions predicted by the optimum use of information from a sample of 10 and a sample of 50 auditory never fibers. The curves are replotted from Viemeister (1988).

of hearing. Despite the low numbers and shallow rate-level functions of the low spontaneous rate fibers, these neurons seem to be sensitive enough to take over the responsibility for intensity coding at levels above the saturation point of the high spontaneous rate fibers (see Fig. 6.7). The pertinent question may not be: Why are we so good at intensity discrimination at high levels? but: Why are we so bad at intensity discrimination at low levels? It appears that intensity discrimination is limited by processes central to the auditory nerve (brainstem? cortex?) that do not make optimum use of the information from the peripheral auditory system.

A final point: The remarkable dynamic range of human hearing is dependent on the compression of the basilar membrane. The compression is a consequence of the level-dependent action of the outer hair cells, which effectively amplify low-level sounds but not high-level sounds (see Section 5.2.5). The shallow growth in basilar membrane velocity with level means that the low spontaneous rate fibers have shallow rate-level functions and therefore wide dynamic ranges (see Section 5.3.2). The auditory system uses compression to map a large range of physical intensities onto a small range of firing rates.

6.4 COMPARISONS ACROSS FREQUENCY AND ACROSS TIME

6.4.1 Absolute and Relative Intensity

The *loudness* of a sound is related to the *absolute* intensity of the sound: The higher the sound pressure level, the louder the sound appears. Absolute intensity may be a useful measure in some situations, when estimating the proximity of a familiar

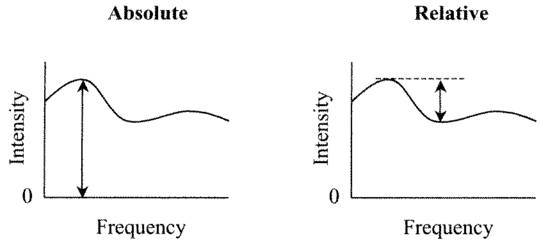

FIG. 6.10. Absolute and relative (in this case, across frequency) measures of intensity.

sound source, for example. When we are *identifying* sounds, however, absolute intensity is not very important. The identity of the vowel /i/ is the same whether the vowel sound is played at an overall level of 50 dB SPL or 100 dB SPL. To identify sounds, it is much more important for us to be sensitive to the *relative* intensities of features within the sound (see Fig. 6.10), for instance, the spectral peaks and dips that characterize a vowel. The overall level may vary between presentations, but as long as the spectral shape stays the same, so will the identity of the vowel.

6.4.2 Profile Analysis

A comparison across the spectrum is sometimes called *profile analysis*, because the auditory system must perform an analysis of the spectral envelope, or profile. Many of the important early experiments were conducted by Green and colleagues (see Green, 1988). In a typical profile analysis experiment, a listener is presented with a complex spectrum consisting of a number of pure tones of different frequencies. In one observation interval the tones all have the same level. In the other interval the level of one of the tones (the "target") is higher than that of the other tones. The listener's task is to pick the interval containing the incremented tone, equivalent to picking the spectrum with the "bump." To prevent listeners from just listening to the target in isolation, the overall level of the stimulus is randomly varied between observation intervals (Fig. 6.11), so that the level of the target in the correct interval is higher than that in the incorrect interval on almost half the trials. To perform well on this task, the listener must be able to compare the level of the target to the level of the other frequency components. In other words, the listener is forced to make *relative* comparisons of level across frequency.

Green's experiments showed that listeners are able to make comparisons of spectral shape, detecting bumps of only a few dB, even when absolute level is not a useful cue. That is not particularly surprising, since if we couldn't do that we wouldn't be able to detect formants in vowels. Of greater interest, perhaps, is the finding that performance was almost independent of the time delay between the two observation intervals (at least, up to 8 seconds: Green, Kidd, & Picardi, 1983). If level is randomized between *trials* (so that listeners cannot build up a long-term

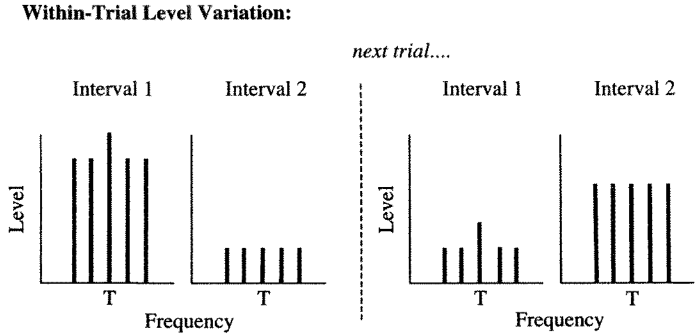

FIG. 6.11. A schematic illustration of the spectra of the stimuli in two trials of a typical profile analysis task. The target component is denoted by the "T." The listener's task is to pick the observation interval containing the incremented target (interval 1 in both trials here, or it could equally well be interval 2, determined at random). The overall level of the components has been randomized between each interval so that the listener is forced to compare the target level to the level of the other components across frequncy. If the listener just chose the interval contaning the most intense target, then he or she would pick interval 1 on the first trial (correct) but interval 2 on the second trial (incorrect). Based on a similar figure in Green et al. (1983).

memory representation of the pedestal), pure tone intensity discrimination performance drops off as the time interval between the two tones to be compared is increased. It appears that we have difficulty holding a detailed representation of absolute intensity in our memories over several seconds. That performance is much less affected by the time interval for comparisons between two spectral shapes suggests that relative intensity is stored in a more robust way, perhaps because we can simply categorize a particular spectrum as "bumped" or "flat," for example. That information is much easier to remember than a detailed representation of the target.

6.4.3 Comparisons Across Time

A static image can provide a great deal of visual information. Think about the amount of information contained on the front page of a newspaper for example. However, a static, or constant, sound provides comparatively little auditory information. I can say "eeee" for as long as I like but you wouldn't learn very much from it, except, perhaps, about the state of my mental health. When there is a bit of variability in the sounds I am producing, for example, "evacuate the building," you have been given much more information. We must be able to make intensity comparisons across time, as well as across frequency, so we can determine how the spectrum is changing. We see in Chapter 11 that this is very important for speech perception.

Most auditory intensity discrimination experiments measure comparisons across time, because listeners are required to compare the levels of two sounds presented one after the other. However, the time intervals between observation intervals (see Fig. 6.5) are often much larger than those that usually occur between the temporal features that help identify sounds. *Increment detection* experiments, on the other hand, measure listeners' ability to detect a brief intensity "bump" on an otherwise continuous pedestal. It has been shown that listeners can detect smaller increments in these cases than when discriminating between two tones, with different levels, separated by several hundred milliseconds (Viemeister & Bacon, 1988). It would appear, therefore, that comparisons of intensity over short time intervals are highly accurate. The temporal resolution experiments described in Chapter 8 also demonstrate that the auditory system is very sensitive to brief fluctuations in intensity.

6.5 SUMMARY

The ability to represent sound intensity, and to make comparisons of intensity across frequency and time, is crucial if the auditory system is to identify sounds. It seems likely that sound intensity is represented in terms of the firing rates of auditory nerve fibers. Because these fibers respond to the vibration of the basilar membrane, the loudness of sounds is dependent on the characteristics of the basilar membrane, in particular, *compression*.

1. We are most sensitive to sounds in the mid-frequency region (1000–6000 Hz). In this region, absolute threshold for normally hearing listeners is about 0 dB SPL. The *dynamic range* of hearing (the range of levels over which we can use the ear to obtain information about sounds) is about 120 dB in the mid-frequency region and decreases at low and high frequencies.
2. *Loudness* is the sensation most closely associated with sound intensity. Increases in intensity increase the loudness of a sound, as do increases in bandwidth and duration. Increases in bandwidth can increase loudness even when the overall level of the sound is kept constant.
3. Loudness scales show that subjective magnitude is a power function (exponent 0.2–0.3) of intensity over most levels. The shape of the loudness function suggests that, at a given frequency, loudness is roughly proportional to the square of basilar membrane velocity. The auditory system may effectively sum this quantity across characteristic frequency to produce the loudness of a complex sound.
4. There seems to be enough information in the auditory nerve to represent the intensity of sounds over the entire dynamic range of hearing in terms of *firing rate* alone, despite the fact that most auditory nerve fibers are saturated at 60 dB SPL. The wide dynamic range is dependent on the low spontaneous rate fibers, whose shallow rate-level functions are a consequence of basilar-membrane compression. In other words, basilar-membrane compression is the basis for the wide dynamic range of human hearing.
5. Intensity is also represented by the spread of excitation across characteristic frequency as level is increased, for a sound with a narrow bandwidth, and possibly by the increased *synchronization* of spikes (phase locking) to a sound as its level is increased in a background sound.
6. Intensity comparisons across frequency and across time are needed to identify sounds. It appears that these *relative* measures of intensity may be more robust than the absolute measures of intensity we experience as loudness.

6.6 READING

The following chapters are useful introductions:

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). London: Academic Press. Chapter 4.

Plack, C. J., & Carlyon, R. P. (1995). Loudness perception and intensity coding. In B. C. J. Moore (Ed.), *Hearing* (pp. 123–160). New York: Academic Press.

Green provides an excellent account of intensity discrimination, including profile analysis:

Green, D. M. (1993). Auditory intensity discrimination. In W. A. Yost, A. N. Popper & R. R. Fay (Eds.), *Human psychophysics* (pp. 13–55). New York: Springer-Verlag.