5

Frequency Selectivity

The term *frequency selectivity* is used to refer to the ability of the ear to separate out the different frequency components of a sound. This spectral analysis is absolutely crucial for mammalian hearing. Chapter 4 describes the basic mechanisms underlying this ability. This chapter goes into much more depth about the nature of frequency selectivity in the cochlea, including the role of the outer hair cells, and describes measurements of tuning at higher centers in the auditory system. It concludes with a discussion of behavioral measures in human listeners.

5.1 THE IMPORTANCE OF FREQUENCY SELECTIVITY

The first thing the visual system does is focus the light coming from each point in space onto a particular place on the retina, so a spatial arrangement of light sources and reflective surfaces in the world around us is mapped onto a spatial arrangement of photoreceptors. The visual system performs a *place-to-place* mapping of the visual world. One of the first things the auditory system does is separate out the different frequency components of the incoming sound on the basilar membrane. The basilar membrane performs a partial spectral (or Fourier) analysis of the sound, with each place on the basilar membrane being most sensitive to a different frequency component. In other words, the auditory system performs a *frequency-to-place* mapping of the acoustic world. The visual system has only three different patterns of sensitivity to the spectral information in light, and the patterns correspond to the spectral sensitivities of the three different cones in the retina. These cells behave like band-pass filters for light, but there are only three different center frequencies. Despite our vivid perception of color, we only get a very limited picture of the variety of wavelengths of light that are reflected (or produced) by objects. The auditory system has, arguably, several hundred different spectral sensitivities. The auditory system extracts quite detailed information about the spectral composition of sounds.

There is a good reason for this difference between the senses, of course. Visual objects are characterized mainly by their shapes, and so spatial visual information is very important to us. Auditory objects are characterized mainly by their spectra, and by the way their spectra change over time. For instance, different vowel sounds in speech can be identified by the positions of their spectral peaks (formants). The way in which the frequencies of the formants change over time helps identify preceding consonants. Similarly, different musical instruments can be identified by the spectral distribution of their harmonics. Indeed, the sound quality, or *timbre*, that characterizes most sounds we hear is largely dependent on spectral information.

As well as being important for sound identification, frequency selectivity enables us to *separate out* sounds that occur together. To offer a crude example, we can easily "hear out" a double bass in the presence of a piccolo. Most of the energy of the double bass is concentrated in low-frequency regions. Most of the energy of the piccolo is concentrated in high-frequency regions. When the two instruments are playing simultaneously, the sound waves are mixed together in the air to produce a sound wave that is a combination of the waves from the two sources. However, because they cover different frequency ranges, the basilar membrane can separate out the sounds originating from each instrument. As we see in Chapter 10, we can even separate two complex tones with harmonics distributed over the *same* frequency region, as long as the fundamental frequencies of the tones, and, hence, the frequencies of the individual harmonics, are different. Without frequency selectivity, we would find it very hard to separate simultaneous sounds.

In short, frequency selectivity can be considered just as important to hearing as spatial sensitivity is to vision. Frequency selectivity is *fundamental* to the way in which we perceive sounds, and that is why the topic is given an entire chapter in this book.

5.2 FREQUENCY SELECTIVITY ON THE BASILAR MEMBRANE

5.2.1 Recent Measurements

Von Békésy reported fairly broad tuning in the cochlea, what we now think of as the "passive" response of the basilar membrane (Section 4.2.3). The bandwidths of the filters, and, consequently, the spatial spread of the traveling wave, were much greater than they would be in a healthy ear at moderate sound levels (von Békésy used levels as high as 140 dB SPL!). It is now known that cochlear tuning is highly dependent on the physiological state of an animal. Even a slight deterioration in the condition of an animal can have large effects on tuning, and therefore on the ability of the ear to separate out different frequency components. Modern experiments are often conducted on anaesthetized chinchillas or guinea pigs. The cochlea is opened up, usually near the base, so that the basilar membrane can be observed. The motion of the membrane can be measured by bouncing laser light off a reflective surface (for example, a tiny glass bead) that has been placed on the membrane. This technique is usually used to measure the response of a single place on the basilar membrane, rather than the entire traveling wave.

The left panel of Fig. 5.1 contains a set of *iso-level* curves, each of which shows the velocity of a *single place* on the basilar membrane (with a characteristic frequency of 10 kHz) as a function of the frequency of a pure tone of a particular level. Basilar membrane velocity is expressed in dB relative to 1 μm per second: A velocity of one millionth of a meter per second would be represented by 0 dB on this scale. The plots use data from a widely cited paper by Ruggero, Rich, Recio, Narayan, and Robles (1997), and show measurements from a chinchilla. Think of these curves as representing filter shapes. The closer the frequency of the tone is to the best frequency of a place on the basilar membrane, the higher is the velocity of that place. Conversely, the more remote the frequency from the best frequency, the lower is the response. Each place on the basilar membrane behaves like a bandpass filter that attenuates frequency components remote from its best frequency. The curves in Fig. 5.1 show that the basilar membrane displays a high degree of tuning at low levels (good ability to separate out different frequency components). At high levels, the width of the iso-level curve (or auditory filter) at each place on the basilar membrane is broad, so each place will respond to a wide range of frequency components, and a single pure tone will stimulate a wide region of the basilar membrane (i.e., the traveling wave will cover a wide region of the basilar membrane).

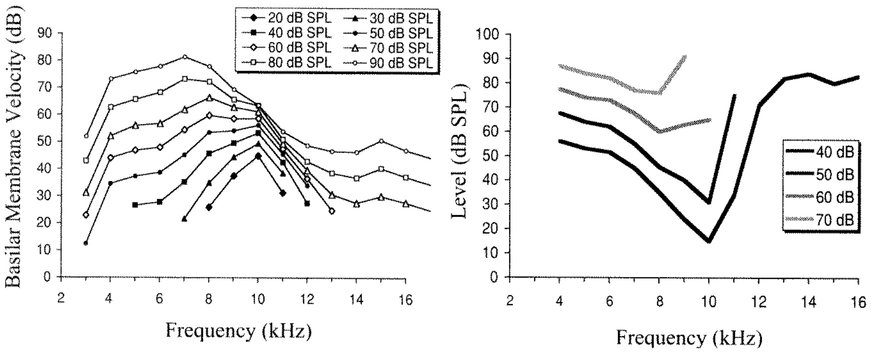

FIG. 5.1. Iso-level curves (left) and tuning curves (right) for a single place at the base of the basilar membrane of a chinchilla. The iso-level curves show basilar membrane velocity (in dB relative to 1 μm/s) as a function of the frequency of a pure tone, for various levels of the tone. The tuning curves show the level of a pure tone needed to produce a criterion velocity of the basilar membrane (shown in the legend in dB relative to 1 μm/s), as a function of frequency. The plots are based on data from Ruggero et al. (1997).

Note also in Fig. 5.1 that the best frequency of this place on the basilar membrane *decreases* as level is increased, from 10 kHz at low levels to 7 kHz at high levels. Because of this, the term characteristic frequency is usually used to refer to the best frequency in response to a *low-level* sound. A consequence of the reduction in the best frequency with increasing level is that the peak of the traveling wave moves toward the base of the cochlea as level is increased. This is called the *basalward shift* of the traveling wave. Think about it in this way. For a pure tone at a low level, the place that responds best to the tone has a characteristic frequency equal to the frequency of the tone (by definition). A more basal place with a slightly higher characteristic frequency does not respond as vigorously. At a higher level, however, the best frequency of the original place is now lower than the frequency of the tone, and the best frequency of a more basal place may have moved down so that it is now *equal* to the frequency of the tone. The result is that the vibration of the membrane will be stronger at a more basal place than it was at low levels.

The frequency selectivity of a place on the basilar membrane can also be measured by playing a tone at a particular frequency to the ear of an animal, and finding the sound level needed to produce a criterion velocity or displacement of the membrane. When the frequency of the tone is close to the best frequency of the place being measured, the level needed will be low. When the frequency of the tone is remote from the best frequency of the place, the level needed will be high. A plot of the level required against the frequency of the tone describes a *tuning curve* for that place on the basilar membrane.

The right panel of Fig. 5.1 shows tuning curves for a place on the basilar membrane with a characteristic frequency of 10 kHz. (These data are actually interpolated from the same set of data that are used to construct the iso-level curves on the left.) Note that at low levels, the tuning curves have very steep high-frequency sides and more shallow low-frequency sides. A tuning curve can be regarded as an inverted filter shape. If the tuning curve were flipped upside

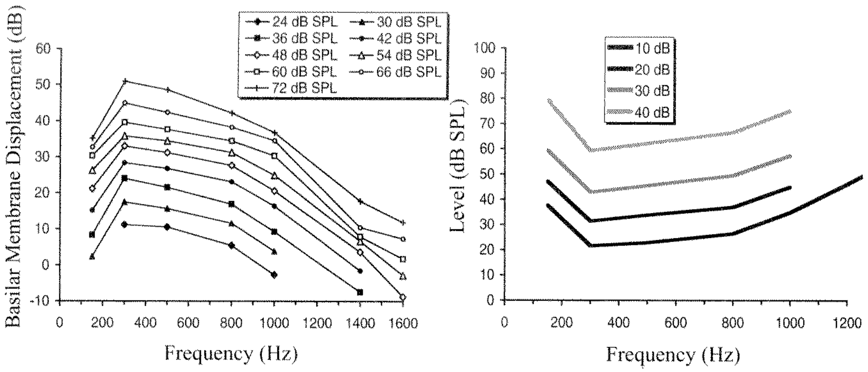

FIG. 5.2. Iso-level curves (left) and tuning curves (right) for a single place at the apex of the basilar membrane of a chinchilla. The iso-level curves show basilar displacement (in dB relative to 1 nm) as a function of the frequency of a pure tone, for various levels of the tone. The tuning curves show the level of a pure tone needed to produce a criterion displacement of the basilar membrane (shown in the legend in dB relative to 1 nm), as a function of frequency. The plots are based on data from Rhode and Cooper (1996).

down, the plot would represent the attenuation of different frequencies relative to the best frequency. If a high level is needed to produce the criterion response, then it is implied that the frequency component in question is being *attenuated* by the filter. Any pure tone level and frequency within the V of the tuning curve will produce *at least* the criterion velocity, although the higher the level, and the closer the frequency to the best frequency, the greater is the response.

Figure 5.2 shows iso-level curves and tuning curves measured at an apical site (from the tectorial membrane, rather than from the basilar membrane, but this shouldn't make much difference), with a characteristic frequency of 500 Hz. The absolute bandwidths of the filters (in Hz) are less at the apex than they are at the base. However, measured as a *proportion* of characteristic frequency, the bandwidths *decrease* from apex to base. The Q_{10}s of the filters (a measure of the sharpness of tuning, see Section 3.3.2) increase as the characteristic frequency increases, from about 2 at 500 Hz to about 5 at 10 kHz.

5.2.2 Ringing

In Section 3.3.3 it is shown how the response of a filter to an impulse depends on the *absolute* bandwidth of the filter. If the filter is very narrow, then the filter will ring for a long time after the impulse. If the filter is broadly tuned, the ringing will be brief. The same is true for the basilar membrane. If a click is played to the ear, places near the apex with low characteristic frequencies and narrow bandwidths may vibrate for several tens of milliseconds, whereas places near the base with high characteristic frequencies and wider bandwidths may only vibrate for a millisecond or so. Figure 5.3 shows a simulation of the pattern of vibration at three different places on the basilar membrane in response to a click.

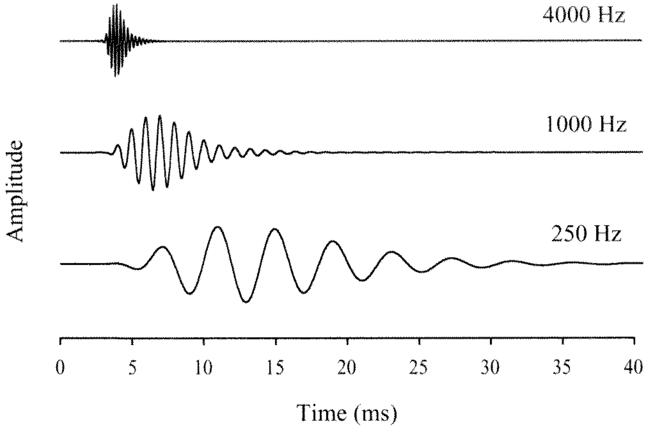

FIG. 5.3. A simulation of the pattern of vibration at three places on the basilar membrane in response to a brief impulse (an instantaneous rise and fall in pressure). These curves represent the *impulse responses* of three places on the membrane. The characteristic frequency of each place is indicated on the right. Note that the period of vibration of the basilar membrane is equal to the period of a pure tone at the characteristic frequency (i.e., one divided by the characteristic frequency).

5.2.3 Non-linearity

A quick observation of the iso-level curves and the tuning curves in Fig. 5.1 is enough to tell us one crucial thing about the response of the base of the basilar membrane: It is highly *non-linear* (see Section 3.3.4). In a linear system, the output amplitude should be a constant multiple of the input amplitude, irrespective of the level of the input. On a dB scale, this means that the output level should be a constant number of dB greater or smaller than the input level. If the system is a filter, then the filter attenuation characteristics should not change with level. The tuning curves in Fig. 5.1 tell us that this is not the case for a healthy cochlea. The filters become *broader* as the criterion (and, therefore, the overall level) is increased. Furthermore, the frequency to which the place on the basilar membrane is most sensitive (effectively, the center frequency of the filter) shifts *downward* as level is increased.

The non-linearity is made even more apparent when we look at the growth in the response of the basilar membrane as input level is increased. The data in

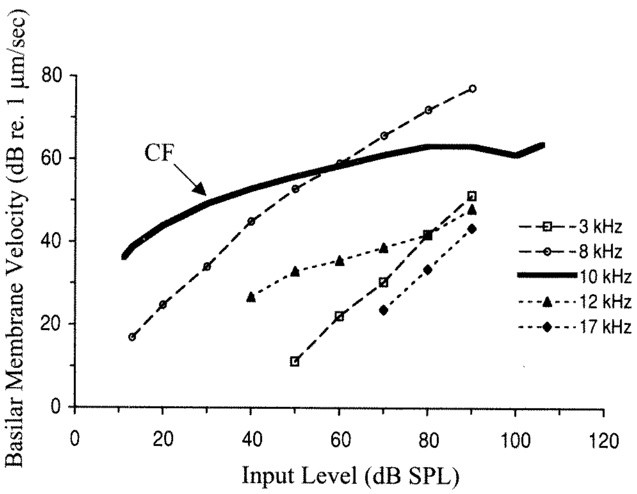

FIG. 5.4. The velocity of vibration as a function of input level for a single place on the basilar membrane of a chinchilla, in response to pure tones of various frequencies. The characteristic frequency of the place was 10 kHz. Note that the response to the tone at characteristic frequency is highly *compressive* at moderate to high input levels (linear growth has a slope of one on these coordinates, as exemplified by the response curves for 3-kHz and 17-kHz pure tones). The data are selected from a study by Ruggero et al. (1997).

Fig. 5.4 are a subset of those that were used to derive the iso-level curves and tuning curves in Fig. 5.1. The data show the basilar membrane response expressed in dB, as a function of the sound pressure level of a stimulating tone at different frequencies. As described in Section 3.3.4, a linear system should show a straight line, with a slope of one, on these coordinates. Note that for input frequencies lower than the characteristic frequency, the basilar-membrane response is roughly linear. However, the slope of the response function is very different around characteristic frequency. At low levels the response is almost linear, but at medium-to-high levels the slope is very shallow. This is indicative of a very *compressive* system: A 10-dB increase in input level may only produce a 2-dB increase in the output level. This compression is very important because it enables us to use acoustic information over a wide range of sound levels (see Chap. 6).

Like the frequency selectivity of the basilar membrane, the non-linear properties of the basilar membrane are dependent on the physiological condition of an animal. This is illustrated in Fig. 5.5, also from the paper by Ruggero et al. The curve on the right shows the characteristic frequency response of the basilar membrane after the animal had died. Note that the response now is nearly linear, and that the basilar membrane is much less sensitive than it was before death (i.e., a much higher input level is needed to produce the same output level).

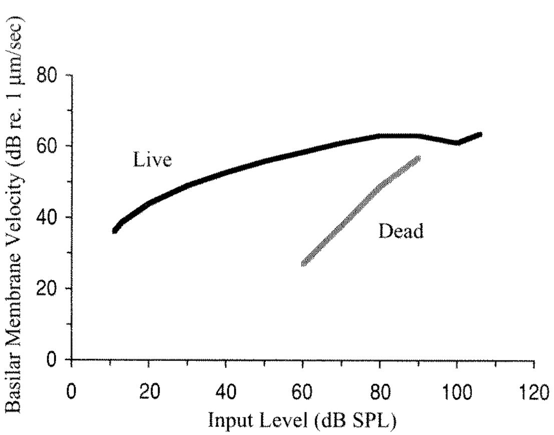

FIG. 5.5. The velocity of vibration as a function of input level for a single place on the basilar membrane of a chinchilla, in response to a pure tone at the characteristic frequency of the place (10 kHz). The curves show the response function before and after the death of the animal. In the latter case, the response is almost linear (slope equal to one). Data are from a study by Ruggero et al. (1997).

Most direct measurements of basilar membrane tuning have been made in the base of the cochlea, near the oval window, mainly because the surgical procedure is more difficult in the apex. Those measurements that have been taken near the apex suggest that the basilar membrane is much more linear here then it is near the base. As shown in Fig. 5.2, the auditory filters with low characteristic frequencies have a low Q and do not change their shape or best frequencies substantially as the input level is increased (Rhode & Cooper, 1996). While there may be some compression, the measurements suggest that it is less than at high characteristic frequencies, with a maximum value of about 2 : 1 (a 10-dB increase in input level produces a 5-dB increase in basilar membrane displacement). In addition, compression at a place in the apex does not just affect stimulus frequencies close to the characteristic frequency of the place (as in the base), but affects a wide range of input frequencies (which is why the filter shapes do not change with level).

5.2.4 Suppression and Distortion

There are two other consequences of cochlear non-linearity that are worth mentioning here. The first is *suppression*. Suppression refers to the reduction in the response to one frequency component when another frequency component is added. If I am playing to my ear a 40-dB SPL tone at 1000 Hz, for example, the response of the place on the basilar membrane tuned to that tone may actually *decrease* when I add a 1300-Hz tone at 60 dB SPL. This is clearly very non-linear behavior: In a linear system, adding an extra component that on its own is excitatory will *never* cause a decrease in the output of the system. We look at measurements of two-tone suppression in the auditory nerve in Section 5.3.3.

A second consequence of non-linearity is distortion. Recall that a non-linear system produces frequency components that were not present in the input (Section 3.3.5). The healthy ear is very non-linear and produces loads of distortion, particularly inter-modulation distortion when two or more components interact at a particular place on the basilar membrane. The components have to be fairly close together so that they both fall within the range of frequencies that are compressed by a single place on the basilar membrane. When they do, however, distortion products called *combination tones* are produced. These distortion products may include the difference tone, and other inter-modulation products with frequencies *lower* than the frequencies of the original components. Combination tones propagate from the place of generation to excite the places on the basilar membrane tuned to the frequencies of the combination tones. They can be clearly audible in some situations.

Suppression and distortion are characteristic of ears in good condition, and are absent when the cochlea is severely damaged. Somewhat ironically, a healthy ear distorts much more than an unhealthy one.

5.2.5 The "Active" Mechanism

We have seen that, for an ear in poor condition, the response of the basilar membrane is linear and the tuning curves are broad. The frequency selectivity of an unhealthy ear is similar to that of a healthy ear at high levels (in particular, the tuning curves are broad). In a *healthy* ear, the tuning curves at low-to-medium levels are sharp, and the response to a tone with a frequency near the characteristic frequency is almost linear at low levels, but highly compressive at higher levels. Furthermore, the tuning of a place near the base of the basilar membrane has a higher best frequency (i.e., higher resonant frequency) at low levels than it does at high levels.

How do we make sense of all this? It seems that the response of the unhealthy ear reflects the passive response of the basilar membrane, as measured by Von Békésy. You can not get much more unhealthy than a cadaver. These broad tuning characteristics are the result of the basic mechanical properties of the cochlea, particularly the variation in stiffness along the length of the basilar membrane. In the healthy ear, however, something else seems to be contributing to the motion of the basilar membrane in the base. That "something else" seems to provide the equivalent of a *level- and frequency-dependent amplification* of the basilar-membrane response. Low-level sounds are amplified, but high-level sounds are not, and this amplification, or *gain*, only takes place for frequencies close to the characteristic frequency of each place on the basilar membrane. In the basal region of the cochlea, the characteristic frequency is *higher* than the best frequency of the passive response, so that, as the amplification goes away at high levels, the tip of the tuning curve shifts to lower frequencies.

I will try and explain this more clearly: Imagine that you have a broad filter and you want to make it sharper, so that it is better at separating frequency components close to the center frequency of the filter from frequency components remote from the center frequency of the filter. One of the ways to do this is to increase the attenuation of the remote components, so less energy from these components is passed by the filter. Another way, however, is to *amplify* frequency components close to the center of the filter, so the relative output from these components is greater. The second method is the one used by the cochlea. However, the frequencies that are amplified are higher (by perhaps half an octave) than the center frequency of the original broad filter (the passive basilar membrane), so as the amplification goes away at high levels, the center frequency of the auditory filter decreases as it returns to its passive state. Figure 5.6 shows how the amplification enhances frequency selectivity (and sensitivity) to low-level tones, and how the reduction in amplification at high levels leads to compression.

It is thought that the *outer hair cells* are involved in this amplification process. The dominant theory, at present, is that outer hairs respond to vibration of the basilar membrane by stretching and contracting at a rate equal to the stimulating frequency (see Møller, 2000, page 80). Like pushing on a swing at the right time,

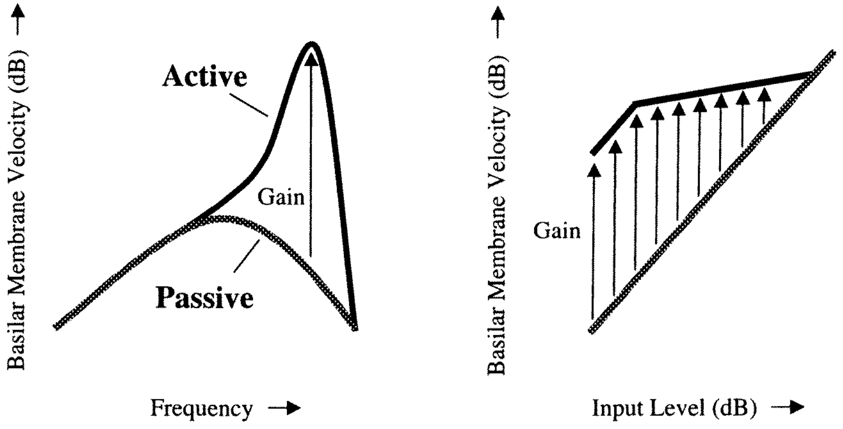

FIG. 5.6. An illustration of how activity by the outer hair cells may change the basilar membrane filtering characteristics (left) and the response to a tone at characteristic frequency (right). The shaded lines show the passive response of the basilar membrane (unhealthy ear, or healthy ear at high stimulation levels), the thick black lines show the healthy, active response of the basilar membrane. Arrows indicate amplification or gain. The left diagram shows how gain over a limited range of frequencies can sharpen the frequency selectivity of the membrane. The right diagram shows how gain at low levels increases sensitivity to a tone at characteristic frequency (a low-level tone now produces a larger response). Note that as level increases, the gain decreases. The result is a shallow response function (compression).

this motion may amplify the vibration of the basilar membrane in response to a tone near the characteristic frequency. The maximum gain may be about 50 dB, but it is not the same for all input levels (i.e., it is not linear). At higher levels, the outer hair cells can not respond sufficiently and the gain decreases. The result is a compressive response to a tone at characteristic frequency, and a broadening in the tuning curve at high levels. In an exciting recent development, the very molecule responsible for the length changes in outer hair cells has been identified (Zheng et al., 2000). The protein *prestin* is found in the membrane of outer hair cells, and it changes its shape in response to changes in the electric potential across the cell membrane. The electric potential changes are caused by the influx of positively charged potassium ions when the stereocilia are bent by the motion of the basilar membrane (see Section 4.3.1). Shape changes in the prestin molecules result in length changes in the outer hair cell, and hence—perhaps—a greater deflection of the basilar membrane. The motion of the basilar membrane may be amplified by a "positive feedback" loop in this way.

The outer hair cells are very sensitive to physiological trauma, which is why experimental animals need to be in good condition to show the healthy cochlear response. Even mild drugs, like aspirin, can temporarily impair outer hair cell function. Outer hair cells are also susceptible to damage by loud sounds, and when they are damaged, they do not grow back. As we get older, we tend to lose outer hair cell function. Dysfunction of the outer hair cells results in a loss of sensitivity *and* a reduction in frequency selectivity and is believed to be the main cause of hearing impairment in humans. The action of the outer hair cells is also influenced by activity in the efferent fibers from the brainstem described in Section 4.5.2. Activity in the efferents may reduce the gain of the outer hair cells.

5.2.6 Magical Sounds From Your Ear

Before we leave the discussion of basilar membrane physiology, a quick word about *otoacoustic emissions*. At the end of the 1970s, Kemp (1978) made the remarkable discovery that the ear can actually *emit* sounds. Virtually no one believed him at the time, but now it is a well-documented phenomenon. For instance, if an impulse or click is played to the ear, the ear may emit a sound containing certain frequency components. These otoacoustic emissions are also called "cochlear echoes," and they are generated by processes in the cochlea. If more than one pure tone component is present in the input, the emission may contain combination tone distortion products called *distortion product* otoacoustic emissions. The energy emitted may even exceed that of the original stimulus, which provides evidence that some sort of amplification is happening. Indeed, sometimes ears can produce pure tones without any input at all. These sounds are called *spontaneous* otoacoustic emissions, and may result from spontaneous activity of the outer hair cells at a particular place on the basilar membrane. Very occasionally, spontaneous emissions are intense enough to be heard by another individual. Strong otoacoustic

emissions are a characteristic of a healthy ear with functioning outer hair cells, and the emissions are now used to screen babies for hearing loss.

5.3 NEURAL FREQUENCY SELECTIVITY

5.3.1 Tuning in the Auditory Nerve

A physiologist can insert a microelectrode (a very thin electrode) into the auditory nerve of, say, an anaesthetized guinea pig or a chinchilla, and record the activity of a single auditory nerve fiber. Each fiber shows tuning properties very similar to those of the place on the basilar membrane to which it is attached. In other words, the fiber will respond with a high rate of firing to a pure tone at its characteristic frequency, and at a lower rate of firing to a pure tone with a frequency higher or lower than its characteristic frequency.

Frequency threshold tuning curves can be obtained by finding the level of a pure tone required to produce a just-measurable increase in the firing rate of a neuron, as a function of the frequency of the pure tone. These curves are equivalent to the tuning curves on the basilar membrane described earlier: The closer the frequency of the tone to the characteristic frequency of the neuron, the lower is the level required. Figure 5.7 shows tuning curves for five neurons from the auditory nerve of the chinchilla, representing a range of characteristic frequencies. The curves are plotted on a linear axis (left graph), and on a logarithmic axis in

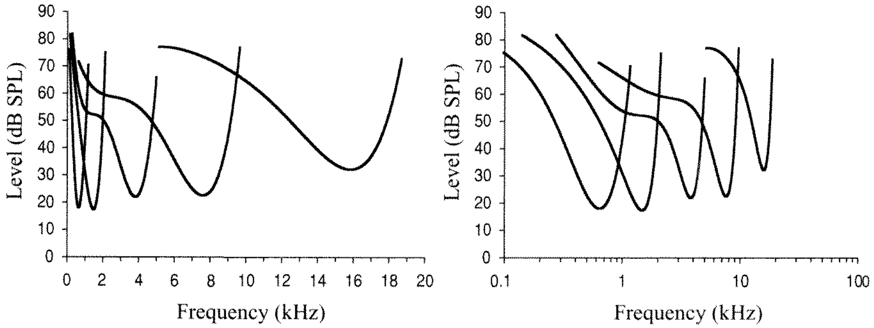

FIG. 5.7. Frequency threshold tuning curves recorded from the auditory nerve of a chinchilla. Each curve shows the level of a pure tone required to produce a just detectable increase in the firing rate of a neuron, as a function of the frequency of the tone. Five curves are shown, illustrating the tuning properties of five neurons with characteristic frequencies ranging from about 500 Hz to about 16 kHz. The curves are plotted on a linear frequency axis (left) and on a logarithmic frequency axis (right). The curves are smoothed representations of recordings made by Ruggero and Semple (see Ruggero, 1992).

which equal distances along the axis correspond to equal *frequency ratios* (right graph). The graphs illustrate that although the *absolute* bandwidths of the filters increase with characteristic frequency (see left graph), the bandwidths *relative* to the characteristic frequency *decrease* with characteristic frequency (see right graph).

The important point in this section is: Because each auditory nerve fiber innervates a single inner hair cell, the frequency selectivity in the auditory nerve is very similar to that on the basilar membrane. Because of the difficulties involved in measuring the vibration of the basilar membrane directly, much of what we know about frequency selectivity has been derived from auditory-nerve recordings.

5.3.2 The Effects of Cochlear Non-linearity on Rate-Level Functions

I mentioned in Section 4.4.2 that, in response to a pure tone at characteristic frequency, high spontaneous rate fibers have steeper rate-level functions, which saturate at much lower levels, than do low spontaneous rate fibers. I promised that I would explain this in Chapter 5. The important difference between the fiber groups is that the high spontaneous rate fibers are more sensitive than the low spontaneous rate fibers and, thus, they respond to the motion of the basilar membrane at levels for which the response function of the basilar membrane is nearly linear (i.e., the *steep* low-level portion of the function). In the low-level region, the vibration of the basilar membrane grows rapidly with input level, and, hence, the *firing rate* of the fiber grows rapidly with input level. Because of this, the firing rate at which the neuron saturates is reached at a low stimulus level.

The low spontaneous rate fibers are less sensitive than the high spontaneous rate fibers. This means that the range of levels to which a low spontaneous rate fiber is sensitive falls within the *compressive* region of the basilar membrane response function. For these fibers, a given change in input level will result in a much smaller change in the vibration of the basilar membrane, and, hence, a smaller change in the firing rate of the fiber. The result is a shallow rate-level function that saturates at a high input level.

If the shapes of the rate-level functions for low spontaneous neurons are dependent on cochlear non-linearity, then we should expect that the functions would be steeper for tones below characteristic frequency, because the basilar membrane response to a tone below characteristic frequency is roughly linear (see Fig. 5.4). This is, indeed, the case. Figure 5.8 shows recordings from a guinea pig auditory nerve fiber with a characteristic frequency of 20 kHz. The rate-level function for the pure tone at the characteristic frequency is quite shallow, because of the shallow slope of the basilar membrane response. The rate-level function for the tone below the characteristic frequency is much steeper, because the basilar membrane response to this tone, at the place tuned to 20 kHz, is much steeper. By assuming that the response to a tone below characteristic frequency is linear, Yates, Winter, and Robertson (1990) were able to derive the basilar membrane response to a tone at characteristic frequency from a comparison of the auditory nerve rate-level functions in response to tones at and below the characteristic frequency of the fiber. Without delving into the details of their procedure, the important general point is that *the rate-level functions of auditory nerve fibers reflect the response function of the basilar membrane*. This has to be the case, because the firing rate of an auditory nerve fiber is determined by the magnitude of vibration of the basilar membrane at the place in the cochlea where the dendrite of the nerve fiber synapses with an inner hair cell.

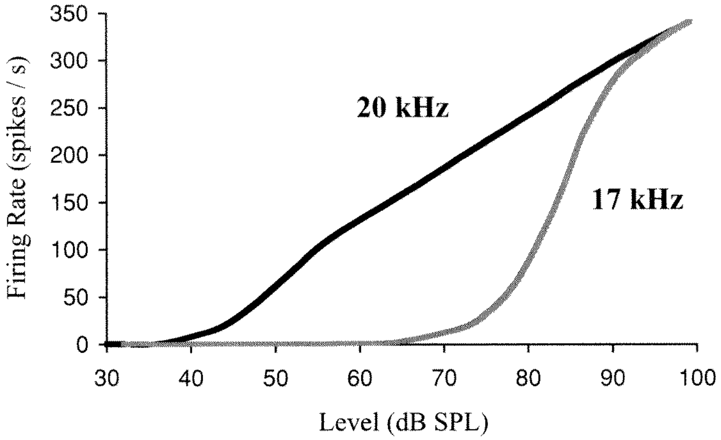

FIG. 5.8. Firing rate as a function of stimulus level for a low spontaneous rate auditory nerve fiber with a characteristic frequency of 20 kHz. The frequency of the pure-tone stimulus was either 20 kHz (at the characteristic frequency) or 17 kHz (below the characteristic frequency). The curves are based on guinea pig recordings reported by Yates et al. (1990).

5.3.3 Suppression in the Auditory Nerve

As I describe in Section 5.2.4, one of the consequences of the non-linearity in the cochlea is suppression. Suppression can be measured in the response of the auditory nerve. For instance, the firing rate of an auditory nerve fiber can be measured in response to a low-level pure tone at the characteristic frequency of the fiber. A second tone is then added. When the second tone falls well within the tuning curve of the fiber, the firing rate will increase. However, for certain levels and frequencies of the second tone outside the tuning curve, the firing rate in the fiber will *decrease* when the second tone is added. These suppression regions are shown in Fig. 5.9. What this figure does not show is that, for a given increase in the level of the second

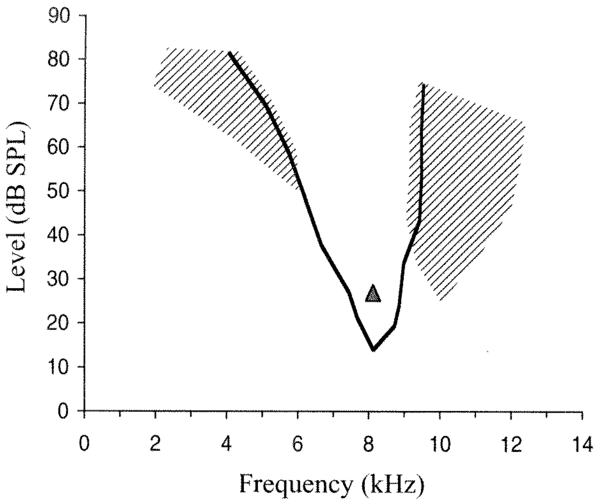

FIG. 5.9. Two-tone suppression in the auditory nerve. The continuous line shows the tuning curve of an auditory nerve fiber, with a characteristic frequency of 8000 Hz. The triangle indicates the level and frequency of a probe tone. When a second (suppressor) tone was added, with a level and frequency within the shaded regions, the firing rate of the fiber was *reduced* by at least 20%. Data are from Arthur, Pfeiffer, and Suga (1971).

tone, the reduction in firing rate is greater for a low-frequency suppressor than for a high-frequency suppressor (see Delgutte, 1990).

5.3.4 Tuning in the Central Auditory System

The tonotopic organization seen on the basilar membrane and in the auditory nerve, with different neurons tuned to different frequencies, continues up to higher auditory centers. In the cochlear nucleus, for instance, there are three separate regions, each with its own tonotopic map. Think of the information being passed up the auditory system in an array of parallel *frequency channels*. Neurons in the brainstem nuclei show frequency selectivity, but the tuning properties are not always the same as those seen in the auditory nerve (see Møller, 2000, Chap. 6). The diversity of tuning curve shapes seems to increase with distance up the ascending auditory pathways, and reflects the convergence of excitatory and inhibitory inputs from several neurons at a lower level in the pathway. Some neurons have very broad tuning properties, which perhaps reflects the convergence of excitatory inputs from several neurons with a range of different characteristic frequencies. Tuning curves with multiple peaks are also observed, which again reflects input from neurons with different characteristic frequencies. Some neurons may even exhibit *sharper* tuning than neurons in the auditory nerve. This can arise because excitatory input from one neuron is accompanied by inhibitory input from neurons with characteristic frequencies on either side, effectively adding greater attenuation to frequencies away from the characteristic frequency of the neuron.

Neurons in the auditory cortex also display a diversity of tuning shapes. Most have sharp tuning, but some have broad tuning curves or multi-peaked tuning curves, and many display patterns of inhibition (see Pickles, 1988, Chap. 7). These diverse response patterns reflect the ways in which different neural inputs are combined to process the auditory signal. It would be very surprising if the tuning properties of neurons did not change along the ascending pathways, as it would suggest that the auditory system is not interested in comparing the activity in different frequency channels. Because sound segregation and identification depend on across-frequency comparisons, complex tuning properties are expected, even if it is not obvious exactly what an individual neuron may be contributing to hearing as a whole.

5.4 PSYCHOPHYSICAL MEASUREMENTS

5.4.1 Masking and Psychophysical Tuning Curves

The basilar membrane is able to separate out the different frequency components of sounds, and this frequency selectivity is preserved throughout the auditory system. What are the consequences of this for our *perceptions*? One of the things it enables us to do is to "hear out" one frequency component in the presence of other frequency components. Imagine that I play to you a noise that has been band-pass filtered to only contain frequency components between 1000 and 1200 Hz. I now add a pure tone with a frequency of 2000 Hz and a level 20-dB below that of the noise. You will easily be able to hear the tone, because it is separated from the noise on the basilar membrane: The two sounds excite different *places* on the basilar membrane. If I now change the frequency of the tone to 1100 Hz, however, you will not be able to hear the tone. The tone is said to be *masked* by the noise, because the noise is effectively obscuring the tone. Masking occurs whenever the activity produced on the basilar membrane by one sound (the *masker*) obscures the activity produced by the sound you are trying to hear (the *signal*). If the masker and the signal are far apart in frequency, then the masker will have to be much more intense than the signal to mask it. If the masker and the signal are close together in frequency, then the masker may only have to be a few dB more intense than the signal to mask it.

Physiological techniques can be used to measure frequency selectivity in non-human mammals. *Psychophysical* techniques, such as masking experiments, can be used to measure the frequency selectivity of the *human* auditory system. In one technique, the pure-tone signal is fixed at a fairly low level, say 10-dB above the level at which it is just audible in quiet (the *absolute threshold*). This is sometimes called 10-dB *sensation level*. A narrowband noise or pure-tone masker is presented at one frequency, and its level increased until the listener can no longer detect the signal. The procedure is then repeated for a number of different masker frequencies. The results can be used to produce a *psychophysical tuning curve*, which is a plot of the level of a masker needed to mask a signal as a function of the frequency of the masker.

A psychophysical tuning curve describes the shape of a band-pass filter (auditory filter) that has a center frequency equal to the frequency of the pure-tone signal. The technique of measuring psychophysical tuning curves assumes that we have the ability to "listen" selectively to the output of a single auditory filter, that is, to a single place on the basilar membrane. When the masker is remote in frequency from the center of the filter, it receives a lot of attenuation. To overcome this, the masker level has to be high to mask the signal. When the masker is close to the center of the filter, it receives less attenuation and the level required to mask the signal is lower. The psychophysical tuning curve is therefore directly comparable to the basilar membrane tuning curve and neural tuning curve described earlier.

Figure 5.10 shows psychophysical tuning curves measured in my laboratory. In this experiment, the masker was presented before the signal in a design known as *forward masking*. We will discuss forward masking in more detail in Section 8.1.2. For the moment, all we need to know is that as the gap between the masker and the signal was increased, the masker level needed to make the signal undetectable also increased. Because the signal was fixed at a low level, a family of psychophysical tuning curves, corresponding to different levels, was produced by simply varying the masker-to-signal gap. Note that as the level increases, the tuning curves broaden (less frequency selectivity) and the tip of the tuning curve (i.e., the best frequency) shifts downward in frequency, as do the equivalent basilar membrane tuning curves (Fig. 5.1).

5.4.2 The Notched-Noise Technique

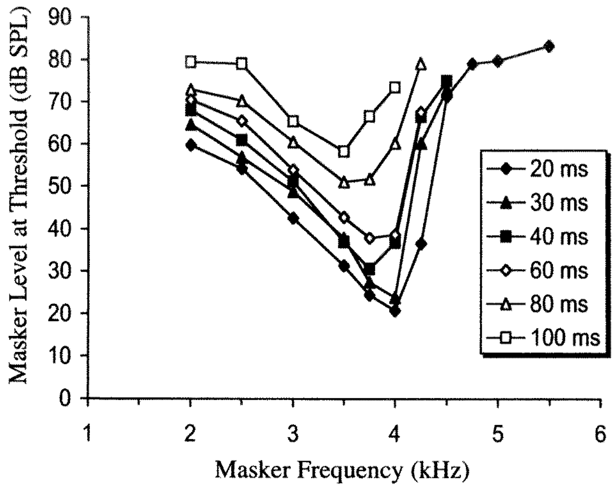

FIG. 5.10. Psychophysical tuning curves at 4000 Hz. The curves show the level of a pure-tone masker needed to mask a 4000-Hz pure-tone signal, as a function of masker frequency. The signal was fixed at 10-dB above absolute threshold, and was presented after the masker (forward masking). The masker-signal interval was varied to produce a set of psychophysical tuning curves covering a range of masker levels. As level increases (filled diamonds to open squares) the tuning curve becomes broader (less frequency selectivity, larger ERB) and the tip of the tuning curve shifts down in frequency. This implies that the place on the basilar membrane tuned to 4000 Hz at low levels is tuned to *lower* frequencies at high levels. Data are from Yasin and Plack (2003).

The tuning-curve experiment described in the previous section is only one of many different psychophysical procedures that have been used to measure the frequency selectivity of the human auditory system (see Moore, 1995 for a review). Thousands of masking experiments have been tried, tones on tones, noise on tones etcetera. Although some techniques may be more reliable than others, the overall conclusions regarding frequency selectivity are consistent with what I report here. However, a popular way of estimating the shape of the auditory filter, worth mentioning in this section, is the notched-noise technique developed by Patterson (1976). In his technique, a pure-tone signal is presented with two bands of noise, above and below the signal frequency, that act as maskers. The signal is in a *spectral notch* between the noise bands (Fig. 5.11). The signal is usually presented simultaneously with the noise bands, but the signal can be presented after the noise in a forward masking design. The lowest detectable level of the signal (the signal *threshold*) is determined as a function of the spectral gap between the signal and the edge of each of the noise bands. If the noise bands are close to the signal, then a large amount of noise energy will be passed by the filter centered on the signal, and the signal threshold will be high. As the width of the spectral notch is increased, threshold decreases. By measuring the way in which signal threshold changes as the spectral notch is changed, it is possible to estimate the shape of the auditory filter.

One of the advantages of presenting the signal between two noise bands is that this limits the effectiveness of *off-frequency listening*. Off-frequency listening describes a situation in which the listener detects the signal by using an auditory filter tuned lower or higher than the frequency of the signal. For example, if only the lower-frequency band of noise were used, the listener might detect the signal through an auditory filter tuned slightly higher than the signal frequency (equivalent to listening to vibration of the basilar membrane at a more basal location). Because

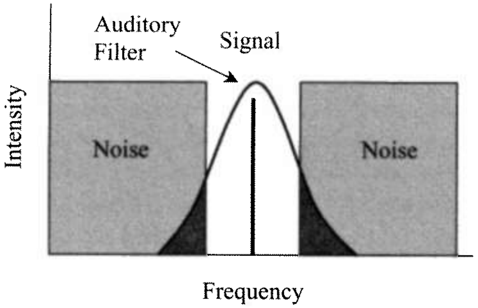

FIG. 5.11. A schematic illustration of the spectrum of the stimulus used by Patterson (1976). The area of the dark shading is proportional to the noise energy passed by the auditory filter centered on the signal frequency. As the width of the spectral notch is increased, the noise passed decreases, and the signal becomes easier to detect (hence, threshold decreases). Based on Moore (2003, Figure 3.3).

the auditory filter has a slightly flat tip, the reduction in signal level at the output of the filter may be less than the reduction in masker level, and so the signal may be easier to detect. Off-frequency listening can lead to overestimates of the sharpness of tuning. Adding a noise above the signal frequency means that such a listening strategy is much less beneficial because a shift in center frequency away from one noise band will be a shift toward the other noise band.

5.4.3 Variation With Center Frequency

As I describe in Section 3.3.2, the ERB of a filter is the bandwidth of a rectangular filter, with the same peak output and the same area (in units of intensity) as that filter. Based on notched-noise masking experiments with human listeners, Glasberg and Moore (1990) estimated that the ERB for the auditory filter (in Hz) follows the relation:

$$\mathrm{ERB} = 24.7(0.00437 f_c + 1) \tag{5.1}$$

where f_c is the center frequency of the filter in Hz. According to this equation, for frequencies above about 1000 Hz, the ERB is approximately proportional to the center frequency (constant Q), and has a value of about 11% of the center frequency at high frequencies (Q_{10} of about five). At high levels this equation is not valid—we know that the filters broaden considerably. In addition, recent evidence has suggested that the filters may be sharper than previously thought at high frequencies and at low levels. The ERB equation of Glasberg and Moore was based on measurements using *simultaneous* masking (i.e., the masker and the signal were presented at the same time). In simultaneous masking, part of the masking may be caused by the masker *suppressing* the signal, and this may broaden the apparent tuning curve. As shown in Fig. 5.9, the suppression tuning curve (the frequency regions for which one tone will suppress a tone at the characteristic frequency) is broader than the excitation tuning curve (the frequency regions that cause excitation at the same place). In short, when tuning is measured using forward masking and the masker cannot suppress the signal because the two are not physically present on the basilar membrane at the same time, you get sharper tuning curves at low levels, with ERBs that may be as little as 5% of the center frequency at 8000 Hz (Shera, Guinan, & Oxenham, 2002).

Of course, the million-dollar question is whether the tuning properties of the basilar membrane are similar to the tuning properties measured psychophysically. The answer is a qualified "yes." The similarity between the psychophysical tuning curves in Fig. 5.10 and the basilar-membrane tuning curves in Fig. 5.1 is obvious. The degree of frequency selectivity observed in the masking experiments is roughly the same as that observed by direct measurements of the basilar membrane response in other mammals. This has to be qualified somewhat, because the recent measurements of Shera et al. (2002) described previously suggest that humans may have better frequency selectivity than chinchillas or guinea pigs, and so a direct comparison between species may be problematic. In addition, it is possible that the human cochlea is more compressive at low characteristic frequencies, compared to some other mammals (Plack & Drga, 2003). Nevertheless, the correspondence between species is close enough to suggest that the frequency selectivity of the auditory system is determined by the tuning properties of the basilar membrane. This is a very important point, so it is worth repeating: The frequency selectivity of the *entire auditory system* is determined by the tuning properties of the *basilar membrane*. What happens on the basilar membrane is the limiting factor for our ability to separate sounds on the basis of frequency.

5.4.4 Excitation Patterns

As mentioned earlier, it is common to regard the cochlea as a bank of overlapping band-pass filters. Because we have a reasonable understanding of the characteristics of these filters, and how these characteristics change with frequency and level, we can produce an estimate of how an arbitrary sound is represented in the cochlea by plotting the output of each auditory filter as a function of its center frequency (equivalent to the characteristic frequency of each place on the basilar membrane). This plot is called an *excitation pattern*, and it is very useful for visualizing the effects of peripheral frequency selectivity on the representation of sounds in the auditory system. The excitation pattern is the auditory system's version of the *spectrum*: It shows the degree to which the different frequency components in complex sounds are resolved by the cochlea.

The top panel of Fig. 5.12 shows excitation patterns for a 1000-Hz pure tone as a function of the level of the tone. These plots were derived using a sophisticated

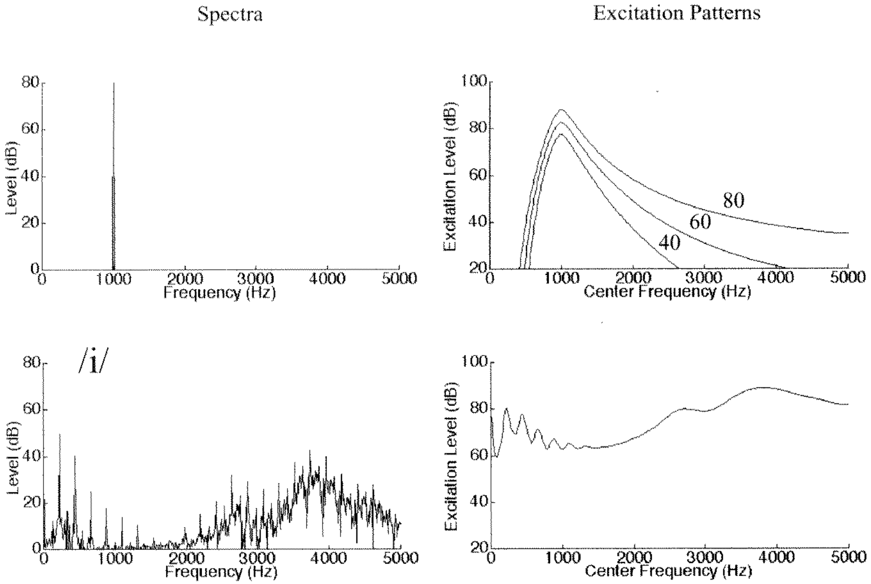

FIG. 5.12. Spectra (left) and excitation patterns (right) for a 1000-Hz pure tone (top) and for the vowel /i/ (bottom). Excitation patterns for the pure tone were calculated at three levels, 40, 60, and 80 dB SPL. Only the 80-dB tone is shown in the spectrum.

model of the basilar membrane that takes into account the non-linear characteristics described in Section 5.2.3. Let us try and understand the basic shape of the excitation pattern of a pure tone. If the center frequency of the auditory filter matches the frequency of the tone, then the auditory filter output has a high value, and there is a peak in the excitation pattern at that frequency. For center frequencies higher or lower than the frequency of the tone, the output of the auditory filter is less (the tone is attenuated), and, hence, the excitation level is not as high. Because the filters are broader at high frequencies, and have steeper high-frequency slopes than low-frequency slopes, a filter that has a center frequency *below* the frequency of the tone will let less energy through than a filter with a center frequency the same deviation *above* the frequency of the tone. Therefore, on a linear center frequency axis such as this, the excitation pattern has a shallower high-frequency slope than low-frequency slope.

Note also that the peak of the excitation pattern does not increase by as much as the level of the input tone. This is because, for a characteristic frequency equal to the frequency of the tone, the tone is *compressed* by the basilar membrane. For a characteristic frequency above the frequency of the tone, however, we know that the response of the basilar membrane is roughly linear (see Fig. 5.4). At center frequencies higher than the frequency of the tone, therefore, the excitation pattern

grows *linearly* (a 20-dB increase in the level of the pure tone produces a 20-dB increase in excitation level). The overall effect is that the *high-frequency* side of the excitation pattern becomes broader as level is increased, as we would expect from the fact that the individual auditory filters are getting broader on the *low-frequency* side (see Fig. 5.1 and Fig. 5.10).

The characteristics of the excitation pattern are reflected in the effects of a masker with a limited bandwidth, for instance, a pure tone or a narrowband noise (see Fig. 5.13). For a signal frequency equal to the masker frequency, the growth of signal threshold with masker level is roughly linear, so that a 10-dB increase in masker level produces about a 10-dB increase in the lowest detectable level of the signal (see also Section 6.3.2). However, more rapid growth of excitation on the high-frequency side of the excitation pattern contributes to a phenomenon called the *upward spread of masking* (the other contributor, at low signal levels, is suppression of the signal by the masker). Simply put, maskers lower in frequency than the signal become *relatively* more effective as level is increased. For an 80-dB SPL masker at 2000 Hz, the lowest detectable signal level at 4000 Hz might be about 30 dB SPL. For a 90-dB SPL masker at 2000 Hz, the lowest detectable signal level at 4000 Hz might be as much as 70 dB SPL. The effect can be explained as follows. When someone is listening to the place on the basilar membrane tuned to

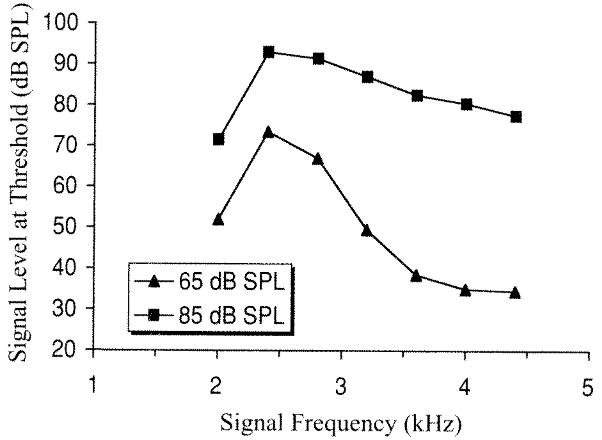

FIG. 5.13. An illustration of the upward spread of masking. The curves show the threshold level of a pure-tone signal in the presence of a 2.4-kHz pure-tone masker, as a function of signal frequency. Thresholds are shown for two levels of the masker (see legend). Threshold is highest (masking is greatest) when the signal frequency is equal to the masker frequency. As the masker level increases the signal threshold increases, although the increase is greatest on the high-frequency side of the masking pattern. Data are from an unpublished study by Oxenham, reported by Oxenham and Bacon (2004).

the signal, the signal is compressed. If the masker has the *same* frequency as the signal, then the masker is also compressed at that place, and the growth of excitation level with physical level for the two is the same (hence, a roughly linear growth in signal threshold with masker level). When the masker frequency is *below* the signal frequency, however, the excitation produced by the masker grows linearly, and, hence, more rapidly with level than the excitation produced by the signal. If the masker level is increased by 10 dB, then the signal level may have to be increased by 40 dB to produce the same 10 dB increase in excitation level.

The bottom panels of Fig. 5.12 shows the spectrum and excitation pattern for the vowel /i/ ("ee"). Notice that only the first few harmonics of the vowel form separate peaks or bumps in the excitation pattern. This is because the spacing between the harmonics is constant, but the auditory filters become broader (in terms of absolute bandwidth) as frequency is increased. At low center frequencies, the output of an auditory filter centered on a harmonic is dominated by that harmonic. An auditory filter centered *between* two harmonics has a lower output, because the harmonics are attenuated by the filter. The result is a succession of bumps in the excitation pattern. At high center frequencies, several harmonics fall within each auditory filter, and variations in center frequency have little effect on the excitation level. The higher formants appear in the excitation pattern as broad peaks, rather than as a succession of bumps. The auditory system can *separate out* the lower harmonics in a complex tone, but not the higher harmonics. This is of great significance for pitch perception (see Chap. 7).

As well as being a representation of the pattern of activity on the basilar membrane, the excitation pattern can also be considered a representation of the pattern of activity in the *auditory nerve*. Center frequency in that case would refer to the characteristic frequency of an auditory nerve fiber. Neural activity should really be measured in terms of firing rate, rather than as a level in dB, and it is common to plot excitation patterns in terms of neural firing rate as a function of characteristic frequency. However, the firing rate at each neural characteristic frequency is related to the basilar membrane vibration at that characteristic frequency. The *information* in the auditory nerve is broadly equivalent to that described by the excitation pattern expressed as excitation level.

We come across more excitation patterns like this over the course of this book, so it is worth taking some time to be certain you understand how they are produced and what they signify.

5.5 SUMMARY

Frequency selectivity is one of the most important topics in hearing, because the nature of auditory perception is largely determined by the ear's ability to separate out the different frequency components of sounds. Frequency selectivity can

be measured at all stages of the auditory system, from the basilar membrane to the auditory cortex, as well as in our perceptions. Arguably, more is known about frequency selectivity than about any other aspect of auditory processing.

1. The tuning properties of the basilar membrane can be measured directly in non-human mammals. At the base of the cochlea (high characteristic frequencies), a single place on the basilar membrane shows a high degree of tuning at low levels (narrow bandwidth), but broader tuning at high levels (particularly on the low-frequency side). In addition, the best frequency of each place on the basilar membrane shifts downward by about half an octave from low levels to high levels.
2. In the apex of the cochlea (low characteristic frequencies) the tuning curves are narrower than in the base, when measured in terms of absolute bandwidth in Hz, but broader as a proportion of characteristic frequency (i.e., the Q_{10}s are smaller in the apex).
3. Frequency selectivity in the base is enhanced by an *active mechanism*, dependent on the motion of the outer hair cells, which effectively amplifies the response to low- and medium-level frequency components close to the characteristic frequency. The active mechanism sharpens the tuning at low-to-medium levels and, because the gain is greatest at low levels and absent at high levels, leads to a shallow growth of basilar membrane velocity with input level (compression).
4. Two side effects of the non-linearity are *suppression* and *distortion*. Suppression refers to the situation in which one tone *reduces* the response to another tone at its best place on the basilar membrane. Distortion is observed when two or more frequency components interact at a place on the basilar membrane, creating lower-frequency inter-modulation products called *combination tones*.
5. The tuning properties of the basilar membrane are reflected in the tuning properties in the auditory nerve and at higher centers in the auditory system. At higher centers, neurons with different characteristic frequencies can converge, and produce neurons with complex tuning curves.
6. Frequency selectivity in *humans* can be measured using masking experiments: Our ability to detect a signal in the presence of a masker depends on how close the two sounds are in frequency. The tuning properties are consistent with those seen on the basilar membrane and in the auditory nerve, which suggests that the frequency selectivity of the entire auditory system is determined by the properties of the basilar membrane.
7. An *excitation pattern* is a plot of the outputs of the auditory filters as a function of center frequency, in response to a given sound. An excitation pattern is a representation of the overall activity of the basilar membrane as a function of characteristic frequency, or of the overall activity in the auditory nerve as a function of characteristic frequency. The excitation pattern is the auditory system's version of the spectrum.

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