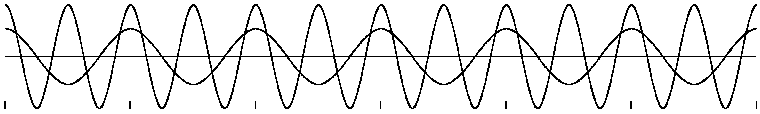
Introduction

What is it about intervals such as an octave and a perfect fifth that makes them more consonant than other intervals? Is this cultural, or inherent in the nature of things? Does it have to be this way, or is it imaginable that we could find a perfect octave dissonant and an octave plus a little bit consonant?

The answers to these questions are not obvious, and the literature on the subject is littered with misconceptions. One appealing and popular, but incorrect, explanation is due to Galileo Galilei, and has to do with periodicity. The argument goes that if we draw two sine waves an exact octave apart, one has exactly twice the frequency of the other, so their sum will still have a regularly repeating pattern



whereas a frequency ratio slightly different from this will have a constantly changing pattern, so that the ear is 'kept in perpetual torment'.

Unfortunately, it is easy to demonstrate that this explanation cannot be correct. For pure sine waves, the ear detects nothing special about a pair of signals exactly an octave apart, and a mistuned octave does not sound unpleasant. Interval recognition among trained musicians is a factor being deliberately ignored here. On the other hand, a pair of pure sine waves whose frequencies only differ slightly give rise to an unpleasant sound. Moreover, it is possible to synthesize musical sounding tones for which the exact octave sounds unpleasant, while an interval of slightly more than an octave sounds pleasant. This is done by stretching the spectrum from what would be produced by a natural instrument. These experiments are described in Chapter 4.

The origin of the consonance of the octave turns out to be the instruments we play. Stringed and wind instruments naturally produce a sound that consists of exact integer multiples of a fundamental frequency. If our instruments were different, our musical scale would no longer be appropriate. For example, in the Indonesian gamelan, the instruments are all percussive. Percussive instruments do not produce exact integer multiples of a fundamental, for reasons explained in Chapter 3. So the western scale is inappropriate, and indeed not used, for gamelan music.

We begin the first chapter with another fundamental question that needs sorting out before we can properly get as far as a discussion of consonance and dissonance. Namely, what's so special about sine waves anyway, that we consider them to be the 'pure' sound of a given frequency? Could we take some other periodically varying wave and define it to be the pure sound of this frequency?

The answer to this has to do with the way the human ear works. First, the mathematical property of a pure sine wave that's relevant is that it is the general solution to the second order differential equation for simple harmonic motion. Any object that is subject to a returning force proportional to its displacement from a given location vibrates as a sine wave. The frequency is determined by the constant of proportionality. The basilar membrane inside the cochlea in the ear is elastic, so any given point can be described by this second order differential equation, with a constant of proportionality that depends on the location along the membrane.

The result is that the ear acts as a harmonic analyzer. If an incoming sound can be represented as a sum of certain sine waves, then the corresponding points on the basilar membrane will vibrate, and that will be translated into a stimulus sent to the brain.

This focuses our attention on a second important question. To what extent can sound be broken down into sine waves? Or to put it another way, how is it that a string can vibrate with several different frequencies at once? The mathematical subject that answers this question is called Fourier analysis, and is the subject of Chapter 2. The version of the theory in which periodic sounds are decomposed as a sum of integer multiples of a given frequency is the theory of *Fourier series*. Decomposing more general, possibly nonperiodic sounds gives rise to a continuous frequency spectrum, and this leads to the more difficult theory of *Fourier integrals*. In order to accommodate discrete spectra into the theory of Fourier integrals, we need to talk about *distributions* rather than functions, so that the frequency spectrum of a sound is allowed to have a positive amount of energy concentrated at a single frequency.

Chapter 3 describes the mathematics associated with musical instruments. This is done in terms of the Fourier theory developed in Chapter 2, but it is really only necessary to have the vaguest of understanding of Fourier theory for this purpose. It is certainly not necessary to have worked through the whole of Chapter 2. For the discussion of drums and gongs, where the answer does not give integer multiples of a fundamental frequency, the discussion depends on the theory of Bessel functions, which is also developed in Chapter 2.

Chapter 4 is where the theory of consonance and dissonance is discussed. This is used as a preparation for the discussion of scales and temperaments in Chapters 5 and 6. The fundamental question here is: why does the modern western scale consist of twelve equally spaced notes to an octave? Where does the twelve come from? Has it always been this way? Are there other possibilities?

The emphasis in these chapters is on the relationship between rational numbers and musical intervals. We concentrate on the development of the standard Western scales, from the Pythagorean scale through just intonation, the meantone scale, and the irregular temperaments of the sixteenth to nineteenth centuries until finally we reach the modern equal tempered scale.

We also discuss a number of other scales such as the 31 tone equal temperament that gives a meantone scale with arbitrary modulation. There are even some scales not based on the octave, such as the Bohlen–Pierce scale based on odd harmonics only, and the scales of Wendy Carlos.

These discussions of scale lead us into the realm of continued fractions, which give good rational approximations to numbers such as $\log_2(3)$ and $\log_2(\sqrt[4]{5})$.

After our discussion of scales, we break off our main thread to consider a couple of other subjects where mathematics is involved in music. The first of these is computers and digital music. In Chapter 7 we discuss how to represent sound and music as a sequence of zeros and ones, and again we find that we are obliged to use Fourier theory to understand the result. So, for example, Nyquist's theorem tells us that a given sample rate can only represent sounds whose spectrum stops at half that frequency. We describe the closely related z-transform for representing digital sounds, and then use this to discuss signal processing, as a method both of manipulating sounds and of producing them.

This leads us into a discussion of digital synthesizers in Chapter 8, where we find that we are again confronted with the question of what it is that makes musical instruments sound the way they do. We discover that most interesting sounds do not have a static frequency spectrum, so we have to understand the evolution of spectrum with time. It turns out that for many sounds, the first small fraction of a second contains the critical clues for identifying the sound, while the steadier part of the sound is less important. We base our discussion around FM synthesis; although this is an old-fashioned way to synthesize sounds, it is simple enough to be able to understand a lot of the salient features before taking on more complex methods of synthesis.

In Chapter 9 we change the subject almost completely, and look into the role of symmetry in music. Our discussion here is at a fairly low level, and one could write many books on this subject alone. The area of mathematics concerned with symmetry is *group theory*, and we introduce the reader to some of the elementary ideas from group theory that can be applied to music.

I should close with a disclaimer. Music is not mathematics. While we're discussing mathematical aspects of music, we should not lose sight of the evocative power of music as a medium of expression for moods and emotions. About the numerous interesting questions this raises, mathematics has little to say.

Why do rhythms and melodies, which are composed of sound, resemble the feelings, while this is not the case for tastes, colors or smells? Can it be because they are motions, as actions are also motions? Energy itself belongs to feeling and creates feeling. But tastes and colours do not act in the same way.

(Aristotle, Prob. *xix. 29)*

1

Waves and harmonics

1.1 What is sound?

The medium for the transmission of music is sound. A proper understanding of music entails at least an elementary understanding of the nature of sound and how we perceive it.

Sound consists of vibrations of the air. To understand sound properly, we must first have a good mental picture of what air looks like. Air is a gas, which means that the atoms and molecules of the air are not in such close proximity to each other as they are in a solid or a liquid. So why don't air molecules just fall down on the ground? After all, Galileo's experiment at the leaning tower of Pisa tells us that objects should fall to the ground with equal acceleration independently of their size and mass.

The answer lies in the extremely rapid motion of these atoms and molecules. The mean velocity of air molecules at room temperature under normal conditions is around 450–500 meters per second (or somewhat over 1000 miles per hour), which is considerably faster than an express train at full speed. We don't feel the collisions with our skin, only because each air molecule is extremely light, but the combined effect on our skin is the air pressure which prevents us from exploding!

The mean free path of an air molecule is 6×10^{-8} meters. This means that on average, an air molecule travels this distance before colliding with another air molecule. The collisions between air molecules are perfectly elastic, so this does not slow them down.

We can now calculate how often a given air molecule is colliding. The collision frequency is given by

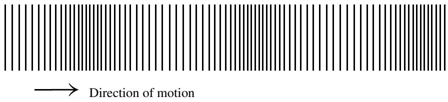
$$\text{collision frequency} = \frac{\text{mean velocity}}{\text{mean free path}} \sim 10^{10} \text{ collisions per second.}$$

So now we have a very good mental picture of why the air molecules don't fall down. They don't get very far down before being bounced back up again. The effect of gravity is then observable just as a gradation of air pressure, so that if we go up to a high elevation, the air pressure is noticeably lower.

So air consists of a large number of molecules in close proximity, continually bouncing off each other to produce what is perceived as air pressure. When an object vibrates, it causes waves of increased and decreased pressure in the air. These waves are perceived by the ear as sound, in a manner to be investigated in the next section, but first we examine the nature of the waves themselves.

Sound travels through the air at about 340 meters per second (or 760 miles per hour). This does not mean that any particular molecule of air is moving in the direction of the wave at this speed (see above), but rather that the local disturbance to the pressure propagates at this speed. This is similar to what is happening on the surface of the sea when a wave moves through it; no particular piece of water moves along with the wave, it is just that the disturbance in the surface is propagating.

There is one big difference between sound waves and water waves, though. In the case of the water waves, the local movements involved in the wave are up and down, which is at right angles to the direction of propagation of the wave. Such waves are called *transverse waves*. Electromagnetic waves are also transverse. In the case of sound, on the other hand, the motions involved in the wave are in the same direction as the propagation. Waves with this property are called *longitudinal waves*.



Longitudinal waves

Sound waves have four main attributes which affect the way they are perceived. The first is *amplitude*, which means the size of the vibration, and is perceived as loudness. The amplitude of a typical everyday sound is very minute in terms of physical displacement, usually only a small fraction of a millimeter. The second attribute is *pitch*, which should at first be thought of as corresponding to frequency of vibration. The third is *timbre*, which corresponds to the shape of the frequency spectrum of the sound (see Sections 1.7 and 2.15). The fourth is *duration*, which means the length of time for which the note sounds.

These notions need to be modified for a number of reasons. The first is that most vibrations do not consist of a single frequency, and naming a 'defining' frequency can be difficult. The second related issue is that these attributes should really be defined in terms of the perception of the sound, and not in terms of the sound itself. So, for example, the perceived pitch of a sound can represent a frequency not actually present in the waveform. This phenomenon is called the 'missing fundamental', and is part of a subject called psychoacoustics.

Attributes of sound

Physical	Perceptual
Amplitude	Loudness
Frequency	Pitch
Spectrum	Timbre
Duration	Length

Further reading

Harvey Fletcher, Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure, *J. Acoust. Soc. Amer.* **6** (2) (1934), 59–69.

1.2 The human ear

In order to get much further with understanding sound, we need to study its perception by the human ear. This is the topic of this section. I have borrowed extensively from Gray's *Anatomy* for this description.

The ear is divided into three parts, called the outer ear, the middle ear or *tympanum* and the inner ear or *labyrinth*. See Figure 1.1. The outer ear is the visible part on the outside of the head, called the *pinna* (plural *pinnæ*) or *auricle*, and is ovoid in form. The hollow middle part, or *concha*, is associated with focusing and thereby magnifying the sound, while the outer rim, or *helix*, appears to be associated with vertical spatial separation, so that we can judge the height of a source of sound.

The concha channels the sound into the auditory canal, called the *meatus auditorius externus* (or just *meatus*). This is an air filled tube, about 2.7 cm long and

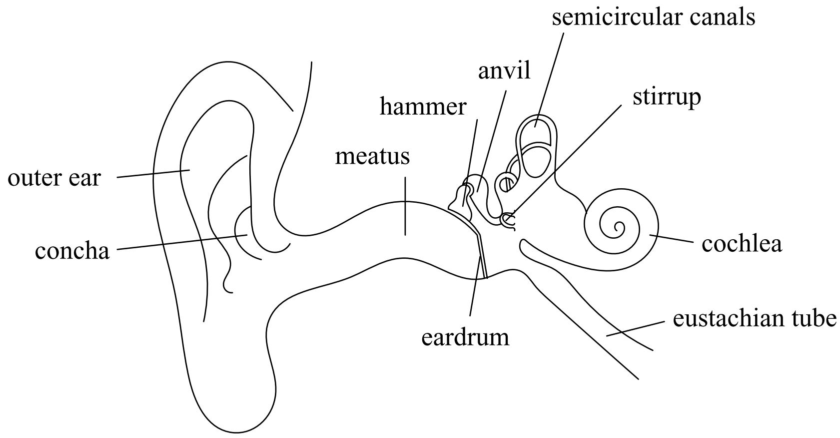


Figure 1.1 The human ear.

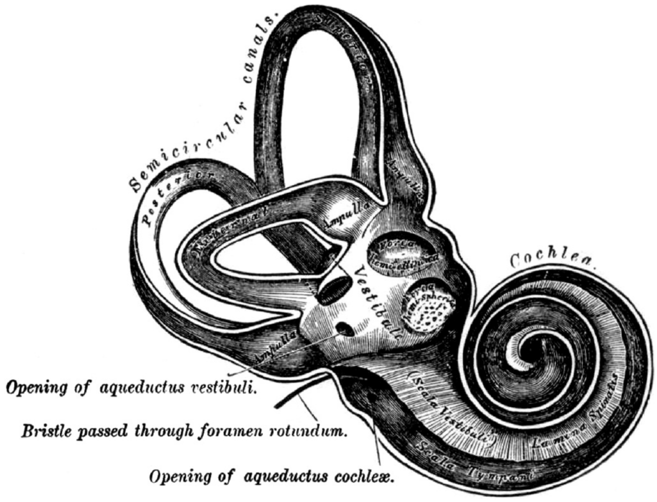


Figure 1.2 The osseous labyrinth laid open. (Enlarged.) From Gray (1901).

0.7 cm in diameter. At the inner end of the meatus is the ear drum, or *tympanic membrane*.

The ear drum divides the outer ear from the middle ear, or *tympanum*, which is also filled with air. The tympanum is connected to three very small bones (the *ossicular chain*) which transmit the movement of the ear drum to the inner ear. The three bones are the hammer, or *malleus*, the anvil, or *incus*, and the stirrup, or *stapes*. These three bones form a system of levers connecting the ear drum to a membrane covering a small opening in the inner ear. The membrane is called the *oval window*.

The inner ear, or *labyrinth*, consists of two parts, the *osseous labyrinth*,[1] see Figure 1.2, consisting of cavities hollowed out from the substance of the bone, and the *membranous labyrinth*, contained in it. The osseous labyrinth is filled with various fluids, and has three parts, the *vestibule*, the *semicircular canals* and the *cochlea*. The vestibule is the central cavity which connects the other two parts and which is situated on the inner side of the tympanum. The semicircular canals lie above and behind the vestibule, and play a role in our sense of balance. The cochlea is at the front end of the vestibule, and resembles a common snail shell in shape. See Figure 1.3. The purpose of the cochlea is to separate out sound into various frequency components (the meaning of this will be made clearer in Chapter 2) before passing it onto the nerve pathways. It is the functioning of the cochlea which is of most interest in terms of the harmonic content of a single musical note, so let us look at the cochlea in more detail.

[1] Illustrations taken from the 1901 edition of Henry Gray, F.R.S. *Anatomy, Descriptive and Surgical*, reprinted by Running Press, 1974.

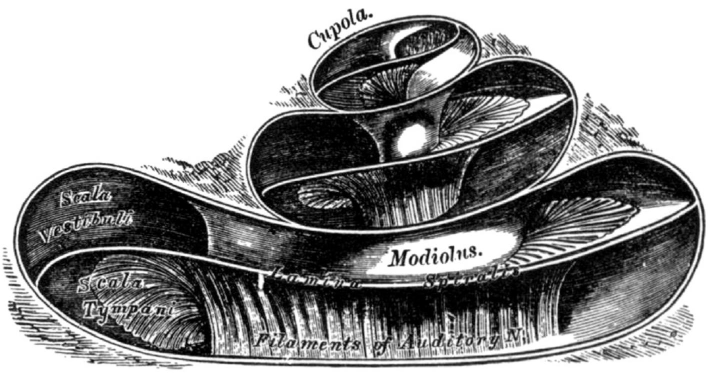


Figure 1.3 The cochlea laid open. (Enlarged.) From Gray (1901).

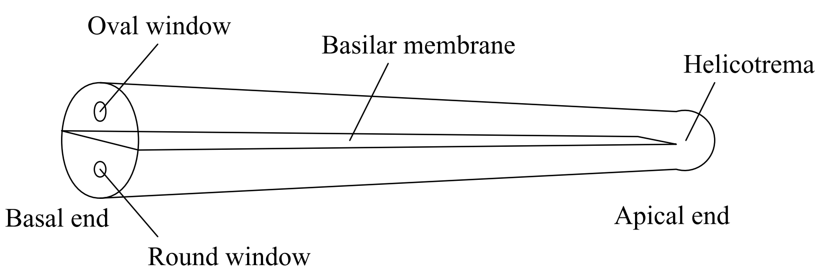


Figure 1.4 The cochlea, uncoiled.

The cochlea twists roughly two and three quarter times from the outside to the inside, around a central axis called the *modiolus* or *columnella*. If it could be unrolled, it would form a tapering conical tube roughly 30 mm (a little over an inch) in length. See Figure 1.4.

At the wide (*basal*) end where it meets the rest of the inner ear it is about 9 mm (somewhat under half an inch) in diameter, and at the narrow (*apical*) end it is about 3 mm (about a fifth of an inch) in diameter. There is a bony shelf or ledge called the *lamina spiralis ossea* projecting from the modiolus, which follows the windings to encompass the length of the cochlea. A second bony shelf called the *lamina spiralis secundaria* projects inwards from the outer wall. Attached to these shelves is a membrane called the *membrana basilaris* or *basilar membrane*. This tapers in the opposite direction to the cochlea (see Figure 1.5), and the bony shelves take up the remaining space.

The basilar membrane divides the interior of the cochlea into two parts with approximately semicircular cross-section. The upper part is called the *scala vestibuli* and the lower is called the *scala tympani*. There is a small opening called the *helicotrema* at the apical end of the basilar membrane, which enables the two parts

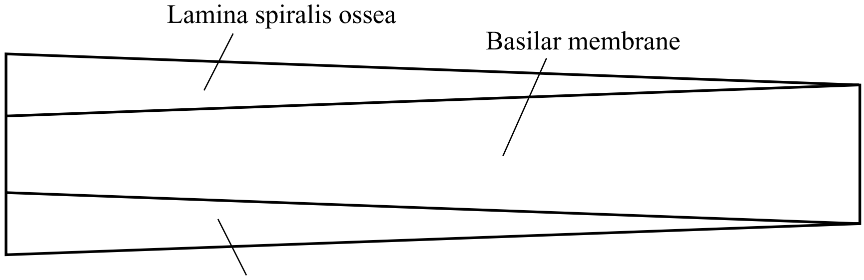


Figure 1.5 The basilar membrane.

to communicate with each other. At the basal end there are two windows allowing communication of the two parts with the vestibule. Each window is covered with a thin flexible membrane. The stapes is connected to the membrane called the *membrana tympani secundaria* covering the upper window; this window is called the *fenestra ovalis* or *oval window*, and has an area of 2.0–3.7 mm^2. The lower window is called the *fenestra rotunda* or *round window*, with an area of around 2 mm^2, and the membrane covering it is not connected to anything apart from the window. There are small hair cells along the basilar membrane which are connected with numerous nerve endings for the auditory nerves. These transmit information to the brain via a complex system of neural pathways. The hair cells come in four rows, and form the *organ of Corti* on the basilar membrane.

Now consider what happens when a sound wave reaches the ear. The sound wave is focused into the meatus, where it vibrates the ear drum. This causes the hammer, anvil and stapes to move as a system of levers, and so the stapes alternately pushes and pulls the membrana tympani secundaria in rapid succession. This causes fluid waves to flow back and forth round the length of the cochlea, in opposite directions in the scala vestibuli and the scala tympani, and causes the basilar membrane to move up and down.

Let us examine what happens when a pure sine wave is transmitted by the stapes to the fluid inside the cochlea. The speed of the wave of fluid in the cochlea at any particular point depends not only on the frequency of the vibration but also on the area of cross-section of the cochlea at that point, as well as the stiffness and density of the basilar membrane. For a given frequency, the speed of travel decreases towards the apical end, and falls to almost zero at the point where the narrowness causes a wave of that frequency to be too hard to maintain. Just to the wide side of that point, the basilar membrane will have to have a peak of amplitude of vibration in order to absorb the motion. Exactly where that peak occurs depends on the frequency. So by examining which hairs are sending the neural signals to the brain, we can ascertain the frequency of the incoming sine wave.