him immediate fame and an academic post as Professor of Physiology, where he devoted his efforts to studying the eye and the mechanism of seeing.

In 1858 he was offered the post of Professor of Anatomy at that most beautiful of German university towns, Heidelberg, some 150 miles from Bayreuth. In the thirteen years he was there, his work on the eye led him to study the ear and the way we hear. He applied his results to the theory of music, which he published in a monumental work *On the Sensations of Tone*. In this book, for the very first time, the psychological perception of music was opened up for investigation—the field we now call **psycho-acoustics**. It was a turning point in the scientific study of our subject.

After Heidelberg he moved to Berlin as Professor of Physics, where much of his efforts were devoted to electromagnetism: and from there his influence spread, not least through the students he taught. Heinrich Hertz was one of these, and the experiments on radio waves were undertaken at Helmholtz's suggestion. Another was Max Planck who would soon be responsible for the revolutionary new Quantum Theory; but more of that later. Helmholtz died in 1894, almost 11 years after Wagner, the most respected scientist in his own country and acknowledged throughout the whole of Europe.

These two intellectual giants, so influential in different spheres of German life, existed side by side for a span of two generations. They were occasional friends. Helmholtz was in the audience for the first performance of the *Ring* at Bayreuth, at the composer's invitation, and wrote indignantly to the press about what he called the critics' 'icy non-recognition' of the work's importance. Likewise, when they were in Berlin, the Wagners were guests at the Helmholtzs' evenings-at-home, at which most of the fashionable intellectuals gathered. At one time they even found themselves on opposite sides of a public argument about vivisection. But they didn't seem to impinge on one another professionally. It is only in retrospect that we can see how much they had in common, and how similar were the effects they had on German intellectual life.

The field where they might have interacted is, of course, the theory of music; and I want now to examine what Helmholtz contributed to that subject. But before I can do that I will need to talk about a bit of physics I have avoided so far—a more complete description of resonance.

The theory of resonance

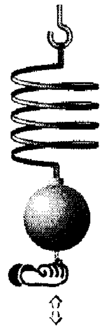
Up till now I have been rather cavalier in talking about resonance. I've discussed it only in the following terms. You have some system which can oscillate with its own natural frequency; and if you disturb it (periodically) at exactly this frequency, then energy will keep going in, and a very size-able amplitude of vibration will build up. The simplest model, you will remember, was of pushing a child on a swing. But I've never addressed the question: what happens if you don't do it at the right frequency?

To answer that I'll have to go back to my discussion (in Chapter 3) of why systems oscillate—of how energy changes periodically between potential and kinetic forms. However, the argument I want to go through is a bit involved, and if you don't want to follow it in detail, you can skip the next five paragraphs.

The problem is one of putting energy into a system, and therefore it makes sense to discuss it in terms of impedance—though of a slightly different kind, called **mechanical impedance**. This is the property of a system which determines how great an oscillating force has to be, in order to get the system moving at a certain speed. Strictly it is the ratio of force applied to velocity produced. If you apply a big force and only produce a small movement, you say the mechanical impedance is large. But if a small force results in large movement, you say it is small. It may not look like the same sort of impedance I talked about earlier, but it is clearly a related concept.

Now for clarity, I will talk about one particular oscillating system—a mass on the end of a spring. First I want you to consider the mass in isolation, and to think about how it responds if you try to get it moving by *shaking* it at a constant frequency. Obviously, if it is heavy, it will be difficult to make it move very fast. Its mechanical impedance must be directly related to its inertia. But even a light mass won't respond if you try to shake it too rapidly — just imagine, for example, trying to shake anything at more than about ten or twenty times a second. This means that inertial impedance must depend on frequency: it gets bigger as the frequency increases.

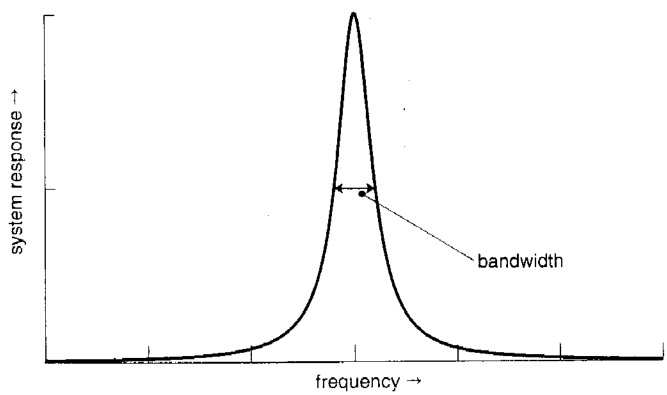
Secondly, think of the spring by itself. If you apply an oscillating force to it, you simply stretch or compress it; and the amount of stretch or compression depends only on the amount of force you apply, not on the frequency. Therefore the velocity with which the end of the spring moves *will* depend on the frequency, since it has to move through a fixed distance in each period of the oscillating force. So the kind of impedance involved here gets *smaller* as the frequency increases. Furthermore, because the spring pushes in the opposite direction (i.e. against you) when you push on it, elastic impedance is, in a sense, the negative of inertial impedance.



Therefore when you analyze a system consisting of both a spring and a mass, you must think of it as having a total mechanical impedance equal to the *sum* of these two (actually the *difference*, since one is negative). So, if the frequency is *either* very high *or* very low, the total impedance will be large, because one of its components is large, even though the other is small. However there is one particular frequency, somewhere in between, where the two parts of the impedance have exactly the same value and *cancel one another out*. At that frequency even a tiny applied force will produce a huge response. And that is of course just what we mean by the term **resonance**.

But one further point. Even at resonance, the total impedance can never go exactly to zero. There will always be friction, or some other means by which energy can leak away; and these will contribute another kind of mechanical impedance (just like electrical resistance) which can't be compensated for. So the exact impedance at resonance, and therefore the magnitude of the system's final response, will depend on how small this 'resistance' is. The actual reason for this concerns the *time* it takes for the energy to dissipate. The vibration will settle down only when the rate at which you put energy *in* just balances the rate at which it leaks *out*. Therefore if the system has a small resistance and loses energy slowly, the amplitude at resonance will be high; whereas if its resistance is large, it will lose energy quickly, and the resonant amplitude will be low.

Let me summarize the conclusions of this argument by means of a graph. I will plot how the response of the system (measured by the velocity of its motion) varies as I change the frequency at which the driving force is applied.

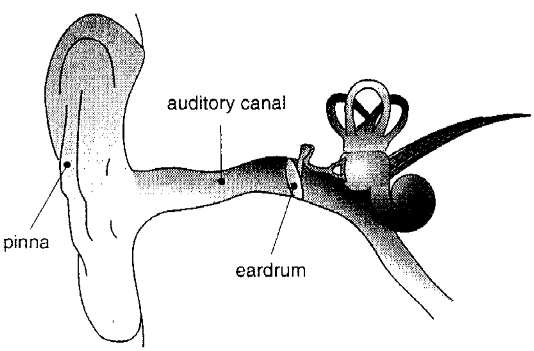


The main feature of this graph is what we already knew—that at one particular frequency, the **resonant frequency**, the system responds much more violently than at others. But what is also shown is that there is a *band* of frequencies around resonance at which the system still shows quite a large response. This range is called the **bandwidth**.

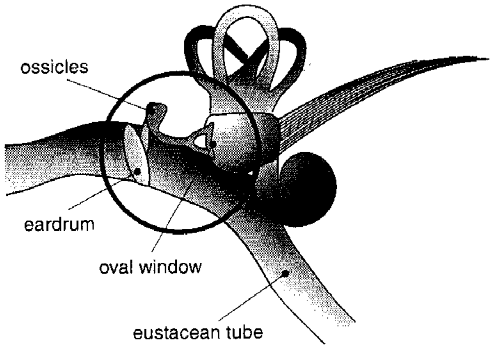
If you want to get a simple intuitive feeling for this graph, think about tuning a radio. As you turn the knob you are altering the circuit in such a way that the frequency at which it oscillates changes. You sweep this past the frequency of the radio station, and at one point they match. Then the tiny signal in the air is able to set up a big electrical oscillation inside and the audio message will come through loud and clear. But you can still hear it even if you are not quite 'on the station'. This is what I mean—this range of 'almost acceptable tuning'—when I talk about bandwidth. There are important applications of this principle in many musical instruments (especially in wind instruments, when players are able to correct for slight inaccuracies in tuning); but above all, it is vital in understanding how the ear behaves.

The ear

The ear can be considered in three distinct parts. The first, the **outer ear**, is the most obvious, consisting of the bits you can see and feel: the large shell-shaped lobe (called the **pinna**) which leads down through a narrow tube (the **auditory canal**) to the **eardrum**.



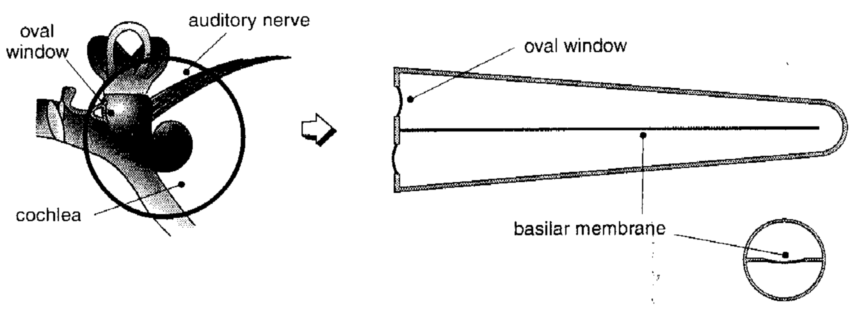
We have already noted that this is essentially the collector of sound energy; the narrowing shape provides enough of an impedance match so that a reasonable fraction of the energy falling on it ends up in a vibration of the surface of the drum. However, recent research has shown that the pinna does a bit more than this. The convolutions of its shape actually enable us to locate how high above the ground is the source of a sound. Similarly the auditory canal has its own acoustic properties. Being a tube about 3 cm long, closed at one end, it has a fundamental resonance mode at about 3 000 Hz; and that accounts for the fact that human hearing is most acute around that frequency. But, so far as *understanding* a message is concerned, we must look to the other parts of the ear.



In the **middle ear** the energy of vibration of the drum is transferred, by a mechanical lever system made up of small bones (called the **ossicles**), to where it sets into vibration a second membrane (the **oval window**). The function of this stage is primarily amplification—the lever system translates the small pressure variations on the (comparatively large) drum, to considerably larger vibrations of the (much smaller) oval window. But a secondary function is that of a buffer—if the pressure variations of the drum are too large, it doesn't amplify them nearly so much, thereby offering some protection to the more sensitive workings of the inner ear.

It is necessary for the middle ear to be unencumbered at all times, and so it is connected (via the **eustacean tube**) to your throat. When, for example, you go up in an aeroplane, or dive deeply under water, and a pressure difference builds up across the eardrum, you can equalize this difference by swallowing. There are several things that can go wrong with the workings of this part of the ear, which can lead to varying degrees of deafness—the drum can be punctured, the small bones can seize up, the chamber can become filled with mucus. Luckily most of these complaints are, at least in principle, treatable to some degree.

The next stage, the **inner ear**, is the most interesting from our point of view. Ignore the strange loops at the top of the diagram—the so-called **semi-circular canals**, which are concerned with the body's balancing mechanism—and concentrate on the bit that looks like a snail shell. It is a helically coiled bone cavity, filled with fluid, and is called the **cochlea**. Its internal structure is easiest to understand by imagining this coil to be 'unwound'; in which case it might look like this:



The chamber, about 3 cm long, is divided most of the way down the middle by a narrow strip of taut skin called the **basilar membrane**, which separates the fluid in the upper and lower halves. This membrane is narrowest and tautest at the front end (near the oval window) , and widest and slackest at the other. When the ossicles set the oval window vibrating, this motion is communicated, via the fluid in the cochlea, to this membrane. And this is where the sound wave is actually 'detected'.

Underneath, in the lower half of the chamber, there are millions of tiny hair-like nerve cells which respond to any movement of the membrane by firing tiny electric currents. These current surges are conducted out of the cochlea, by the **auditory nerve**, to the brain. If anything goes wrong with this part of the ear, it is obviously extremely serious; and unfortunately this is where damage resulting from excessive loudness occurs. When the nerve cells are subjected to too much stress, they are destroyed one by one. Under a microscope they look like a bomb site. And this can be disastrous because, as we will see, they are responsible for the ear's ability to distinguish pitch. Therefore musicians who play in very loud rock bands are often putting at risk their most valuable asset—their musical ear.

Just in passing, it is worth mentioning a particularly impressive example of the marriage of technology and medicine—the **bionic ear**, or if you prefer, the **cochlear transplant**. This device was developed by Australian scientists in the 1970s, and it collects sounds with a a tiny microphone sited just behind the ear. These are transmitted to a receiver buried under the skin, where they are converted into coded electrical signals by a speech processor. They are then passed on to 22 electrodes which have been surgically implanted into the cochlea at particular points along the basilar membrane. The procedure is still very expensive, and requires a lot of rehabilitation for the patient's brain to learn to interpret the unfamiliar signals it is receiving; but they do work and many, many thousands of deaf people have been fitted with them in the past decades.

Anyhow, in this discussion I've left out a lot of detail, and my description of the essential function of each part of the ear is grossly oversimplified; but it highlights an important consideration in thinking about the ear as a instrument for interpreting music. Every sound we hear is processed twice: once by the cochlea, where it is coded into an electrical signal, and then by the brain, where its message is extracted. It is currently fashionable to think of the human brain as a kind of electronic computer, and the processing of information to be carried out by some kind of computer program. In so far as this is valid, it is clear that the second stage of processing is more or less under our control—we can *learn* to change the way we think about music. But that which is done by the cochlea we are stuck with; and a lot of our response to musical sounds must be tied up with exactly what it is that the cochlea does. So we have to look at that next.

Pitch recognition

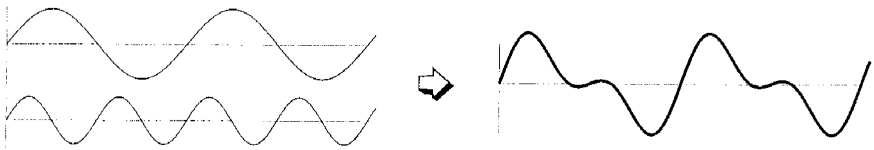
Clearly what is important is the way the basilar membrane responds to vibrational energy falling upon it. Remember that it is roughly triangular in shape; and up till now, I have said nothing about how standing waves are set up on a two-dimensional surface. But the basic principles are the same as those we have met before, and a good deal of insight can be gained by thinking about the membrane as though it were made up of a whole lot of short strings, like some long, thin, many-stringed dulcimer. It has very short, taut 'strings' at the front end, near the oval window; while those further down are longer and slacker. If the fluid around it vibrates at a pure frequency, then there will probably be one 'string' somewhere along the line which will resonate with it—near the front end if the pitch is high, further along if it is lower. This resonant vibration will, in turn, cause the nerve cell beneath it to fire, and therefore the brain will be able to recognize the frequency by noting *which* 'string' resonated.

This description is a bit simplistic, and many researchers prefer to talk about the process which gives energy to a particular part of the membrane in terms of a travelling wave, rather than a vibration (i.e. a standing wave). You see, the process has to happen quickly, so there can't be any significant 'build-up' time for the resonance. Instead they describe what happens like this. A wave of displacement travels down the basilar membrane. As it does so, it continually reaches parts of the membrane (the 'strings') where the elastic properties are different, and the speed of the wave gets slower (how much depends on the frequency). Eventually there will be a point at which the wave stops and dumps all its energy, causing the membrane to oscillate strongly at that point. (The process is exactly the same as a surf wave breaking when the depth of the water gets too shallow.)

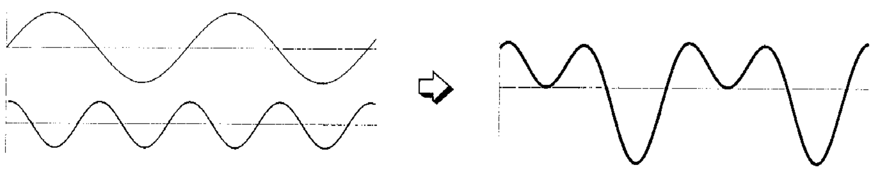
However, I have made the point many times that there isn't much conceptual difference between a vibration and a wave, and therefore my model will let you intuit what is going on. Certainly that was how Helmholtz imagined the ear working—as a row of graded resonators: and the process of recognizing frequency as being equivalent to *locating* where on this row the resonance occurred.

It certainly explains very simply how your ear assesses the *timbre* of different notes. Since a complex periodic vibration is entirely equivalent to the summation of pure tones of harmonically related frequencies, then, depending on which overtones are present, more than one part of the basilar membrane will respond at the same time. The cochlea therefore performs the kind of harmonic analysis I described in Chapter 4, and the message it sends to the brain consists of a number of electrical signals along different fibres of the auditory nerve, one for each overtone. The brain can then be 'programmed' to identify them.

There is good evidence that this is a very useful model of the cochlea's function. It concerns a particular relationship between tones which are added together. You will recall (see page 95) that, if I add a fundamental oscillation to one of its harmonics, I generate a complex shape, like so:



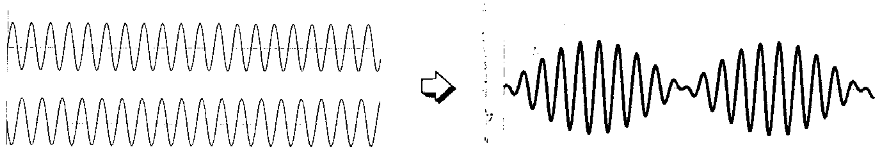
However, when I (or rather, my computer) drew these figures, I started off both oscillations exactly in step. I didn't have to do this. I could equally have started one of them at a different part of its cycle (i.e. with a different **phase**); and the result would *look* different.



But if these shapes were pressure waves in the air (provided they were not too loud) they would *sound* the same. Your cochlea can tell that there are two tones present, and what their amplitudes are; but it can't tell anything about their relative phase.

This observation has been known for over a century, and is usually given the name **Ohm's law of acoustics** (after the same Ohm who did all that work on electrical circuit theory). I'm sure you will appreciate how strongly it supports the 'place theory' of pitch recognition. The nerve cells in the cochlea can tell that two different parts of the basilar membrane are vibrating, and how strongly; but, because they are physically separated, they have no way of telling whether or not they are going up and down in step with one another.

There is another piece of evidence which supports this same model. In Chapter 4, I talked about **difference tones**. You will recall that, if I add two high frequency oscillations, I characteristically get an oscillation whose amplitude fluctuates with time.



If the two frequencies are close to one another, you can hear this fluctuation as a beat; but if they are far apart you can *sometimes* hear the fluctuation as a tone. Now I want to ask the question: why only sometimes? Why can't you hear it all the time?

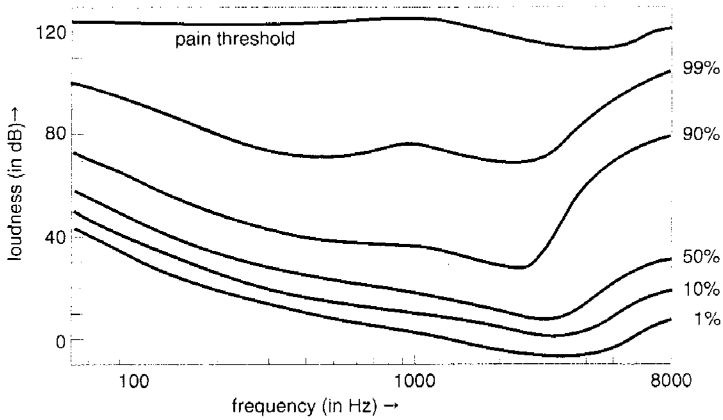
The answer lies in a detail which I have avoided saying much about so far. Whenever I talked about wave motion, I implicitly assumed the properties of the system which control how the wave moves are not affected by the wave's being there. I took it for granted that two waves can travel through the same medium, and the motion of each is unchanged by the other. Now when you come to think of it, this requires the medium to have some pretty specific properties. When it is stretched because of the passage of one wave, it must still be able to stretch the same amount extra as a second wave goes through. (This property is called **linearity** by mathematicians, in case you ever come across the term.) If your ear, for example, really does behave like this, then you can see that it will not hear the fluctuations we were talking about, as an independent tone. It will simply register that there are two pure tones present, because its response to each one is unaffected by the presence of the other.

But of course, very few materials are absolutely, absolutely linear. There is a limit to how far any elastic material can stretch, and if you get near that limit, then it's not much use trying to make it stretch any more. A second wave will not be able to travel through properly. And there are lots of points in your ear which behave elastically and which may not be able to respond absolutely faithfully to a large wave coming through—the eardrum, the ossicles, the oval window, the basilar membrane itself. So when your ear encounters a *loud* signal which fluctuates, the signal will be distorted. Then when your cochlea tries to harmonically analyze the signal it will still detect the two pure tones, but it will also detect a distortion which fluctuates with a frequency equal to the difference of the two pure frequencies. You will hear the difference tone. Likewise you may be able to detect a change of phase in two very loud tones, because they stretch the elastic materials in your ear differently.

Therefore I hope you can see that this model of how the cochlea works allows us to understand a lot of what we know about hearing in a simple and straightforward manner. I should point out here that this whole field of research is still changing. For example, there is evidence that the brain receives *some* direct information about how rapidly the basilar membrane vibrates—in the rate at which the nerve cells fire. As a result, some researchers have proposed a so-called 'time theory' of pitch recognition to complement the 'place theory'. Obviously a definitive understanding of the mechanism of hearing is not yet with us. Nonetheless, most of the features I want to talk about can be understood from the simple picture, even if it isn't complete. And that's all we want.

Range of pitch

Let me now turn to the question of what range of frequencies the ear is sensitive to. Everyone's ears are a little different; and over the years many experiments have been done, getting volunteers to listen to tones of different frequencies, trying to determine the lowest intensity they can detect. The average results of countless such tests are usually summarized in a graph like this:



Notice that loudness is measured in its own special unit, called the **decibel** (abbreviated as **dB**). If you are interested, I have included a brief discussion of this unit in Appendix 4.

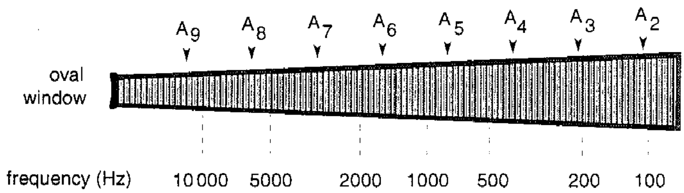
You interpret the graph as follows: 1% of people can hear any sound whose intensity is above the 1% curve; 10%, a sound above the 10% curve; and so on. The topmost curve represents the intensity at which most people start feeling pain.

The graph is not extended below 60 Hz, because of a rather strange observation. When you listen to a repetitive pressure wave of very low frequency (say around 10 Hz) you don't hear it as a tone at all. It just sounds like a series of clicks; and this is true up to about 20 Hz. After that the clicks run together, but you don't start hearing them as a tone until about 50 Hz. In between, the sound is fuzzy and not particularly pleasant.

The upper frequency limit is even less well defined. Some people can hear as high as 20 000 Hz; but for most of us, the threshold is much lower. It is one of those sad facts of life that this figure decreases as we grow older—after age 40 at the alarming rate of about 80 Hz every six months. There are a lot of reasons for this, the most straightforward being that all skin loses resiliency with age, and none of the membranes respond as well as they should.

One of the starkest examples of this is in the background squeal of a TV set. In television, the image is built up in lines, by the bright spot moving across the screen about 600 times for each frame. The illusion of motion is achieved by the picture changing 25 times a second. So in every TV signal there is a pulse telling the spot to start a fresh sweep, which occurs about 15 000 times each second. This gets into the audio system, and comes out as a tone of 15 000 Hz. Children and young adults can hear this clearly; and, I am told, find it very annoying. The rest of us, alas, have long since sunk below the level at which we can even hear it.

Between the upper and lower frequency limits, our ears respond with varying degrees of sensitivity. It is greatest around 3 000 Hz; and one factor accounting for that is, as I mentioned earlier, the physical size of the outer ear. But an equally important factor involves the way the resonators are distributed along the basilar membrane. In this diagram, I have indicated which 'strings' respond to various musical pitches:



Notice firstly that the musically most important range of frequencies (from about 100 Hz to about 4 000 Hz) occupies roughly 2/3 of the length of the membrane, while the rest of the scale (up to nearly 20 000 Hz) is squeezed into the remaining 1/3. Secondly, notice that the frequencies are spaced *logarithmically* (a word which you will find defined more carefully in Appendix 2); that is, whenever the frequency is doubled (and the pitch rises an octave) the resonant point moves roughly the same distance (some 4 mm) to the left.

The importance of this last observation cannot be overemphasized. It gives a straightforward explanation of that intriguing fact I mentioned in the very first chapter of this book—that the natural way to express the 'difference' between the pitch of two notes involves forming the *ratio* of two frequencies, rather than *subtracting* those frequencies. Once you know how the ear works, it seems perfectly reasonable that the apparent interval between two notes should depend on the number of auditory resonators separating them; and that depends on the ratio of their frequencies.

Of course, this leaves unanswered the question of *why* our ears should have evolved in this way. So we haven't solved everything yet. But it is interesting that all members of the animal kingdom who employ sounds which we consider to be musical—like birds or whales or dolphins—have ears whose structure is very similar to our own. Surely there can be no doubt that a lot about music is determined by the way our ears are put together.

Pitch discrimination

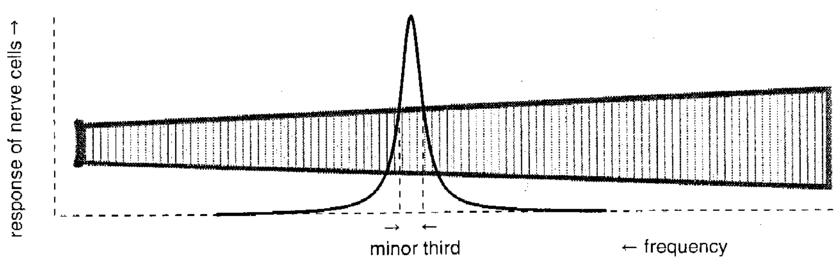
The next question to be addressed is: how good is the ear at telling frequencies apart? Or, if you prefer, what is the smallest pitch interval you can have between two notes and still hear that they are different?

The most useful evidence bearing on this comes from the kind of experiment in which you sound two pure tones of different frequencies together and listen to what they sound like. The results of such experiments, again the average of many listeners, can be summarized like this:

- When the two frequencies are very close together you hear **beats**—a regular pulsation in loudness at a single pitch somewhere between the two. (I talked about this in Chapter 4). This effect persists up to a frequency difference of around 20 Hz.
- On the other hand, when the notes are widely separated, you hear them as two clearly distinct pitches, and this is true for any two tones which are more than about a minor third apart (i.e. three semitones or about 20% difference in frequency).
- In the in-between region, what you hear is a bit uncertain. The sensation is often described as 'roughness', a term which is used by a remarkably wide range of listeners.

A lot of this can be explained from what we know about the way that all resonating systems behave. Let us think first of all how we would expect the ear to respond to *one* tone. When a single pure tone is sounded, there is only one auditory resonator, one 'string' in the cochlea, which exactly matches it in frequency. This is the one which will respond most strongly. But there are other resonators nearby which *nearly* match, and these will also be set vibrating. Whether or not they do so with an appreciable amplitude, depends on the *bandwidth*. If this is large, i.e. if each resonator will respond to a wide range of frequencies, then many neighbouring resonators will respond to the tone. A large *area* of the membrane will vibrate.

It is useful to represent this conclusion diagrammatically, by imagining that the response of the nerve cells along the basilar membrane follows a resonant curve, like the one on page 226.



Now consider what happens when the ear hears *two* tones together.

- If they are very close together in frequency, these areas will overlap and the part of the membrane in the overlapping region will be performing two independent oscillations. It will have a big amplitude when they are in phase, and a small amplitude when they are out of phase; and will change regularly from one to the other. You will hear **beats**.

- If the two tones correspond to parts of the membrane well outside one another's bandwidth, then this effect should be entirely absent—the two vibrations should proceed independently of one another. You will hear two distinct pitches.

- But if they are not too widely separated and their bandwidths partially overlap, it is more difficult to say what will happen. Perhaps some of the signal going to the brain would say "two distinct frequencies", while another part of the same signal would say "a pulsating single frequency". It seems plausible that such a message might be described as 'rough'. It is equally plausible therefore, that we can identify the bandwidth of the auditory resonators to be the range over which this roughness is known to be detected—i.e. a minor third.

That seems a plausible explanation for the observations, but it does raise another question. If each resonator will respond to any note within three semitones of its natural frequency, how is it that we can identify pitch as accurately as we can? Most people with minimal training can pitch a note correctly at least to within a semitone. And a trained ear can do much better. There are plenty of choir conductors who expect their singers to be able to distinguish between a Pythagorean tone (a ratio of 9/8) and a just tone (10/9). The difference between these is a ratio of 81/80, or about 1/5 of a semitone.

What is clear, I think, is that the cochlea cannot, of itself, make such distinctions. Once again, it is just like tuning your radio. You can get to within the bandwidth of a particular station, just by listening to how loud the signal is. But if you want to get it 'right on', you've got to wait till you hear something you can recognize and then try to judge whether it is distorted or not. The point is that you need more information to work on. It must be the same with the auditory system. The brain needs more information than it gets from the simple observation of *which* resonators are moving. Just what this extra information is, is still the subject of investigation—it probably has something to do with the regularity of nerve firing—but I don't think it is important right now. It is enough to know that it is a secondary process, under the control of the brain. So you can learn accurate pitch recognition; but the broad, general features of relating tones to one another is built in.

I cannot leave this section without making some brief mention of the fascinating topic of **absolute** (or **perfect**) **pitch**. Most people, if they hear one tone, and then another, can tell whether the second is higher or lower than the first. Those with musical training can usually recognize the standard intervals between two tones, or can sing these intervals after having heard a reference tone. Many with well trained ears can detect a frequency shift of as little as 1% (or a sixth of a semitone), and sometimes even smaller intervals. This is called **relative pitch**, and it is, when you come to think of it, a quite extraordinary sensory ability. It is difficult to think of any evolutionary advantage which could have caused our ears to develop like this.

But even more extraordinary are those 0.01% of the population (or even fewer) who have absolute pitch—who can recognize or sing a given note, without referring to any other tone as a reference. Psychologists have been studying absolute pitch for nearly a century, but there is still no agreement about why some people have it and others don't. Some researchers claim there is evidence that it is inherited. There have been some very recent studies reported which suggest that many babies are born with this ability, but quickly lose it as they listen to the way the people in their world sing and play music without much need for absolute tuning. But there are just as many studies which suggest that it is an acquired characteristic, and can be learned (most successfully while you are young).

Possessing absolute pitch can obviously be advantageous for a professional musician—as a singer you don't need an accompaniment to sing in the correct key, or as a conductor can more easily determine what notes should be played. However it also has disadvantages. It is a reasonably common complaint among choral singers with absolute pitch that they get put off when the rest of the choir, blissfully unaware, drifts out of tune. And it is certainly not an essential prerequisite for a musician. Many composers have been reported to have had perfect pitch, including Mozart and Beethoven, but there are even more who didn't.

Loudness

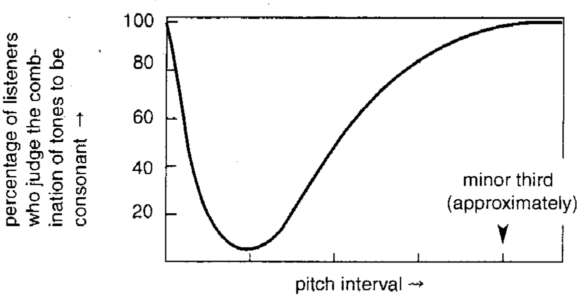
The other job that your ear has to do is to recognize how loud a sound is. And here you can see what a truly remarkable instrument it is, because the range of intensities it will respond to is enormous.

The **intensity** of a sound is measured as the amount of energy which falls each second on an area of standard size (usually taken to be 1 metre square). Hence its unit of measurement is the watt/square metre, or W/m^2

The psycho-acoustical theory of consonance

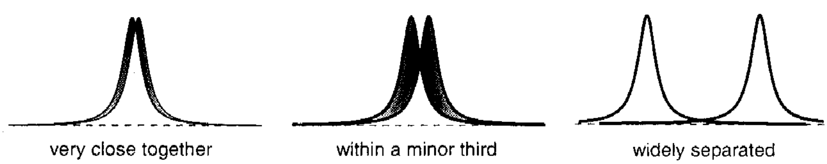
Let us return to the problem that we have kept coming back to many times throughout this book, the centuries old **puzzle of consonance** which I last mentioned when I was talking about Galileo's contributions to musical theory: why do some pairs of musical notes sound well together, while others do not? There is information bearing on this to be got from a series of experiments that were carried out in the 1960s, by a pair of Dutch scientists, Reinier Plomp and Willem Levelt (among others).

In rough outline, these experiments consist of the following. You play two pure tones of different frequencies together and, this time, ask listeners whether or not the combination sounds pleasant. (You must use untrained listeners, because you don't want them to have any preconceived notions of which intervals *ought* to sound good). Your results will probably look like this graph—which actually comes from many such experiments over the years.



Obviously nearly everybody agrees that two tones less than about a minor third sound dissonant, and at about a semitone, extremely so. But what is interesting is that intervals greater than this are *all* judged more or less equally pleasant. There seems to be no preference at all for the musically significant intervals—the fifth, fourth, third and so on.

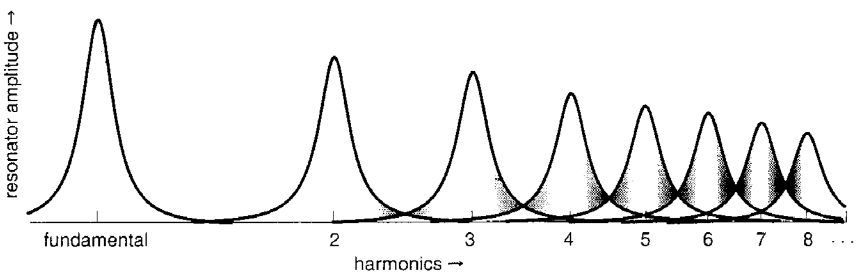
This observation can be reconciled with what I said before, when I tentatively identified the resonant bandwidth of the auditory nerve cells as a minor third. Let me try to represent on a diagram the nerve cells' response to two pure tones at various intervals apart.



Now it seems plausible to identify on these diagrams the areas where the nerve cells respond erratically—the regions of 'roughness'—as those areas where the two curves have large amplitudes at the same frequency but are different from one another.

A conclusion we would draw from these diagrams is that, when the two tones are far apart in frequency, there is no physiological reason why we should perceive any dissonance at all. But we all know that there *is* a widespread preference for the classical musical intervals when ordinary musical notes are heard together. The only significant difference between pure tones and real notes is that the latter contain overtones; so the solution must be sought in how our ears respond to these.

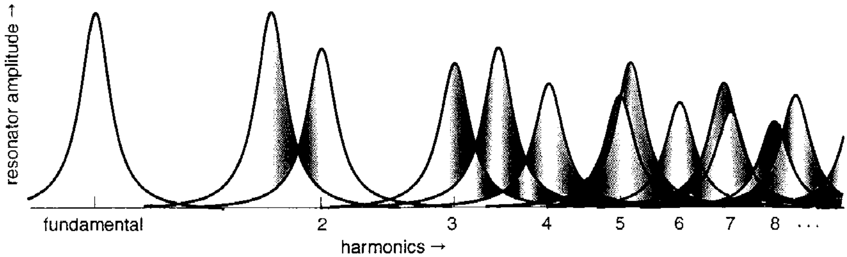
If you bear in mind that, when a single pure tone is sounded, the resonators within the **critical bandwidth** all respond; then when a real note is played, resonators within many such bands will start resonating. The *total* response can then be represented by plotting the vibration amplitude of each of the resonators against their natural frequencies—in other words, by drawing the **frequency spectrum** of what the basilar membrane detects.



Each of these resonant peaks is centered on a harmonic of the fundamental. These should be equally spaced along the frequency axis, but I have used a logarithmic scale (because that's what the ear prefers). On that scale the higher harmonics get closer and closer together. However the critical bandwidth stays the same apparent width—it's always just under a minor third remember—so for the higher harmonics there is considerable, and increasing, overlap. And that kind of overlap implies 'roughness'.

This must mean that, in any real note, there is actually a fair bit of dissonance, especially if the high harmonics are strong. It doesn't follow that they should sound unpleasant: 'rough' doesn't necessarily mean 'nasty'. Nonetheless the effect is noticeable. We use the adjective 'brassy' for any note with very prominent upper harmonics, like those of a trumpet; while those of a flute, which has very few, are often described as 'gentle'. It is as though a little bit of dissonance is a kind of spice—too much is to be avoided, but food tastes bland without it.

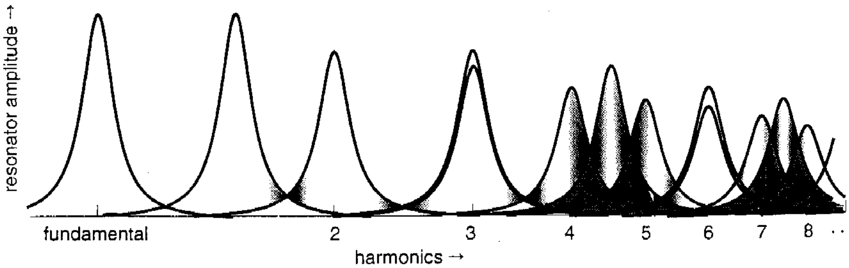
But now think what happens when *two* notes with overtones are sounded together. If their frequencies are randomly chosen, even if their fundamentals are separated by more than a minor third, it's likely that there will be a great deal of overlap among the harmonics. (I can't keep using this way of drawing things, but it will serve to indicate what I mean here.)



It is difficult to believe that your ear won't register this as pretty 'rough'. To continue the culinary metaphor, it is surely a bit too highly spiced. Though again, you could *learn* to like it—there is such a thing as an acquired taste.

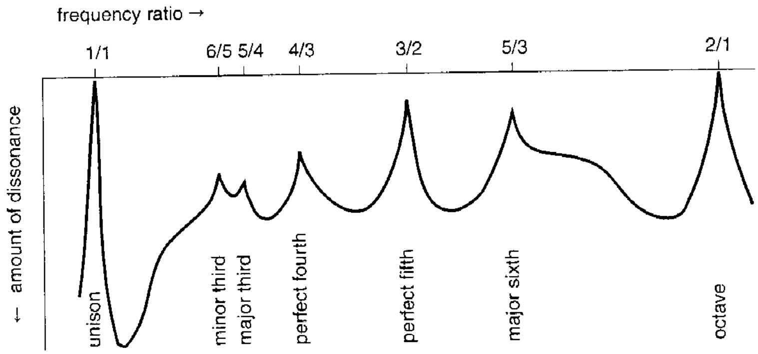
However, there are some special intervals between the two notes for which this won't happen. The most obvious is when they are an *octave apart*. Then the resonant peaks of the higher note will exactly coincide with every second peak of the lower. So adding the former to the latter will produce *no increase in roughness at all*. That seems to me to go a long way towards explaining why two notes an octave apart are so completely harmonious that they can almost be considered the same note. In other words, the absence of any extra roughness must be what we mean by 'perfect consonance'.

A similar claim can be made if the two notes are a perfect fifth apart (with frequencies in the ratio 3/2). Then every second peak of the higher will coincide completely with every third one of the lower.



There is clearly more roughness here than for either note singly, but much less than in the preceding diagram. This is because the ratio of fundamental frequencies is just what is needed to put some of the peaks completely on top of one another, and cut down on the total amount of overlap roughness.

Much the same will be true for other pairs of notes whose frequencies are in the ratio of two small whole numbers. Therefore it is possible to calculate the degree of overlap from *any* pair of notes, and to predict how much dissonance they should generate when sounded together. The result of this calculation, as first carried out by Plomp and Levelt, is as follows. (Note that the graph is plotted so as to look like the results of the experiment at the start of this section.)



Clearly the traditional musical interval ratios stand out from others around them as being particularly free of dissonance. It would appear therefore, that we have found a truly basic explanation, in terms of the properties of the ear, for why these intervals should be pleasing to listen to. This is an important change in the theory of harmony, because it suggests that consonance is a negative feature—an absence of dissonance—rather than a positive quality in its own right. It is also important because it shows that the property of consonance is not absolutely dependent on the exact value of the frequencies involved. There is room for a little inaccuracy, and therefore the intervals will sound much the same no matter what *musical scale* (i.e. just or equal tempered) they are played in.

These insights were largely developed by Helmholtz during those years in Heidelberg—although some of the results I called upon came from more recent research. In his book *On the Sensations of Tone*, he went a great deal further than this. He devoted a lot of time to discussing combination tones, and pointed out that, when two notes sound together, there will be many *difference tones* at the frequencies separating the various harmonics. These will only be heard faintly (as we discussed earlier) but unless the fundamental frequencies are in simple ratios (again), they will be dissonant with the primary tones. Hence he was led to a theory of chords and an understanding of the role of the **fundamental bass**, exactly as Rameau had been over a century earlier.

To me, there is a paradox about Helmholtz's place in the history of the theory of harmony, as seen by musicians. If you look under his name in most of the standard musical encyclopedias, you will find only the briefest mention (if any). I suppose from the point of view of those interested in *aesthetic* questions, he didn't do much that was new. But from the viewpoint of someone like me, his contribution was immense. He supplied an answer to the great question: "Why?" Whereas Rameau had said that the rules of harmony had to be as they were because the consonances on which they were based sprang from a kind of cosmic 'rightness', it was Helmholtz who firmly showed that the answer lies—to coin a phrase—not in our stars, but in ourselves.

Envoi

Helmholtz died in 1894 and Wagner eleven years earlier, in 1883. With their deaths, a chapter of the history of both music and physics seemed to close. Almost immediately both went through a period of such great change as can only be described as a revolution. I will talk about the new music later, but now let me concentrate on what happened to physics.

In the 1880s, James Clerk Maxwell had announced, with typical Victorian complacency, that essentially all of physics had been solved. To use his metaphor, the scientific sky was perfectly clear, except for one or two small clouds on the horizon. These 'clouds' were a couple of obscure observations about the way that light reacted with electricity, and the newly discovered phenomenon of radioactivity. They were to prove precursors of a cyclone.

Both in chemistry and physics, the really exciting area of research in the second half of the 19th century, was into the structure of matter. Experiments had finally confirmed that all substances were made up of atoms; and that electricity was also carried by small particles (called **electrons**). On a fine enough scale, all of nature seemed to be 'grainy'. It seemed reasonable that these electrons were a part of the atoms, and therefore electricity was a fundamental property of all matter. Because light was also intimately connected with electrical effects, the source of all light waves must be electrons oscillating inside atoms.

But just as matter and electricity was 'grainy', experiments seemed to be pointing to the conclusion that energy was also. The first to realize this was one of Helmholtz's ex-students, Max Planck, in 1900. He proposed calling these 'grains' of energy, **quanta**; from which the whole subject came to be known as the **quantum theory**. Within five years, (in 1905, the same year in which he published his work on relativity) Einstein showed that Planck's hypothesis would be perfectly understandable if it was assumed that light were made up of particles (**photons**), just as Newton had said. But this really created a paradox, because Young's results were still