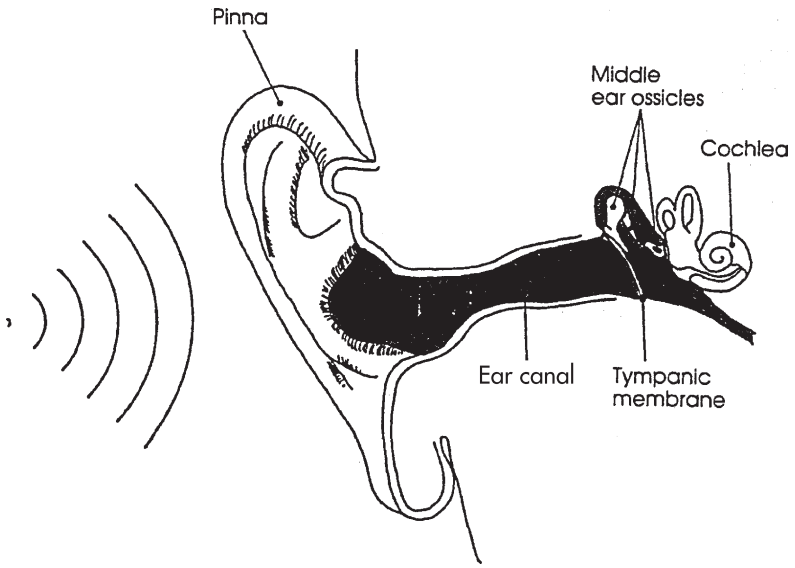
CHAPTER 12

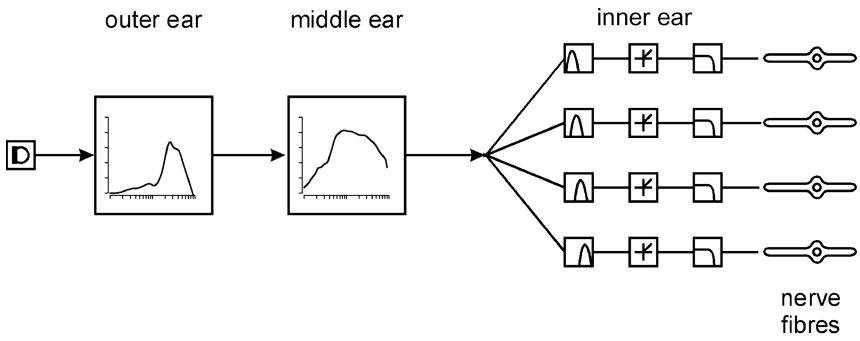
Applications to Hearing

So far we've been considering LTI systems in a fairly abstract way, although using specific examples from speech and hearing to illustrate many of the ideas. In this chapter we'll discuss in greater detail how the concepts we've developed can be applied to better understand the functions of the peripheral auditory system.

From Chapter 4, you are already familiar with the anatomy of the peripheral auditory system and its three major subdivisions—the outer, middle and inner ear:



What we will do here is to present an alternative way of thinking about this set of organs—not as 'wet' biological tissue, but as a collection of 'black box' systems. Each of these transforms the signals passing through it in a similar way to the transformations imposed by various structures in the auditory periphery:



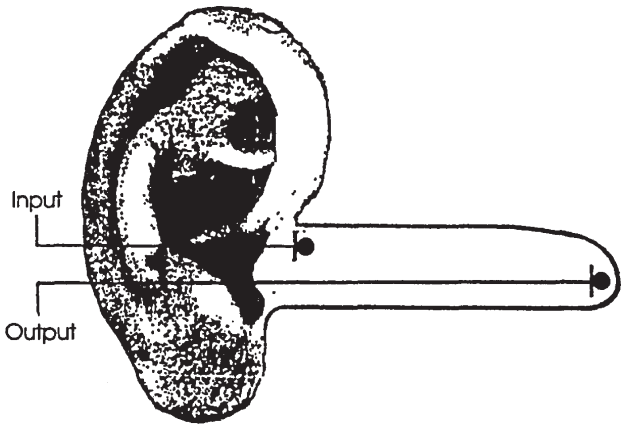
The original acoustic wave in the environment is first transduced by the microphone depicted at the left-hand side of the figure. The resulting electrical wave can then be processed by the rest of the systems in turn, resulting in the representation of the sound that is sent to the brain by the auditory nerve fibres at the far right.

Some of the systems in this chain can be well characterized as LTI, but others cannot. However, even for systems that are not LTI, the concepts of LTI signals and systems analysis can often still be usefully employed. We'll now look at each one of these systems in turn.

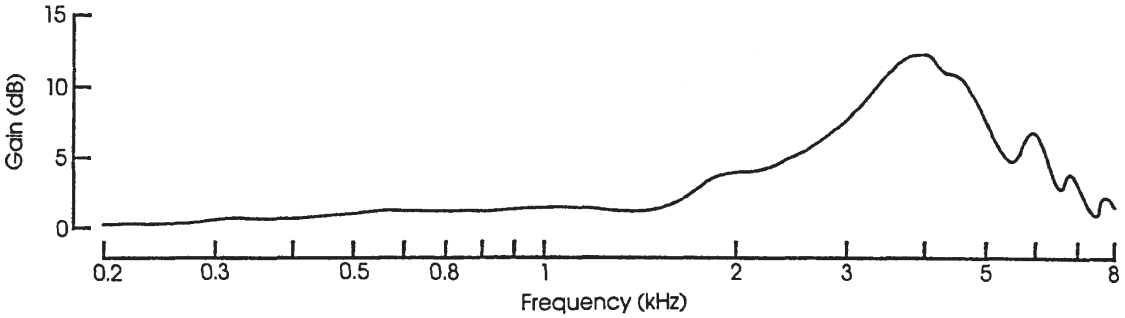
Outer ear

We'll begin with the outer ear, which includes the *pinna* and ear canal (also known as the *external auditory meatus*). The head itself also alters the sounds we hear, so we will need to include its effects. As you might expect, the effects of the pinna and head depend upon the direction the sound comes from. Therefore, it will be simpler for us to first describe the acoustic effects of the ear canal, as these do not depend upon sound direction.

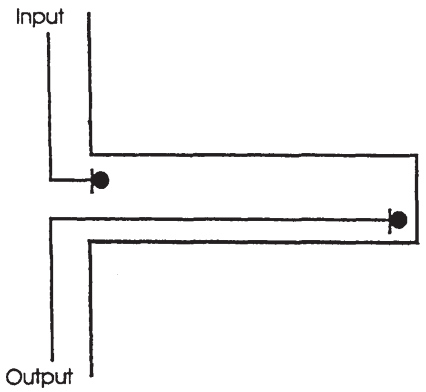
The ear canal is a sort of tube stretching from the surface of the head to the *tympanic membrane* (eardrum). Having the tympanic membrane at the bottom of a short canal, rather than at the surface of the skull, significantly changes the sound that acts upon it. More than 60 years ago, Wiener and Ross reported measurements of the amplitude response of the ear canal in several ears (average length 2.3 cm). They delivered sound to the open end of the ear canal and measured the output at the tympanic membrane with a microphone. The input to the system was defined as the sound pressure at the entrance of the ear canal, and its output as the sound pressure at the tympanic membrane as shown on the next page:



The mean amplitude response that Wiener and Ross obtained was characterized by a single resonant peak near 4 kHz. At low frequencies (below about 1500 Hz), there was little or no effect of the ear canal:

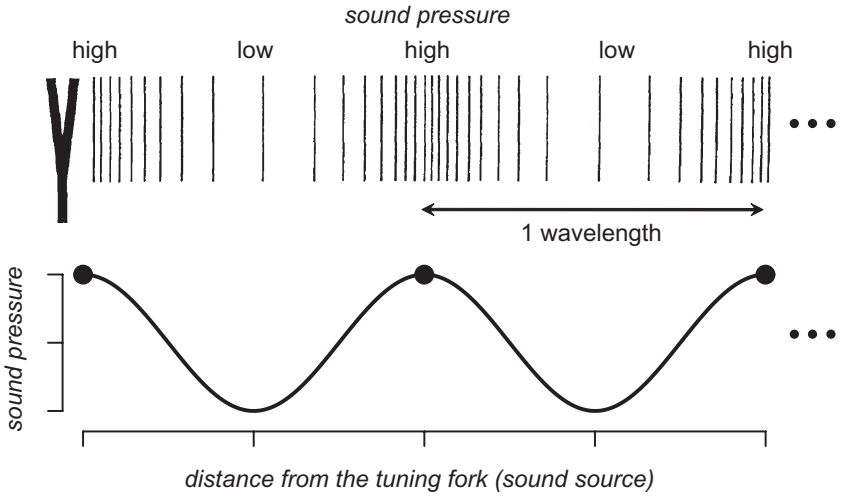


Some understanding of this response can be gained by imagining the ear canal to be a short cylindrical tube closed at one end, as shown below. The input signal is the sound at the opening of the model ear canal, the output is the sound at the closed end of the tube, with the system being the tube itself.



It turns out that the main factor that determines what frequencies are transmitted best in such a system is its *length*.

Before seeing how this is so, we must first introduce a term that you haven't encountered up until now—*wavelength*. The wavelength of a sinusoid is the *distance* the wave travels during one cycle of vibration. This is easiest to understand in a diagram that shows the pattern of acoustic pressure set up by a tuning fork, at some point in time after 'twanging' the fork:



As was explained for the figure on page 8, regions of high pressure are indicated by lines that are close together, while regions of low pressure are indicated by lines spaced more widely apart. The sinusoid at the bottom of the figure is a 'snapshot' of the instantaneous pressure across space at a particular moment of time, with positions of maximum pressure marked with solid circles. This is *not* a waveform, so the x-axis is not time. The stretch between the two places of peak pressure represents a particular distance—one *wavelength* (typically indicated by the Greek letter lambda λ). A wavelength is, thus, equal to the distance between two points in space that are in the same position in their sinusoidal cycle of pressure variations—that is, one period apart.

We want to use a measure of wavelength to relate the dimensions of a system like the ear canal (here its length) to the frequencies that it passes best. It would be useful, therefore, to have a formula that translates between wavelength and frequency. Clearly, the distance a wave travels in a particular time period must depend upon the speed at which it is moving through its medium. For air (the medium), the speed of sound (conventionally symbolized as c) has a value of about 340 m/s (about 770 miles/h).

We can now calculate wavelength (λ) as a function of the frequency (f) of the sinusoidal sound, by considering how far the wave travels in one period. In general, the distance anything

travels is simply the product of its speed and the time its journey takes. In a simple formula:

$$\text{distance} = \text{speed} \times \text{time} \tag{1}$$

We can readily rewrite the equation above in symbols as:

$$\lambda = c \times \text{time} \tag{2}$$

All we need now is the time the wave has to travel. Because we have defined wavelength as the distance travelled in one period, this is simply given by $1/f$. Therefore:

$$\lambda = c \times (1/f) = c/f \tag{3}$$

So, in order to calculate the wavelength of a sinusoid at 1 kHz, we simply substitute the appropriate values into the equation above:

$$\lambda = 340 \text{ m/s} \times (1/1000 \text{ Hz}) = 0.34 \text{ m} \tag{4}$$

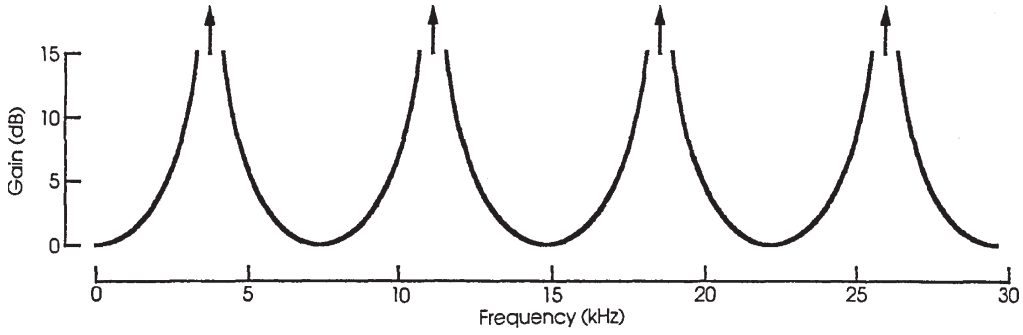
or a little more than one foot.

You should be able to see from these equations that lower frequencies (because they have longer periods) will have longer wavelengths, whereas higher frequencies (with shorter periods) will have shorter ones. However, the wavelength of a sound is determined not only by its frequency, but also by the speed of propagation of the wave, which in turn depends upon the medium in which the sound is presented. So, for example, if the role of the external ear in the auditory perception of scuba divers was being studied, we would need to know the speed of sound in water—about 1450 m/s. The wavelength of a 1-kHz sinusoid in water would then be:

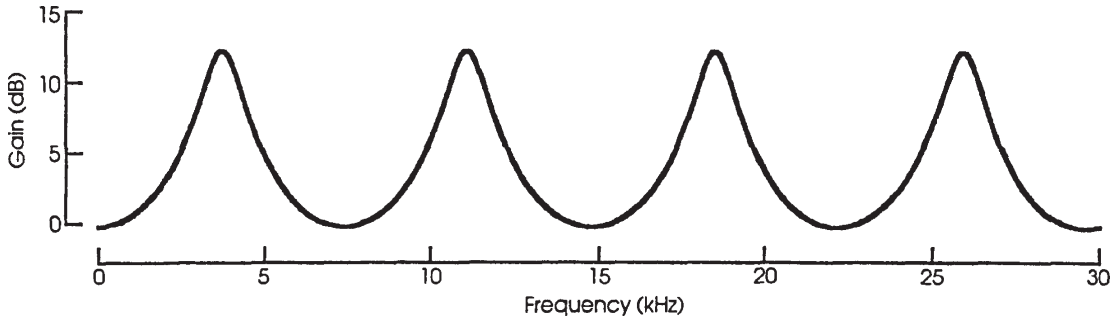
$$\lambda = 1450 \text{ m/s}/1000 \text{ Hz} = 1.45 \text{ m} \tag{5}$$

or nearly 5 feet.

We can now use the notion of wavelength to characterize features of the amplitude response of our model ear canal. If its walls were infinitely rigid, the response seen below would be obtained. You can see that the amplitude response consists of a series of valleys separated by resonance peaks. Because the system is idealized and has no losses



(damping), the amplitude response shoots up to an infinitely-high value. In a real system, such as the ear canal, frictional forces (primarily at the canal walls) damp the amplitude response. This broadens the resonances, and also prevents them getting infinitely high:



Thus, over the 8-kHz frequency range measured by Weiner and Ross, which includes only one resonance, the amplitude response has a shape like a band-pass filter.

It turns out that the position of these resonances is related in a simple way to the length of the ear canal. The lowest resonant frequency has a wavelength four times the length of the ear canal (called a *quarter-wavelength resonance*). Before working out what frequency this corresponds to, we need to re-arrange equation (3) from above to obtain:

$$f = c/\lambda \tag{6}$$

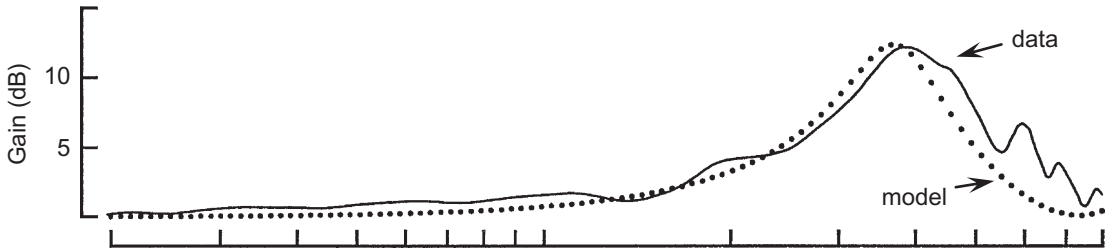
Now, if the length of the ear canal is L (expressed in metres), its first resonant frequency f_1 has a wavelength of $4L$. Thus, from equation (6), the lowest resonant frequency is:

$$f_1 = c/4L \tag{7}$$

This indicates that ear canals of different length have different lowest resonant frequencies. More specifically, the lowest resonance is inversely proportional to the length of the canal. Thus, longer canals have a lower first resonance, as you might expect from the general physical principle that the bigger something is, the lower its 'frequency'. (Consider the difference between the length of the strings on a violin and those on a double bass). Because frequency is equal to one over the period, this is equivalent to saying that the period of the lowest resonant frequency is directly proportional to the wavelength (again, longer tubes have lower first resonant frequencies). It turns out that the higher resonant frequencies are odd integer multiples (3, 5, 7 and so on) of the lowest resonant frequency.

Let's see how well our simple formula predicts the peak in the amplitude response of the ear canal. The first resonance for an

open tube that is 2.3 cm (0.023 m) long should occur at a frequency of $340/(4 \times 0.023) = 3696$ Hz. The next resonant frequency should occur at 3×3696 Hz $= 11{,}088$, the next at 5×3696 Hz $= 18{,}480$ Hz, and so on. As Wiener and Ross only performed measurements up to 8 kHz, we can only compare the model results with the actual results in the vicinity of the first resonance. Note first that the resonant frequency that Wiener and Ross found (near 4 kHz) corresponds well with the first resonant frequency we've just calculated. Moreover, with an appropriate choice for the damping characteristics, the model (dotted line below) can predict the shape of the measured curve (solid line) quite well. An even better fit of model to data could be obtained by assuming a slightly different length for the ear canal:

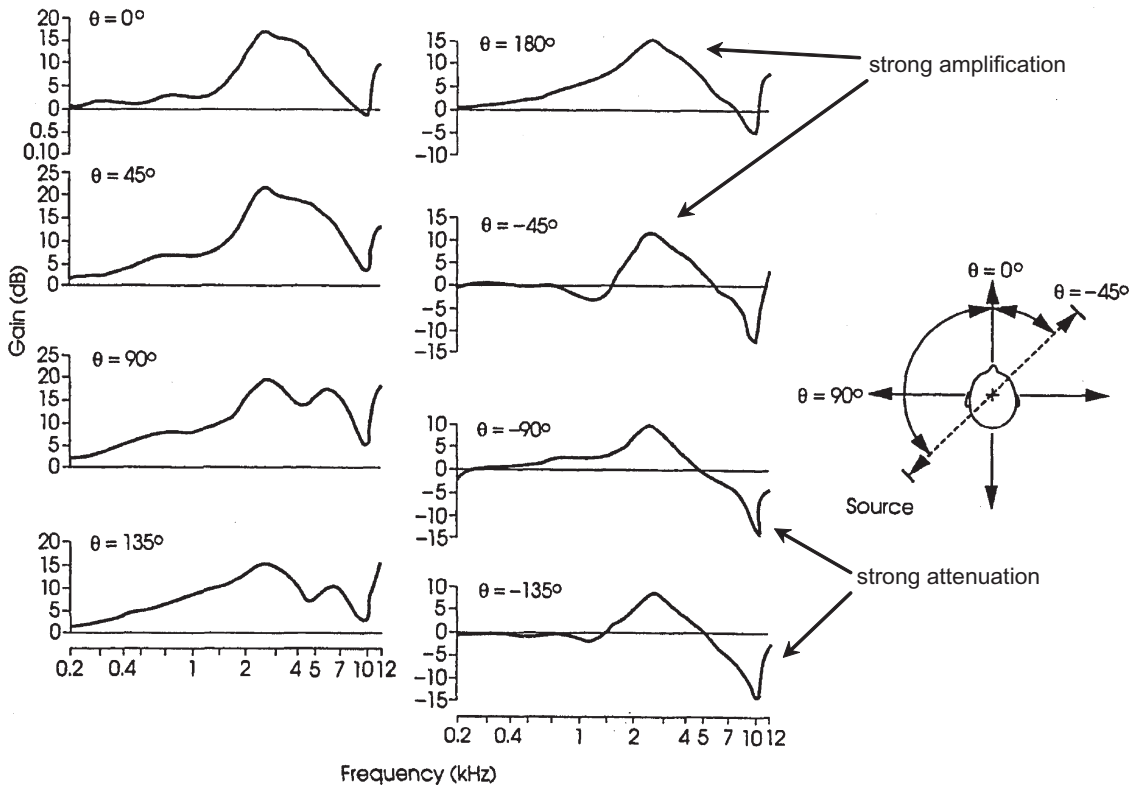


Of course, the acoustic effects of the external ear are not limited to those caused by the ear canal. The head and pinna cause acoustic 'shadows', so that the amplitude response depends upon the orientation of the sound source relative to the ear. In order to get a more complete view of the acoustic effects of the external ear and head, consider the system head-plus-pinna-plus-ear-canal. The input is a sinusoid delivered from different orientations relative to the head (its amplitude determined at the centre of where the listener's head would be) and the output will be taken as the sound pressure level at the eardrum. Defining the input sound pressure level as that arising in the sound field *without* the head present lumps together the effects of the head with those of the pinna and ear canal.

If heads and ears were symmetric, we would get the same result no matter which ear we measured. That is clearly not the case, so a full understanding of the sound field presented to a particular person at both ears would require measurements at *both* eardrums. Here we will only present measurements made at the *left* eardrum.

Shaw has summarized diagrammatically the amplitude responses obtained in several studies. The coordinate system used to denote the angle of the sound source relative to the head (known as the *azimuth*, θ) is represented schematically in the inset. In this set of measurements, the sound source is always presented at the same height, level with the opening of the ear canal (in the

so-called *horizontal plane*). So, for example, when $\theta = 0°$, the sound is straight ahead of the listener. When $\theta = 90°$, the sound is directly opposite the left ear and when $\theta = -90°$, the sound is directly opposite the *right* ear. Varying the elevation of the sound would add an extra complicating factor which, although important for a thorough understanding of the effects of the outer ear, will not concern us here:



As you can see, the amplitude responses vary greatly depending upon the position of the sound relative to the ear measured. However, these variations only occur when the input sinusoid has a wavelength that is comparable to, or smaller than, the head and pinna. This principle applies generally—the transmission of a sinusoid in a sound field is only affected by objects that are comparable in dimension to, or larger than, the wavelength of the sound. So, the acoustic effects of the head are most marked at frequencies above about 1.5 kHz, equivalent to wavelengths smaller than 22.7 cm (the approximate diameter of an adult male head). These changes in the amplitude response are most dramatic for negative values of θ, where the sound source is on the opposite side of the head from the ear being measured. For these angles, the amplitude response can show a strong attenuation of sound (for frequencies near 10 kHz or so).

Similarly, the measured amplitude responses have gains that are small for low frequencies (near 200 Hz and below), no matter what the value of θ. In other words, the sound field is not affected in this frequency region. This is not surprising, as a sinusoid of 200 Hz has a wavelength of about 1.7 m (about 5½ feet), dimensions not approached by any part of the ear or head, at least of a person!

Note too the resonant peaks at about 2.5 kHz that can be seen for all angles of presentation. These result from a quarter-wavelength resonance arising from the combined ear canal and *concha* (the shallow bowl in the pinna that leads directly to the ear canal). This resonance is at a lower frequency than that seen for the ear canal alone, because the concha effectively adds to the length of the ear canal. Because the rest of the pinna-plus-head response doesn't have sharply defined features between 2 and 4 kHz (peaks or valleys), the effect of the combined concha and ear canal always shows up in the output of the entire system.

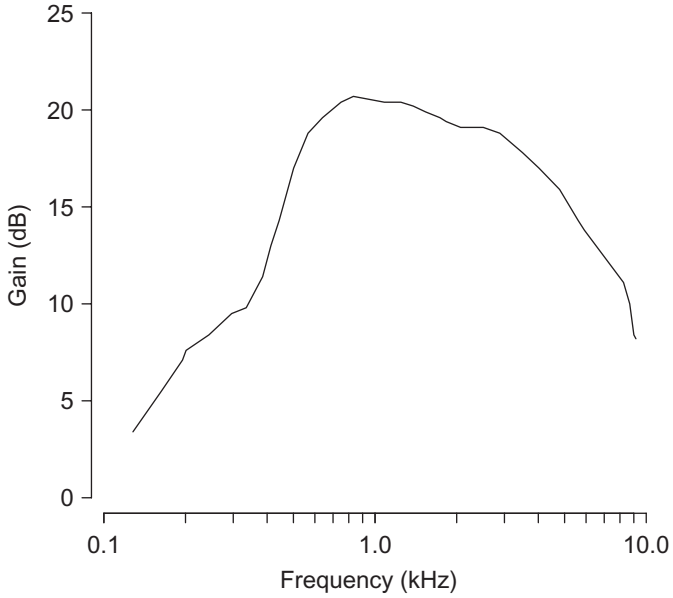
Of course, this is only the first stage in the chain that leads to perception of a sound. In a normal listening situation, the next system (the middle ear) would be presented with a signal that has already been affected by the ear canal, pinna and head in a way which depends on the position of the sound source and the frequency content of the sound presented. When we measure transmission by the middle ear, however, we don't normally use signals that have been modified by the outer ear. Rather, we apply a reference signal that is constant at the input of the middle ear (say, at the eardrum) so as to determine its transmission properties alone.

Middle ear

We've already examined the middle ear system in Chapters 4 and 6, describing investigations of the displacement of the stapes for sinusoidal sound pressure variations applied to the tympanic membrane in anaesthesized cats. You'll recall that this system (tympanic membrane-to-ossicles) was found to be LTI. Here, we'll be looking at the amplitude response of the middle ear in humans. These experiments, by Puria, Peake and Rosowski, were performed on temporal bones obtained from cadavers. Although many aspects of peripheral auditory function are very different in living and dead people, it turns out that important aspects of middle ear function are pretty much the same.

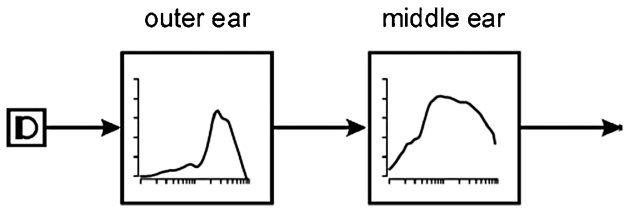
The amplitude response was determined on the basis of the same input as was defined for cats (sound pressure near the tympanic membrane) but with a different output. As you know (and which we will discuss more fully in the next section) the stapes moves in and out of the fluid-filled *cochlea* (inner ear), setting up pressure

variations which are an essential stage in hearing. Puria and his colleagues decided to use these pressure changes as the output of the middle ear, measured in the cochlear fluids near the stapes footplate using a *hydropressure transducer* (a kind of underwater microphone). Therefore, the 'gain' referred to in the *y*-axis of this figure reflects, on a dB scale, the pressure level in the cochlear fluids relative to the pressure level at the tympanic membrane:



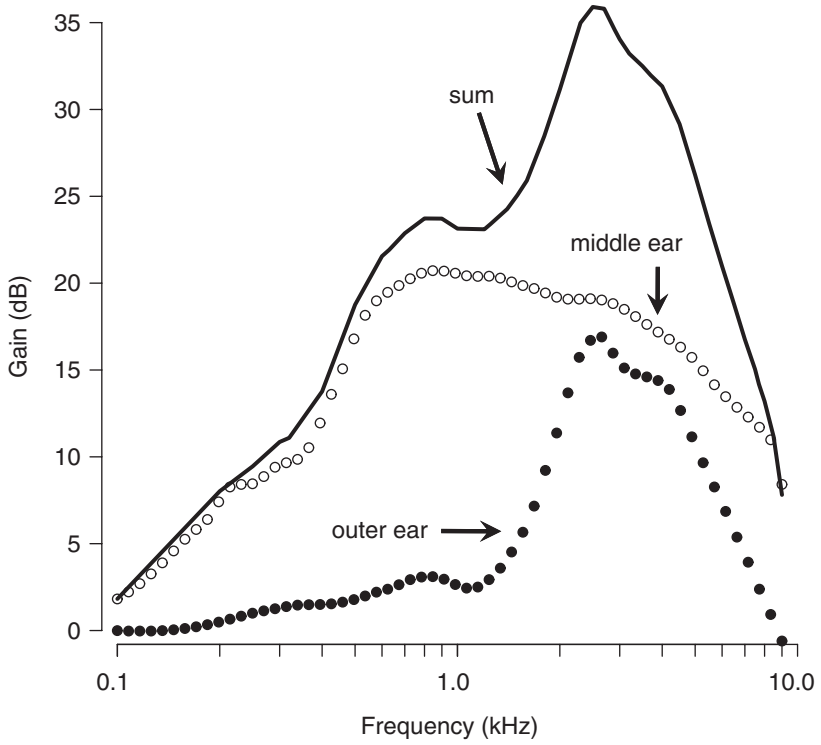
As you can see, the amplitude response is of the form of a band-pass filter centred near 1 kHz, and with quite a broad bandwidth. In other words, spectral components of sounds near 1 kHz are transmitted extremely well through the middle ear, with little or no relative attenuation of spectral components varying from about 500 Hz to 5 kHz.

Let's see how far we have come in our journey through our model of the auditory periphery.



The sound has been picked up by a microphone, transduced into an electrical wave and then filtered by two systems in cascade. You already know from Chapter 6 that it is readily possible to calculate the total amplitude response of these two systems from the individual responses. Since they are expressed

on dB scales, it is a simple matter to add together the gains at each frequency:



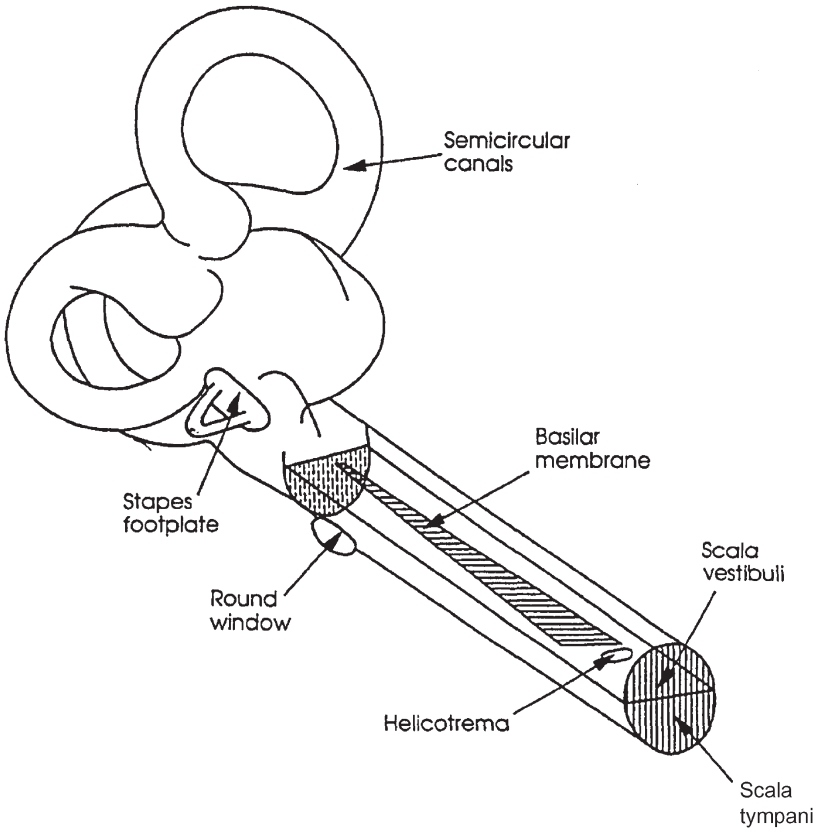
As you can see, the combination of the outer and middle ear leads again to a band-pass response, with its peak dominated by the response of the ear canal plus concha, near 3 kHz. This peak has a direct impact in determining the frequencies of sounds we are most sensitive to. Broadly speaking, it is perhaps not too surprising that sinusoids transmitted effectively through the auditory periphery can be detected at lower levels than those transmitted less effectively.

The movement of the basilar membrane

We now come to the inner ear, a crucial part of the auditory system. Not only is the input signal radically transformed in its structure here, it is also converted—or *transduced*—from a mechanical signal into an electrical one. This signal can then be handled by other parts of the nervous system. Before embarking on a detailed analysis, let's first describe its anatomy and general functioning.

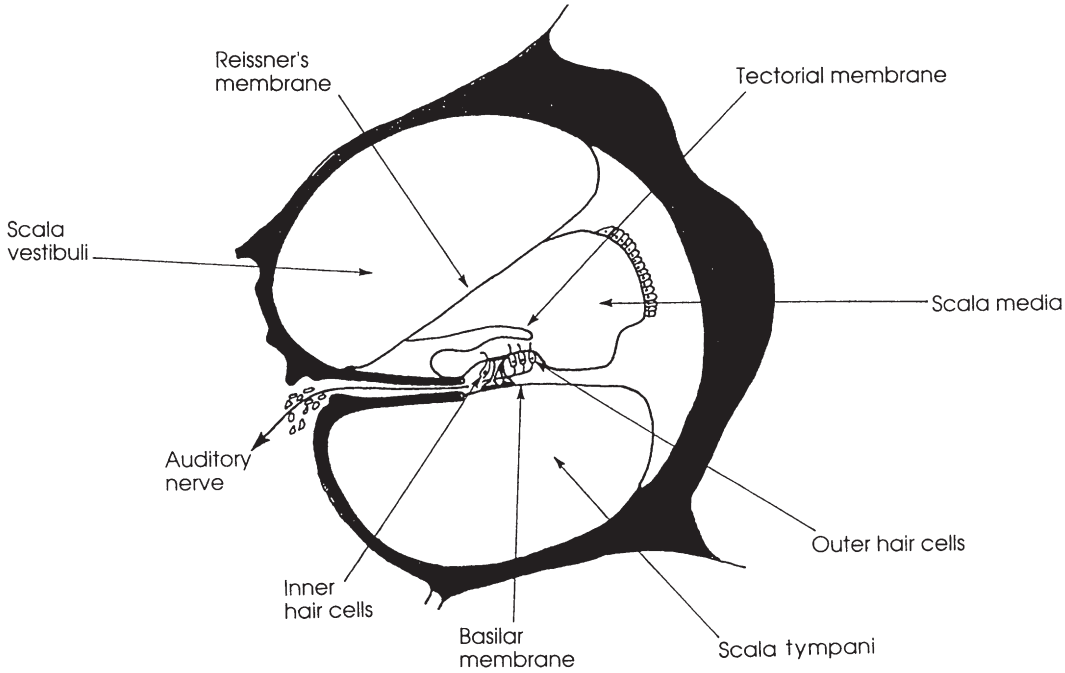
Although the inner ear also includes the organ of balance (*the semicircular canals*) the main structure that will concern us is

the *cochlea*. The cochlea is a fluid-filled tube coiled in the shape of a snail. This coiling is pretty much irrelevant to what the cochlea does, so we'll visualize it unrolled to give a clearer picture:



The *cochlear partition* runs down the length the cochlea, dividing it into two chambers (scala vestibuli and scala tympani). Because the partition does not quite reach the end of the cochlea, the two chambers connect through a space known as the *helicotrema*. When the stapes is pushed into the oval window by sound at the tympanic membrane, the incompressible fluid causes the membrane covering the round window to bulge outwards. Similarly, when the stapes is withdrawn from its resting position, the round window membrane moves inward. It, thus, acts as a sort of pressure release, or 'give', to allow the inward and outward movement of the stapes. Because the cochlea is surrounded by rigid bone, the stapes would not be able to move without this 'give'.

The cochlear partition is not simply a barrier but is itself a complex array of structures, as this cross-section shows:



As you can see, the cochlear partition is also a tube that, in a rolled up cochlea, spirals along its length. It is separated from the two scalae by membranes: *Reissner's membrane* seems only to serve as a dividing wall. Much more important functionally is the *basilar membrane*, upon which are found the *hair cells*. Hair cells are so named because they have cilia (which look like hairs) sticking out from their tops. These bundles of cilia come near or touch the *tectorial membrane* which lies across the top of all the hair cells.

There are two types of hair cell. The *inner hair cells* (IHC) form a single row running along the inner part of the cochlear spiral. At the base of the IHCs are the endings of the fibres of the auditory nerve which make synaptic contact with the hair cells. Note that this is the first time in the system that we've encountered any neural elements. Also important to cochlear function are the *outer hair cells*, which are found in three rows, along the outer part of the cochlear spiral.

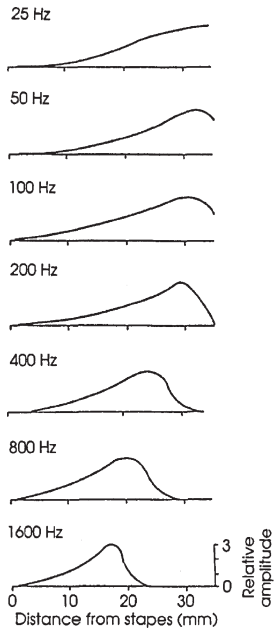
Although the precise nature of the chain of events that leads from sound to neural firing is still the subject of much controversy, there is general agreement about the major stages. Roughly speaking, this is what happens. The movement of the stapes in and out of the oval window sets up a pressure wave in the cochlea, which in turn causes the basilar membrane to vibrate. As a result of this vibration, the cilia on the IHCs are bent (perhaps due to a sliding motion between the basilar membrane and the tectorial membrane), causing neurotransmitters to be released from the base of the hair cells into the synapse.

The transmitter diffuses across the synaptic gap and causes the nerve to fire. The hair cells, thus, serve as transducers, transforming mechanical vibrations into electrical pulses. The neural 'spikes' are then relayed to other parts of the nervous system.

Because IHCs only cause nerves to fire in those places where the basilar membrane is set in motion (ignoring for the moment the spontaneous firing that goes on even in the absence of sound), the characteristics of this motion are crucial to an understanding of the firing patterns eventually seen on the auditory nerve. Again, there is much controversy, with agreement on certain major principles.

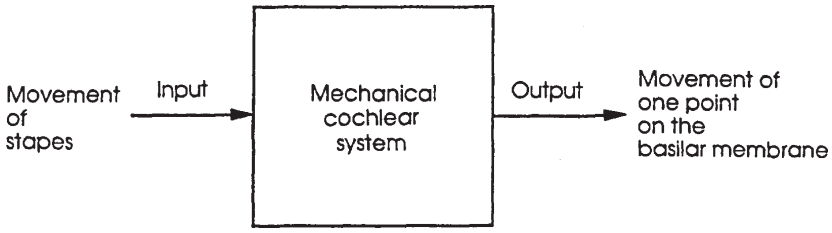
The most important characteristic of the basilar membrane is that it is *selectively resonant*. Not all parts of it vibrate equally well to sinusoidal inputs of a particular frequency. Put the other way round, different frequency sinusoids cause maximum vibration at different places along the membrane. Because the basilar membrane is narrower and stiffer at its basal end (near the stapes) than it is at its apical end (near the helicotrema), the basal end vibrates more to high frequencies than does the apical end. Conversely, a low-frequency movement of the stapes causes the apical end of the basilar membrane to vibrate more than the basal end.

This was first observed directly by von Békésy (pronounced BEH-kuh-shee) using a light microscope. He was able to measure the amplitude of the vibration of the basilar membrane over a significant portion of its length. At right are the results that von Békésy obtained when he presented a sinusoidal input of constant amplitude, at various frequencies, to the stapes of an excised cochlea. You can see that as the sinusoidal input increases in frequency, the peak amplitude of vibration occurs more basally. Note that these curves represent the *maximum* displacement undergone by any particular point on the membrane—the details of the temporal aspects of the vibration have been left out. In fact, every single point on the basilar membrane that moves would be moving in a sinusoidal way at the stimulating frequency.

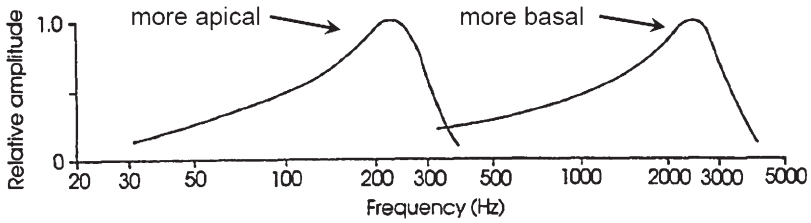


It is important not to mistake these graphs for the amplitude responses we've discussed so frequently. Measuring an amplitude response would necessitate the presentation of a number of sinusoids of different frequency. Because each of these curves is a measure of the motion resulting from a *single* input frequency, they cannot be amplitude responses. They are often known as *excitation patterns* because the response pattern of the entire basilar membrane to a single sound (or excitation) is shown. We can, however, combine together the information from these and a number of other similar curves to obtain the amplitude response of a single point on the basilar membrane. This now familiar way

of thinking about a system (here, the basilar membrane) may be schematized as shown here:

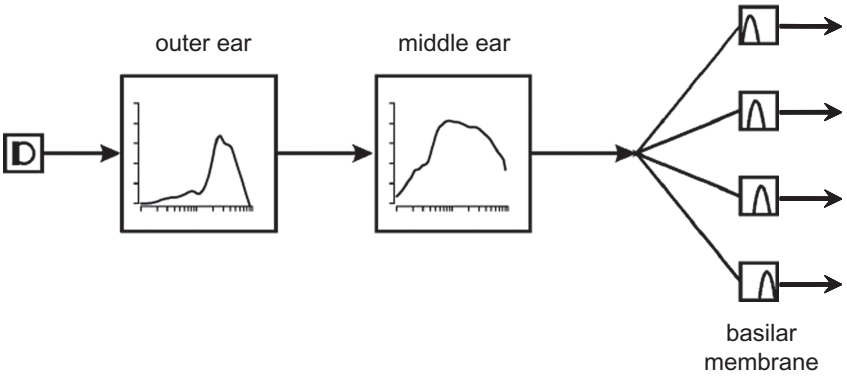


All that needs to be done is to move the stapes sinusoidally at a variety of frequencies and measure the amplitude of the response at a single point on the basilar membrane. This results in the familiar amplitude response. (We'll ignore phase in this discussion although such information can be important.) Of course, for a full understanding of the basilar membrane it would be necessary to measure the frequency response at a number of different places. Here are two such curves that von Békésy measured:



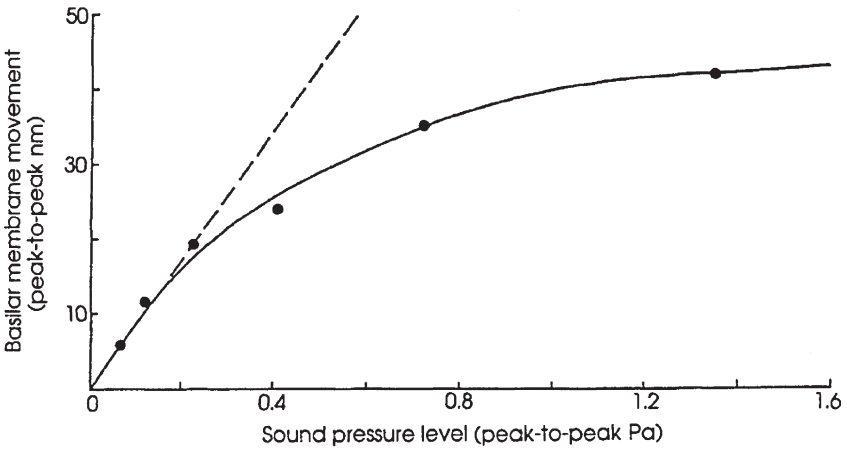
As the curve on the left was obtained from a place on the basilar membrane more apical than that associated with the curve on the right, it is more responsive to low frequencies.

Curves such as these should look very familiar to you—they are nothing more than band-pass filters. One way to think of the basilar membrane, then, is as a sort of filter bank (like those described in Chapter 11). Each point on the basilar membrane corresponds to a band-pass filter with a different centre frequency. As one goes from the base to the apex in the cochlea, the centre frequency of the band-pass filter decreases. This adds a further stage in our model of the auditory periphery:



A realistic model would mean an enormous number of filters, of course. Here, for practical purposes, we only show four. Using this model, we can predict the response of any point on the basilar membrane to any input, in the same way we would do for any LTI system. Essential, of course, is the assumption that the transformation between stapes and basilar membrane movement is, in fact, linear. von Békésy claimed that it was in his preparations, and gave as supporting evidence the fact that '... the amplitude of vibration of the cochlear partition, as observed at a particular point, increased exactly in proportion to the amplitude of the vibration of the loudspeaker system ...' which drove the stapes. In other words, he showed that the system was homogeneous.

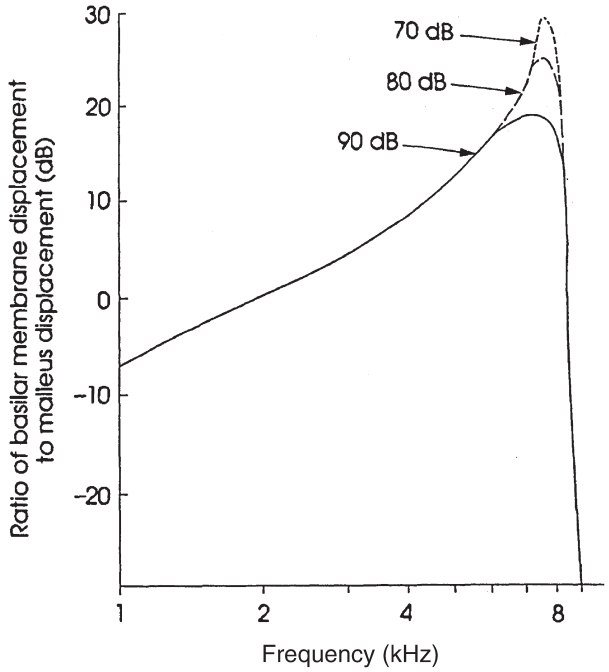
Some decades after von Békésy's work, however, Rhode showed that the movement of the basilar membrane is highly nonlinear, at least in squirrel monkeys. He also tested homogeneity, in essentially the same way as von Békésy, but with quite different results. Here are Rhode's data for 7.4 kHz, which show the amplitude of the movement of the basilar membrane at a fixed point, as a function of the input sound pressure level. (You have already seen these data in the figure on page 54:



It's easy to see that the amplitude of basilar membrane movement does not grow linearly with the amplitude of the input sound. In other words, the system is not homogeneous. The dashed line shows what would be expected if it were. This nonlinearity can be ascribed to some mechanism in the cochlea since Rhode (and previous researchers) demonstrated that the middle ear system *is* linear (see below).

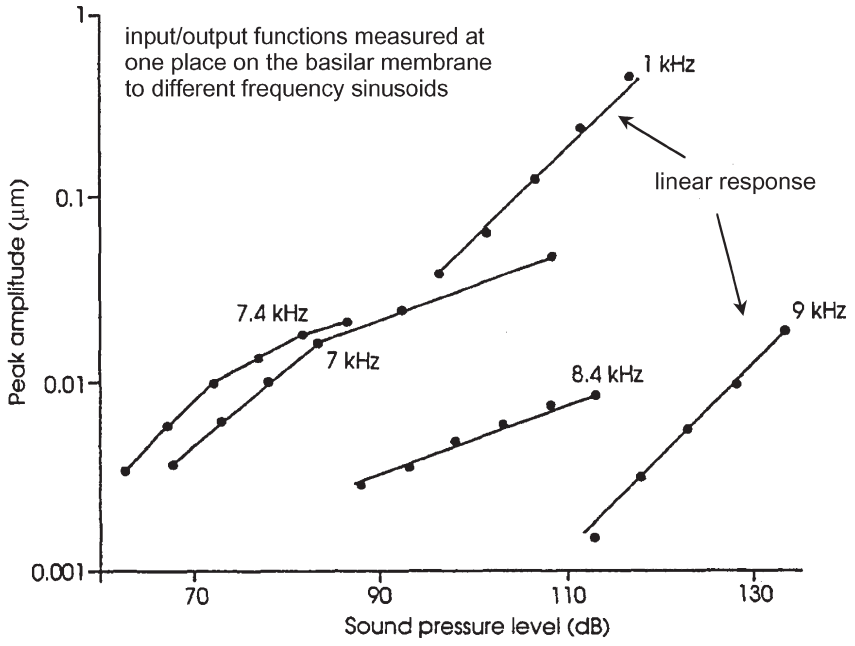
This result makes the task of characterizing the movements of the basilar membrane much more difficult than it would be for an LTI system. In the simplest case, suppose we only wanted to know the amplitude of the response of the basilar membrane to single sinusoidal inputs. If the system were homogeneous, we would only need one amplitude response, measured at an arbitrary input level for each frequency. We could then use homogeneity to determine the response to any sinusoid. Since the system under consideration here is not homogeneous, we need to look at its amplitude response at a number of levels to know what it will do. Rhode did just this and showed that the shape of the amplitude response did in fact depend on the input level used in the measurement, as shown on the next page.

The measurements were made at 70, 80 and 90 dB SPL but are normalized to the amplitude of the malleus displacement. (Rhode used this instead of the stapes displacement, as it was more convenient to measure.) Note that the three curves are distinct in the frequency region where the particular place measured on the basilar membrane responds best; they overlap outside this region. If the system were linear, all three curves would overlap completely.



A nonlinearity like this makes it difficult to apply in a straightforward way many of the concepts we've developed. Take bandwidth, for example. In a linear system it is possible to define a single value for a filter's bandwidth because it does not depend on the level of the input signal. If we tried to estimate the bandwidth of cochlear 'filters' from Rhode's data, however, a single number would not do. Here the bandwidth of the filter increases with increasing level. In such a system, 'bandwidth' would have to be a function of level and not a single value.

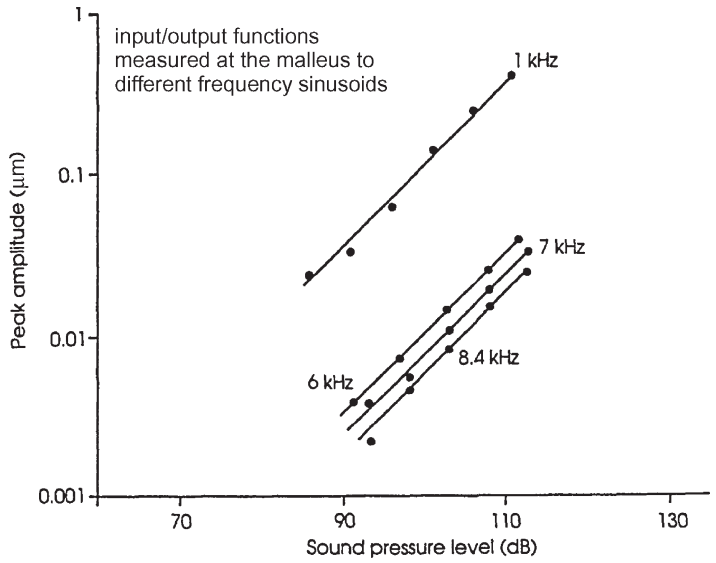
It's interesting to note that the cochlear 'filters' seem to operate linearly for frequencies relatively remote from their centre frequencies, but are highly nonlinear for signals near their centre (in the so-called 'pass-band'). This can also be seen in the following diagram where the peak amplitude of vibration for one place on the basilar membrane (most sensitive to 7.4 kHz) is plotted as a function of level for a number of different frequencies:



As we saw previously, when the stimulating frequency is near the centre of the pass-band, the amplitude of the basilar membrane vibration grows at a much smaller rate than it would in a linear system. The data for 7.4 kHz are in fact the same used in constructing the figure on page 274, which used linear scales because it is easier to understand homogeneity that way. Here we use logarithmic scales for both axes (log amplitude versus dB SPL).

If this system were linear, the amplitude of movement of the basilar membrane would increase proportionately with sound pressure level for all frequencies. In other words, a factor of 10 increase in the sound pressure level (20 dB) would lead to a factor of 10 increase in the amplitude of movement of the basilar membrane (again 20 dB). Therefore, on an input–output graph like this, all LTI systems would be characterized by a straight line with a slope of 1 (meaning the output grows by 1 dB for each 1 dB increase in the input). Although the system is homogeneous for frequencies of 1 and 9 kHz, it is not homogeneous for the other three frequencies—those in the centre of the pass-band of the 'cochlear filter'. This is another reflection of the finding that the amplitude response curves taken at different levels only overlap outside the 'pass-band'.

Contrast these data with those Rhode obtained when he tested a part of the middle ear for homogeneity. Here he measured the peak amplitude of the movement of the malleus as a function of sound pressure level, and found homogeneity at all frequencies (see over). Note how each 20-dB change in input level leads to a factor of 10 change in the measured amplitude of the malleus motion. In short, the amplitude of the malleus movement is proportional to sound pressure level.



The causes of the discrepancies between von Békésy's and Rhode's results on basilar membrane vibration arise from crucial methodological differences between the two studies. Almost certainly the essential one is that von Békésy always used preparations from cadavers. Rhode used live animals and found the nonlinearity to disappear very quickly once the animal had died. Also, von Békésy's technique necessitated the use of extremely high signal levels (up to 140–150 dB SPL) in order to

make the movements visible, whereas Rhode used levels that would be encountered in everyday life. Nonlinearities in basilar membrane movements have since been found many times by other groups of experimenters in cats, chinchillas and guinea pigs, and there is now a general agreement that the system is nonlinear.

The details of this controversy are, for our purposes, less important than the way in which the concepts of linear systems analysis pervade the entire discussion. LTI systems serve as a benchmark against which other systems can be compared; hence much effort goes into determining the exact nature of the departures from linearity found in a nonlinear system. Therefore, when Rhode claimed that basilar membrane motion is nonlinear, he did so on the basis of what would be expected of a linear system. This is yet another reason why an understanding of linear systems analysis is crucial to appreciate discussions of even nonlinear systems.

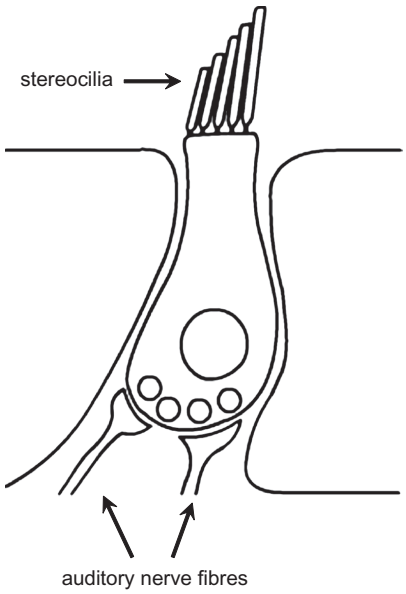
In terms of our model then, we would have to implement nonlinear filters in the filterbank meant to represent basilar membrane movement. In fact, this is not as complicated as it might appear, and many appropriate algorithms for doing nonlinear filtering are available. The details of those won't concern us here. We still need to develop at least one more stage in the model to get to auditory nerve firing.

Transduction by the inner hair cells

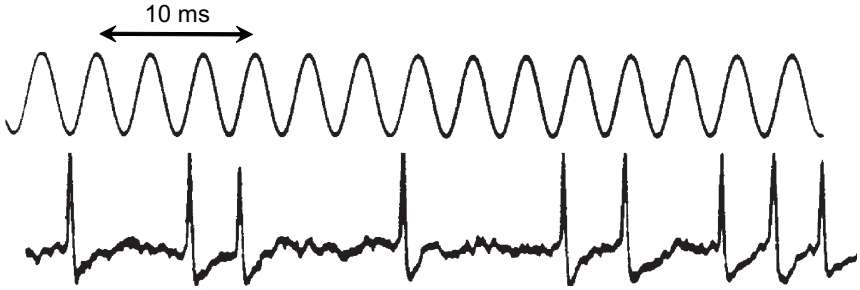
Up until this point, we have only been talking about *mechanical* signals, which is to say, those concerning either movement or changes in pressure. In order for any information about sound in the outside world to be relayed to the brain, it needs to be converted into an *electrical* code, as firings on the auditory nerve. This transduction, as mentioned above, is carried out by the IHCs. We will not describe the outer hair cells, although there are three times more of them than the IHCs (three rows compared to one). What has become clear over the last 40 or so years is that the outer hair cells are active and can move, and thus amplify the response of the basilar membrane, especially at low sound levels. In other words, they are responsible for the crucial nonlinearities in basilar membrane movements. However, they play no direct role in transduction, so we will not discuss them further here.

The vast majority of afferent nerve fibres (that is, those carrying information from the ear to the brain) synapse on a single inner hair cell (IHC), with each hair cell having about 10–30 nerve fibres synapsing to it. Therefore, in order to understand the firing on a particular auditory nerve fibre, we only need to consider a single IHC and the movement of the basilar membrane where that IHC lies. Here you can see a

schematic of a single IHC, with two auditory nerve fibres synapsing to its base:

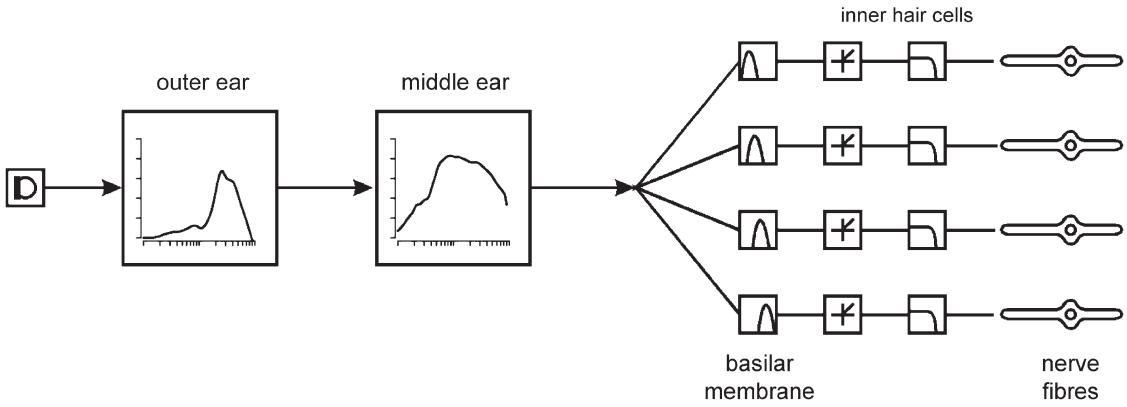


Imagine now presenting a sinusoid at the tympanic membrane. This would create a sinusoidal basilar membrane motion at the frequency of the stimulating sinusoid, with maximum vibration at a particular place on the basilar membrane, as determined by its resonant properties. This, in turn, causes the stereocilia at the top of an IHC in the appropriate place on the basilar membrane to vibrate back-and-forth at the same frequency. When the stereocilia move towards the tallest part of the hair bundle, neurotransmitter is released into the synaptic *cleft* (the tiny gap between the hair cell and auditory nerve fibre ending), making the nerve more likely to fire. When the stereocilia move the other way, neurotransmitter is taken up out of the cleft, making the nerve less likely to fire. Therefore, as long as there is time for the neurotransmitter to be injected and removed from the synaptic cleft, the nerve will tend to fire in synchrony with the stereocilia movements—at the same phase of the stimulating sinusoid. Here, for example, is the genuine firing pattern of an auditory nerve fibre to a section of a 300 Hz sinusoid:

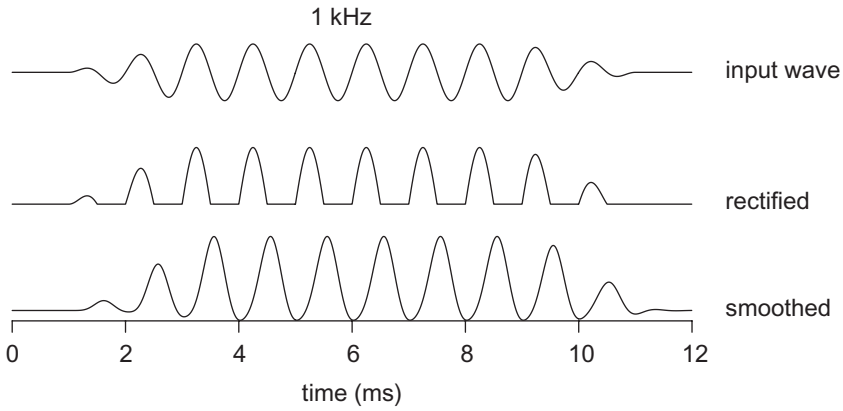


Nerves don't fire on every cycle of the stimulating wave, but when they do fire, it is at a similar point in the cycle of stereocilia movement (here when the sinusoid is at its lowest value). Because the movement of neurotransmitter in and out of the synaptic cleft takes some time, this synchrony is only present for sinusoids up to certain frequencies. Roughly speaking, synchrony is strong up to about 1.5 kHz, and then tails off, becoming undetectable for frequencies above about 5 kHz. At that point, the nerve fires without regard for the phase of the stimulating waveform, because the neurotransmitter does not have time to be released and cleared. Therefore, there is a more or less constant amount of it in the synaptic cleft through the time corresponding to one period.

We somehow need to account for these processes (the synchrony of nerve firing and its dependence on frequency) in the IHC portion of our model of the auditory periphery. It turns out to be easy to do this with a combination of rectification and a low-pass 'smoothing' filter, a concept you have met before when discussing the construction of spectrograms. For making spectrograms, we used full-wave rectification, because we wanted to account for *all* the energy in the wave. In order to model the way the IHC only releases neurotransmitter when the stereocilia bend in one particular direction, we use *half-wave* rectification. Following the rectification with a low-pass filter with a cut-off of about 1.5 kHz will filter out any fluctuations that would lead to synchrony at high frequencies. Our model is now complete, and looks like this:

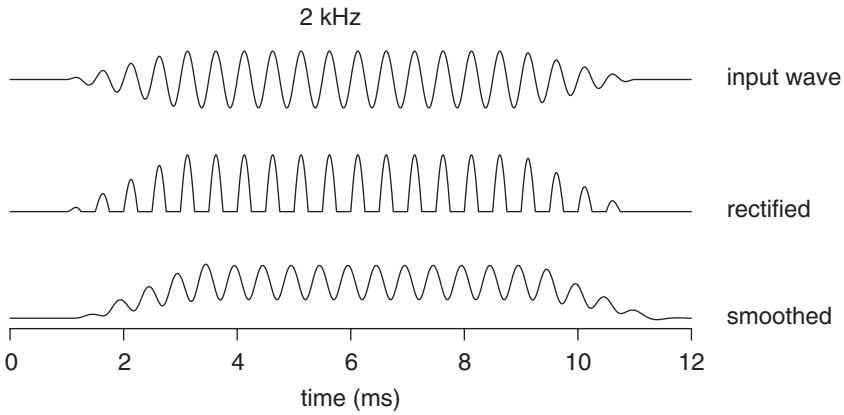


Before discussing how this model could be used, let's try to get a better feel for how the rectification and smoothing would simulate what happens in the IHC. Consider putting a 1 kHz sinusoid into the model. Assuming it is sufficiently intense, this will appear strongly as a sinusoid at the output of one of the band-pass filters simulating basilar membrane filtering with a centre frequency near 1 kHz. Half-wave rectification and smoothing would lead to the waveforms here:

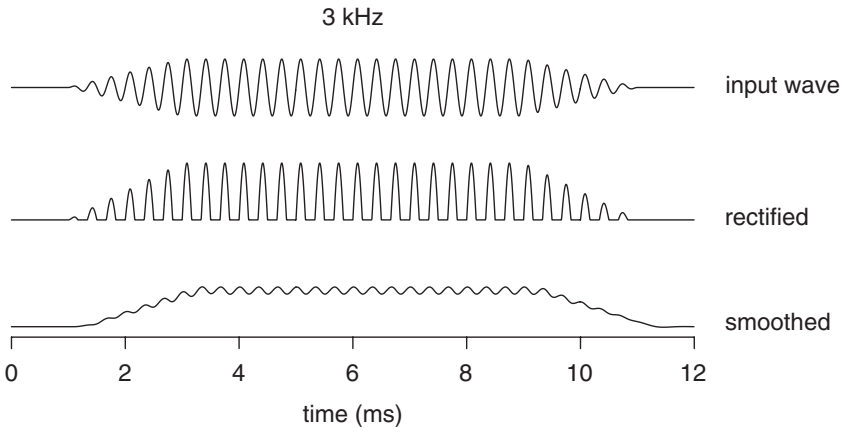


You can think of the final waveform at bottom as representing the amount of neurotransmitter in the synaptic cleft, or the probability that the nerve will fire at any particular moment. When the wave representing the amount of neurotransmitter is high, the nerve is likely to fire. When it is low, it is very unlikely to fire. At the 1-kHz frequency used here, the strongly fluctuating amount of neurotransmitter means that the nerve firings would be highly synchronized to the input wave. Perhaps the best measure of synchrony is the extent to which the peaks and valleys in the amounts of neurotransmitter vary. Here, the valleys go right down to the amount of neurotransmitter present in the absence of any sound, so this represents the maximal degree of synchrony.

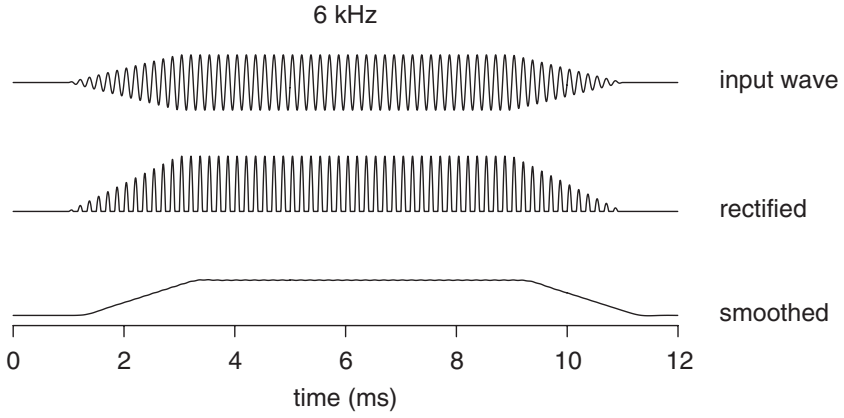
Let's now consider an input wave of 2 kHz, going through a channel tuned near 2 kHz. As you can see, there are still strong fluctuations in neurotransmitter although the valleys don't go right down to what you get with no sound at all. Therefore, you would expect some synchrony between the neural firings and the stimulating waveform, but not as strong as that found at 1 kHz:



For an input wave at 3 kHz, there is even less evidence of fluctuations in the amount of neurotransmitter. We would, therefore, expect little synchrony of nerve firing with the stimulating wave:



Finally, at 6 kHz, the amount of neurotransmitter in the synaptic cleft increases as the sound is turned on, but there are no fluctuations related to sound frequency at all. Note again that it is not the half-wave rectification that is responsible for this loss of synchrony—it is the low-pass filtering that matters.



Making an auditory spectrogram

It probably has not escaped your attention that the model of the auditory periphery that we have developed in this chapter is structurally very similar to the collection of systems we described in Chapter 11 for making spectrograms (p. 225). In both cases, an input signal is fed to a filter bank, with each

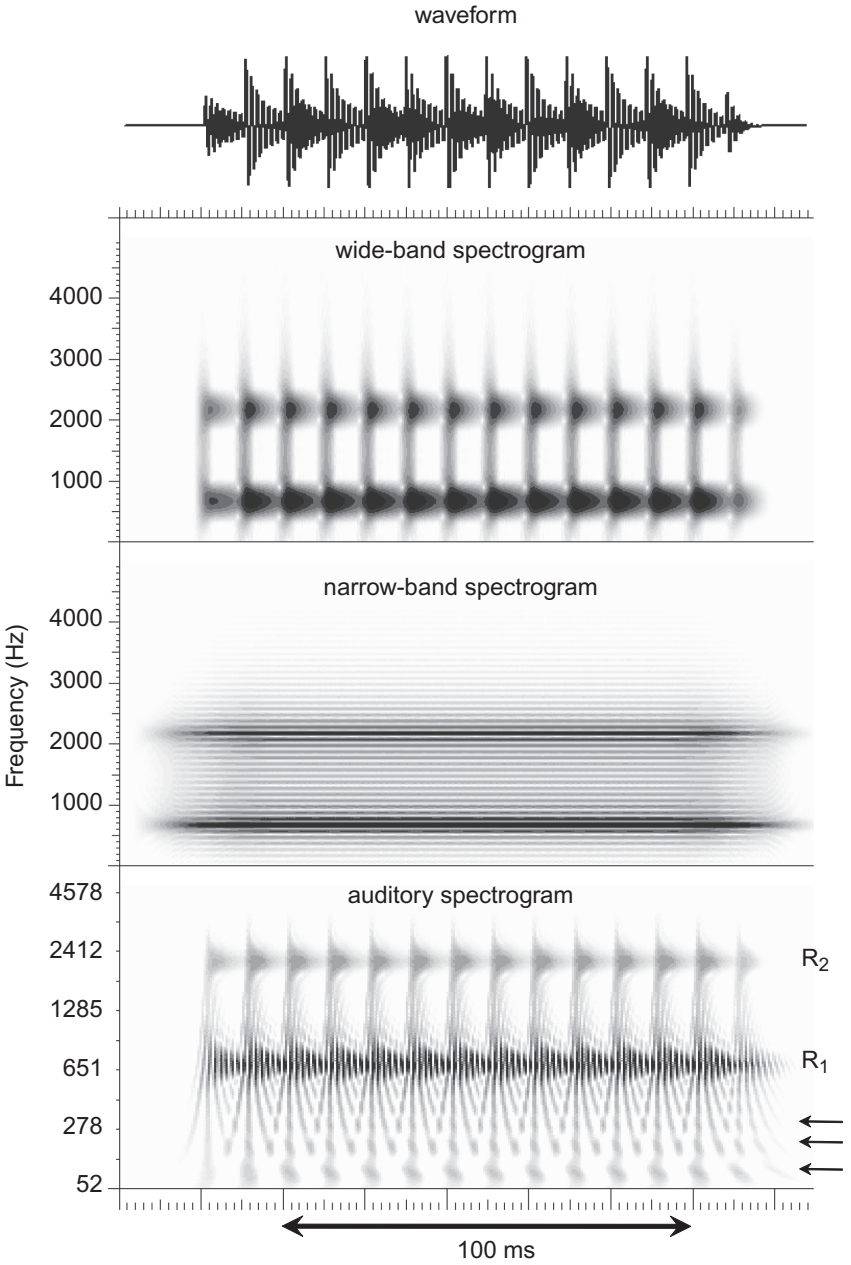
channel output being rectified and smoothed. The only significant difference in overall structure is that the auditory model has linear band-pass filters to represent the effects of the outer and middle ear. These serve only to amplify or attenuate various frequency regions but otherwise do not have a large effect on the representation of the information within most of the audible range of frequencies.

Given this similarity then, it might not be too surprising for you to learn that a special kind of spectrogram can be made using an auditory model, a so-called *auditory spectrogram*. All we need do is to take the model outputs and convert the amplitude variations into the darkness of a trace, exactly as was done for ordinary spectrograms. Before showing you an auditory spectrogram however, let us clarify two important differences in detail (apart from those we have already mentioned) between the processing that goes on for the two kinds of spectrogram.

- All the filters in a filter bank used to make an ordinary spectrogram have the same bandwidth, whether that is a wide or narrow band. In the auditory periphery, bandwidths increase as we move from the apex to the base of the cochlea—in other words, they increase with increasing centre frequency. We need to include this aspect in our auditory spectrograph. For frequencies of about 1 kHz and above, the bandwidth of an auditory filter is approximately a fixed percentage of its centre frequency. For frequencies below this, the percentage changes, but the absolute bandwidth still always increases with increasing frequency. As it turns out, at low frequencies, the bandwidths of the auditory filters are similar to the typical narrow band analysis in a standard spectrogram. But they increase steadily, becoming wide band at frequencies of about 2 kHz and above. This means that harmonics in a complex periodic wave are resolved and hence visible at low frequencies. At high frequencies, the harmonics are unresolved, and beat together, so result in striations, just as we saw in wide-band spectrograms previously.
- The spacing of the filters in a filter bank used to make an ordinary spectrogram is linear, whereas an auditory filter bank has a spacing that corresponds to the way in which sinusoidal frequency maps onto place on the basilar membrane, a so-called *tono-topic* map. For frequencies of about 1 kHz and above, this mapping is logarithmic. For frequencies below 1 kHz, the mapping is somewhere between linear and logarithmic. So, for example, if you make a spectrogram over the frequency range from 20 Hz to 20 kHz (the audible range of frequencies), half of an ordinary

spectrogram is taken up by frequencies above 10 kHz, which hardly matter at all to us. On an auditory spectrogram, however, the midpoint would be at about 1.8 kHz, which is a reasonable reflection of the relative importance of these two bands.

We'll only look at the auditory spectrogram for one particular wave, one for which you have already seen ordinary spectrograms (p. 241). This is a periodic train of narrow pulses with a fundamental frequency of 100 Hz, which has been put through a cascade of two resonators, one at 700 Hz and one at 2200 Hz:



Going from top to bottom, you can find the waveform, and three different kinds of spectrogram. The time axis is the same for all the four panels, but look at the frequency axes on the spectrograms. The ordinary spectrograms have, of course, a linear scale, but the auditory spectrogram is more or less logarithmic, at least for the frequency range above about 1 kHz. You can confirm this by taking the ratios of successive numbers on the axis, which are about equally spaced in distance. So, $4578/2412 = 1.90$ and $2412/1285 = 1.88$. But the scale is *not* logarithmic at low frequencies, because $278/52 = 5.35$, which is not close to 1.90 or 1.88.

Let's look first at the low frequencies, where we would expect the auditory spectrogram to look like a narrow band one. In fact, you can see evidence of 3–4 separate harmonics resolved, at the frequencies indicated by the arrows at the bottom right. But what is unlike the narrow band spectrogram is that you can see evidence of strong phase locking. So in the frequency region near 100 Hz (bottom arrow), you can see one 'pulse' per period. At 200 Hz (middle arrow), you can see two pulses per period (corresponding to a frequency of 200 Hz) and so on.

Round about the 7th harmonic, the trace gets darker, which results from the spectral prominence in the wave there, labelled R_1 (resonance 1). At higher frequencies, especially near 2.2 kHz, where the other spectral prominence is (R_2), the harmonics are no longer resolved, so features very similar to those in the wide-band spectrogram can be seen. This corresponds to the striations normally associated with a periodic wave.

Obviously, there is much more we could do in terms of understanding the representation of various acoustic features in an auditory spectrogram. For the moment, the most important aspect of this exercise is to show how many of the concepts you have learned with regards to systems and signal analysis can clarify processing in the auditory periphery.

Exercises

1. Here is a table of the speed of sound in various media. Calculate the wavelengths corresponding to frequencies of 500 Hz, 2 kHz and 10 kHz. In which medium would the first resonant frequency of the auditory ear canal of a particular person be lowest and highest?

Medium (m/s)	Speed of sound
Air	340
Water	1450
Oxygen	317
Hydrogen	1286

2. Calculate the first resonant frequency of the auditory ear canal (length = 2.3 cm) of an adult female scuba diver while in air and under water. Compare the two values.

3. How much would the frequency of the first resonance of the ear canal change over the course of life if it was 2 cm long at birth, and reached 2.4 cm in adulthood?

4. Discuss all the reasons you can think of for the advantages of saturating nonlinearities in natural systems.

5. On page 264, in attempting to model the amplitude response of the ear canal, we noted that an even better fit of model to data could be obtained by assuming a slightly different length for the ear canal. Would a better fit be obtained if the assumed length of the meatus was longer or shorter, and why?

6. Sound waves are affected by the head when they have a wavelength that is comparable to, or smaller than, the head and pinna. For an adult, the acoustic effects of the head are most marked at frequencies above about 1.5 kHz, equivalent to wavelengths smaller than 22.7 cm (the approximate diameter of an adult male head). What frequencies would be most affected for a child with a head size of 15 cm?

7. The motion of the middle ear ossicles can be modelled approximately as a pendulum (a weight suspended from a pivot so it can swing freely). The period of swing of a pendulum (T) is given by the formula:

$$T = 2\pi\sqrt{L/g}$$

where L is the length of the pendulum and g is the local acceleration due to gravity (9.8 m/s^2). What is the length of suspension of the middle ear ossicles for a frequency of 1 kHz (roughly the centre frequency of the middle ear system)?