Physics of Music Lecture Notes

Instructor: Guy D. Moore

July 14, 2006

Preface

These are the lecture notes provided for Physics 224, The Physics of Music, taught at McGill University in Fall 2006. They were written by the instructor in fall 2005 and represent a fairly accurate guide of what will be covered in lecture. The notes are broken into 29 lectures; in most cases each corresponds to what will be covered in a single 1-hour lecture, but some are intended to run over two or more lectures on the same topic. A detailed schedule of which lecture will be used when will be provided on the course webpage, which will also provide all problems and some other course materials, such as sound files which accompany some of the lectures.

While you will only be held accountable in this course for material presented in these lectures and in the homeworks, I still highly recommend attendance at the lectures, since the presentation there will not be exactly the same as the notes, there will be demonstrations which are sometimes quite valuable to understanding (in a course about sound, you can really experience some of the demos), and you can get your questions answered in class.

The expected background knowledge for these notes is very limited. I try not to assume that you know musical notation (even scales) and I do not assume calculus. I do assume basic trigonometry and algebra, as well as unit analysis. Sometimes these notes provide supplementary material for interested people with more advanced background; for instance, there are asides for people with some physics background, more music background, more math background, and so forth. These are enclosed in square brackets [like this]. When you see these, it means that you do not have to read if you don't have the prescribed background and you should not worry if you don't understand–these asides are provided because I find that it can be very helpful for people who -do- have the added background to be presented with another way of thinking about certain issues.

Contents

Chapter 1

Introduction

[[Banging sound of a popping mini-balloon]]

What happened?

In about 1/1000 of a second, a bubble about 0.5 cm across popped. That released about 0.2 milligrams of air, to expand into the air around it. The energy released was about 0.02 Joules, about the amount released when a paper clip hits the ground, if dropped from head height. (Alternately, your body releases 5000 times that much heat every second.)

- The air in the bubble did not cross the room.
- The air from the bubble did not even cross the table.
- The air at your ear moved back and forth about 5 microns (5 μm), which is 1/10 the width of a human hair.

Yet everyone in the room turned around!

This is the phenomenon of sound.

As we will learn, the air in the bubble pushed the air next to it. This air pushed the air next to it, and so on. A "ripple" phenomenon went across the room, and that is what people responded to.

People far from the bubble actually received more sound from reflections off the wall, floor, ceiling, people, and so forth, than they received directly from the bubble. Nevertheless, no one had any trouble telling where the sound came from. This reflects the remarkable properties of your ear.

This course has three main goals:

1. To study sound;
    - what it is,

- how it is characterized,
- how it travels.

2. To study the ear;
   - its anatomy,
   - how it works,
   - how the ear-brain system determines (some of) our notions of music.

3. To study musical instruments;
   - how they work,
   - why their sounds are "musical,"
   - why they sound different from each other.

Some definitions we will encounter:

Physics: The study of nature using the scientific method and mathematical tools.

Psychophysics: The study of the physical basis of perception and the functioning of the nervous system (though we will concentrate on the physical basis of hearing and will not go into the nervous system very much).

Let me answer ahead, some questions you may have about the course.

Will this be a physics course?

Yes – but – I will *not* assume more than high school background in physics or mathematics (though you may find it helpful if you have more). In particular, I *will* assume you know algebra and have seen logarithms, though you may be rusty with them. I will *sort of* assume trigonometry. I will *not* assume calculus or Freshman physics. **Don't panic.**

Will this be an anatomy / physiology course?

No - I am not qualified to teach one – but – we will discuss anatomy of the ear in some detail, and of the throat in rather less detail, and particularly the ear anatomy will be important in the course.

Will this be a music course?

No – but – many musical concepts will come up, and a good background in Western musical theory and a familiarity with some musical instruments will come in handy. In particular I will *almost* assume a familiarity with rhythmical marking, and will *hope for* familiarity with musical scales and musical intervals.

Will this course be "hard"?

I don't think so, but I am biased. It may depend on your background. It is not intended to be a hard course, though it is also not intended to be an easy "A". I expect everyone to learn something new, though, including myself!

What is the instructor's background?

I have an advanced formal education in physics. In music, I played the 'cello through college and sat 3'rd, 2'nd, and 1'st in the Pomona College Symphony Orchestra, but I am now hopelessly out of practice. I have some familiarity with musical theory (I know what a Dominant VII chord is, but not what a French VI chord is). I have no real expertise in human anatomy or neurology. I have been at McGill since September 2002, and have taught this course since September 2003.

Chapter 2

Air and Air Pressure

The purpose of this lecture is to review or introduce the properties of the air, air pressure, and some basic physics concepts which we will need repeatedly in this course.

On the large scale, air is described by

- temperature
- density
- velocity
- pressure

It might help to think about these, to see how they arise from the microscopic description in terms of individual air molecules.

Air is made up of a vast number of tiny, freely flying atoms and molecules. These are

- Tiny–typically $3 \times 10^{-10} = 0.0000000003$ meters across, 1/200000 the width of a hair.
- Fast moving–each in a random direction, typically around 450 m/s = 1600 km/h, twice as fast as an airplane and nearly as fast as a bullet.
- Packed together loosely, around $3 \times 10^{-9} = 0.000000003$ meters apart from each other.
- Huge in number. 1 liter of air has about 3×10^{22} molecules in it.

Because they are smaller than their separation, they fly a (comparatively) long ways between collisions, roughly $10^{-7} = 0.0000001$ meters. But that is still tiny, only 1/500 of the thickness of a hair. That means that each separate air molecule "runs back and forth" very fast but never gets anywhere.

Since there are so many air molecules, what we feel is the total effect of a huge number, that is, an average. The relation between the properties of the air molecules and the large scale quantities we observe is as follows:

TEMPERATURE: Temperature corresponds to how fast the air molecules are moving; basically it measures how much energy each air molecule is carrying. When the temperature in Kelvin is doubled, the typical speed of an air molecule is increased by $\sqrt{2} = 1.41$, so the energy $mv^2/2$ is doubled. Hot air means air with faster moving molecules.

DENSITY: This is the mass of the air per unit volume, that is, how many molecules there are in each cubic centimeter of air, times the mass of a molecule. Air does weigh something, but because the molecules are so widely spaced from each other, it is very light compared to materials we are used to:

(material)	g/cm^3	ton/m^3	pound/foot^3
iron	7.8	7.8	490
rock	~ 2.5	~ 2.5	~ 160
water	1	1	62.5
wood	~ 0.6	~ 0.6	~ 37
air	0.00120	0.00120	0.0750

VELOCITY: By this I mean, wind–the air moving, in the normal sense we experience it. Each molecule is moving at around 500 m/s, each in a different direction. If we average the speeds, we get some huge number like 450 m/s. But average the velocities–that is, speed *and direction*–and you typically find the air is at rest, or maybe moving at some slow speed. In other words, in air at rest, for each molecule going at 400 m/s to the right, there is one going at 400 m/s to the left. If the air is moving to the right, it means that slighly more of the molecules are moving to the right than to the left. For instance, if the air is moving at 10 m/s, it means that for each molecule going to the left at 390 m/s, there is one going to the right at 410 m/s. As the air molecules bounce off each other, each individually changes direction and speed, but the average stays the same.

Conversion: to get from m/s to mi/hr, multiply by 2.23.

To get from m/s to km/hr, multiply by 3.6.

PRESSURE: This is the most important concept of this lecture, so I will explain it slowly.

First, recall some physics concepts:

- Inertia: "resistance to moving," same as mass. Heavy things are harder to get moving than light things. Measured in kg.
- Velocity: how fast something is moving, distance/time. Measured in m/s. INCLUDES the direction the thing is moving in [it is a vector].

- Force: "what makes things move," measured in Newtons, N = kg m/s^2. One Newton gets one kilogram to move at 1m/s in 1 second (hence, seconds squared in the denominator).
- Acceleration: how fast velocity is changing, measured in (m/s)/s or m/s^2. You are accelerating when you speed up, when you slow down, *and* when you go around a curve (because your direction is changing).

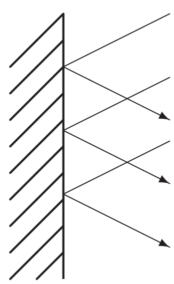
Two of Newton's laws will be very important to us through the course. The first is the force law,

$$F = ma$$

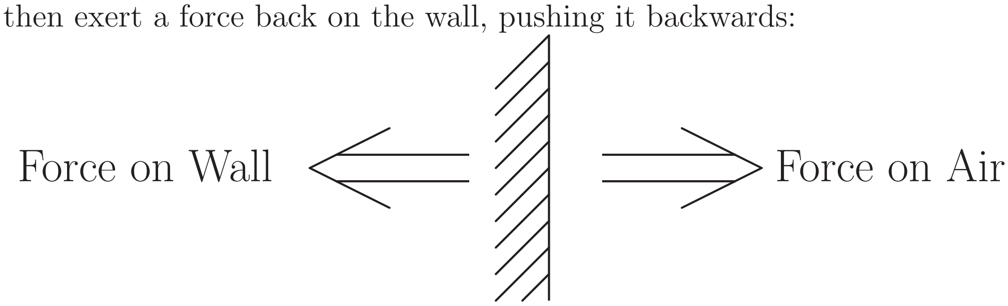
which tells how a force makes something accelerate: F is the size of the force (say, in Newtons), m is the mass (in kg), and a is the acceleration (in m/s^2).

The second is Newton's third law, which basically says, if I push on you with force F, you must be pushing on me with force $-F$ (same size, opposite direction).

Now think about what happens when air meets a wall. The air molecules which are moving at the wall will bounce off:



(The shaded thing is the usual way to indicating a solid object.) Obviously the air molecules bouncing off the wall must have a force act on them, pointing out from the wall. They must then exert a force back on the wall, pushing it backwards:



The bigger the area of wall, the more molecules will bounce off it. Therefore the force is proportional to the area of wall (or whatever). When this is the case, it is convenient to define

pressure P

which is the force per area,

$$F = PA$$

with P the pressure and A the area. The MKS unit of pressure is the Newton/m^2, or kg/ms^2, called the Pascal. One Pascal is tiny: the pressure a piece of paper exerts on a table, because of its weight, is about 0.7 Pascal.

The more molecules in the air, the bigger the pressure, because there will be more molecules to bounce off the wall or solid object. Higher temperature also increases pressure, since the molecules are moving faster–and therefore hit harder and more often. Therefore pressure obeys

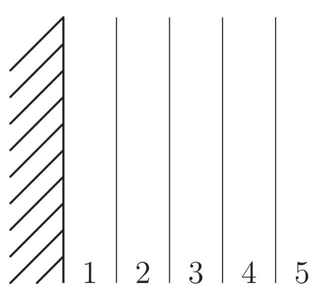
$$P = nRT$$

with n the density of molecules (number/m^3) and T the temperature, and R some boring constant. Just remember that denser, or hotter, air has a higher pressure.

The pressure exerted by typical air is about 101,000 Pascal, or the force per area exerted by about 150,000 pages thick of paper. This enormous number comes about because, even though the air is very light, the molecules are moving with such enormous speed. (100,000 Pascal is called a "Bar.")

The reason the air does not push us around is because it exerts force from both sides–our fronts are pushed backward, our backs are pushed forward, so we balance out. Our ribs don't collapse because the air in our lungs pushes out with the same force as the air pushing in on the ribcage. The pressure inside our bodies pushing out, balances the pressure of the air. If the air were suddenly removed from one side of your body, the air on the other side would push you up to a remarkable speed very quickly, like someone in a space movie when they open the airlock. The air pushes down on the ground with the equivalent of about 10 tons/m^2 weight, or 14.7 pounds/square inch. The atmosphere is being pushed up by the same sized pressure. The reason it doesn't fly off is that this is exactly the weight per unit area of the atmosphere. (Weight means mass times the acceleration of gravity $g = 9.8$ m/s. A kilogram of weight means 9.8 kgm/s^2 or 9.8 Newtons of force.)

The last comment in this lecture: why does air stand still? Because it is being pushed with the same force from all sides. Imagine cutting the air into thin slices and ask what happens to each slice:



Region 1 is getting pushed with pressure P, and force PA, by the wall. But it is getting pushed with force $-PA$ by region 2 The total force, which tells us whether the region will accelerate, is $PA - PA - 0$, so it stays at rest. Region 2 is getting pushed with force PA by region 1, and force $-PA$ by region 3. Again, the forces balance and it stays at rest. Region 3 is getting pushed with force PA by region 2, and force $-PA$ by region 4 . . . .

Chapter 3

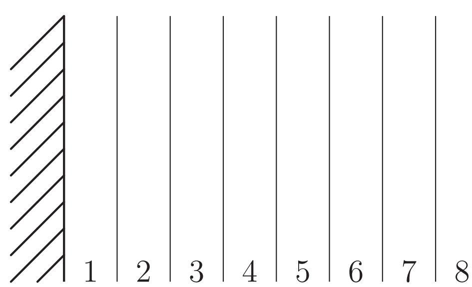
From Pressure to Sound

The main goal of this lecture is to explain what a sound wave, in air, "is," by thinking about pressure and velocity in the air.

At times in this and future lectures I will make asides intended for some students with more background knowledge in a particular area. I will always put these in square brackets [], beginning with a note of who might find it useful, eg, [musicians: this interval is a perfect fifth]. In this lecture there will be a long aside for chemists and physicists.

I will work with the example of the little "siren" I demonstrate in class. A disk with holes in it spins in front of a disk with matching holes. The holes alternately line up and don't line up. A source (my mouth, for the one shown in class) presents compressed air on one side. We want to understand what happens when the air on the other side suddenly feels the higher pressure of the compressed air. To understand what is happening, mentally divide the air into a series of layers. Remember that air molecules only travel a tiny distance between scatterings, so molecules generally stay within their layer.

The starting picture looks like this:



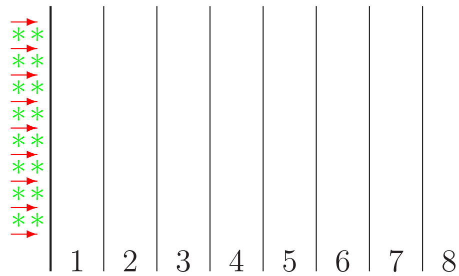
Recall that

1. Denser air has higher pressure, that is, it pushes harder.

2. When the forces on an object do not balance, the object accelerates.

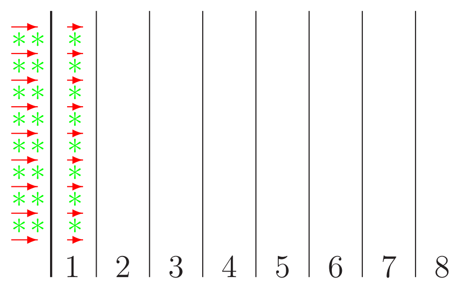
3. Once something is moving, it *keeps* moving until forces on it cause it to stop.

Now consider what happens when the wall is suddenly replaced by compressed, forwards moving air:

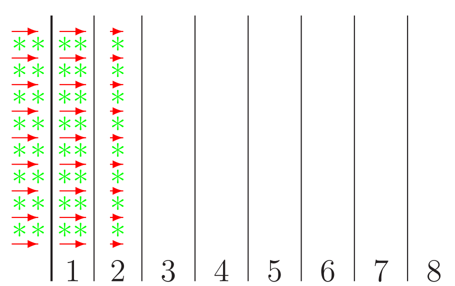


The green * are supposed to indicate that this air is compressed, that is, under higher pressure. The red arrows indicate what direction air is moving.

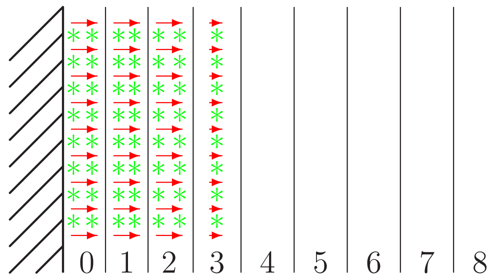
What happens to region 1? There is a larger pressure behind it than in front of it. Therefore the forces do not balance, and so it will start to move forward. Since the air behind it is moving forward, it will also get squeezed into a smaller space. Therefore, a moment later, the situation will look like this:



The first layer of air has become compressed, and has started to move forward. Since it is now compressed, it pushes harder on the second layer, than it did before. Since it is moving forward, it compresses the second layer. Therefore, a moment later the situation will look like,

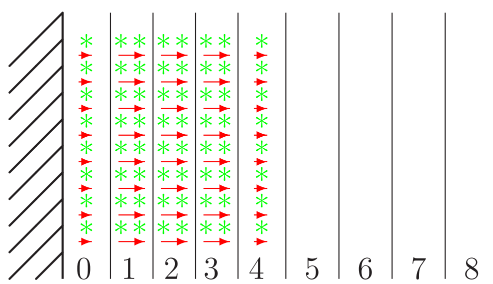


Now suppose a moment later that the barrier re-appears. Since the air moved forward, I will put it behind the new layer that got brought in. The situation looks like this:

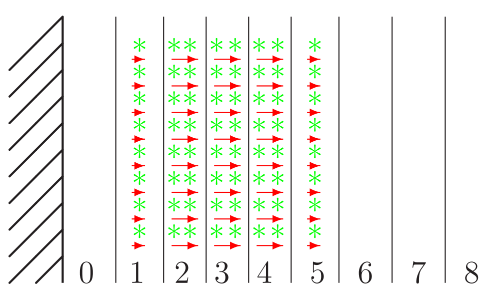


Notice that the air has only moved forward a bit–but the region where the air is compressed, is pretty big now.

Next what? The air in region 1 is feeling extra pressure in front but also behind. The pressures balance on it. But once something is moving forward, it keeps moving forward. Similarly, regions 0 and 2 keep moving forward, and regions 3 and 4 are being pushed forward and move faster. Since the wall is back in place, region 0 will now be getting stretched out, which means its pressure will fall. So we will get,



Now regions 4 and 5 will get compressed and pushed forward, but regions 0 and 1 feel more pressure on their fronts than their backs, and will keep moving forward (and hence, will de-compress) and will be slowed down. Therefore we will get,



From now on, the compressed region has nothing to do with the presence of a wall behind it. The front of the region sees a higher pressure behind than in front; so it speeds up. The back region sees a higher pressure behind than in front; so it slows down. Since the middle is moving, the front gets compressed and the back gets de-compressed. That is just the right thing to keep moving the pattern forward.

This phenomenon is called a sound wave.

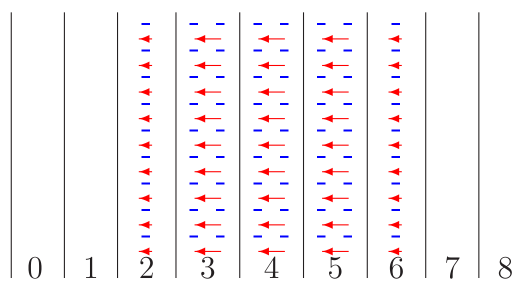
A *wave* is any phenomenon like this, where a disturbance–a phenomenon–moves forward, but the things which are causing the disturbance move very little. An ocean wave is the same thing. The wave is the phenomenon that there is a spot where the sea water is high and moving forward. If you swim in the sea, the wave carries you forward and the trough between waves carries you backwards. You, and the water, end up in the same place; but the wave certainly goes forward.

Another analogy is when you are waiting in a line. Someone at the front of the line gets to go to the counter. The person behind moves up, the person behind them moves up, so forth. The "moving up" is the wave. Each person moves a short distance, but the "moving up" phenomenon quickly goes the whole length of the line.

Let's clear up some easy to make mistakes about waves:

- The speed that the air moves, and the speed that the wave moves, need not be related and are usually very different, just as you may shuffle very slowly in the line analogy above, but the spot in the line where people are shuffling may move much faster.

- The direction the air moves, and the direction the wave moves, need not be the same. Think about what would happen here (blue dashes - represent pressure below normal):



  The regions in the middle are stretched out and moving backwards. The area in front, 6 and 7, feel *less* pressure behind and more in front, so they are accelerated backwards and start to move backwards. They are also stretched out, since the area behind them is moving away, and therefore become low pressure. At the other end, 2 and 3 are being pushed forward and squeezed, and so they regain normal pressure and stop moving. So a moment later, 7 will look like 6 looks now, 6 like 5 looks now, ... That means that the wave will move to the *right*, even though the motion of the air is to the *left*.

  What is true is that in a sound wave or part of a sound wave where pressure is higher, the wave and air move in the same direction. Where the pressure is low, the wave and air move in opposite directions.

How fast is a sound wave? That depends on two things:

- How hard does the air push or get pushed? The harder the air can push, that is, the higher the pressure, the faster the air can be made to move, and the faster the wave will propagate.

- How heavy is the air? That is, what is its density? The heavier the air, the more pushing it takes to get it to move, and he slower the wave will propagate.

From this argument, and just looking at the units, one can guess that the speed of sound is

$$v_{\rm sound} = C\sqrt{\frac{\rm Pressure}{\rm Density}} = C\sqrt{\frac{P}{\rho}}$$

(pressure and density are usually written with the letters P and ρ.) Here C is some constant which the argument I just gave is not quite enough to determine, and which turns out to be nearly but not quite 1. [Scientists: $C = \sqrt{c_P/c_V}$ with c_P and c_V being the fixed pressure and fixed temperature specific heats. These appear because the speed of sound is not $\sqrt{P/\rho}$, but actually $\sqrt{\partial P/\partial\rho}$ and the derivatives should be performed under adiabatic, not isothermal, conditions.]

Recall that the pressure of air is very large but the denisty is very small. That means $v_{\rm sound}$ will be large. It turns out that, at room temperature and for normal air,

$$v_{\rm sound} \simeq 344 \text{ m/s}\,.$$

If you prefer, this is 770 miles/hour or 1240 km/hour, and is also defined to be "Mach 1". The Mach number is just how many times faster than sound something is going; "Mach 3" means 3 times the speed of sound.

The speed of sound is about 1 km per 3 seconds or 1 mile per 5 seconds, hence the rule for lightning followed by thunder,

Five seconds, one mile.

Three seconds, one kilometer.

Recall that $P = nRT$. The density is $\rho = mn$, with m the mass of one molecule. That means that

$$v_{\rm sound} = C\sqrt{\frac{RT}{m}}$$

where R is another constant, as you recall. That means that

- Hot air has faster sound speed, cold air has a slower sound speed. (This causes problems in wind instruments, where pitch goes as sound speed.)

- High and low pressure air have the *same* sound speed, because the density of molecules n doesn't matter.

- Gas made of light atoms or molecules, like Helium, has a fast sound speed. Gas made of heavy atoms has a slow sound speed. Hence, people sound funny when their lungs are filled with Helium.

In a normal sound wave, the speed of the air and distance the air moves are tiny. For instance, at the threshold of hearing, the air speed is only 70 nm/s (nanometer/second), or .00000007 m/s. At the threshold of pain, the air moves a whopping 7 cm/s or .07 m/s. Similarly, the pressure changes only a tiny amount from the normal pressure in the environment (without the sound).

Sound can also occur in other materials besides air. For solids and liquids, since the atoms are pressed against each other, a compression leads to a much larger pressure than in a gas. Therefore the speed of sound is typically much larger in solid materials. For instance,

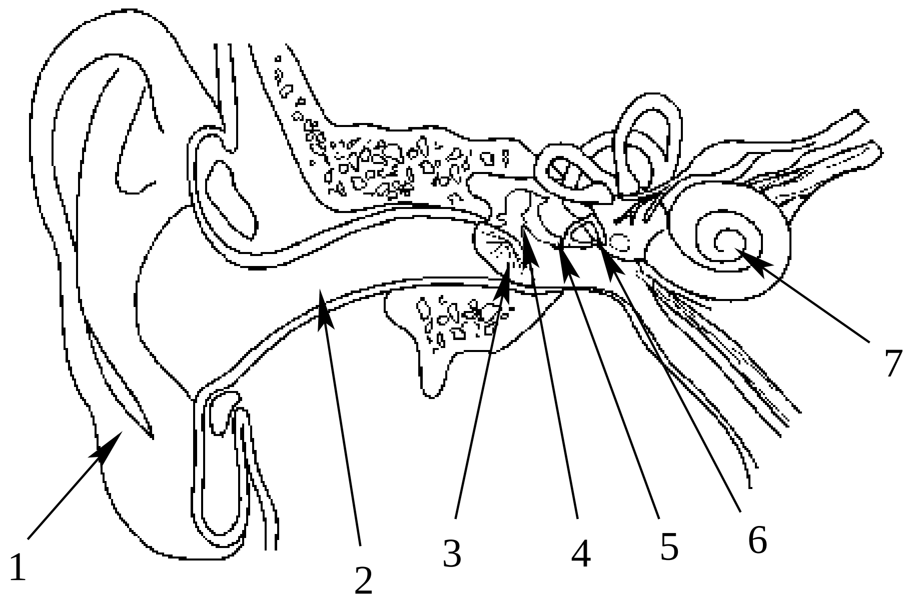
- v_{sound} in water: 1400 m/s
- v_{sound} in glass: 5000 m/s
- v_{sound} in steel: 5100 m/s

Because solids are so dense, when a sound wave in the air reaches the surface of a solid, only a tiny amount of the wave enters the solid and the rest is reflected. We will talk (much) more about this in later lectures.

Chapter 4

The Ear

The ear is divided, semi-artificially, into three regions:



The Outer Ear consists of

1. the **Pinna**, or **external ear**,
2. the **Meatus**, or **ear canal**, and
3. the **Tympanic membrane**, or **eardrum**.

The eardrum separates outer and middle ears, so it counts as being in both.

The Middle Ear consists of

3. the **Tympanic membrane**, or **eardrum**,

4. the **Malleus**, or **hammer**,
5. the **Incus**, or **anvil**, and
6. the **Stapes**, or **stirrup**. 4-6 together are the **ossicles** or **earbones**.

The Inner Ear consists of

7. the **Cochlea**.

We will talk about the cochlea in another lecture, this lecture is just about the outer and middle ears.

Notice that in anatomy and medicine, everything has a Latin name. Most things also have common names. In every case here, the Latin name is just Latin for the common name; that is, Incus is Latin for anvil, etc.

What are these things? I assume everyone knows what the external ear and ear canal are. The eardrum is a membrane which forms an airtight seal between the ear canal and the middle ear. The middle ear is an air filled cavity containing the earbones, the hammer, anvil, and stirrup (named for their appearance). The cochlea is a snail-shaped soft tissue organ, embedded in the skull bone, touching the stirrup on one end and with a nerve (the aural nerve) emerging on the other.

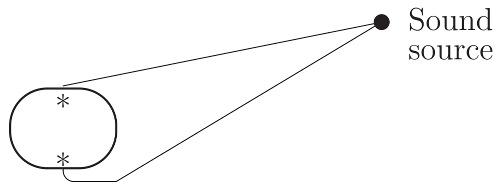
Now I go into more detail about each element of the ear.

4.1 Pinna

The pinna, or external ear (what we usually call "the ear") serves several purposes. Obviously, it has a decorative role, often accentuated with earrings or other jewelry (in most cultures). In several animals, usually tropical ones (such as elephants and African wild dogs), the ears are "radiators" used for heat loss.

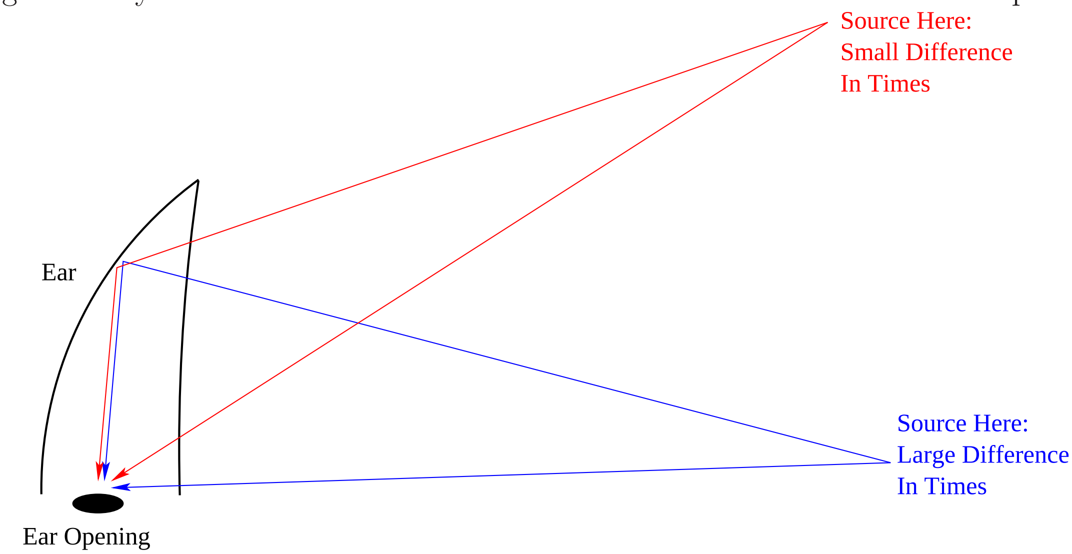
The main use of the pinna in hearing, though, is in the capture and localization of sound. To understand sound capture, think about what happens when you cup your hand to your ear, or use an ear horn. (Try sticking the small end of a funnel in your ear and see how it affects your hearing.) Doing so increases the amount of sound–especially sounds of certain frequencies–which enters your ear canal. Your pinna is a permanent device to do that, designed around a specific range of frequencies, presumably in the range 1000 to 5000 Hertz (frequencies and Hertz are explained in a later lecture).

The pinna is also important in our ability to localize sound–that is, to tell where a sound is coming from. A clear example of this is the pointy ears on animals like bats and cats. First, let's see that having two ears lets us tell which side of us a sound comes from. Here is a view from above, of sound arriving at your two ears (the *) from a source:



The distance sound has to go, to get to one ear, is shorter than to get to the other ear. Since it takes sound time to travel, it will take longer to get to the further ear. You can sense the difference in time. The further to one side the source goes, the bigger the difference: that lets you judge the angle to the sound source. We will talk about this more in a future lecture.

This doesn't help us judge whether a sound comes from high up or low down, though. If the source is in front of you and above, it arrives at the ears at the same moment, just as if it is in front of you and below. However, a bat or cat ear can tell the height of the source by timing the delay between the sound and the reflection of the sound off of the cup of the ear:



the delay between the sound arrival, and the echo off the ear, depends on whether the sound source is above, even with, or below.

Our ears are not shaped like cat ears. Cat and bat ears work great for sources in front of them. Ours work reasonably for sources in all directions, what our ancestors needed in trees. They let us tell if a sound is from in front or behind, above or below, and so forth, but not with enormous accuracy.

4.2 Meatus

The ear canal is about 2.5 cm = .025 meters deep, basically a narrow tube into the side of your head. It also has a slight bend. The main purpose is to recess the delicate components of your ear into your head, so that they are not vulnerable if you get hit on the side of the head. It also plays a role in amplifying sounds at frequencies close to 3500 Hertz, a key hearing range. We will return to this later in the course.

The meatus is lined with hairs and wax producing glands. The hairs actually move the wax out along the walls of the canal to the opening. Any junk that gets inside becomes stuck in the wax and is carried out. It is a self-cleaning organ and it is not actually a good idea to clean it, except around the opening of the meatus. Shoving a q-tip in the meatus can remove some of the wax, but push the rest down against the eardrum, which can interfere with the proper function of the eardrum. If you have excessive wax buildup, there are gentle drops you can put in your ear which loosen the wax and help it to get out–ask your doctor.

4.3 Eardrum

The tympanic membrane is a thin, somewhat stiff membrane separating the outer and middle ears. The area of the membrane is about 0.2 cm^2, or .00002 m^2. The middle ear is an air filled cavity in the skull, with a tube, called the Eustatian tube, which connects it to the throat. This lets air in and out, and can drain any fluid which forms in the middle ear (when you are sick, for instance). The pressure in the middle ear should be the same as outside. When it is not, air should pass though the Eustatian tube. When the air pressure changes (takeoff and landing in an airplane, say) the tube can become closed and a pressure difference builds up. You yawn or swallow to force open the tube and "pop," air flows through and equalizes the pressure. Such pressure differences change the tension on the eardrum, affecting hearing and sometimes causing pain.

The role of the tympanic membrane in hearing is to transmit sound–pressure variations–at the back of the ear canal, to the bones of the middle ear. The changing air pressure bends the eardrum in and out, pulling the malleus back and forth.

4.4 Ossicles

The malleus touches the eardrum on one side. Then the three bones form a chain, with the stapes touching a soft spot on the cochlea called the **oval window** on the far side. Each bone is also anchored by ligaments, so it can pivot but is not loose to move freely. The ossicles' role is the following:

The malleus is pushed on one side by the eardrum. It moves back and forth, pushing the incus on the other side.

The incus is pushed back and forth by the malleus. It moves back and forth, pushing the stapes.

The stapes is pushed back and forth by the incus. It moves back and forth, pushing the oval window.

This sounds like a Rube Goldberg device. Why not just have the eardrum push straight on the oval window? This is actually what happens in fish. However, for animals which live in the air, it is not a good method, because the cochlea is full of fluid, whereas the sound arrives down the meatus through air. Air is "thin"–it takes a small pressure to move it a large distance. The pressures involved in hearing are not large. Fluid is "thick"–it takes a large pressure to move it a small distance. Without the earbones, the sound reaching the ear would barely be able to move the fluid in the cochlea. This means that almost all of the sound energy would bounce off the eardrum and go back out the ear. The role of the ossicles is to make an efficient conversion of the sound energy, from a small force on a big area to a large force on a small area. The motion of the oval window is about 1 1/2 times the motion of the eardrum, and the pressure on the oval window is about 14 times the pressure on the eardrum. This difference is made up for by the oval window being much smaller than the eardrum. An air pressure of 30 Pascals in a sound wave is the threshold of pain. The limit of hearing is about $3 \times 10^{-5} = .00003$ Pascals. Therefore this factor of 14 is important!

[Technically, the ossicles are an impedence matching device. Impedance matchers are more efficient as they include more components or a more gradual taper. They become inefficient at very low frequency, which is also the case for the ossicles–we will discuss this in future asides.]

Fish have no ossicles, but 7 jawbones. Birds, reptiles, and amphibians have a single ossicle in each ear, the stapes, and 5 jawbones. ($5 + 2 \times 1 = 7$.) Mammals have 3 ossicles per ear, and 1 jawbone. ($1 + 2 \times 3 = 7$.) Evolution preserved the number of bones, but changed their role. This allows the reptilian jaw to perform feats impossible for a mammal. However, mammalian ears yield superior hearing because of the ossicles.

4.5 Safety mechanisms of the Ear

There is a set of muscles attached to the eardrum, and another set attached to the malleus. When these muscles contract, they tighten the eardrum, reducing the efficiency of sound transmission, and draw the malleus back from the eardrum, reducing the efficiency of transmission between the eardrum and the malleus. That is, these muscles can partially "turn off" the propagation of sound from the outside into your ear.

When there is a loud sound, a reflex called the **acoustic reflex** activates these muscles.

This reduces the danger that a loud sound will cause damage to your ear. For future reference, the reflex usually occurs for sounds louder than about 85 deciBels (dB).

However, like all reflexes, there is a delay before the reflex starts. In this case, it is about 30–40 milliseconds, or 0.03 to 0.04 seconds. This delay means that sudden loud sounds go through before the reflex can act. This means that sudden loud sounds (gunshot) are much more likely to damage hearing than steady loud sounds (jackhammer). Also, sudden loud sounds are more likely to be in the upper register of hearing, which is part of the reason that deafness usually progresses from high to low pitch.

Chapter 5

The Cochlea

The cochlea is a beautifully complex organ and the most important component of the ear, since it is what converts the mechanical vibration into a nervous impulse.

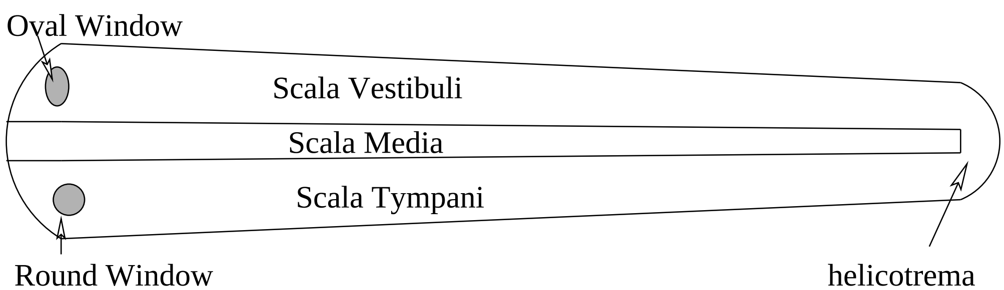
The cochlea is shaped like a snail–a long tube which is wound around in a spiral. On the big end (where the snail head would be), it has two soft spots, the **oval window** and the **round window**, which both open onto the air cavity of the middle ear. The stapes touches and pushes on the oval window, but nothing pushes on the round window. The cochlea is embedded in the bone of the skull.

The fact that the cochlea is wound up is similar to a tuba or certain other wind instruments being wound up–it is simply a way for a long, thin tube to fit inside a smaller space. The full length, uncoiled, of the cochlea would be about 3.5 cm, whereas its actual diameter is about 1 cm. It is easier to picture how the cochlea works if we think about the unwound version, so from here on we will think about it as a long, narrow tube, round in cross-section and slightly tapered towards the point.

The cochlea's tube is divided by two membranes into three fluid-filled regions running the length of the cochlea. Figure 5.1 shows a creepy photograph of a cochlea in which two of the regions are filled with a red gel and the third is filled with a blue gel. Figure 5.2 shows a cross-section (actual photograph) of what you would see if you cut the tube of the cochlea open, that is, a cross-section of the tube. The top region is called the **Scalar Vestibuli**. A membrane (Reissner's membrane) separates it from the **Scala Media**, which is separated from the **Scala Tympani** by a membrane (the basilar membrane) which carries a gelly-like body on its top, called the **organ of Corti**.

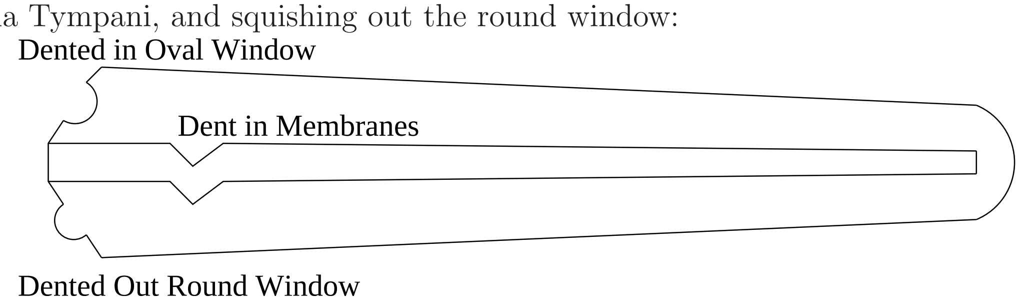
So here is a bad cartoon of the cochlea, as it would look if it were unwound. The oval window touches the Scala Vestibuli and the round window touches the Scala Tympani. The two are connected at the back by a tiny hole called the **heliocotrema**.

Figure 5.1: Photograph of the cochlea with each region filled with a colored gel.



When the stapes pushes in or pulls out the oval window, it squishes or stretches the fluid in the Scala Vestibuli. That fluid has to go somewhere. The whole cochlea is embedded in bone, except the round window. Therefore, the fluid has to cause the round window to bulge out. How can it?

There are two ways, the "fast way" and the "easy way." The "fast way" is for the extra fluid in the Scala Vestibuli to "dent in" the membranes between the 3 chambers, squishing the Scala Tympani, and squishing out the round window:



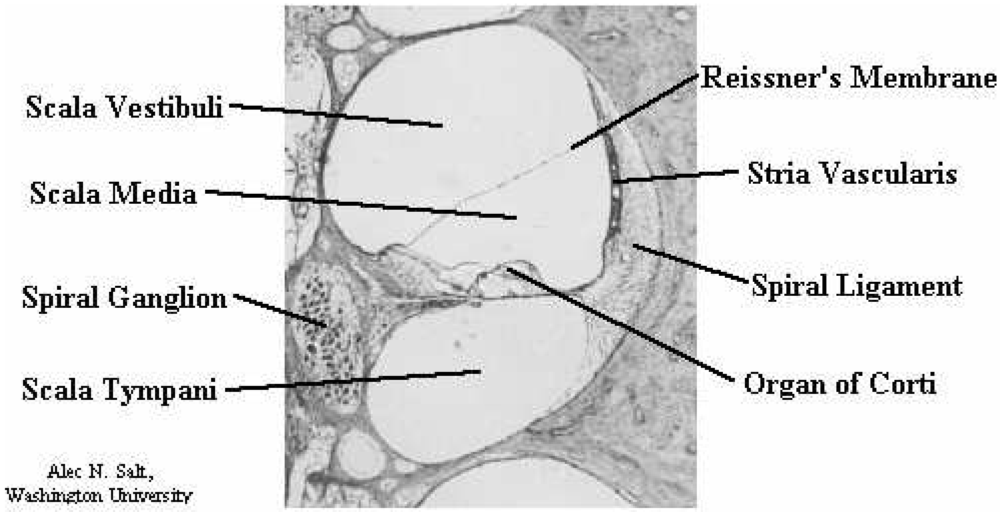


Figure 5.2: Cut open cochlea, showing its internal structure.

The "easy way" is for the fluid to run the length of the cochlea, through the helicotrema, and back.

The trick is that the membranes across the cochlea are stiff near the beginning and get softer and softer as you go the length of the cochlea. Therefore, it is faster to make the dent near the beginning, but easier to make it near the end. The faster you pull back and forth the oval window, the nearer the beginning of the cochlea the "dent" will appear, because there is not time for the fluid to go further. The slower you pull back and forth the oval window, the nearer the end the "dent" will appear, because there is time for the fluid to go to where the membrane is more flexible. Therefore the location, along the cochlea, where the membrane "dents in" will vary according to how fast the oval window is being pulled in and out. We will see that this is one of the two ways that we can tell the pitch of a sound.

We still have to see how that motion is converted into a nerve impulse. To understand that, we need to look closely at the organ of corti, on the membrane between the Scala Media and the Scala Tympani. We just saw that, when there is a sound, these membranes will get squished (or stretched, if the oval window gets pulled out by the stapes).

Figure 5.3 shows a zoom-in of the cross-section of the Organ of Corti. You see that there is a membrane hanging into the Scala Media, called the **Tectoral Membrane**. It separates a little bit of the Scala Media from the rest. When the Scala Media is pushed on from above, the tectoral membrane moves back and forth with respect to the membrane below it, essentially because it is "tied to the wall" while the body below it is "tied to the floor". The mass of tissue underneath that channel contains rows of cells (rows stretching along the length of the cochlea, so when you look at the cross-section you see just one cell), called **hair cells**. Each cell has about 100 **stereocilia**, which are like tiny hairs, projecting off its

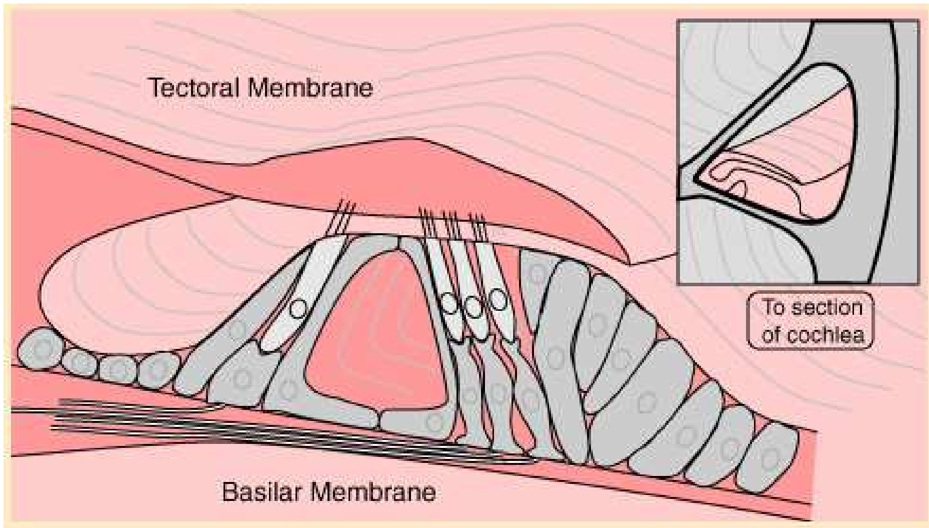


Figure 5.3: Zoom-in picture of the cross-section of the Organ of Corti.

top into the fluid of the channel. They either touch, or are actually anchored in, the tectoral membrane. When the tectoral and basilar membrane move with respect to each other, it pulls these cilia (hairs) back and forth. The hair cells are nerve sensors. That means that the back end of the cell is attached to nerve cells, and when the hairs move, it triggers an electrical impulse to be sent to the nerve cells it attaches to. So the order of cause and effect is,

- Sound–pressure–on the eardrum pushes the ossicles, which push the oval window;
- Fluid in the Scala Vestibuli flows partway along the cochlea, where it pushes in the Reissner's Membrane and makes the fluid in the Scala Media flow;
- Fluid in the Scala Media moves, stretching the basilar and tectoral membranes in different directions;
- Shearing motion between the membranes pushes back and forth the stereocilia (hairs) on the hair cells;
- Hair cells respond by sending an electical signal to the nerve cells, which will communicate it down nerves to the brain.

There are about 16 000 to 20 000 hair cells along the length of the cochlea, in 4 rows; one row of "inner" hair cells, which are the only ones attached to nerves, and 3 rows of "outer" hair cells, whose function we will discuss later.

It is a subject of future lectures, how the properties of the sound (pitch, loudness, timbre) are related to the ways they stimulate the hair cells and send a signal to the brain.

Chapter 6

Frequency

Musicians describe sustained, musical tones in terms of three quantities:

- Pitch
- Loudness
- Timbre

These correspond to our perception of sound. I will assume you have an intuitive understanding of what each of these things "means."

A physicist would describe the same tone in terms of three quantities:

- Frequency
- Intensity
- Overtone or Harmonic Structure

These correspond to a physical description of the sound. These three things are in close, but not exact correspondence to the musical descriptions:

- Pitch = Frequency (+ small loudness correction)
- Loudness = Intensity (+ substantial frequency and overtone corrections)
- Timbre = Overtone Structure (+ frequency, intensity corrections)

By far the simplest of these three to explain is pitch and frequency, so that is where we will start.

The first observation is that musical tones are generally (almost) **periodic** . This is an important concept so we should illustrate it.

Four sounds are played in the first sound file accompanying this lecture. The sounds, in the order I made them, were

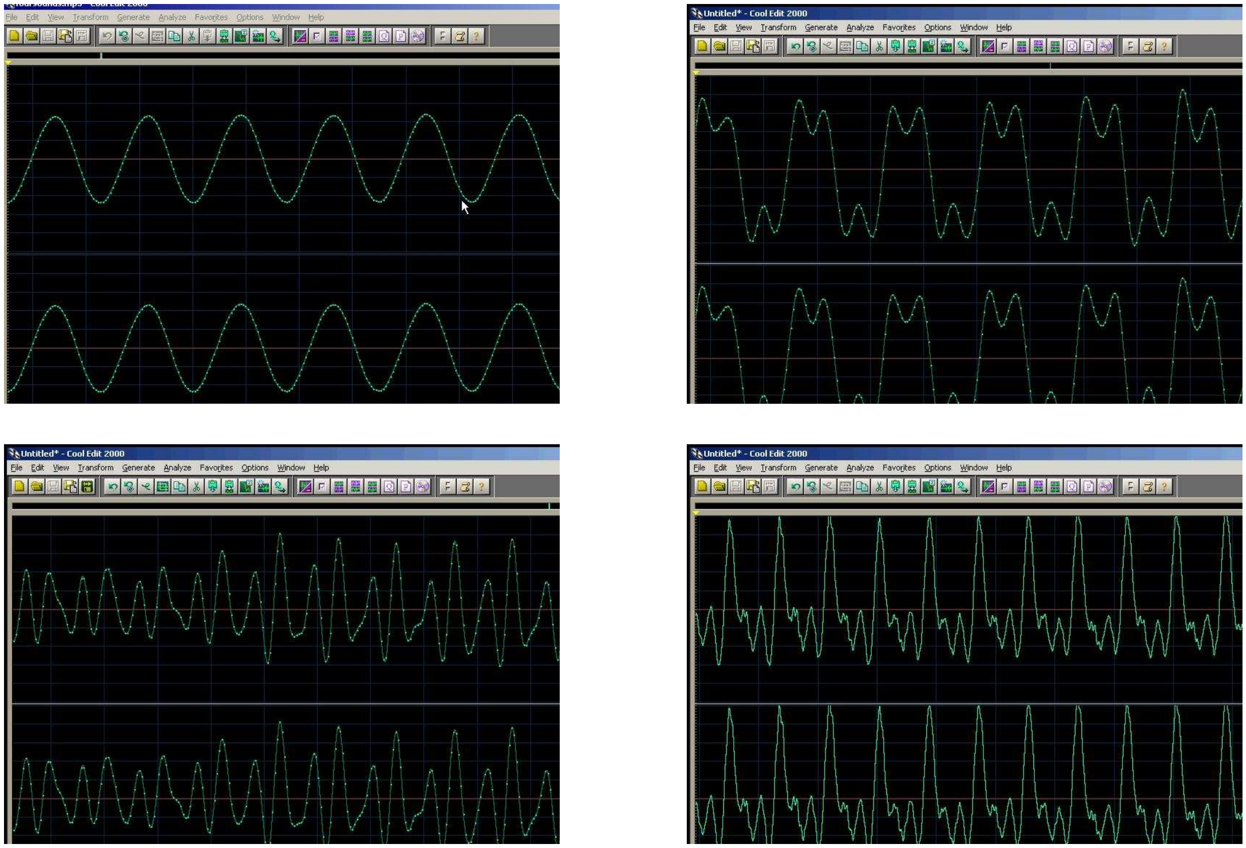


Figure 6.1: Pressure (vertical axis) against time (horizontal axis) for a finger on the rim of a wine glass (upper left), a wooden whistle (upper right), a toy siren (lower left), and the voice (lower right). Each plot appears twice because it was recorded in stereo but the microphones were next to each other.

1. a finger being rubbed around the rim of a wine glass
2. a wooden whistle
3. the little siren I demonstrated in the second lecture
4. the human (my) voice, singing “aah”

Each of these sounds has a definite pitch, though the siren’s pitch moves around during the sound.

Each of these sounds was sampled with a microphone. The pressure, as a function of time, is shown in Figure 6.1.

These recordings were made using software called Syntrillium Cool Edit; Syntrillium has been bought out by Adobe, which renamed it Adobe Audition.

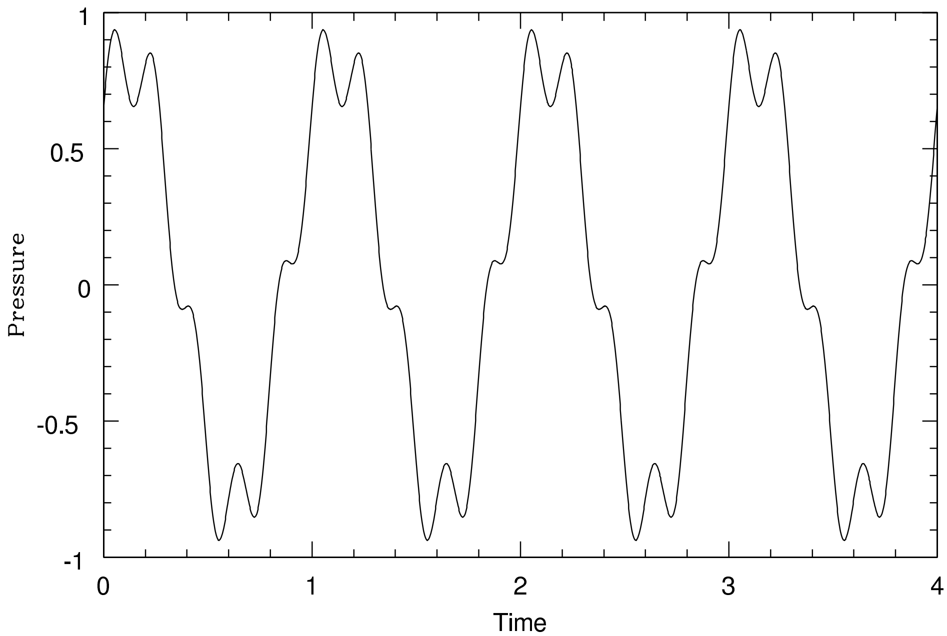


Figure 6.2: A periodic function, with period 1.

The thing to notice about all four pictures is that the pressure pattern repeats over and over. In physics we call something which repeats like this **periodic** . The fact that these sounds are periodic, is the key to their having definite pitches. In fact, we can even say,

Sounds which are periodic have a definite pitch. Sounds with a definite pitch are periodic.

Of course nothing is perfectly periodic. However, musical sounds come surprisingly close.

Obviously we need language to describe periodic functions. The most basic property is the **period** , which is just the time that it takes for the pattern to repeat. This is the same as the time difference between when some pattern appears, and when it appears again. For instance, for the plot in Figure 6.2, the period is 1 because every 1 on the horizontal axis, the pattern repeats.

Frequency is defined as the inverse of the period. Period is usually written as T, and is measured in seconds. Frequency is

$$f = \frac{1}{T}$$

and is measured in 1/seconds, which are called **Hertz**. Think of Hertz as "somethings per second," specifically, how often something repeats per second. The more often something repeats, the shorter the period, but the larger the frequency.

The most familiar example to musicians is metronomic markings. At a *moderato* tempo of 100 beats per minute, the period is (60 seconds/100 beats) = 0.6 seconds. The frequency is 1/period = 1/(0.6 seconds) = 1.67 Hertz. Since there are 2 eighth notes in a quarter note, the frequency of 8'th notes is twice as high; sixteenth notes are twice as high again, so sixteenth notes in moderato tempo occur at a frequency of about 4/(0.6 seconds) = 6.67 Hertz.

If something makes a series of sharp sounds, like staccato notes or like something bouncing in bicycle spokes, your ear can distinguish them as separate sounds if they are spaced less than about 12 or 15 per second. Beyond that, your perception tends to blur them together. [This is why musicians can get away with replacing a series of very fast notes with a slur sometimes without anyone noticing.] At about 20 to 30 Hertz, you instead perceive them as forming a low, dull sound. As the frequency goes up, the interpretation as a sound of definite frequency becomes clearer.

You can test this by putting a bicycle in a high gear, flipping it upside down so you can spin the back wheel with the pedal, and sticking a pen or piece of cardboard into the spokes. Depending on the speed you spin the wheel, you can make the object strike the spokes anywhere from a few times to a few hundred times per second. When it goes above 40 or 50 times a second, it is easily recognizable as a pitch. The faster you spin the wheel, the higher the pitch, eaching us that

Pitch corresponds to the frequency of a sound. Higher frequency is higher pitch.

There are many ways to convince yourself of that.

The limits of the range of Human hearing are about 20 Hertz to about 17 000 Hertz. At the bottom end, your ear's sensitivity becomes very poor, and sound becomes more of a sensation than a hearing experience. It is hard to demonstrate because most things which make a low pitched sound are also making higher pitched sounds at the same time [overtones, the subject of future lectures], and what you hear is the higher pitched tones. For instance, for the deepest instruments, like the contrabassoon, the bottom range of the piano, or the E string on the string bass, your perception is based almost entirely on these higher frequencies [overtones] rather than the fundamental note itself.

That is not true of the top range of hearing. Again, you lose sensitivity above some frequency, but in a sharper way. Sounds outside your range of hearing are simply not perceived at all. To see this, listen (on headphones or speakers) to the sound file accompanying this lecture, which plays in succession, notes of

- 11 000 Hertz
- 12 000 Hertz
- 13 000 Hertz
- 14 000 Hertz
- 15 000 Hertz

- 16 000 Hertz
- 17 000 Hertz
- 18 000 Hertz
- 19 000 Hertz
- 20 000 Hertz

Chances are you hear clearly the first 5 or 6, and barely notice the next one , but cannot hear the ones past 18 kiloHertz (18 000 Hertz). This varies between individuals, but it varies much more with age; the highest frequency you can hear will fall as you age.

Higher frequency is higher pitch, but we also have a definite notion of how much higher pitched one note is than another. Is this determined in a simple way in terms of the periods or frequencies? The answer is yes, but not as simple as you were hoping.

The next sound file plays, in succession, notes of frequency

- 500 Hertz
- 1 500 Hertz
- 2 500 Hertz
- 3 500 Hertz
- 4 500 Hertz
- 5 500 Hertz
- 6 500 Hertz

which are **evenly spaced in frequency**. You will perceive them to be getting closer together in frequency instead. So even spacing in frequency is not the natural perception of even spacing in pitch.

Is it even spacings in period? The next sound file plays, in succession, notes of

- period 1/2000 sec, or frequency 2000 Hertz
- period 3/2000 sec, or frequency 667 Hertz
- period 5/2000 sec, or frequency 400 Hertz
- period 7/2000 sec, or frequency 285.7 Hertz
- period 9/2000 sec, or frequency 222 Hertz

- period 11/2000 sec, or frequency 181.2 Hertz
- period 13/2000 sec, or frequency 153.8 Hertz

which are evenly spaced in period. You will perceive them as getting closer in spacing again. So the answer is something in between.

Let us try equal **ratios** of frequency, say, each note exactly twice the frequency of the note before. The next sound file plays notes of frequency

- 220 Hertz
- 440 Hertz
- 880 Hertz
- 1760 Hertz
- 3520 Hertz
- 7040 Hertz

which are of equal ratio, namely, the ratio of successive frequencies is 2. These **do** sound like they are evenly spaced. [Musicians will instantly recognize them as being spaced in octaves. Now you know what an octave is.]

To confirm this, we play notes spaced by factors of 3/2: the next sound file plays notes spaced by this ratio, namely,

- 440 Hertz
- 660 Hertz
- 990 Hertz
- 1485 Hertz
- 2227 Hertz
- 3341 Hertz
- 5012 Hertz

These are also evenly spaced perceptually [and musicians will recognize them as being spaced by perfect fifths. Now you know what a perfect fifth is.]

This tells us that our perception of pitch is logarithmically related to frequency, and since the octave is the most important musical interval, we should study this using logarithms base 2. (More in the next lecture)

Chapter 7

Frequency and the Ear

The last lecture ended by showing that our perception of pitch is determined by the frequency of the sound. The two natural questions this leaves are, why and how?

The answer to "why" must be evolutionary. We encounter a wide range of frequencies in our lives. There is important sound information throughout that frequency range. Telling apart a 150 Hertz tone from a 200 Hertz tone can tell you if someone is nervous or scared. Much of the information in speech is in the 3000 Hertz range. We need good sensitivity and need to make perceptual distinctions through this range.

The answer to "how" lies in the cochlea. Sine wave pressure variations excite different spots in the cochlea depending on the frequency of the sine wave. (Sine waves are one special periodic function, which turn out to play a special role. We will talk about why and how in future lectures.) The unwound cochlea is about 34 millimeters long. Every factor of 2 in frequency moves the spot that gets vibrated by 3.4 millimeters along the cochlea. Thus, our sense of frequency difference corresponds to how far along the cochlea a sound appears. A cartoon form is presented in Figure 7.1. The highest frequencies cause fluctuations at the beginning of the cochlea near the oval and round windows, the lowest frequencies cause fluctuations near the end.

Musicians might prefer Figure 2, which is the same idea but with the frequencies presented as pitches in musical notation. [Technically these are where the fundamentals of the pictured notes are perceived, but that is a future lecture.]

A sound of a definite frequency does not actually vibrate the membranes in the cochlea at one point. Rather, there is one point where the vibration is biggest, and a range around that value where the vibration is big enough to be important. This range on the cochlea is called the **critical band** excited by a sound. It covers frequencies about 10% to 15% higher and lower than the note which is played. The flexing of the cochlea falls away as you move from the center of the band. The picture to have in your mind is something like that shown in Figure 7.3. Note that the critical band does not have clear edges. There is

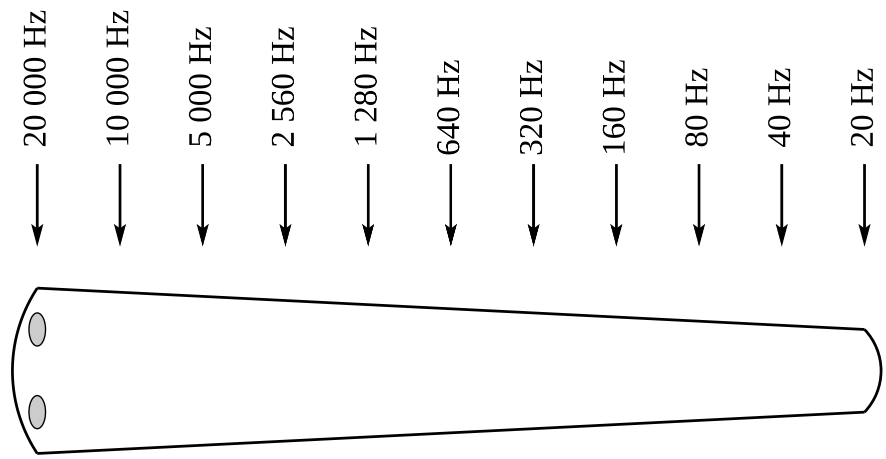


Figure 7.1: Cartoon of where on the unwound cochlea different frequencies are perceived.

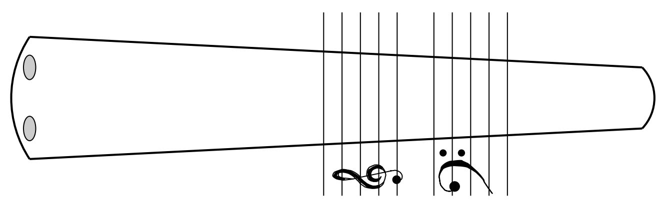


Figure 7.2: (Rotated) cartoon of how the previous figure corresponds to notes in standard musical notation.

a little vibration some distance away on the cochlea, even where the nominal frequency is almost a factor of 2 different. The spread is higher on the high frequency side than on the low frequency side.

The size of this region turns out to be smaller in living than in dead cochleas. (Don't ask how they did the measurements.) This has to do with the outer hair cells. Apparently, they respond to stimulus, not by sending a nerve signal, but by flexing or moving their stereocilia in a way which changes the tension and motion of the cochlear membranes. This enhances the motion at the center of the critical band and narrows the size of the band.

Although a region covering about a 10% change in frequency makes a nervous response, you are able to distinguish a much smaller change in frequency than this. That is because your brain can determine the center of the excited region much more accurately than the width of that region.

The size of a change in frequency which you can just barely distinguish as different, is called the **Just Noticable Difference** (because you can just barely notice the difference). Listen to the MP3 file in the HTML version of the lecture to see how sensitive your ear is to a small change in frequency. It plays ten tones. Each one starts out at 600 Hertz, but shifts in the middle of the tone. For the first two, this shift is by 4%, that is, by 24 Hertz; in one

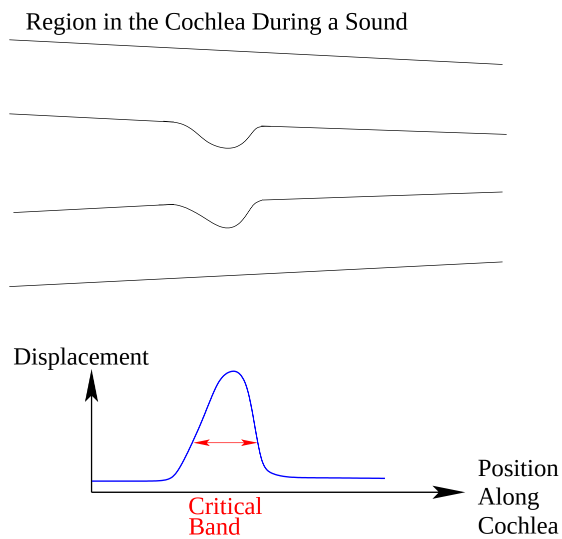


Figure 7.3: Cartoon of how the amount of excitation on the cochlear membranes varies along the membrane. There is a spot where the vibration is largest, but there is a range around it, called the *critical band*, where the excitation is also substantial.

it shifts up, in one it shifts down. The next pair, the shift is 2%, or 12 Hertz; then 1% or 6 Hertz, then 0.5% or 3 Hertz, then 0.25% or 1.5 Hertz.

The challenge is to tell which of each pair has the "up" shift and which has the "down" shift. For the 4% and 2% changes, you will find it very easy. For the 1% it is not hard. For the 0.5% you can probably just barely tell, and for the 0.25% you probably cannot tell. Therefore, your JND at 600 Hertz is around 0.5%.

The JND (Just Noticable Difference) is frequency dependent. Your ear is less accurate at low frequencies, especially below about 200 Hertz. It is also somewhat individual dependent. However, the percent change you can sense is about the same 0.5% across most of the range of hearing.

We saw above that two pairs of notes sound the same distance apart if they are in the same ratio of frequencies; so your ear considers frequencies which are 1,2,4,8,16,32,64 times some starting frequency to be evenly spaced. A mathematician would say that your sense of frequency is **logarithmic**, that your sensation of pitch depends on the **logarithm** of the frequency.

People who know and understand logarithms should skip the rest of this lecture.

Since a factor of 2 in frequency has a deep musical meaning (in fact, if two notes differ by a factor of 2 musicians write them using the same letter), it makes the most sense to use logarithms base 2 to describe frequencies.

I will write log base two of a number x as $\log_2(x)$. What $\log_2(x)$ means is, "how many 2's must I multiply to get the number x?" For instance,

- $\log_2(16) = \mathbf{4}$ because 16 is $2 \times 2 \times 2 \times 2$ and that is **4** 2's
- $\log_2(4) = \mathbf{2}$ because 4 is 2×2 and that is **2** 2's.
- $\log_2(1/8) = -\mathbf{3}$ because 1/8 is $1/(2 \times 2 \times 2)$. dividing by 2's is the opposite of multiplying by them, so it lowers the log by one.
- $\log_2(2) = \mathbf{1}$ because 2 is 2, that is, you need **1** 2 to make 2.
- $\log_2(1) = \mathbf{0}$. Careful. It doesn't take any factors of 2 to get 1.

The most important property of logs is that **the log of a product is the sum of the logs**

$$\log_2(x \times y) = \log_2(x) + \log_2(y)$$

To see that this makes sense, look at some examples:

- $\log_2(4) = \log_2(2 \times 2) = 2$, since you can see that it took 2 2's.

  Using the rule, $\log_2(2 \times 2) = \log_2(2) + \log_2(2) = 1 + 1 = 2$. It works.
- $\log_2(4 \times 8) = \log_2(32)$ which is 5, since $32 = 2 \times 2 \times 2 \times 2 \times 2$ has 5 2's in it. But $\log_2(4 \times 8) = \log_2(4) + \log_2(8) = 2 + 3$, which is indeed 5.
- $\log_2(2 \times (1/2)) = \log_2(2) + \log_2(1/2) = 1 - 1 = 0$. $\log_2(2 \times (1/2)) = \log_2(2/2) = \log_2(1) = 0$. It works again.

Also, if you are presented with an exponential, like 2 to the power n (2^n), its log is,

$$\log_2(2^n) = n \qquad \text{and} \qquad \log_2(x^n) = n \times \log_2(x)$$

These actually follow from the first one.

You can also take logs of numbers which are not some power of 2. In fact, insisting that $\log_2(x \times y) = \log_2(x) + \log_2(y)$ is enough to tell uniquely what $\log_2(x)$ should be for any (positive) x.

Fractional powers follow the rule I showed. For instance, $\log_2(\sqrt{2}) = 0.5$. [This will be important, especially when we learn that a half-step represents a 1/12 power of 2 in frequency, and therefore $12 \times \log_2(f_1/f_2)$ tells how many half steps there are between the notes corresponding to frequencies f_1 and f_2.

You can also define $\log_{10}(x)$ as the number of factors of 10 it takes to get x; for instance, $\log_{10}(1000) = \mathbf{3}$ because $1000 = 10 \times 10 \times 10$, which took **3** 10's. We will use this more when we do intensitites.

If your calculator will only do $\log_{10}$ and not $\log_2$, you should know that

$$\log_2(a) = \frac{\log_{10}(a)}{\log_{10}(2)}$$

which is exactly true. Also, very close but not exactly true are,

$$\log_{10}(2) = 0.3$$

and

$$\log_2(10) = 3\tfrac{1}{3} = 3.33$$

These are "good enough for homeworks."

Chapter 8

Superposition and Fourier Theorem

Sound is **linear**. What that means is, if several things are producing sounds at once, then the pressure of the air, due to the several things, will be

$$P_{\rm air} = P_{\rm atmospheric} + P_{\rm source\ 1} + P_{\rm source\ 2} + P_{\rm source\ 3} + \ldots$$

and the air velocity will be

$$v_{\rm air} = v_{\rm wind} + v_{\rm source\ 1} + v_{\rm source\ 2} + v_{\rm source\ 3} + \ldots$$

How each sound wave propagates, is independent of what all the other ones do. Therefore it is fully legitimate to solve for the behavior of the sound wave for each source, one at a time, and add them up. Further, if there is a sensible way of "breaking up" the sound wave from *one* source into pieces, you can treat each piece separately and then add them back together.

These statements are true so long as

$$v_{\rm air} \ll v_{\rm sound} = 340 \text{ m/s}$$

and

$$P_{\rm air} - P_{\rm atmospheric} \ll P_{\rm atmospheric}$$

that is, so long as the velocity and pressure being caused by the sound wave are small compared to the speed of sound and ambient pressure. Even for a sound at the threshold of pain, these conditions hold by a cool factor of 10 000. [They break down at about 200 decibels.] Only violent explosions would violate these conditions, so we will always assume they are true in this class.

What linearity means is that a complex sound–say, an orchestra–can be described as the **sum** of simple sounds–say, each instrument. Sometimes, your ear can make the separation between these sounds (though not always, as we discuss when we talk about masking).

We have already seen that musical sounds, at least the ones of definite pitch, are periodic. But they are *not* generally sinusoidal. This matters because everything we said last lecture, about how a sound causes vibration on the cochlea, is true of sine wave sounds. The purpose of this lecture is to start seeing why that understanding is still good enough to describe complicated sounds, and what goes on when a complicated sound enters your ear.

The sound files accompanying the lecture (in the HTML version) play a sine wave, triangle wave, square wave, and sawtooth wave, as well as a voice "singing" (quotes, because it was me) "ooo" and "eee". The wave forms of "ooo" and "eee" are shown in Figure 8.4 at the end of these notes. The computer generated waves have the same frequency, 440 Hertz, and the sung notes have the same frequency, 256 Hertz. They sound very different, that is, they have different **timbre**.

- Pitch: determined by a periodic sound's frequency.
- Timbre: determined by the "shape" of one period of the wave.

Can we describe timbre more precisely? Yes. That is because of a powerful piece of mathematics, called **Fourier analysis**, which says that any periodic function–say, the pressure pattern from an instrument–can be viewed as the sum of many sine waves.

We saw last time that a sine wave which repeats with frequency f is generally of form,

$$P - P_{\rm atmos} = \Delta P \sin(2\pi t f + \phi)$$

Here the 2π is needed because a "standard" sine wave repeats every 2π. The ϕ in there just tells whether the sine wave starts at zero right at time zero, or a little sooner or later. This wave repeats with period $T = 1/f$ because $\sin(2\pi T f) = \sin(2\pi)$ is the same as $\sin(0)$.

A sine wave with exactly twice the frequency of the first one, repeats every $T/2$. But that incidentally means that it also repeats every T; where it is at a time t is the same as where it is at time $t + T$ because it repeats exactly twice in between. The same is true of a wave with 3 times the frequency, 4 times, and so on. This is illustrated in Figure 8.1.

If we add together one sine wave which repeats with period T, and another which repeats twice or three times as often, the resulting sum will repeat with period T. It will not repeat with period $T/2$ or $T/3$, since one of the waves which built it does not. However, all of them repeat with period T, so the sum will too.

To understand how to add together two waves, look at Figure 8.2, which illustrates the summation of two waves, one with 3 times the period of the other. You see how much the smaller wave moves up and down with respect to the axis, and you move up and down the same amount with respect to the larger wave.

Now for Fourier analysis:

- Any function of form $\sin(2\pi n f t + \phi)$, with $n = 1, 2, 3, \ldots$, has period $T = 1/f$.
- Any periodic function with period $T = 1/f$ can be expressed as a sum of such sine waves!

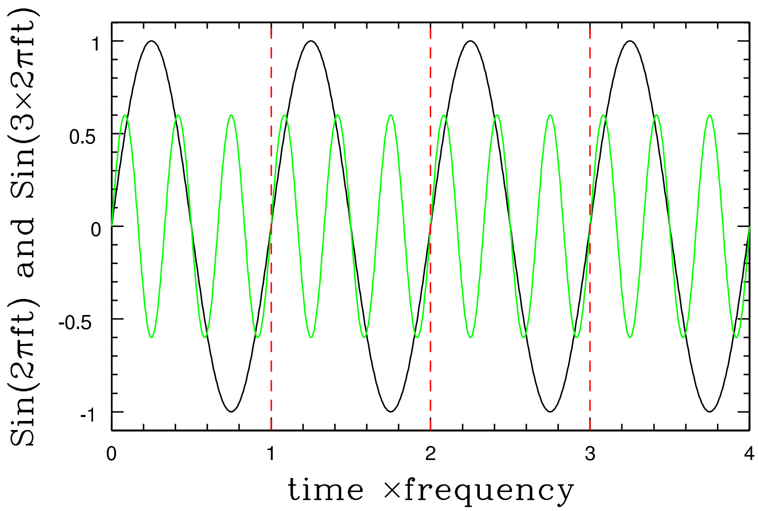


Figure 8.1: Four periods of a wave with a certain frequency, and a wave with exactly 3 times the frequency. Both waves certainly repeat at each point where the red line is drawn.

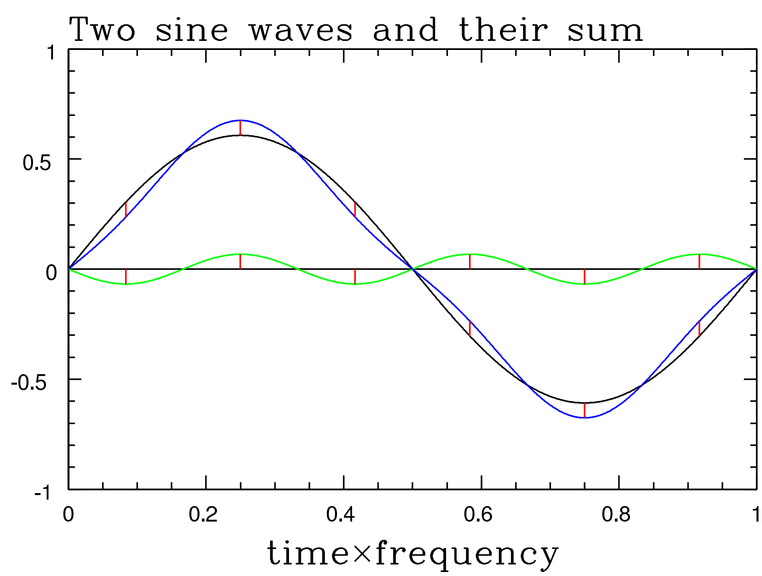


Figure 8.2: Addition of two sine waves, of period T and $T/3$, which form a wave of period T which is not a sine wave.

It is this last point which makes the subject useful. It's a pretty tall claim, so I should try to "prove it to you with examples."

First, let me explain the claim. Suppose you have the pressure pattern for an instrument or something, which is periodic with period $T = 1/f$. The claim is that there are coefficients P_1, P_2, P_3, etc, and phases ϕ_1, ϕ_2, ϕ_3 etc, such that

$$\begin{aligned} P - P_{\rm atmospheric} \quad = \quad & P_1 \sin(1 \times 2\pi f t + \phi_1) \\ & + P_2 \sin(2 \times 2\pi f t + \phi_2) \\ & + P_3 \sin(3 \times 2\pi f t + \phi_3) \\ & + P_4 \sin(4 \times 2\pi f t + \phi_4) \\ & + P_5 \sin(5 \times 2\pi f t + \phi_5) \\ & + P_6 \sin(6 \times 2\pi f t + \phi_6) + \ldots \end{aligned}$$

This certainly builds a function which is periodic with period T, since each sine wave repeats every T. That any curve can be built in this way is illustrated in Figure 8.3, which shows the construction of a triangle wave as the sum of sine waves.

The individual sine waves which build up the complex tone are called the **harmonics** of the tone. Everyone agrees that the sine wave with the same frequency as the tone is called the **fundamental** of the tone. There are two nomenclatures for the higher multiples. Either they are called:

- $2\times$ the frequency: 2'nd harmonic
- $3\times$ the frequency: 3'rd harmonic
- $4\times$ the frequency: 4'th harmonic

or sometimes,

- $2\times$ the frequency: 1'st harmonic
- $3\times$ the frequency: 2'nd harmonic
- $4\times$ the frequency: 3'rd harmonic

I think calling them the latter way is confusing, so in this class I will use the former–so the 7'th harmonic has 7 times the frequency of the fundamental. Be aware that some people will use the second nomenclature.

You will also hear people referring to the fundamental and the **overtones** (with the same confusiong about nomenclature). For a periodic sound, "overtone" and "harmonic" are synonymous. The difference is that some non-periodic sounds (many percussion sounds,

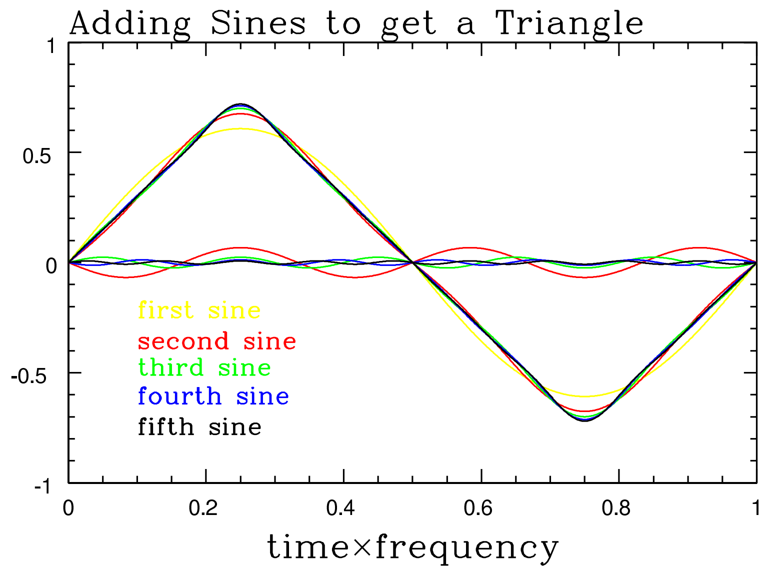


Figure 8.3: How a triangle wave can be built by adding up a series of sine waves. Each color shows a sine wave and the total wave after that sine has been added.

for instance) are composed as the sum of sine waves, but with the frequencies of the higher components not integer (counting number) multiples of the fundamental. In such case we call the other things overtones, and reserve the word "harmonic" for the case where they are integer (counting number, $1, 2, 3, 4, 5 \ldots$) multiples of the fundamental.

A more nontrivial example is shown in Figure 8.4, which shows how successive Fourier components build even a very peculiar and complicated wave form, the sounds "ooo" or "eee" of the voice.

I should also show what it sounds like as you add one after another Fourier component to produce a complex wave. This is done in the sound files found in the HTML version of this lecture.

What happens when a complex tone arrives at your ear? The answer is that you can use linearity to think about what each harmonic does, separately. The behavior of the cochlea is *almost* linear. We will see later that it is not perfectly linear, but don't worry about that now. Therefore, for a complex tone, the excitation of the cochlea will be the excitations of the points corresponding to each harmonic, with sizes corresponding to how loud each harmonic is. This is illustrated in Figure 8.5. Your ear automatically recognizes this tower of spots of excitations as arising from a single complex tone. Your brain extracts the fundamental and gives this to you as "the" frequency, and you sense or feel the extra information about the relative loudnesses of the harmonics as the sound's timbre.

Note that the harmonics get closer and closer together in terms of their pitch. This is because what is important to perception is the log of the ratio of frequency. The ratio of the

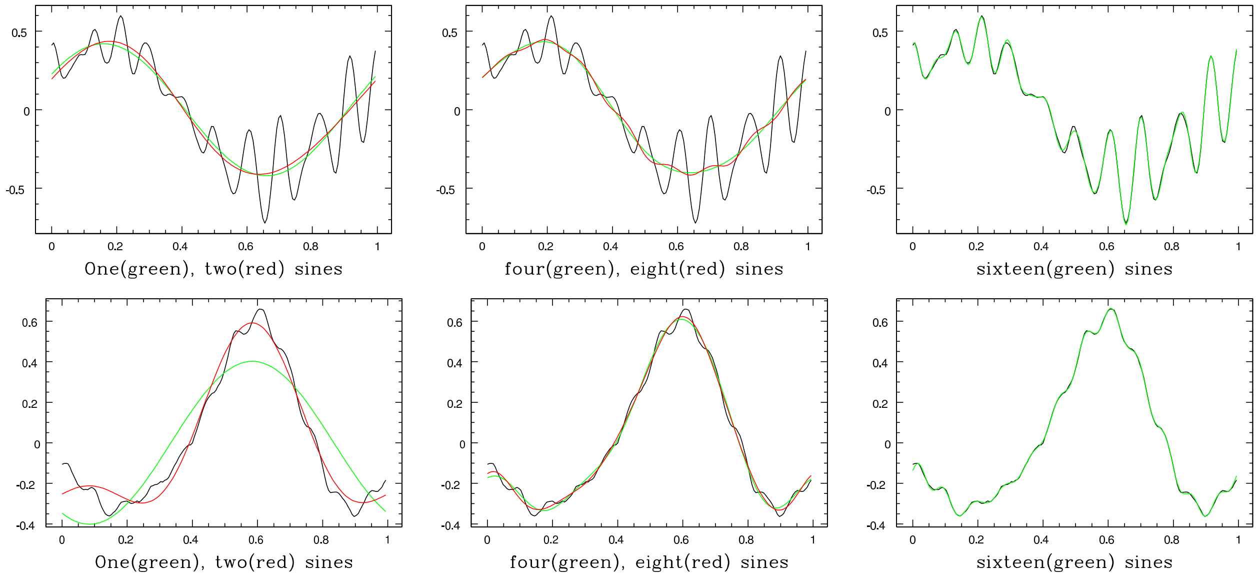


Figure 8.4: Top row: the sound "eee" fitted using different numbers of Fourier components. Bottom row: the same for "ooo."

2'nd to 1'st harmonic is 2/1=2; of 3'rd to 2'nd is 3/2=1.5; of 4'th to 3'rd is 4/3=1.33, and so on. [Musically, the separations are: 2/1=octave, 3/2=perfect fifth, 4/3=perfect fourth, 5/4=major 3'rd, 6/5=minor 3/rd, but the thirds are not tempered by more than the JND, as we will discuss in future.]

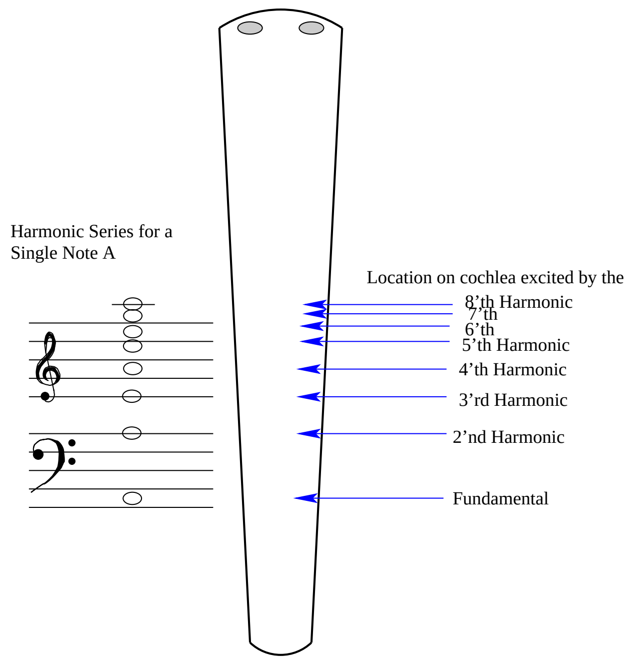


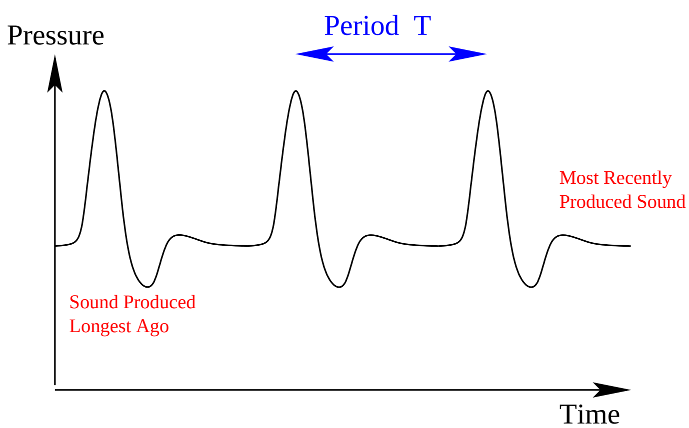
Figure 8.5: How the harmonic series of a complex note causes vibration in a series of points on your cochlea, at definite separations which your brain can interpret as a harmonic series.

Chapter 9

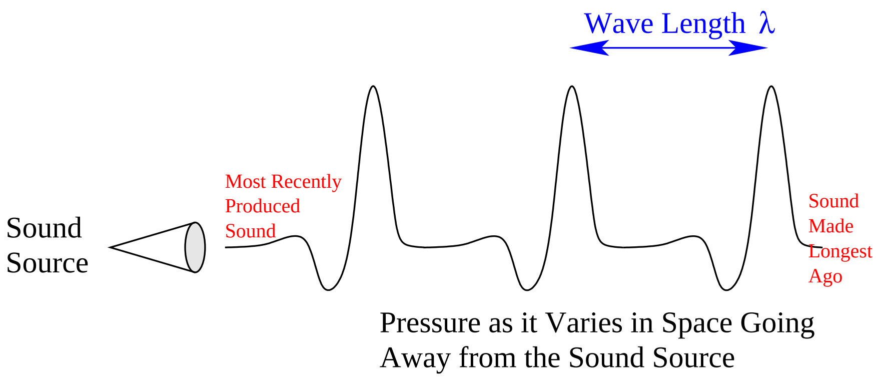
Wavelength and Diffraction

9.1 Wavelength

Suppose that the pressure at the source of a sound looks like this as a function of time:



Then as the sound moves away from the thing that makes it, the pressure will vary in the air. The pressure furthest from the source is the pressure signal the source made the longest ago. Therefore, the distribution in *space* of the pressure, will be backwards from the production in *time* at the source, like so:



If the pressure produced at the source varies periodically in time, the pressure pattern in the air will vary periodically in space (ignoring the way the sound spreads out, reflects, etc). We can define the **wave length** to be the distance over which the sound wave is periodic.

It is easy to figure out the relation between the wave length and the period of the sound. If the source produces a pressure spike at time 0 and at time T (the period), then the wave length will be how far away the first pressure spike has made it by the time the second one was made. That is just

$$\lambda = v_{\text{sound}} T = \frac{v_{\text{sound}}}{f} .$$

The funny looking thing, λ, is the Greek lower-case letter lambda, which is always used by physicists to denote wavelength.

Complex tones are produced of multiple sounds of different wave lengths λ, just as they are of different frequencies f. For a periodic tone, the wave lengths of the sine wave (Fourier) components are 1, 1/2, 1/3, 1/4, ... times the wavelength of the sound, just as the periods are 1, 1/2, 1/3, 1/4, ... of the sound's period and the frequencies are 1, 2, 3, 4, ... times the sound's frequency.

Light also has a wave length. The difference is that the wave length of light is smaller than you can see, 0.4 to 0.7 μm (micrometers or microns), which is about 1/70 to 1/250 of the width of a human hair. The wave lengths of sound waves are in the range of common lengths we encounter in our everyday experience. For instance, sound at the lower threshold of our hearing has

$$\lambda_{\text{Lowest}} = \frac{344 \text{ m/s}}{20 \text{ Hz}} = \frac{344 \text{ m/s}}{20 \text{ /s}} = 17.4 \text{ m}$$

which is actually pretty huge. Sound at the upper threshold of our hearing has a wave length of

$$\lambda_{\text{Highest}} = \frac{344 \text{ m/s}}{17000 \text{ Hz}} = 0.02 \text{ m} = 2 \text{ cm}$$

which is just under an inch–about the width of one of my fingers. Musical sounds have wave lengths varying from about the width of the palm of your hand, to the height of a room.

To understand how sound propagates, there are two important things to realize. First, we are used to how light propagates, which is in straight lines. The reason light does that is, that its wave length is so much smaller than the things it is propagating through. Lots of cool effects happen when things are the same thinkness as the wavelength of light. For instance, the film of oil you sometimes see on water, making rainbow patterns, does that because it is around 1 μm thick. Since sound's wave length is comparable to everyday objects, it will *not always* move in straight lines and cast shadows like light does–but it also will sometimes, depending on the frequency and the size of the objects involved.

The other thing to realize is that, if the behavior depends on the wave length, then the different harmonics in a complex tone will behave differently. That means that the tone color you hear (relative loudness of the harmonics) can get affected by the propagation of the sound to reach you. To understand what you will hear, you have to

1. break up the sound which is produced into its harmonics,
2. determine how each harmonic propagates through its environment,
3. put the harmonics back together at the place where the sound will be heard (ear or microphone) to understand what sound will be received there.

One case where different frequencies behave differently is sound absorption. Most sound absorbing materials give a different fraction of absorption and reflection depending on frequencies. The general tendency is that most things absorb high frequencies more effectively than low frequencies. Very high frequencies are also absorbed as they propagate long distances through the air, which is why nearby lightning has thunder with a lot of high frequencies, but distant thunder only has the low frequency rumbling.

9.2 Diffraction

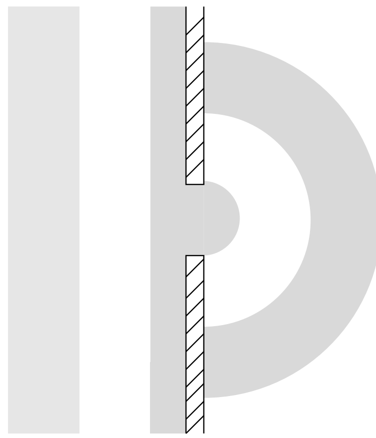
When sound moves around objects, it can "bend" to go around obstacles. This is called **diffraction**.

We are familiar with how light casts shadows and generally does not "bend around" objects. It can be reflected, but we rarely see it being diffrated. This will also be true for sound when the wave length is short compared to the objects it is moving around. In the opposite limit, it will bend around objects without any trouble:

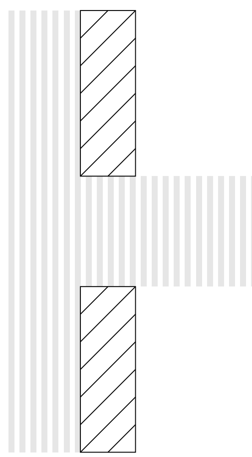
Wavelength $\ll$ Size of Objects: Straight Lines and Shadows, like Light

Wavelength $\geq$ Size of Objects: Bends Around Objects Without Difficulty

To see why, think about sound approaching a doorway, window, or other opening in a wall or other reflecting object. First, consider sound with a wave length larger than the size of the opening. Then, for a long time, all the room on the other side of the opening sees is that the pressure is high there. That will cause air to rush out in all directions. This leads to a sound wave which moves out from the opening in all directions. Showing high pressure in gray, this looks like,



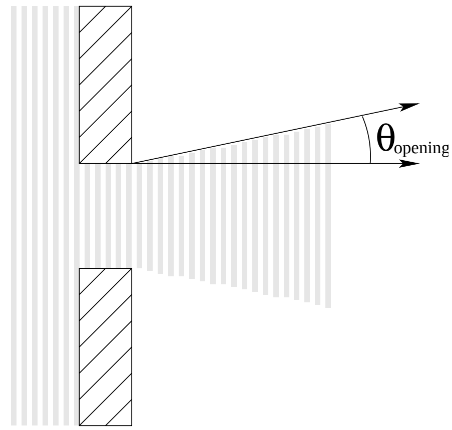
On the other hand, if the wave length of the sound is much shorter than the size of the opening, then the pressure peaks and troughs, the pattern of the sound wave, "barely sees" that there is wall so far away, and proceeds to move through the opening in the same direction it was moving in:



Actually, that was too quick. The wave which goes through the opening *will* spread out, but only gradually. The angle it spreads out by, is given roughly by

$$\theta_{\text{opening}} = \frac{\lambda}{L} \qquad \text{with } L \text{ the opening size}$$

which is, in pictures,



The truth is that at least a little sound escapes into the region that this says should be silent–but only a little, the sound is attenuated [by roughly $(\theta/\theta_{\text{opening}})^2$].

Chapter 10

Sound Localization

Localization refers to the process of using the information about a sound which you get from your ears, to work out where the sound came from (above, below, in front, behind, left, right, ...).

It is essential to distinguish from the start, the problem of localizing a steady sound from the problem of localizing a sudden sound. The steady sound is much harder (as we experience daily!).

Steady sounds are localized mainly by two things:

- relative loudness in the two ears,
- timing in the two ears.

10.1 Relative loudness

Relative loudness is fairly straightforward. If one ear is in the direction of the sound, and the other ear is on the far side of the head from the sound, then the "sound shadow" of the head will make the sound softer on the far side than on the near side.

However, last lecture we saw that sound tends to bend around obstacles. Whether it does so or not, depends on the relative sizes of the object and the wavelength of the sound. The head will give a good shadow provided that,

$$\lambda_{\text{sound}} < D_{\text{head}} \,,$$

with D_{head} the diameter of your head, which is around 20 cm = 0.2 m. This gives,

$$\lambda_{\text{sound}} < 0.2 \text{ m} \quad \Rightarrow \quad f_{\text{sound}} = \frac{v_{\text{sound}}}{\lambda_{\text{sound}}} > \frac{340 \text{ m/s}}{0.2 \text{ m}} = 1700 \text{ Hz} \,.$$

Of course, this is not an exact statement. The further above this frequency the sound is, the more pronounced the head shadow effect will be. Therefore, it will be a pronounced effect at 5000 Hertz, a modest effect at 2000 Hertz, and nearly nonexistent below 1000 Hertz.

Naturally, the more information you have, the more accurately you can determine the sound's origin. For instance, a complex tone with strong harmonics will let your ears simultaneously determine the size of the head shadow effect at the fundamental and at each harmonic. Above several thousand Hertz, the pinnae (outer ears) also start to provide shadows, giving more forward-backward information.

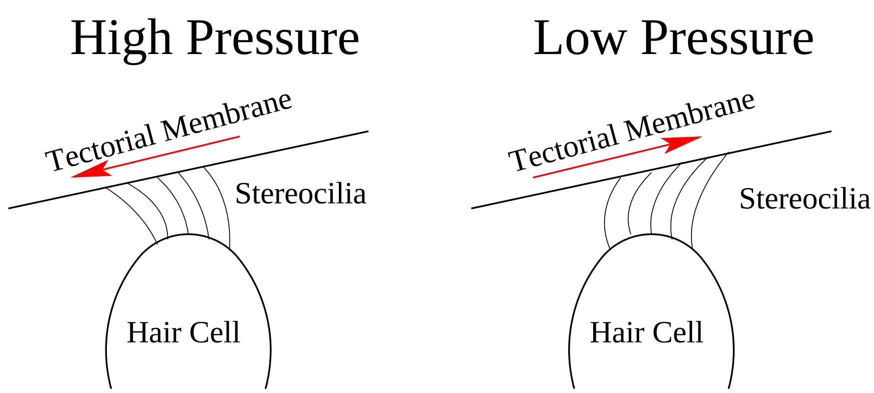
To illustrate this, try the following experiment. Get two speakers, and place them so they will be about 40° degrees from each other as seen by you. Then play the (stereo) first MP3 file provided with this lecture, and see if you can identify which speaker each tone is coming from. The tones are,

- a 2000 Hertz sine wave,
- a 4000 Hertz sine wave,
- a 6000 Hertz sine wave,
- a 2000 Hertz wave with strong harmonics.

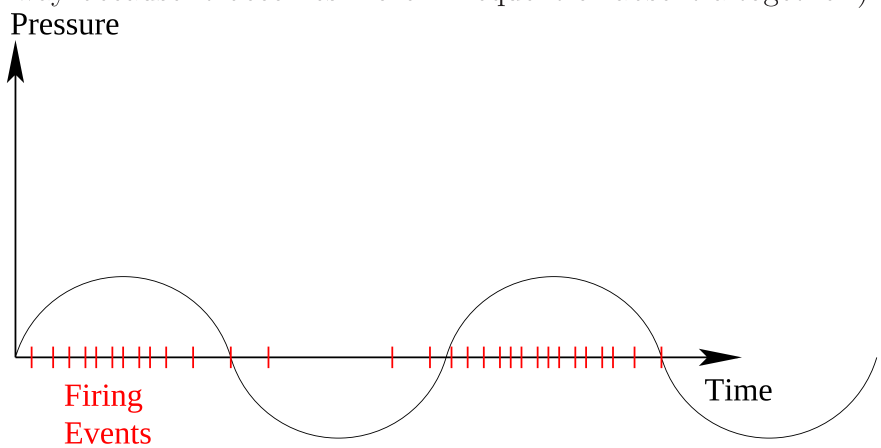
You will agree that localization this way is possible, but not very effective. (Try seeing how close together the speakers can be before you can no longer tell.)

10.2 Timing

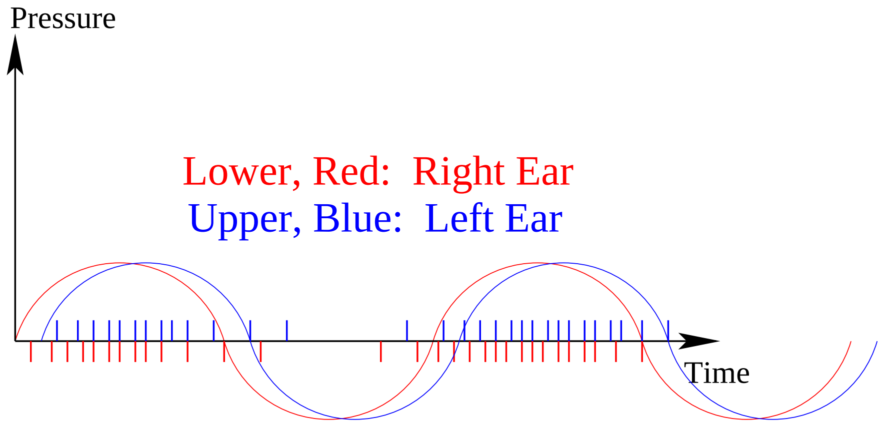
You would think that there would be no way to use timing to tell the origin of a steady tone. However, you would be wrong. To see why, one has to think about the mechanism by which the sound wave gets converted into a nerve signal by the hair cells. Recall the picture. The pressure on the eardrum turns (via the ossicles and the fluid system of the cochlea) into alternating right and left waving of the stereocilia of a hair cell:



The nerve cell "fires" when the stereocilia bend one direction. Bending the other direction actually *inhibits* their firing. (If there is no sound, the hair cells actually fire occasionally, randomly. You can tell the hair cells are bending one way because the firing becomes frequent, and the other way because it becomes more infrequent or absent altogether.)



If there is a sound in each ear, even if they are the same loudness, the high pressure peaks occur earlier in the ear closer to the source of the sound (because the pressure peak reaches that ear sooner). Then the firings of nerves in one ear will start a little sooner than the firings in the other ear:



The figure is for the case where the right ear is nearer the signal.

These signals are then carried by nerves to the brain, where their relative time of arrival is determined in a small lobe near the bottom middle of the brain, called the *Medial Superior Olive* (really it is, honest). From the difference of time of arrival of the "high pressure" signals in each ear, this organ can find the difference of time between arrivals in the two ears.

The limitation of this mechanism is that a nerve cell which has just sent a nerve signal cannot send another signal right away. The nerve requires a **refractory period** before it has recharged and can send another signal. The refractory period is about 1 millisecond (.001 seconds), and there is a further 4 milliseconds or so during which it takes a larger stimulus than usual to cause the cell to fire. This actually does not mean it is impossible to get timing information at higher frequency, but it certainly makes it harder. Therefore, localization through timing becomes ineffective for frequencies above around 1000 Hertz.

(Incidentally, the timing of nerve cell responses is a secondary way that you determine the frequency of a sound, besides determining based on the spot in the cochlea where the excitation occurs. The fact that it works well at low frequencies helps make up for the larger errors in the location method at low frequencies. Most music involves fundamental frequencies lower than 1000 Hertz, which might be related to the limit mentioned above. Of course, it also might not.)

To test how well this works, do the same experiment as before using the second sound file provided, with frequencies of 400, 800, and 1200 Hertz.

10.3 Problems

You will notice two problems with these methods.

First, there is a gap between about 1000 Hertz and 2000 Hertz where neither method is very effective. Steady sounds in this frequency range are hard to localize. In nature, this is not that big of a problem, because most real sources of sound have a lot of overtone structure (many frequencies at once). Some will be above or below this range, and the more frequencies, the more pieces of information you have to determine direction with. Surprisingly many human-made devices for getting your attention (cell phones, emergency sirens) make sounds in this inefficient range, though. Recently manufacturers are starting to get smarter about this.

Second, in many environments, much or most of the sound you get is reflected sound, from the various objects around you. That means that sound will be approaching your head from different directions; the direct line between the sound source and you, the simplest path for a reflected sound, the next simplest path, and so on. Especially inside rooms and in crowded environments, this makes for trouble.

Now let us talk about

10.4 Sudden sounds

which you will remember from the first day of lecture, are quite easy to localize. This is done almost entirely by arrival time of the sound in the two ears, and to a lesser extent by echos involving the pinna and head shadow. To convince yourself of how easy it is to determine direction, play the third sound file in the same experiment as before. See how close together the speakers can be before you cannot tell which speaker a sound comes from.

There is still the problem of reflections. However, for a sudden sound, it can be solved in an interesting way. To see this, play the last sound file and try to figure out which speaker each sound came from. The answer is–it came from *both* speakers *each* time. However, it had a delay of,

1. 10 ms on the right
2. 5 ms on the left
3. 1 ms on the right
4. 50 ms on the left

For the last case, you must have noticed that the sound came from each speaker, but one was late. For the others, you may not have.

The ear *automatically ignores* localization information for about 35 milliseconds after a sudden sound. That is because this is a normal amount of time for echos from nearby objects to reach you. This corresponds to about 10 meters of extra travel distance for the reflected sound. (Think about our ancestors living in trees, or us living in rooms.) This is done subconciously, and prevents the brain from being confused by reflections. This matches our experience–when there was a loud noise in the front of the room, the people at the back turned forward, even though most of the sound they recieved was reflected sound from the walls, ceiling, floor, and back of the room.

Chapter 11

Loudness, Intensity

We have now discussed pitch/frequency and timbre/spectrum. The third attribute of a sound is loudness/intensity. We begin by discussing the physics side, intensity, which we will see is tightly related to (but not quite the same as) the perception of loudness.

The short version of this lecture, which you are not expected to understand at the beginning, is that

$$\textbf{Intensity} = \frac{\text{Power}}{\text{area}}$$

(measured in Watts per square meter, $W/\text{m}^2 = \text{kg /s}^3$) describes the "strength" of a sound wave, and that it is related to the pressure and air movement in the sound wave by,

$$\text{Intensity} = \frac{1}{\rho v_{\text{sound}}} \langle (\Delta P)^2 \rangle = (\rho v_{\text{sound}}) \langle v_{\text{air}}^2 \rangle \,,$$

with ΔP and v_{air} the pressure and velocity change in the air due to the sound wave. Here $\langle\rangle$ means the average over the wave-form. For a sine wave, $\langle (\Delta P)^2 \rangle = (1/2) \Delta P_{\text{peak}}^2$, half the square of the peak pressure difference from atmospheric.

Now let's actually discuss intensity slowly with the intention of explaining.

11.1 Energy and Power

First recall that **energy** is "the stuff that makes things move." Heavy things are harder to get to move–so if they are moving, they contain more energy. Faster moving things also contain more energy. To be more specific, the faster something is already moving, the more energy it takes to increase its speed. Suppose you push on something which is moving. If you push with force F for a distance Δx, then the energy you impart on the object (and therefore use up yourself) is,

$$\Delta E = F \cdot \Delta x \, ,$$

The dot here is instructions on the *sign* of the result: if you push in the direction the thing is moving (speeding it up), the energy is positive. If you push against its motion, the energy is negative. (When you push to slow down a car, you are taking energy out of the car.) [Technically the $\cdot$ means that this is a dot product of two vectors].

If each time you increased v it cost the same energy, then the energy would go as mv. However, since an object with a larger v takes more energy to have its v increased, the correct energy dependence on velocity is,

$$E_{\rm motion} = \frac{1}{2} m v^2 \, .$$

This is not the only kind of energy; energy can also be stored in gravitational potential, chemical energy (think gasoline), heat, a pattern of compression and decompression, and so forth. The units of energy are the

$$\textbf{Joule} : \quad {\rm J} = \frac{\rm kg \; m^2}{\rm s^2}$$

which indeed has the same units as mv^2.

Power is the rate at which energy is delivered. I don't want my wall socket to give some amount of energy—I want it to give some amount of energy every second. The units of power are,

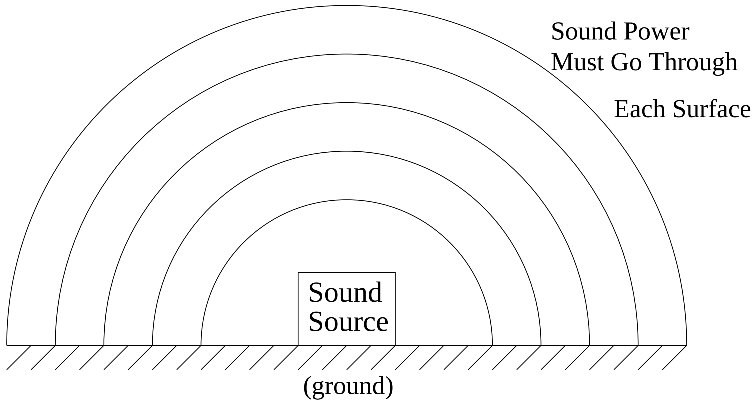
$$\mathrm{Power} = \frac{\rm energy}{\rm time} : \qquad \textbf{Watt} : \quad {\rm W} = \frac{\rm J}{\rm s} = \frac{\rm kg \; m^2}{\rm s^3}$$

For most purposes, 1 Watt of power is not very much. Light bulbs famously absorb around 100 Watts of power (and return maybe 5 Watts of light). You, sitting in a chair, generate 100 Watts of heat, which is why you have to breathe and eat. One horsepower is (defined to be) 735 Watts.

However, **for sound, 1 Watt is a lot!**. In ordinary conversation with one other person, your voice produces 10^{-5} W = .00001 W of sound power. When you breathe quietly, the sound production is more like 10^{-10} Watts. A trumpet, played by a professional at maximum dynamic (say, fff in the New World Symphony) can produce about 1 Watt of sound power–filling a large concert hall. Whatever the sound rating on your stereo equipment, the maximum it can produce, say, on a drumbeat when cranked to the maximum setting, is about 5% to 10% of the rated power (which is how much electricity goes in, not how much sound power goes out).

11.2 Intensity

The power of a sound source does not tell how loud a sound you will hear, because the sound energy spreads out as you move away. For a source on the ground, for instance,



The sound power produced by the source moves outwards, and has to go through each half-sphere. Since the area of the half-spheres gets larger, the sound is being "stretched out" over a larger and larger area, and will not sound as large.

Since what counts to you is how much sound energy enters your ear, and since your eardrum's size does not depend on the distance to the sound source, the relevant way to measure the strength of a sound wave is,

$$\textbf{Intensity} = \frac{\text{Power}}{\text{Area}}\,.$$

Multiply the intensity by the area of your eardrum to find out how much sound power actually enters your ear. The units of intensity are,

$$\text{units of }\;\textbf{Intensity}: \quad \frac{\text{W}}{\text{m}^2} = \frac{\text{J}}{\text{m}^2\ \text{s}} = \frac{\text{kg}}{\text{s}^3}$$

The last set of units looks rather strange, but that is what it turns out to be. The best thing to remember is Watts per square meter.

You might want to remember that the area of a hemisphere (relevant for a sound producer on the ground), and of a sphere (relevant for a sound producer suspended in the air) are,

$$\text{hemisphere}: \quad A = 2\pi R^2 \qquad \text{sphere}: \quad A = 4\pi R^2\,.$$

Here $\pi = 3.141592\ldots$ as usual.

11.3 Intensity, air speed, pressure

Next we need to relate the intensity to the physical description of the air which the sound wave is going through. Consider a bit of air, of length ℓ on a side. The volume of the bit of air is ℓ^3, so the mass, in terms of the density, is $m = \rho V = \rho \ell^3$. Therefore, the energy stored, as motion, in that bit of air is,

$$E_{\text{motion, in a box}} = \frac{1}{2} m v_{\text{air}}^2 = \frac{1}{2} \rho_{\text{air}} \ell^3 v_{\text{air}}^2 \,.$$

In a sound wave, energy is also stored in the fact that the pressure is higher some places and lower other places (similar to how energy is stored in the stretching or compression of a spring). This turns out to be exactly equal to the amount of energy stored in the motion of the air, so I will just double the above:

$$E_{\text{in a box}} = \rho_{\text{air}} \ell^3 v_{\text{air}}^2 \,.$$

So what is the intensity? It is the power per area–that is, the energy which *leaves the end* of the box, per unit time, per unit area of the end of the box. The sound energy moves with the sound wave; so the energy in the box moves out through the end of the box at the speed of sound. Therefore the power leaving the box out the end is,

$$P_{\text{leaving box}} = \frac{E_{\text{in box}}}{\text{time to leave box}} = \frac{E_{\text{in box}}}{\ell / v_{\text{sound}}} = \frac{\rho_{\text{air}} \ell^3 v_{\text{air}}^2}{\ell / v_{\text{sound}}} \rho_{\text{air}} v_{\text{sound}} \ell^2 v_{\text{air}}^2 \,.$$

The intensity is the power per unit area:

$$\text{Intensity } I = \frac{P_{\text{leaving box}}}{\text{Area of box}} = \frac{\rho_{\text{air}} v_{\text{sound}} \ell^2 v_{\text{air}}^2}{\ell^2}$$

or

$$I = (\rho_{\text{air}} v_{\text{sound}}) v_{\text{air}}^2 \,.$$

The (purely notional) box size ℓ has dropped out of the calculation, which is good–that indicates that the intensity I is something which depends on what the air is doing, not on how big a box of air you consider [intensity is an intensive quantity].

We should do the same thing for pressure. There is a shortcut, though. By thinking about how a pressure difference causes the air to move, one can show that

$$P - P_{\text{atmos}} = (\rho_{\text{air}} v_{\text{sound}}) v_{\text{air}} \qquad \text{or} \qquad v_{\text{air}} = \frac{P - P_{\text{atmos}}}{\rho_{\text{air}} v_{\text{sound}}}$$

for a forward moving sound wave. That means that the intensity is related to the pressure via,

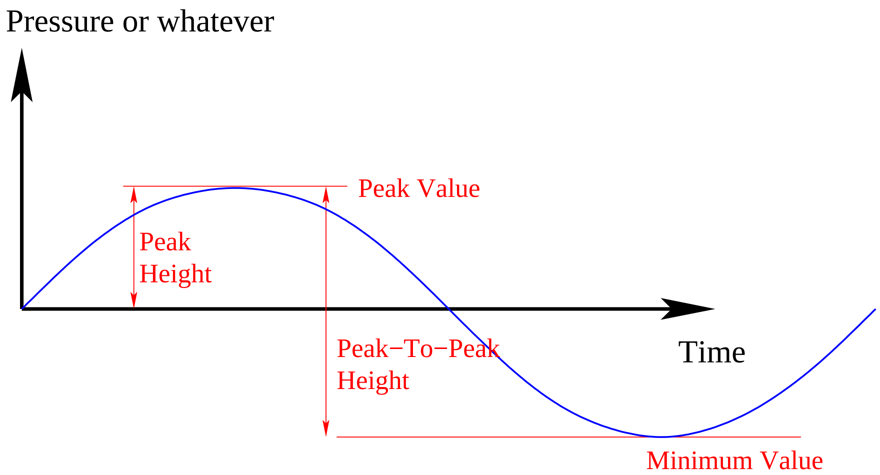
$$I = \frac{1}{\rho_{\text{air}} v_{\text{sound}}} (P - P_{\text{atmos}})^2 \,.$$

These expressions are correct for a sound wave as a whole if we interpret $(v_{\text{air}})^2$ or $(P - P_{\text{atmos}})^2$ to mean the average over the sound wave. Most waves have places where v_{air}

is larger and places where it is smaller or near zero. For a sine wave, it turns out that the average is exactly half of the peak value:

$$(P - P_{\text{atmos}})^2_{\text{average}} = \frac{1}{2}(P - P_{\text{atmos}})^2_{\text{peak}} \quad \text{for a sine wave.}$$

For those unfamiliar with peak values for a sine wave,



The last question for this lecture is, how much does the air itself move back and forth? An estimate is that the air should move back and forth by

$$\Delta x_{\text{air displacement}} = v_{\text{air}} \times t\,.$$

But what should t be? The period? And what should v_{air} be? The peak value? Clearly not, since the air is only moving forward for half the period (in a sine wave), and most of that time it is slower than the peak value. Also, we probably want the peak air movement (the difference between the furthest forward the air gets, and the middle or average location) rather than the peak-to-peak. This shaves off another factor of 2. So the right answer is,

$$\Delta x_{\text{air, peak}} = \frac{v_{\text{air, peak}}}{2\pi f} \qquad \text{for a sine wave.}$$

The peak-to-peak value, which means the furthest forward the air gets minus the furthest backward the air gets, is twice this:

$$\Delta x_{\text{air, peak to peak}} = \frac{v_{\text{air, peak}}}{\pi f}\,.$$

Chapter 12

Perception of Loudness

Last time we saw that a physicist's answer to "how loud a sound is" was,

"What you mean is, what is the sound's intensity, which is power per unit area."

The musician's answer is, "how loud does it sound to my ear?" which is a perceptual question. Some perceptual questions can be answered quantitatively, but others are harder to stick numbers to. Keep this in mind in what follows.

Loudness as perceived by your ear depends on several things. The most prominent is the intensity of the sound, but as we will discuss, the duration and frequency of the sound also have a bearing, and so does the frequency spectrum for sounds which are not sine waves. We should treat these one at a time, starting with intensity for sine waves.

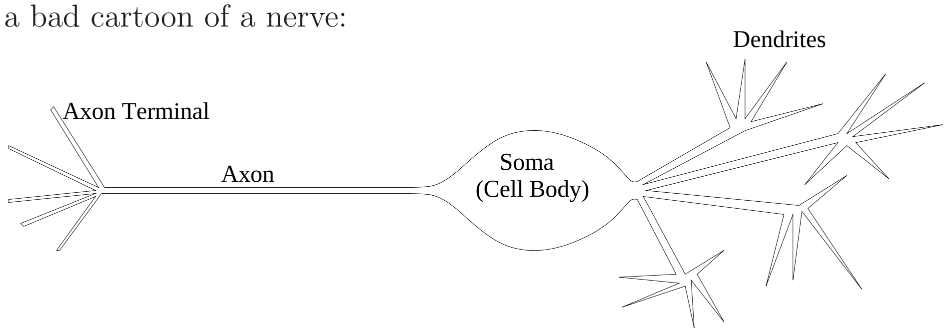
Listen to the first sound file for this lecture, which plays a tone at two loudnesses. How many times louder is the louder one? Everyone will agree about the qualitative answer of "which is louder?" Not everyone will put the same number on how many times louder it is, and no one's answer will be the actual ratio of intensities, which is 30. In fact, most people perceive the louder sound to be 3 to 5 times louder, rather than 30 times louder.

Therefore our loudness perception does not scale linearly with the intensity (sound power). Rather, a louder tone does not sound as much louder to us as it "really" is. Further, relative loudness is a qualitative perception which is a little hard to put a number on at all. Why and how does the nervous system do this?

The "why" is easy. Just as we hear sounds which occur in a wide range of frequencies, we hear sounds in a very wide range of loudnesses. Sometimes it is important to be able to hear that someone is breathing in the same room as you. Sometimes someone will shout in your ear. The sound intensity of the shout may be more than a billion (milliard, 10^9) times larger than the loudness of the breathing. You have to be able to usefully perceive both. (Your eyes have a similar problem. The difference in brightness between a sunny day and a moonless night is more than a factor of a billion. You have to be able to see things at night without being blinded in the day. If your vision perception were linear, you could not possibly do both.) If you really registered a sound 30 times louder than another as sounding 30 times louder, you would lose the quiet sounds behind the loud ones, and you could not handle as wide a range of loudnesses as you will encounter in your environment.

How does your ear do it? To see how, we have to go back and talk more about how nerves function.

Here is a bad cartoon of a nerve:



On one end the nerve has a bunch of stringy attachments called **dendrites**. These join together at a cell body (the fat lump) called the **soma**. Then there is a long cord, the **axon**, and a bunch more stringy attachments at the end, terminating in **synapses**, where they come up to the dendrites of other nerve cells. (They do not quite attach, but come close enough for electrical signals to cross from one to the other).

The dendrites are signal receivers. The cell body receives signals from all the dendrites, and "decides" whether or not (or, when) to send an electrical impulse down the axon. When an impulse is started on the axon, it travels along at about 10 m/s, reaches the axon terminals (all the stringy attachments at the end), and goes down each, to be received by the next nerve cells. The synapse is a connection which lets the signal hop from one nerve cell to the dendrite of the next.

The nerve signal going down the axon is always of the same size: a voltage of tens of millivolts (compare to computer logic, at a few volts, or the wallplug, at 120 volts in North America), with a tiny current, around 10^{-12} Amps. The nerve cell does not send information via the size of the signal, only in the presence or absence of the signal, and in its timing.

The signals coming in from the dendrites are interpreted in different ways. Some dendrites, called **excitory dendrites**, are interpreted by the soma as saying, "Fire!" Others, called **inhibitory dendrites**, are interpreted as saying, "Don't fire!" The soma (cell body) does some kind of "voting" between these inputs to determine whether (and when) to send a nerve pulse down the axon. After such a fire, there is a period of about 1 millisecond when another firing cannot occur, and 2-4 milliseconds when firing is suppressed.

Each nerve cell connects to several hair cells, and possibly other nerve cells. They can serve many different purposes:

- "nervous" nerve cells, which fire whenever a few hair cells give them a signal;

- “normal” nerve cells, which require several hair cells to fire, to give a signal;
- “lazy” hair cells, which need almost all the hair cells to be firing before they will give a signal.

Further, since some signals can be inhibitory, a nerve cell might send a signal when it sees more activity in the lower-frequency hair cells than in the higher-frequency hair cells, or vice versa. This gives extra information to help the brain recognize whether the excitation the hair cells are seeing is on the low or high frequency edge of the critical band (improving the ability to resolve where the critical band’s center is).

For a quiet sound, only the “nervous” nerve cells fire. For a medium sound, the “normal” ones start to fire, and for a loud sound, the “lazy” ones fire too. This means that the number of nerve cells responding is not simply proportional to the intensity. It can be designed, by how many nerves of each type there are, to have whatever relation proves useful.

Roughly, the number of nerves transmitting the signal tells the brain how loud the sound is; twice as many nerve signals, twice the interpreted loudness. This is not a hard and fast rule, because how your perception of loudness works at the brain level also depends on a somewhat qualitative perception. However, averaging over many people answering questions like the one at the beginning of the lecture, one finds the

Rough Rule of Thumb: $10\times$ increase in intensity is perceived as a $2\times$ increase in perceived loudness.

Chapter 13

Loudness, Decibels, Logarithms

Let us continue with how loudness is perceived and how it is characterized.

Last time we saw that a louder sound will stimulate more nerve cells to fire, but not in direct proportion to the sound's intensity, because of the way the nerves attach to the hair cells.

Besides this effect, one must also consider what the outer hair cells do. This is hard, because their role is not fully understood. However, it appears that, rather than *transmit* nerve signals, they actually *receive* nerve signals and react to them by moving their stereocilia (hairs). That is, they actually push and move around the tectoral membrane, which their hair cells are attached to. Their role seems to be,

- to amplify the motion of the membranes in very soft sounds, extending the sound threshold down to quieter sounds,

- to change the tension of the membranes, in a way which narrows the region of the cochlea responding to a given frequency (that is, narrowing the critical band).

[That is, they appear to be an element in an active feedback loop, which increases the resonant quality factor of the membranes, so the resonance caused by a sound wave is narrower and larger at its center.] People have studied this role by doing experiments on living and dead cochleas, and on ones where one or the other type of hair cell has been damaged.

DeciBels

Because the range of intensities which the ear can respond to is so large, it is most practical to describe it using logarithms. This is different than our using logarithms to describe frequencies, where it really directly described the way we perceive the frequencies.

For loudness, it is just a convenient way of compressing the scale into an easy to handle range.

It has become standard to use a unit called the **decibel** to describe sound loudness. The definition and its use are a little confusing, so I will spend some time on this.

One defines a **Bel** to be,

$$\# \text{ of bels } = \log_{10} \frac{\text{something}}{\text{some standard}} \, .$$

Then a **deciBel** (**dB**) is defined to be,

$$\# \text{ of deciBels} = 10 \times \# \text{ of Bels}$$

This definition is used in many applications, which is why I have written it in such a vague way.

In this course we will only use dB to talk about intensities of sounds (and power of sound producers, but I won't emphasize that application). For this context, "something" will be sound intensity, and we need some standard intensity to use as a basis for comparison. People have chosen as a standard against which to compare sound intensities, the fixed intensity of 10^{-12} W/m^2, because it is the softest sound your ear can hear when there are no other sources of sound present. (Actually you cannot quite hear that small a sound, but it is too late to change the standard.) Therefore, in sound, we define

$$\# \text{ of dB for a sound} = 10 \times \log_{10} \frac{\text{Intensity}}{10^{-12}\ \text{W/m}^2} \, .$$

Some examples:

- The softest sound you can hear is $I = 10^{-12}$ W/m^2. This intensity, expressed in dB, is,

$$\text{loudness in dB} = 10 \times \log_{10} \frac{10^{-12}\ \text{W/m}^2}{10^{-12}\ \text{W/m}^2} = 10 \times \log_{10}(1) = 0\ \text{dB} \, .$$

- A very quiet sound you might encounter in music, say *pp* (pianissimo), might have an intensity of $I = 10^{-9}$ W/m^2. The dB rating would then be,

$$10 \times \log_{10} \frac{10^{-9}\ \text{W/m}^2}{10^{-12}\ \text{W/m}^2} = 10 \times \log_{10}(1000) = 10 \times 3 = 30\ \text{dB} \, .$$

- A normal speaking voice for a person-to-person conversation in a quiet environment typically uses an intensity of about $I = 10^{-6}$ W/m^2. In dB, we would describe this as,

$$10 \times \log_{10} \frac{10^{-6}\ \text{W/m}^2}{10^{-12}\ \text{W/m}^2} = 10 \times \log_{10}(1000000) = 10 \times 6 = 60\ \text{dB} \, .$$

Now let's learn how to go backwards, from dB to intensity. If you don't like derivations, skip to the answer at the end. [The derivation is, that since

$$\#\text{ dB's} = 10 \times \log_{10} \frac{I}{10^{-12}\text{ W/m}^2}$$

dividing by 10 and exponentiating both sides gives,

$$10^{\#\text{ dB's}/10} = \frac{I}{10^{-12}\text{ W/m}^2}$$

or,] the answer is,

$$I = 10^{-12}\text{ W/m}^2 \times 10^{\#\text{ dB's}/10}$$

For instance, the acoustic reflex starts to kick in at a loudness (roughly *ff* fortissimo) of about 85 dB. What is the intensity?

$$I = 10^{-12}\text{ W/m}^2 \times 10^{85/10} = 10^{-12} \times 10^{8.5} = 10^{-12+8.5} = 10^{-3.5} = .00031\text{ W/m}^2\,.$$

You should be able (try it!) to take the dB results from the previous examples and work backwards to get the intensities back out.

quick and dirty way: there are 10 dB for each factor of 10 above 10^{-12} W/m^2. For instance, consider 10^{-8} W/m^2. Starting with 10^{-12} W/m^2, you have to multiply by 10 4 times to get up to 10^{-8}; each of those is 10 dB, so you get 40 dB.

It is even easier to find the *difference* between the number of dB of two sounds, of intensity I_1 and I_2:

$$\#\text{ of dB difference } = 10 \times \log_{10} \frac{I_1}{I_2}\,.$$

Other things to know to do dB quickly:

$$\begin{aligned}
10 \times \log_{10}(2) &= 3 \quad \text{(good approximation)}\\
10 \times \log_{10}(5) &= 7 \quad \text{(good approximation)}\\
10 \times \log_{10}(3) &= 5 \quad \text{(OK approximation)}
\end{aligned}$$

Using these, we see that $2 \times 10^{-7}\text{ W/m}^2 = 53$ dB, for instance. The first one means that

Doubling intensity adds 3 dB.

Beware: decibels are used as a simple way to describe the wide range of loudnesses we can hear. They *do not* indicate how loud a sound will "sound," in any simple way.

60 dB does NOT sound twice as loud as 30 dB.

The dB rating is not just "how loud it sounds." Rather, each extra 10 dB means the sound is 10 times as intense. The rule of thumb from last time means, that it is perceived to be ("sounds") roughly 2 times as loud. Therefore, 60 dB is perceived to be about $2\times2\times2 = 8$ times as loud as 30 dB. In intensities, it is 1000 times louder.

To complicate matters further, sound producing power is also often measured in dB, where,

$$\text{Sound Power in dB } = 10 \times \log_{10} \frac{\text{Power in Watts}}{10^{-12}\text{ W}} \, .$$

This does *not* correspond to the intensity in dB that you will hear.

It is a good idea to do an example. Suppose you have a 300 W driven stereo system, and it is turned up "pretty high," so that for a loud sound in the music, it will draw 1/10 of its rated power, or 30 W. It is about 10% efficient (which is actually very good) at converting this into sound. Therefore it is making 3 Watts of sound power.

Power output in dB:

$$\begin{aligned}
&10 \times \log_{10} \frac{3\text{ W}}{10^{-12}\text{ W}} = 10 \times \log_{10}(3 \times 10^{12}) \\
&= 10 \times (\log_{10}(3) + \log_{10}(10^{12})) = 10 \times (0.5 + 12) = 125\text{ dB}\, .
\end{aligned}$$

Suppose you do what you should never do, and you put your head 1 meter from the speaker. What intensity do you hear?

If the speaker is on the ground, the sound is radiating out from it in all directions, over the hemisphere surrounding it. The area of the hemisphere at distance 1 meter is, $2\pi R^2 = 2\pi \times (1\text{ m})^2 = 6.28\text{ m}^2$. Therefore the intensity is,

$$I = \frac{3\text{ W}}{6.28\text{ m}^2} = 0.48\text{ W/m}^2$$

which in dB is,

$$10 \times \log_{10} \frac{0.48\text{ W/m}^2}{10^{-12}\text{ W/m}^2} = 117\text{ dB}$$

near the threshold of pain.

How does the dB rating fall off with distance (in an open field, no reflections or sound absorbers)? Every factor of 2 in distance, the sound spreads over 4 times as much area. Since a factor of 2 in intensity is a reduction of 3 dB, a factor of 2 in distance is a reduction of $3 + 3 = 6$ dB. (Since dB arise as the logarithm of the intensity, dividing the intensity by something involves subtracting from the number of dB.)

Therefore we find,

117 dB at 1 meter

111 dB at 2 meters

105 dB at 4 meters

99 dB at 8 meters

93 dB at 16 meters
87 dB at 32 meters
81 dB at 64 meters
75 dB at 128 meters
69 dB at 256 meters
63 dB at 512 meters

This is the bad way, though it eventually works. It would go faster if I used that a factor of 10 in distance is a factor of 100 in the area, which will bring down the intensity by 1/100. Since $10 \times \log_{10}(1/100) = -20$, this lowers the number of dB by 20. Therefore,

117 dB at 1 meters is
97 dB at 10 meters
77 dB at 100 meters
57 dB at 1000 meters

Where is the intensity that of normal conversation, 60 dB? Where

$$I = 10^{-12} \text{ W/m}^2 \times 10^{60/10} = 10^{-6} \text{ W/m}^2$$

but the intensity is,

$$I = \frac{\text{power}}{\text{area}} = \frac{3 \text{ W}}{2\pi R^2}$$

Equating them,

$$\begin{aligned}
\frac{3 \text{ W}}{6.28 \; R^2} &= I = 10^{-6} \text{ W/m}^2 \\
\frac{3}{6.28} \text{ W} &= R^2 \times 10^{-6} \text{ W/m}^2 \\
\frac{3 \text{ W}/6.28}{10^{-6} \text{ W/m}^2} &= R^2 \\
480000 \text{ m}^2 &= R^2
\end{aligned}$$

which gives $R = 700$ m. Realistically, sound usually encounters obstructions and absorbers over that distance, which is also far enough for the air to absorb the highest frequencies. Still, I have heard drummers at the statue of Cartier in Parc Mount Royal while walking on the trail on the top, at about this distance, and I have heard a bagpipe band in a valley while on a hill overlooking them at a comparable distance. So the answer is reasonable.

To re-iterate:

$$\text{\# of dB's} = 10 \times \log_{10} \frac{I}{10^{-12} \text{ W/m}^2}\,, \qquad I = 10^{-12} \text{ W/m}^2 \times 10^{\text{\# of dB}/10}$$

- Multiplying or dividing the intensity by N adds or subtracts $10 \times \log_{10}(N)$ to the number of dB.
- Doubling or halving the intensity adds or subtracts 3 dB.
- Doubling distance from a source reduces the intensity by a *factor* of 4, subtracting 6 from the number of dB.
- The number of dB's is *not* simply proportional to the perceived loudness. Rather, each difference of 10 dB is an actual change in intensity of a *factor* of 10, and a difference in perceived intensity of *about* a *factor* of 2.

Chapter 14

More on Perception of Loudness

We have now seen that perception of loudness is not linear in how loud a sound is, but scales roughly as a factor of 2 in perception for a factor of 10 in intensity. However, we still have not seen how to compare,

- sounds of different frequency as well as intensity
- sounds of different duration
- sounds composed of several sine waves, such as real musical tones or simultaneous tones from several instruments.

Let's talk about these next.

14.1 Frequency and Loudness

Tones of the same intensity (power per area), but of different frequency, are perceived as being of different loudness. To simplify the discussion, consider just sustained tones where the pressure is a sine wave. If two sounds have the same intensity and their frequencies lie between about 600 and 2000 Hertz, they will be perceived to be about the same loudness. Outside of this range, that is not the case. For sounds near 3000 to 4000 Hertz, the ear is extra-sensitive; these sounds are perceived as being louder than a 1000 Hertz sound of the same intensity. At frequencies lower than 300 Hertz, the ear becomes less sensitive; sounds with this frequency are perceived as being less loud than a sound of the same intensity and 1000 Hertz frequency. The loss of sensitivity gets bigger as one goes to lower frequencies. Also, at very high frequencies sensitivity is again reduced.

Let us very briefly explain why each of these features is present.

- Hearing below about 300 Hertz becomes inefficient partly because the cochlea does not respond as well here, but also largely because the transmission of the vibrations through the ear bones becomes less efficient at low frequencies. [This is a common problem with impedance matching: below some characteristic frequency of the impedance matcher, it stops working efficiently. We might return to this after talking about resonance and impedance, when we discuss brass instruments.]

- The meatus is a tube, roughly cylindrical and of a certain length. As we will see later, such a tube has a resonant frequency, and sound waves at or near that frequency bounce back and forth several times in the tube before leaving, giving the ear a larger sensitivity to capture that sound. As we will see, the resonant wave length is $\lambda = 4L$ with L the length of the tube. This gives a number around 3500 Hertz, and explains why the ear has a region of especially high efficiency there. This explanation may not make sense to you now, but we will return to it when we talk about resonances in tubes.

- At very high frequencies one goes beyond the frequency where the cochlea is designed to work efficiently. Also, the hair cells responsible for the highest frequencies die with age and exposure to loud sounds, so the high frequency cutoff of the ear tends to move to lower frequencies with age.

We saw that it is hard to answer the question, "how many times louder is sound A than sound B?" However, it is much easier to answer, "is sound A louder, softer, or the same as sound B?" Furthermore, different people will generally give the same answer for the same pair of sounds. That is, if I find a 1000 Hertz tone and a 200 Hertz tone which one person finds to be of equal loudness, another person will also find them to be of equal loudness. (The exception is frequencies above 10 000 Hertz, where some peoples' ears lose sensitivity at a lower frequency than others.)

Adopting 1000 Hertz sine waves as a standard, we can then ask, what intensity must a sound at 100, 200, 300, ... Hertz be, to sound as loud as a 60 dB tone at 1000 Hertz? The answer will be a curve in a plot with frequency on the x and intensity on the y axes. We can also make curves for a 50 dB tone at 1000 Hertz, a 40 dB tone at 1000 Hertz, and so forth. The first people to do this were named Fletcher and Munson, so such a curve is referred to as a Fletcher-Munson curve, and is shown in figure 14.1. Having just advertized that different people will give the same answer, I nevertheless found at least 2 Fletcher-Munson curves on the web which give slightly different answers. This may be an issue of how modern the equipment used was. There is a beautiful website, linked off the course page entry for this lecture, which lets you measure your own Fletcher-Munson curve.

Let us take some time to explain what the Fletcher-Munson curve means. We will do so by answering some example questions, using the plot.

Q. What must be the sound intensity of a 100 Hertz sine wave, if it is to sound as loud as a 1000 Hertz sine wave of 60 dB intensity?

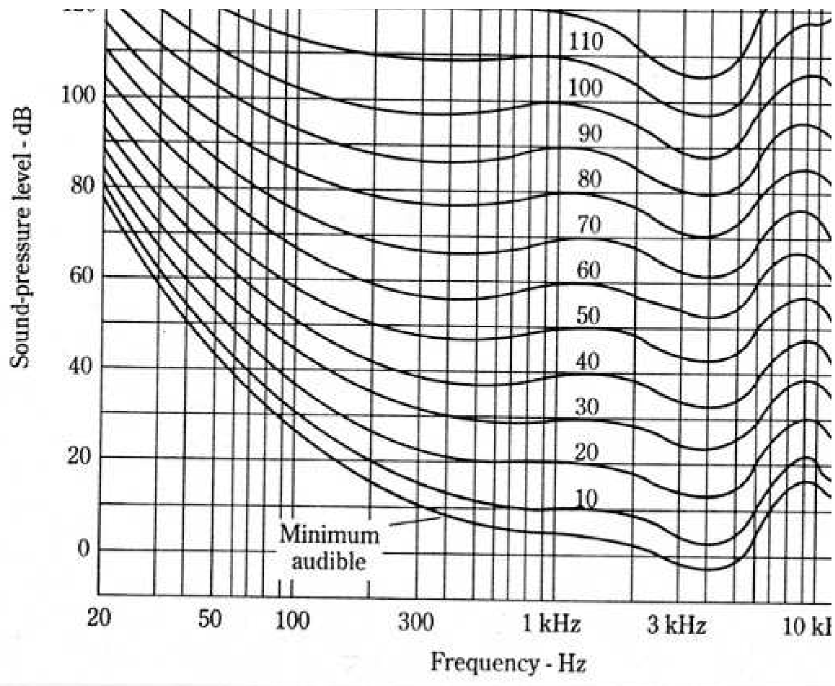


Figure 14.1: Fletcher-Munson curve, showing what sounds will be perceived as equally loud.

A. Look on the plot at the curve labeled "60". All points on this line sound as loud as a 1000 Hertz sine wave of 60 dB intensity. (1000 Hertz because it happens to be the standard, 60 dB because the curve is labeled "60".) Find the place this curve intersects the vertical line going up from 100 Hz on the x axis. Now look up what intensity that is, by going horizontally to the y axis: it is most of the way from the 60dB to the 70dB line, so it is about **67 dB**. The relevant points on the Fletcher-Munson curve are circled in green in figure 14.2.

Q. how loud must a 4kHz sound be, to sound as loud as an 80 dB sine wave at 400 Hertz?

A. First we have to find the point on the graph for 400 Hertz and 80 dB. Find the vertical bar corresponding to 400 Hertz on the x axis, and find where it meets the horizontal bar which is 80 dB on the y axis. We see that the point where they meet is a little above the curve marked "80" but well below the curve marked "90". If we filled in the curves between 80 and 90, it would be about on the "83" curve. A sound of equal perceived loudness must be the same amount up from the 80 curve. So go over to the 4000 Hertz vertical

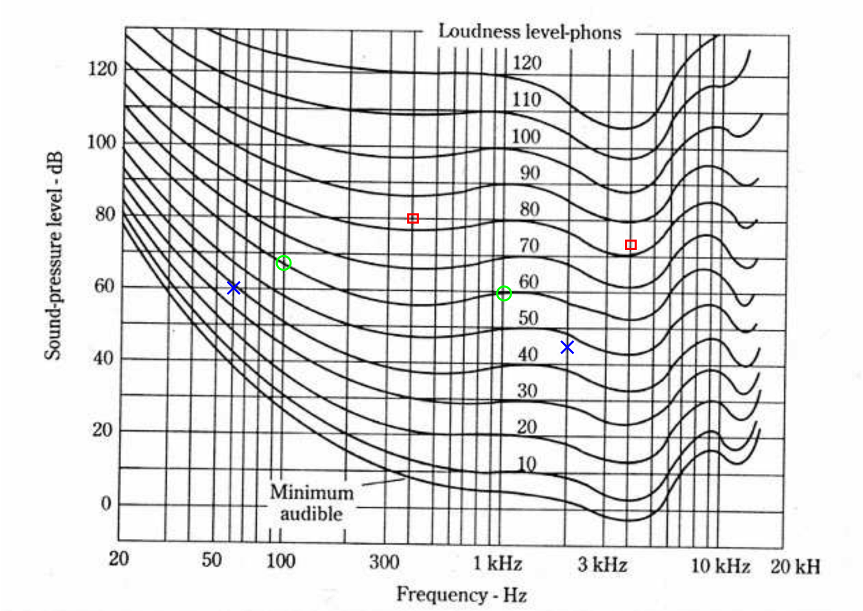


Figure 14.2: Fletcher-Munson curve marked with the points relevant for the questions in the text: green circles, first question; red squares, second question; blue crosses, third question.

line, and see how far up it you must go to be 3/10 of the way from the "80" curve to the "90" curve. The curve marked "80" meets the vertical bar at 4000 Hertz at about 70 dB on the vertical axis. To go 3/10 of the way to the curve marked "90", we go up to about **73 dB** actual intensity. The relevant points on the Fletcher-Munson curve are put in red squares in figure 14.2.

Q. Which sounds louder: a 60 Hertz, 60 dB tone, or a 2000 Hertz, 45 dB tone?

A. The two frequency-loudness points are displayed as blue crosses in figure 14.2. Note that the cross at 60 Hertz and 60 dB lies just below the curve marked "40" meaning that it is quieter than a 40 dB, 1000 Hertz sound. The cross at 2000 Hertz and 45 dB lies above the line marked "40" so it is louder than such a tone. Therefore, the 2000 Hertz, 45 dB tone sounds louder.

14.2 Phons and Sones

It is usual to define a unit of perceived loudness called a "phon", defined as,

Phon: A tone is x phons if it sounds as loud as a 1000 Hertz sine wave of x decibels (dB).

That is, all the points on the Fletcher-Munson curve labeled 70 are 70 phons loud, and so forth. The phon is defined to take into account the differences in ear efficiency at different frequency. While dB tell the intensity, a physical measure of loudness, phons give a scale which honestly compares how loud sounds will "sound."

While two sounds of the same number of phons are perceived as being equally loud, phons have the same problem as decibels in terms of interpreting just how much louder a sound with more phons will sound. That is, 60 phons does *not* sound twice as loud as 30 phons; in terms of perceptions, they are not a linear scale.

One can over-literally interpret the "rule of thumb" we met previously, that 10× as large an intensity sounds like 2× as loud a sound, and define a unit called a "sone" as follows:

Sone: A tone of x phons is $2^{(x-40)/10}$ sones.

That is,
40 phons is 1 sone (choice of starting point)
50 phons is 2 sones (since it should sound 2× as loud as 40 phons)
60 phons is 4 sones (since it should sound 2× louder still)
70 phons is 8 sones
80 phons is 16 sones
so on. In theory, if our "rule of thumb" that 10× the intensity really sounds 2× as loud, the **sone should correspond with how loud a sound actually "sounds."** That is, twice as many sones should really mean, twice as loud sounding. However I emphasize that this is subject to the caveats about how hard it is to really put a number on how many times louder one sound is than another sound.

14.3 Complex sounds

What about sounds which are not sine waves?

Any sound, whether periodic or not, can be broken into sine waves of different frequencies. The linearity of sound in air tells us that:

The intensity of a sound is the sum of the intensities of the sine waves which make up the sound. That is the answer to the "physics question" of how sine waves add up into complex sounds. Let us see how to add up intensities to get the total intensity of a sound.

You can describe quantitatively the timbre of a periodic sound by telling the intensity of each harmonic. For instance, suppose a cello plays a tone at 200 Hertz. You can describe (almost[1]) all the information about its timbre by telling how many dB or W/m^2 loud it is in each harmonic, at 200, 400, 600, 800, ... Hertz. Suppose the answer is, 60 dB at 200, 56 dB at 400, 54 dB at 600, 57 dB at 800, and very small at higher frequencies. To find the total intensity in dB, one must convert them into intensities, add the intensitites, and go back to dB:

To add intensities, you must add in W/m^2, not dB. 40 dB plus 40 dB is *not* 80 dB. For the example above, we convert each dB measure into W/m^2: the 60, 56, 54, and 57 dB from the example are (10, 4, 2.5, and 5)$\times 10^{-7}$ W/m^2. Adding gives 21.5×10^{-7} W/m^2, which is 63.3 dB.

Unless several components have almost the same intensity, the intensity in dB is usually almost the same as the intensity in dB of the loudest component.

What about perception?

Widely different frequencies cause excitation in different areas on the cochlea. Therefore, to the extent that perceived loudness is decided by how many nerves are firing, the perceived loudnesses of the different frequency components *do* add up. But if two frequencies are close enough together that their critical bands overlap, then they are exciting some of the same hair cells. All the mechanisms we saw last lecture, involving the way hair cells are ennervated, mean that the number of nerve signals will be less than the sum of the signals if each sound were separate. Therefore,

Sine waves of very different frequency have their perceived loudnesses (in sones, say) add up. Sine waves of nearby frequencies do not and act more like increasing intensity of a single sound.

Unfortunately there is no simple rule for how to add perceived loudnesses, only this rule of thumb. These rules of thumb have some interesting consequences for music.

- An instrument with a rich frequency spectrum (significant intensity in several harmonics) will sound louder than an instrument which makes something close to a sine wave. This is especially so if the spectrum of frequencies reaches over 1000 Hertz, where the ear is more sensitive–especially if it reaches the 3000 Hertz range.

- Adding different instruments with different timbres, and/or playing different notes, increases perceived loudness faster than adding more of the same instrument playing

[1]The remaining information is the relative phases of the harmonics. The ear gets almost no information from these phases, except perhaps for high harmonics where the critical bands strongly overlap.

the same note. Ten violins playing the same note only sound about twice as loud as one violin, but ten different instruments combined in harmony (playing different frequencies) will sound more than twice as loud as any one of the instruments (though probably not quite ten times as loud).

14.4 Masking

Next, we should ask questions about whether one sound can cover up another so you will not notice it, which is called **masking**.

To warm up, let us ask how much you can change the loudness of a single sound, before you notice that the loudness has changed. That is, how large a change in intensity are you sensitive to? The sound file provided with the lecture contains a tone played 4 times. Each time it changes in loudness (either louder or softer) halfway through the tone. The changes are by 10%, 5%, 20%, and 40%. Which tones get louder, and which get softer? The 10% tone is hard to tell; the 40% tone is easy. The 5% tone is probably impossible. Therefore, your sensitivity to a change in intensity is about 10 or 20 percent in intensity, or around 0.5 dB.

If I play one tone at frequency f, and I turn on another tone at a very slightly different frequency f', I can tell the second tone is there when the intensity is about 1/100 of the first intensity. We will see why it is 1/100 instead of 1/10 (as we just found) when we talk about beats. Notice that if the first tone is, say, 60 dB, that means that a 40 dB tone at close to the same frequency will be completely covered up.

On the other hand, if I play a high frequency at a loud intensity and a much lower frequency at a very small intensity, since the lower frequency sound excites a completely different spot on the cochlea, I should be able to hear the soft tone despite the loud one. Not only should I, but I actually can. Therefore, whether one tone covers up another or not, really depends on both the loudnesses and the frequencies of the two tones.

When a loud tone makes it impossible for your ear to notice a soft tone, we say that the loud tone is **masking** the soft tone, which is **being masked**. Again, this is an issue of perception, so one has to study it on human subjects. Again, whether a soft tone shows up underneath a loud tone can vary from subject to subject, particularly on their level of musical training. Nevertheless, one can make plots of what "typical" listeners can or cannot hear. Such plots are shown in figure 14.3, which requires even more clarification than the Fletcher-Munson curve.

Suppose I play a tone of 500 Hertz and 60 dB. What sounds can I hear at the same time, and what sounds will be drowned out? Find the curve on the 500 Hertz plot (upper right in the figure) which is labeled "60." Everything below this curve is drowned out by the tone; everything above it can be heard. For instance, for a tone of 1000 Hertz, the curve is at the

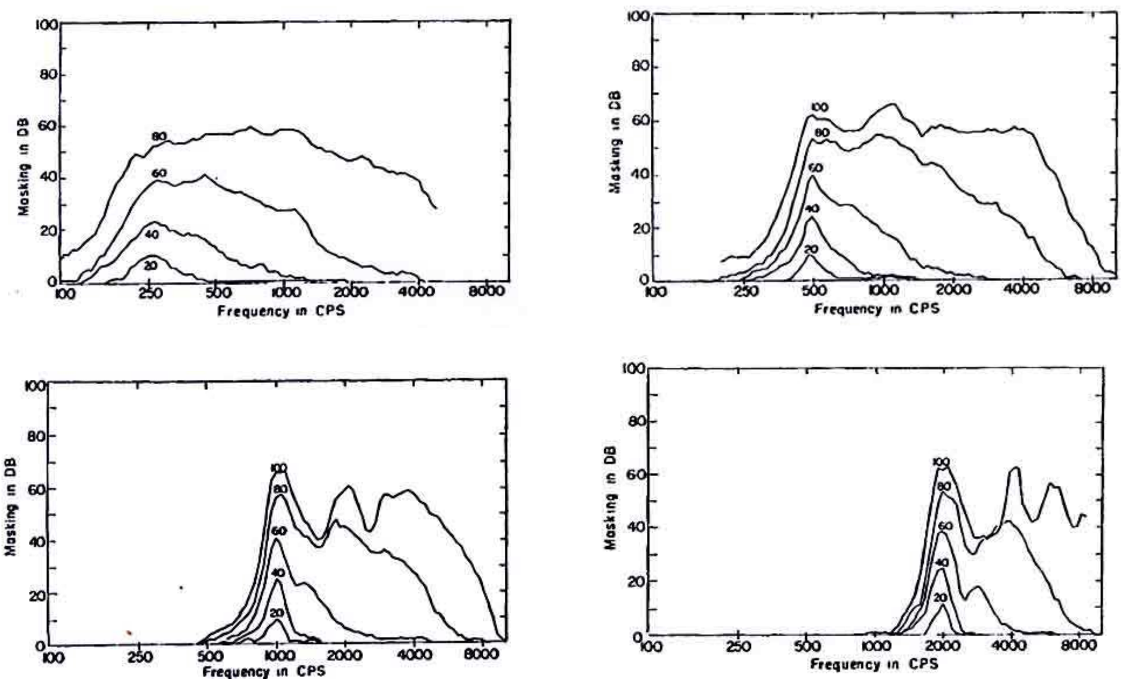


Figure 14.3: Masking curves, where the loud tone is 250 Hz (top left), 500 Hz (top right), 1000 Hz (bottom left), and 2000 Hz (bottom right).

20 dB level. Therefore, we find that 1000 Hz sounds of more than about 20 dB are audible, sounds of less than 20 dB are not.

The 20 and 40 dB curves can be thought of as mapping out the critical band of the masking tone. Wherever the loud tone is vibrating the cochlear membrane, another tiny sound which tries to vibrate it at the same place will not be noticable. The curves for very loud sounds, 80 and 100 dB, look very different; in particular there are high frequency bumps and features, often at multiples of the main pitch. We will learn more about why when we talk about aural harmonics.

Things to know about masking are,

- Generally, playing one tone masks tones of very nearby frequencies if the nearby pitch is more than 20 dB softer.
- Masking is stronger on the high frequency side of the loud tone than on the low frequency side. That is, a deep, loud tone covers up high pitched, soft tones, but a high pitched, loud tone does not cover up deep, soft tones very much.
- Very loud tones have funny effects.
- Noise with many frequencies masks more effectively than pure tones.

There are also some funny and unexpected kinds of masking which tell you that it is only partly in the ear, and partly in the brain:

- A loud sound can mask soft sounds even if they come up to 0.4 seconds after the loud sound turns off.
- A loud sound can mask a soft sound if it turns on within .040 seconds of the start of the soft sound.
- A loud tone played only in one ear can mask a soft tone played only in the other ear.

The last thing to discuss about loudness is very short sounds. The ear seems to "add up" the loudness of a sound over about 1/10 of a second, so a very short sound of a given intensity does not sound as loud as a longer sound of the same intensity. There is a sound file to illustrate this, but it is difficult to make it work well because of speaker transients.

Chapter 15

Beats

As we have seen, when two notes are played at once from different sources, their intensities add. However, they do so in a very interesting way if the two notes are close to the same frequency, because at any point in time, the pressures may be adding or they may be canceling each other off. That means that, rather than a steady intensity, you will hear the combined sound alternate between loud and soft (as the pressures add together or cancel each other out). This will happen at a frequency which is related to the frequency difference of the two tones. This phenomenon is called **beating** or **beats** and is probably familiar already to musicians. It is important in tuning and in our perception of sound, particularly in our sense of intonation (whether sounds seem in tune with each other or out of tune), so it is worth exploring it in detail.

One picture illustrates the concept of beats fairly well; Figure 15.1. This shows two sine waves, plotted in red and blue–think of them as the pressure change caused by each of two different sounds. They have slightly different frequencies, which means that they bounce up and down at almost the same rate, but one (the red one) is constantly "getting ahead" of the other. The sum of the curves is shown in black. When the two curves are exactly in step, the black curve looks like either colored curve, but twice as tall–they are adding up perfectly. Recall that the intensity goes as the square of the pressure. Since the sum has twice the pressure, it has 4 times the intensity. It looks like we have gotten something for nothing–by doubling the number of sound sources, we made the intensity rise by a factor of 4!

However, as the red curve gets ahead of the blue curve, it eventualy is a half-wavelength ahead, which means that the red curve is high precisely when the blue curve is low, and *vice versa*. This makes the two curves cancel each other; the total pressure barely changes, meaning that the intensity we hear is zero. Now it looks like we did something for nothing–with two sound sources, we have done twice the work but we get no intensity out at all! However, the time average of the intensity is what it should be; 4 and 0 average to 2, so the

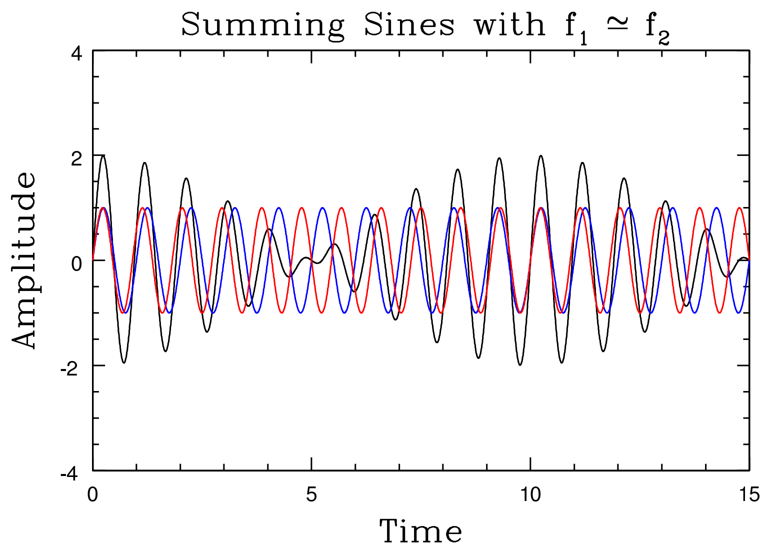


Figure 15.1: Two sine waves, in red and green, at slightly different frequencies, together with their sum, in black.

intensity (averaged over time) is twice as large with two sound sources, as it should be.

Let us try to understand this behavior in detail. This will involve a little trigonometry, which I will present and which you should try to understand, but if you don't, just skip to the underlined result and accept it as true.

Consider, then, two sounds for which the pressure change is a sine wave, with frequency $f + \Delta f/2$ and $f - \Delta f/2$–that is, the average of the frequencies is f and the difference is Δf. The pressure (difference from atmospheric) your ear hears is the sum;

$$\sin\left[2\pi f t + \pi \Delta f t\right] + \sin\left[2\pi f t - \pi \Delta f t\right] \, .$$

Now use the trigonometric identity for the sum of two angles:

$$\sin(a \pm b) = \sin(a)\cos(b) \pm \cos(a)\sin(b)$$

to rewrite this as

$$\begin{aligned}
& \sin\left[2\pi f t + \pi \Delta f t\right] + \sin\left[2\pi f t - \pi \Delta f t\right] \\
= \; & \sin(2\pi f t)\cos(\pi \Delta f t) + \cos(2\pi f t)\sin(\pi \Delta f t) \\
& \qquad\qquad + \sin(2\pi f t)\cos(\pi \Delta f t) - \cos(2\pi f t)\sin(\pi \Delta f t) \\
= \; & \underline{2\sin(2\pi f t)\cos(\pi \Delta f t)} \, .
\end{aligned}$$

How does this behave? The main bit is the $\sin(2\pi f t)$, which is just oscillations at the frequency which is the average of the two notes' frequencies. But this is **modulated** (made

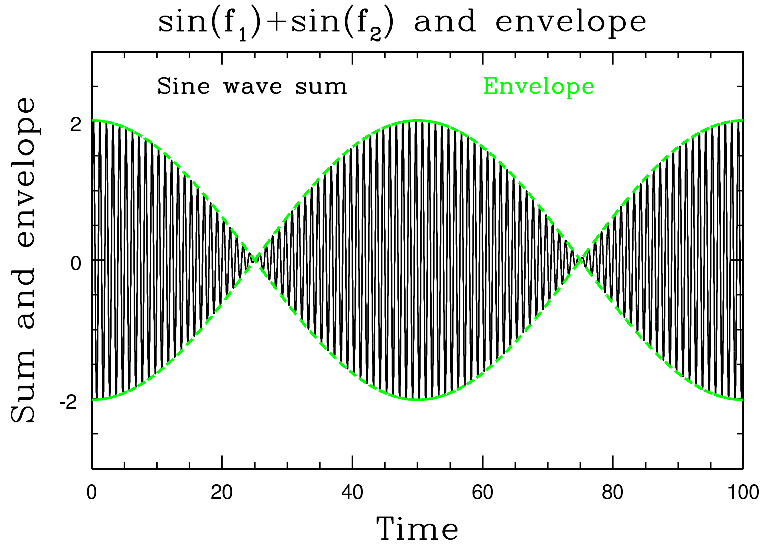


Figure 15.2: The sum of two sine waves is a sine wave inside an envelope.

louder and softer) by being multiplied by the second term, $\cos(\pi \Delta f t)$, which is called an **envelope** for the oscillations. The sum is illustrated in figure 15.2.

The envelope function means that the sine wave will get louder and softer, which is the beating phenomena. We name the frequency between quiet points the **beat frequency**. The quiet points occur wherever the cosine function has a zero, that is, at t such that $\cos(\pi \Delta f t)$ is zero. This happens when $\pi \Delta f t = \pi/2, 3\pi/2, 5\pi/2, \ldots$, that is, every π. That means that the time interval is $\pi \Delta f \Delta t = \pi$, or $\Delta t = 1/\Delta f$. Therefore the beat frequency is precisely the frequency difference:

$$\boxed{\textbf{Beat frequency} = \textbf{frequency difference}}$$

a remarkably simple relation.

For instance, if two instruments are trying to play A_4, and one plays at 440 Hertz and the other plays at 442 Hertz, you will hear a frequency of 441 Hertz (the average) with a beat frequency of 2 Hertz (the difference)–that is, it will get loud and quiet twice per second.

When the two sounds are not of equal loudness (that is, in the real world), they do not cancel completely. Therefore you get a beat pattern of louder and softer sounds, but without the sound becoming perfectly quiet. This is illustrated in figure 15.3. Incidentally, if the second sound is just 1% as loud in intensity, that makes it 1/10 as large in pressure, which means that it will bring the pressure of the first sound up and down by 10%. This is why your loudness sensitivity of about 10% gives you a masking level for very nearby frequencies of 1%.

When is the beat phenomenon audible? If the frequency difference is large enough, you cannot distinguish the loud and soft pattern because they go by more quickly than you can

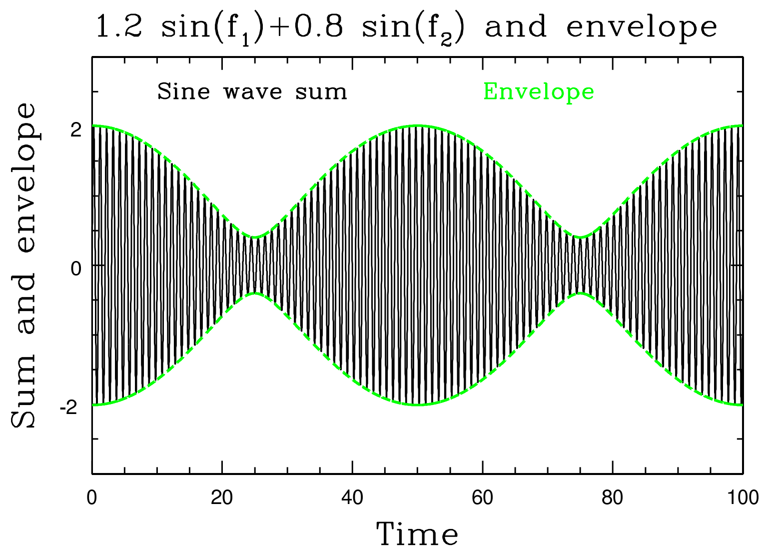


Figure 15.3: Beating between two frequencies of somewhat different amplitude gives incomplete cancellation.

separate them. This is the “movie picture phenomenon”; you never notice that a movie or TV show is a series of still pictures spaced 1/24 second apart because you cannot make out such rapid changes–but if they were spaced 1/10 of a second apart, you would notice. Where you start to notice beats is not a clean line, but very roughly it occurs around 15 Hertz. Faster than that, and you hear a “roughness” but don’t distinguish the sound as getting louder and softer; slower than that, and you do notice that it is getting louder and softer.

How low a frequency can you distinguish? That depends on how long the notes are played! The sound has to go through at least one and maybe more cycles of getting louder and softer for you to notice that it is happening. If the notes are out of tune by 2 Hertz but they are only played for 1/4 second, there is no way to notice that the beat phenomenon is occurring, because less than 1 beat occurs. In fact, the exact frequency of a very short note is not well defined, which is part of why you can get away with poor intonation when playing very rapid passages, but not on long sustained notes or chords.[1]

Of course, real tones in music or life are rarely sine waves. They are very often periodic, though, containing harmonics at integer multiples of the fundamental frequency. It is possible for beat phenomena to occur which involve these, as well as the fundamental.

For instance, if one player performs D_3 at 147 Hertz and another performs D_4 at 293 Hertz, then the second harmonic of the first tone, at 294 Hertz, will beat with the second tone. In this case the beat frequency will be $294 - 293 - 1$ Hertz. (The beat frequency is

[1] [For people who know Fourier transforms: this is related to the fact that the Fourier transform of a short wavetrain involves a finite range of frequencies. Beats only occur when the Fourier transforms do not substantially overlap.]

always the difference of the higher with the lower–you cannot have a negative frequency.)

There can also be beating between the harmonic of one note and the harmonic of another; so one player playing A_3 at 220 Hertz and another playing E_4 at 331 Hertz are both producing a frequency close to 660 Hertz; $220 \times 3 = 660$ and $331 \times 2 = 662$. These will beat with frequency 2 Hertz (even though the person playing E_4 is only off by 1 Hertz). This is the beat phenomenon used by string players to tune their open strings, which are usually spaced by musical 5'ths, which is factors of 3/2. A nice way to visualize this is to write the frequencies on a horizontal axis, spaced according to how far apart they are in frequency. So in the last example, the frequencies can be written,

$$\begin{array}{cccccccc} 220 & & 440 & & 660 & & 880 & \\ & 331 & & & 662 & & & 993 \end{array}$$

Musicians very often use beating as a means to tune. As these examples show, the two notes do not have to be the same; they can differ by an octave, or even by a musical 5'th.

When two frequencies are close, but a little too far apart to hear beats, there is a "rough" sensation in the ear, called **dissonance**. Most people find dissonance mildly unpleasant, though it is certainly musically useful. This occurs between two notes which are too close, too close to being a factor of 2, or too close to having low harmonics overlap with each other. As we will discuss in a few lectures, musical scales are largely designed to avoid too many dissonant intervals.

On the flip side, the ear "likes it" when two tones have harmonics overlap precisely; this is called **consonance**. The most important musical intervals are those for which these overlaps occur in relatively low harmonics. Writing the frequencies as a multiple of the lowest common denominator of the two frequencies, the most important intervals and the places their harmonics overlap are:

$$\text{Octave: } \frac{2}{1} \qquad \begin{array}{cccccc} 1 & \textcolor{blue}{2} & 3 & \textcolor{blue}{4} & 5 & \textcolor{blue}{6} \\ & \textcolor{blue}{2} & & \textcolor{blue}{4} & & \textcolor{blue}{6} \end{array}$$

$$\text{Fifth: } \frac{3}{2} \qquad \begin{array}{ccccccccccc} 2 & & 4 & & \textcolor{blue}{6} & & 8 & & 10 & & \textcolor{blue}{12} \\ & 3 & & & \textcolor{blue}{6} & & & 9 & & & \textcolor{blue}{12} \end{array}$$

$$\text{Fourth: } \frac{4}{3} \qquad \begin{array}{cccccccccccc} 3 & & 6 & & 9 & \textcolor{blue}{12} & 15 & & 18 & & 21 & \textcolor{blue}{24} \\ & 4 & & 8 & & \textcolor{blue}{12} & & 16 & & 20 & & \textcolor{blue}{24} \end{array}$$

$$\text{Major Third: } \frac{5}{4} \qquad \text{Minor Third: } \frac{6}{5} \qquad \text{Major Sixth: } \frac{5}{3} \qquad \ldots$$

The challenge of designing a musical scale is to have many notes which differ by these intervals, but very few which differ by dissonant intervals.

Chapter 16

Aural Harmonics

We have seen that two things that our ears "like" are,

- sounds which are periodic–that is, overtones are in harmonic relation to a fundamental (integer multiples, 1,2,3,4,. . . of a fundamental frequency);
- harmony–that is, multiple sounds where the overtones (usually harmonics) of different tones land exactly on top of each other.

On the other hand, the ear does not like it when two tones have overtones or harmonics which come *close* to overlapping, but miss, such that the critical bands substantially overlap but do not coincide.

Are these preferences related? It turns out that they are!

First recall that sound, in the air, is *linear*: you can take several sounds and play them at once, and all that happens is the ΔP and $v_{\rm air}$ from the sounds add up. However, *devices* are not always linear, or at least it might take a much smaller pressure or air motion to get them to become non-linear. The ear is a device, and this statement is true of the ear.

To explain, consider an analogy. What happens when you use a large driving amp on a small speaker, not designed to take so much power (or just turn up the sound too loud on most speakers)? The sound becomes *distorted*–the timbre changes. Why? The electric signal is trying to push the membrane of the speaker cone further than that membrane "wants" to go. The membrane, pushed too far, becomes stiffer, and does not move as far as it should. The result is that the peaks and troughs of the pressure get cut off. If you were trying to play a sine wave, you get a sort of squashed sine wave, as illustrated in Figure 16.1.

The same phenomenon can occur lots of places:

- When digitally recording, the digitizer cannot accept amplitudes (basically, pressures) larger than some size. Anything larger gets replaced with the maximum value. That would take a wave and crop the tops and bottoms to be flat.

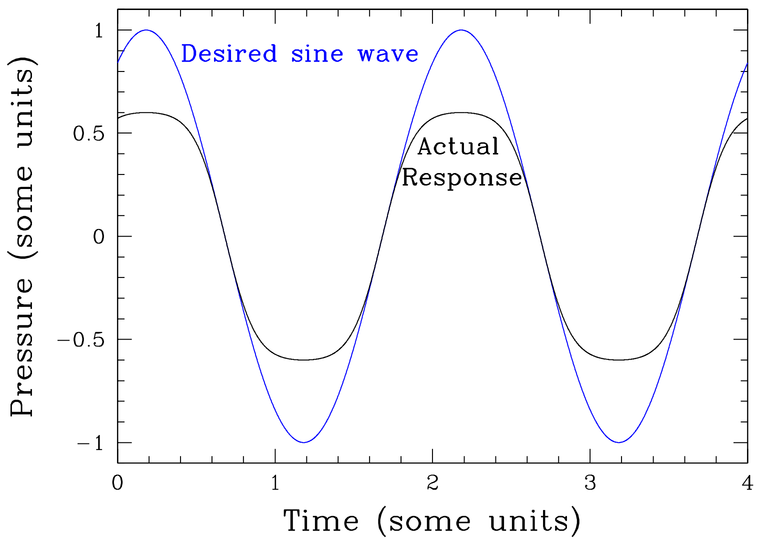


Figure 16.1: Example of how a sine wave gets distorted if a speaker membrane simply cannot move as far as it needs to, to produce the desired sine wave.

- Microphones also do this. Many involve membranes (or piezoelectric devices) which respond linearly for some range of pressures, but not outside of that range. Push them too far, and the electrical signal they give out will not faithfully reflect the pressure signal that went in.

- The membrane in a speaker does this, as mentioned. Pushed too far out or pulled too far in, it becomes stiffer and less responsive. The wave it produces becomes distorted.

- Your ear contains several membranes and moving parts–the eardrum, the ossicles, and the oval and round windows of the cochlea, in particular. Just like the membrane of a speaker, they are designed so that if you push twice as hard, they move twice as far–within limits. Push them too far, and their motion tends to be less than it "should," because they get stiffer.

So what? Well, suppose the input sound is a sine wave. The output sound is distorted, typically by having the tops and bottoms trimmed, not necessarily by the same amount. That is no longer a sine wave. However, it is still a periodic wave. That means that harmonics get added to the wave, which were not present to begin with. This happens in the membranes and bones of the ear on the way to the cochlear fluid. So if you play a sine wave, what arrives in the cochlea is not a sine wave, but distorted. The distortion is *always* the addition of harmonics with integer related frequencies (that is, play 300 Hertz, and the generated extra sine waves have frequencies of 600, 900, 1200, *etc* Hertz). These added harmonics are called **aural harmonics**, since they are harmonics added by the ear.

The louder the original sound, the less linear the response–which means that, in terms of decibels, the aural harmonics get closer to the loudness of the original tone. As usual, the size of the aural harmonics also varies with frequency, though not that strongly. A reasonable approximation is that

for a 70 dB tone, the 2'nd harmonic (twice the frequency) is 25 dB loud, and the 3'rd harmonic (three times the frequency) is 15 dB loud.

Further, every 10 dB added to the fundamental adds 20 dB to the 2'nd harmonic and 30 dB to the 3'rd harmonic:

for an 80 dB tone, the 2'nd harmonic is 45 dB, the 3'rd harmonic is 45 dB.

This helps explain something we saw two lectures ago. When we looked at the masking curves of a tone, say, 1000 Hertz, it was fairly symmetric for quiet intensities, but at very loud intensities it developed big peaks at twice and three times the frequency being played. That is because the aural harmonic of the loud tone can cover up the softer tone.

What does it all have to do with our preference for periodic sounds? Suppose you had some sound which was not periodic. It was built out of sine waves, but the higher frequency sine waves (called **overtones**) did not appear at integer (1,2,3,...) multiples of the lowest tone. (This happens for primitive percussion instruments, for instance.) Then, if the sound were played loudly, the fundamental would also add its aural harmonics, which *would* be at exact multiples of the lowest frequency. For instance, if the instrument played 500 Hertz plus 1050 Hertz sine waves, then the 500 Hertz sine wave would add an aural harmonic at 1000 Hertz. 1000 and 1050 Hertz are dissonant, because they are too close together to excite separate regions on the cochlea, but not close enough together to hit *exactly* the same spot. Therefore, they sound "ugly" or dissonant. An instrument which makes sounds like this would be discarded (or used for special dissonant effects). On the other hand, if the overtone the instrument produced were at 1000 Hertz, that is, exactly twice the fundamental, then it would overlap with the aural harmonic of the fundamental, which your ear likes.

Therefore, while we may not know why the ear/brain system "likes" consonance and "dislikes" dissonance, the fact that it "likes" periodic sounds is related to these facts.

It turns out, luckily, that it is pretty easy to build instruments which produce periodic sounds. Most instruments which make a "sustained" sound (wind instruments, bowed strings, voice) produce periodic sounds. If you look at the design of other instruments (piano, guitar, percussion), it often revolves around getting the overtones to be as close to integer multiples of the fundamental as possible. For instance, as we will discuss when we talk about piano, the longer, thinner, and higher tension a plucked or hammered string is, the more harmonic the overtones. One of the reasons that upright pianos sound bad in their lower registers is that the strings are not producing harmonic overtones. Check to see whether an upright's bottom octave sounds worse to you when played very loud (when the aural harmonics are important) than when played softly (when they are less prominent).

Chapter 17

Intervals, Scales, and Tunings

17.1 Reason behind intervals

Rather than using any available frequency, most musical traditions use several fixed frequencies. These should be chosen (or have been chosen) to ensure the presence of lots of harmonies, that is, pairs of notes with overlapping harmonics. In particular, it is *essential* that if a frequency f is a standard note, that twice the frequency, $2 \times f$, also be a standard note (one **octave** up). It is also very nice that, if f is a standard note, the note at 1.5 times that frequency (up by one **fifth**), also be a standard note. Recall that our frequency sense cares about the logarithm of frequency.

There is an amazing numerical coincidence:

$$2^{7/12} \simeq 1.5 = \frac{3}{2} \qquad \text{(actually it is 1.4983...)}$$

which means that if we divide the octave into 12 evenly spaced (in logarithm) frequency steps, then notes separated by 7 of those steps will be related by a musical fifth, the ratio 1.5. Further, because

$$\frac{2}{3/2} = \frac{4}{3}, \qquad \text{so} \qquad 2^{5/12} \simeq \frac{4}{3} \qquad \text{(actually it's 1.3348...)}$$

we will also find that notes separated by 5 of these steps are spaced by a ratio of 4/3, the musical **fourth**. (We will see where these names come from in a moment). Furthermore, the following approximation, while not quite so accurate, is not too bad:

$$2^{4/12} \simeq 1.25 = \frac{5}{4} \qquad \text{(really 1.2599)} \qquad \text{and} \qquad 2^{3/12} \simeq 1.20 = \frac{6}{5} \qquad \text{(really 1.1892)}$$

Therefore such spacing of notes also gives the two next most important intervals, called the **major third** and **minor third**, fairly well, though not perfectly.

17.2 Musical scale

Because of these numerical accidents, music designed around 12 (approximately) even divisions of the octave, that is, frequencies which are some standard times 1, $2^{1/12}$, $2^{2/12}$, $2^{3/12}$, $\ldots$ $2^{12/12} = 2$, is a good idea, and several musical traditions have converged on this idea: the Western musical tradition, the Eastern pentatonic tradition, and West African tradition (which uses $2^{1/24}$, half the spacing). In Western tradition, the spacing between notes separated in frequency by a factor of $2^{1/12}$ is called a **half-step** and we will use this nomenclature from now on:

Two notes of frequency f_1 and f_2 are said to be separated by a **half-step**, with f_2 a half-step higher than f_1, if $f_2/f_1 = 2^{1/12}$, or $\log_2(f_2/f_1) = 1/12$ (which is the same).

How should these notes be used in music? It turns out that using all of them encounters numerous dissonances, as well as the consonances we designed for. In particular, these intervals are quite dissonant:

- notes separated by one half-step, $f_2/f_1 = 2^{1/12} = 1.05946$. The notes are so close that their critical bands badly overlap. Similarly, 13 half-steps is bad because of the 2'nd harmonic of the lower note.
- notes separated by 11 half-steps, $f_2/f_1 = 2^{11/12} = 1.8877$. The lower note's second harmonic partly overlaps the upper note, causing dissonance.
- notes separated by 6 half-steps, $f_2/f_1 = 2^{6/12} = 1.4142$. Several higher overtones come too close to overlapping.

Therefore it is a good idea to only use some of the 12 available notes, to make these overlaps rare. (Early to mid 20'th century composers called 12-tonalists experimented with using all 12 without prejudice. While the resulting music had some success within the experimental music community, it never gained wide popularity; the public does not want that much dissonance.)

In Western tradition, this problem is solved by using the following notes: some chosen fixed note, call it **Do**, and the notes separated from it by 2, 4, 5, 7, 9, 11, and 12 (and then 12+2, 12+4, 12+5, 12+7, ...) half-steps–or, if you prefer, moving forward by 2 half steps, 2 half steps, 1 halfstep, 2 half-steps, 2 half-steps, 2 half-steps, 1 half-step:

frequency	# half-steps	name
$f_0 \times 2$	12	Do
$f_0 \times 2^{11/12} = 1.8877$	11	Ti
$f_0 \times 2^{9/12} = 1.6818$	9	La
$f_0 \times 2^{7/12} = 1.4983$	7	Sol
$f_0 \times 2^{5/12} = 1.3348$	5	Fa
$f_0 \times 2^{4/12} = 1.2599$	4	Mi
$f_0 \times 2^{2/12} = 1.1225$	2	Re
$f_0 \times 1$	0	Do

By making two of the spacings one half step and the rest of them two, there are lots of pairs separated by 5 or 7 half-steps, and only one combination separated by 6 half-steps. To clarify, at the top, the note one octave above the starting note is also called Do, and one then repeats the spacings of intervals and the names of the notes. That is,

# half-steps	name
...	...
26	Fa
24	Do
23	Ti
21	La
19	Sol
17	Fa
16	Mi
14	Re
12	Do
11	Ti
9	La
7	Sol
5	Fa
4	Mi
2	Re
0	Do
-1	Ti
...	...

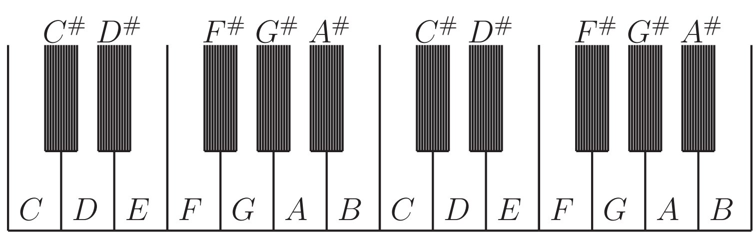
Incidentally, the fastest way to describe pentatonic is to say it is like the above, but never using Mi or Ti. That is, instead of the intervals going 2 2 1 2 2 2 1, they go 2 3 2 2 3. This avoids *ever* encountering 1, 6, or 11 half-steps, the three most dissonant intervals. The result is music which is highly consonant, but considered by Western ears to be a little boring (all resolution, no tension).

I have not said anything about what frequency Do should be. We can choose it to be whatever we want–say, 200 Hertz, or 210 Hertz, or 317.5 Hertz, whatever. However, because instruments like pianos are difficult to tune, it is a good idea to choose one special standard value. Then tune the piano, and allow someone to start with "Do" being either the standard value or anything separated by some number of half-steps from that standard value.

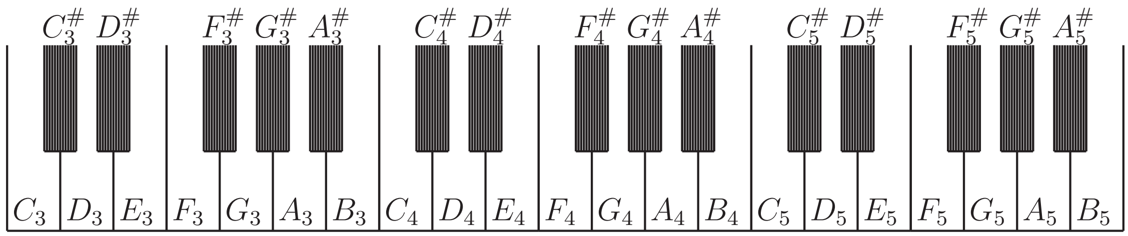
The standard value for Do is named C, and Re etc are named Re=D, Mi=E, Fa=F, Sol=G, La=A, and Ti=B. It has become conventional to tune A to 440 Hertz. In that case we name the notes as follows:

halfsteps above Do	Name
0	C
1	$C^{\#} = D^{\flat}$
2	D
3	$D^{\#} = E^{\flat}$
4	E
5	F
6	$F^{\#} = G^{\flat}$
7	G
8	$G^{\#} = A^{\flat}$
9	A
10	$A^{\#} = B^{\flat}$
11	B
12	C
...	...

On the piano keyboard, it looks like this:

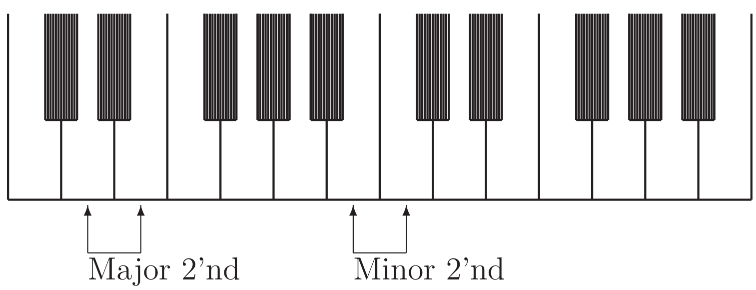


Now the problem is that there are actually several notes named A: the note at 440 Hertz, and all notes separated by it by some number of octaves: going down, 220 Hertz, 110 Hertz, 55 Hertz, 27.5 Hertz, and going up, 880 Hertz, 1760 Hertz, 3520 Hertz, and so forth. We distinguish these by putting little subscripts on them, as follows. The lowest C which appears on a piano keyboard is called C_1; the notes going up from it are D_1, E_1, and so forth, until B_1; then the next C is called C_2. That is, as you go up the keyboard, each time you reach C, you add one to the little index. Middle C is C_4, at about 261 Hertz. A_4 is the one with a frequency of 440 Hertz. **Careful**: the index changes at C, *not* at A. That is too bad, because the lowest note on a piano keyboard is actually an A (at 27.5 Hertz), which is unfortunately called A_0.



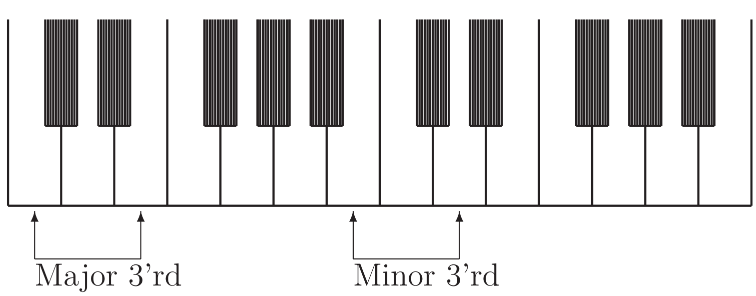
The nomenclature for intervals is as follows: the interval (frequency ratio) between two white keys on the keyboard is the number of white keys, starting with the lower key, counting all the ones between them, and *counting* the top key. That is, if two keys are next to each other, the interval is a **second**; if they are separated by 1 key, it is a **third**; by two keys, a **fourth**; and so forth. Because there are two places where there is no black key between white keys, most intervals can occur in one of two ways; with a larger number of black keys between, or with a smaller number between. The way with more black keys between, which means that there are more half-steps, is called a **major** interval; with fewer, a **minor** interval. Since this is confusing, I will illustrate the first few examples.

Look first at the seconds (neighboring keys):



There are 7 ways of taking pairs of neighboring keys. In 5 of them, there is a black key in between, so they are separated by 2 half-steps. In two cases there is no black key, so they are separated by only 1 half-step. Therefore a separation of 2 half-steps is called a **major second** (also a **whole step**); a separation of one half-step is a **minor second** (also **half-step**).

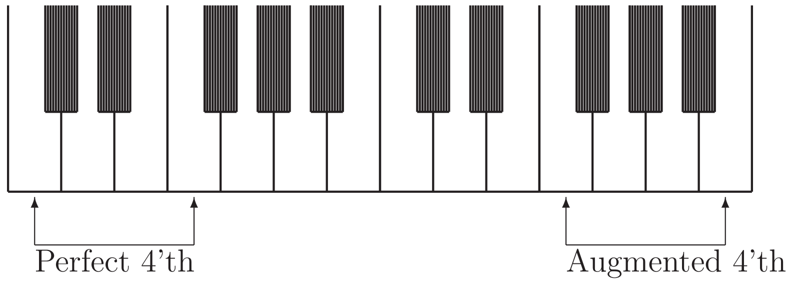
For the thirds:



4 cases involve 3 half-steps, and 3 cases involve 4 half-steps. Therefore a separation of 3

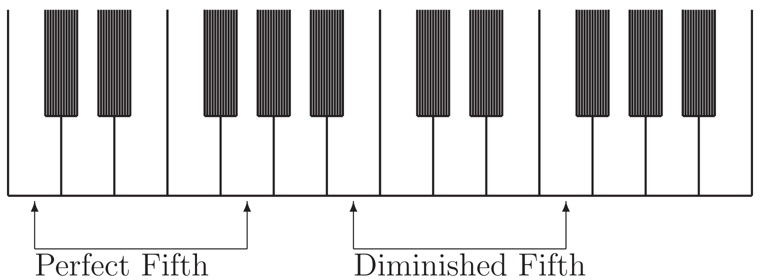
half-steps is called a **minor third**, and 4 half-steps is called a **major third**.

For fourths:



All but one case involves 5 half-steps, which is also an ideal musical interval. Since it is so common and so nice, it is called a **perfect 4'th**. The other case is 6 half-steps, called an **augmented 4'th**.

For the fifths:



All but one case involves 7 half-steps, again a special interval called the **perfect fifth**. The one case with 6 half-steps is called a **diminished fifth**. This is the *same thing* as an augmented fourth. Just for fun, it has an extra name, the **tritone** (I don't know why).

For the sixths, they are either 8 or 9 half-steps; 8 is a **minor sixth**, 9 a **major sixth**. For the sevenths, it is either 10 or 11; a **minor 7'th** or a **major 7'th**. For the eighth, it is *always* 12 steps, called a **perfect 8'th**, an **8'th**, or most commonly an **octave**. One can define 9'ths, 10'ths, and so forth, but they are just 12 half-steps plus the 2'nds, 3'rds, and so forth.

In short:

# semitones	Name
1	minor second (2'nd)
2	major second (2'nd)
3	minor third (3'rd)
4	major third (3'rd)
5	perfect fourth (4'th)
6	augmented 4'th, diminished 5'th, tritone
7	perfect fifth (5'th)
8	minor sixth (6'th)
9	major sixth (6'th)
10	minor seventh (7'th)
11	major seventh (7'th)
12	octave

17.3 Tunings

The way I have described tuning–using $2^{1/12}$ for the half-steps–is called **equal tempered tuning**. It is pretty standard, because keyboard instruments cannot be re-tuned at will, and it is common to play in many different keys. However, it has some disadvantages; none of the perfect 4'th and 5'ths are actually frequency multiples of 4/3 and 3/2, and none of the major and minor thirds are multiples of 5/4 or 6/5. Historically (and in practice!) there are a few other ways of tuning, where one does *not* make all the intervals the same size, in order to make these intervals more perfect.

Pythagorean tuning

One makes all octaves exact factors of 2, and all but 1 of the 5'ths exact factors of 3/2. This is the oldest and most traditional way of tuning, and it has the advantage that it is easy to do; one uses beats to determine if two notes are in tune. If there are beats, they are not in tune, and must be changed until they are in tune.

The tuning works like this: one first tunes A_4 to be 440 Hertz (say). Then tune the notes up by a fifth, E_5, by making it 3/2 the frequency of the A. Tune the note up a fifth from E_5, B_5, to be 3/2 the frequency of E_5. Also, you can tune downwards from A_4; down a fifth is D_4, tuned to have 2/3 the frequency of A_4; down from it is G_3, tuned to have 2/3 the frequency of D_4, and so on. The notes differing by an octave are tuned by using that an octave should be an exact factor of 2.

The series of notes you get, by going up and down by fifths, is called the "circle of 5'ths", because after going 12 fifths you come back to the same letter note you started with:

$$\ldots G^{\#} = A^{\flat} \quad E^{\flat} \quad B^{\flat} \quad F \quad C \quad G \quad D \quad A \quad E \quad B \quad F^{\#} \quad C^{\#} \quad G^{\#} = A^{\flat} \ldots$$

This way of tuning was discovered by Pythagoras (yes, the triangle dude) about 2700 years ago. It would work perfectly, except that the $G^{\#}$ you get to at the top does not differ from the $A^{\flat}$ you get to at the bottom by the 7 octaves that it should. That would happen if

$$\left(\frac{3}{2}\right)^{12} \stackrel{?}{=} 2^7 \qquad \text{or} \qquad 3^{12} \stackrel{?}{=} 2^{19}$$

which it does not quite: they are actually different by about 1.36%. That means that one of the 12 fifths has to be mis-tuned. It is usually chosen to be a fifth between two of the sharped or flatted notes, on the theory that it will not be played very often.

Most of the time one plays in keys where the “wrong” 5’th is not used, and everything is fine. However, when you do play in a key where that 5’th is used, it sounds wrong. There are sound file examples as extras for this lecture.

Just tuning

This is a more ambitious attempt to get the thirds, as well as the fifths, to be in tune. If you tune an instrument this way and play in the “right” key, it sounds great. Play in the “wrong” key, and it sounds painfully bad.

The problem is that the “circle of thirds” runs into trouble much faster than the circle of fifths. Going up by major thirds from C, one encounters $C, E, G^{\#}, C$. If we force each interval to be a factor of 5/4, then the difference from C to C would be $(5/4)^3 = 125/64$, which is *not* $2 = 128/64$. In fact, it is pretty far off. That means that one major third out of every three has to be out of tune, by a big margin. No matter how you decide which ones to get “right,” there are many keys which will require using the ones which are wrong. If you listen to the music accompanying this lecture, you will see that there are several keys where the just tuned music sounds beastly.

For these reasons, usual practice is,

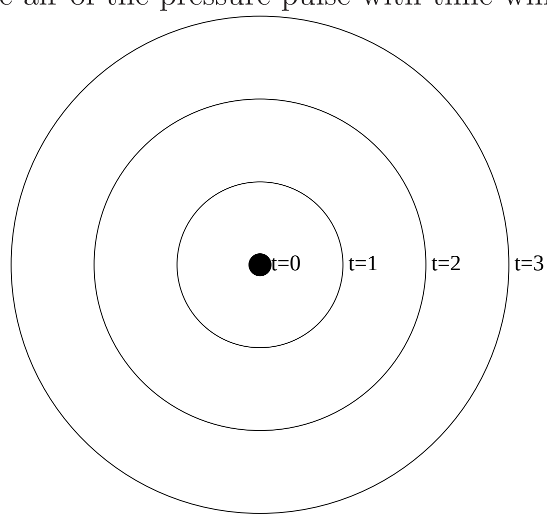
- Tune instruments in equal temperament (all half-steps equal, $2^{1/12}$);
- Cheat whenever your ear tells you to, particularly by playing the higher tone in a major third flat and the higher tone in a minor third sharp.

Chapter 18

The Doppler Effect

The doppler effect is the phenomenon we have all noticed, that a sound produced by a moving source, or which you hear while you are moving, has its perceived frequency shifted. When a car is coming towards you, the car stereo sounds higher pitched and the tempo sounds faster. As it drives away, the sound is lower pitched and the tempo is slower. Let us see why, and by how much.

The first thing to understand and remember is that sound *is a phenomenon in the air*. It is, it is each piece of air pushing on the piece next to it. Once a sound is produced by a source, it will propagate out at the speed of sound in the air–no matter what the source of the sound does. That is, if a source of sound makes a sudden bang (pressure pulse) at time 0, then the location in the air of the pressure pulse with time will look like:



This is what happens *no matter where* the thing which made the bang goes to. That is, the pressure from a sudden banging is a sphere centered *where the sound was made*, *not* at the current location of the sound source. For a moving source, letting out a series of "bangs," the location of pressure peaks will therefore look like what is shown in figure 18.1. In the

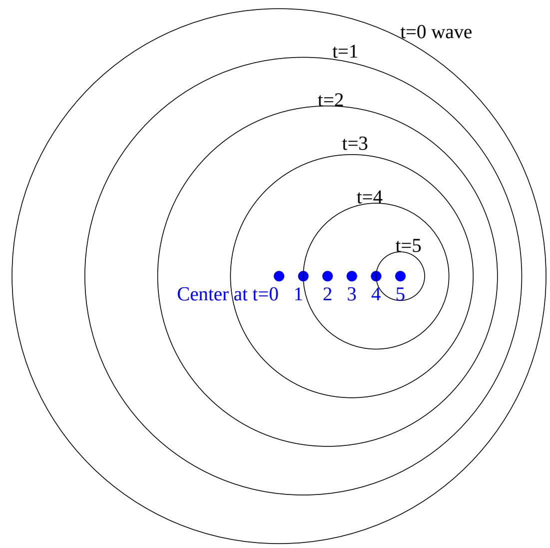


Figure 18.1: Pressure peaks radiated by a moving sound source. The blue dots represent the time-series of locations of the sound source when it made a series of "bangs," the black circles represent the locations of the pressure peaks at a time slightly after the last "bang" was made.

figure, the blue dots are the location of the sound maker when each bang was produced, and the circles are the locations of the pressure peaks radiating out from those bangs. Notice that, because the sound maker is moving to the right, the pressure peaks are closer together on the right and further apart on the left.

Now remember that a sound is just a series of pressure peaks, which are tightly separated in time, by the period of the wave. Therefore, instead of thinking of the sound waves in these two pictures as the locations of pressure peaks due to "bangs," you can think of them as pressure peaks in a periodic sound wave. In front of the moving object in figure 18.1, the peaks are closer together. That means that the wave length is shorter, which means it is a higher frequency wave. Behind, the peaks are farther apart, meaning that it is a longer wavelength sound, at a lower frequency. This is the gist of the Doppler effect.

Let us now actually calculate the size of the effect.

Suppose a sound source is moving right at you, at velocity v. At time 0, it emits a pressure peak. At time Δt, it emits a second pressure peak. If its distance from you at time 0 was x, its distance from you at time Δt was $x - v\Delta t$ (it is nearer, since it is moving towards you).

The times that the two pressure peaks will reach you, are the time they were emitted plus the propagation time. For the first peak, that means,

$$t_{\text{arrival 1}} = 0 + \frac{x}{v_{\text{sound}}}$$

while for the second peak, it is

$$t_{\text{arrival 2}} = \Delta t + \frac{x - v\Delta t}{v_{\text{sound}}} \,.$$

The actual period and frequency of the sound–the period and frequency of the sound which is being produced–are,

$$T_{\text{produced}} = t_{\text{pulse 2}} - t_{\text{pulse 1}} = \Delta t \qquad f_{\text{produced}} = \frac{1}{T_{\text{produced}}} = \frac{1}{\Delta t} \,.$$

The period and frequency *you perceive*, are the time difference of the arrivals of the pulses, and its inverse:

$$\begin{aligned}
T_{\text{observed}} &= t_{\text{arrival 2}} - t_{\text{arrival 1}} \\
&= \Delta t + \frac{x - v\Delta t}{v_{\text{sound}}} - \frac{x}{v_{\text{sound}}} \\
&= \frac{v_{\text{sound}}\Delta t}{v_{\text{sound}}} + \frac{x - v\Delta t}{v_{\text{sound}}} - \frac{x}{v_{\text{sound}}} \\
&= \frac{v_{\text{sound}}\Delta t + x - v\Delta t - x}{v_{\text{sound}}} \\
&= \frac{v_{\text{sound}}\Delta t - v\Delta t}{v_{\text{sound}}} = \frac{v_{\text{sound}} - v}{v_{\text{sound}}}\Delta t
\end{aligned}$$

and

$$f_{\text{observed}} = \frac{1}{T_{\text{observed}}} = \frac{v_{\text{sound}}}{v_{\text{sound}} - v}\frac{1}{\Delta t} = \frac{v_{\text{sound}}}{v_{\text{sound}} - v} f_{\text{produced}}$$

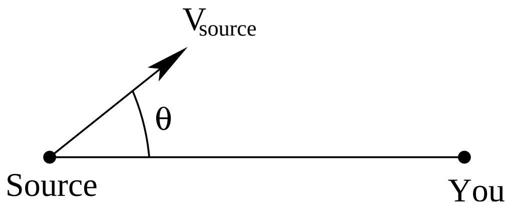
or,

$$\boxed{\text{Approaching Source:} \;\; \frac{f_{\text{observed}}}{f_{\text{produced}}} = \frac{v_{\text{sound}}}{v_{\text{sound}} - v_{\text{source}}} \,.}$$

Note that the time, distance, and Δt all do not matter to the final result.

The quick way to understand (and derive) this result is as follows. The time between when you receive two pressure peaks, is the time between when they were made, plus the time difference in their propagation times. That time difference is the distance away from you the source moved, between the pulses, divided by the speed of sound. The distance is the time times the speed of the source.

What if the source is moving away from you? Then the time the second pulse needs, to reach you, will be longer instead of shorter. We just got the *sign* of v_{source} backwards in the above. What if the source is moving at some general angle θ with respect to the line between you and the source? Then how much closer the source gets to you, gets multiplied by $\cos(\theta)$:



$$\boxed{\text{Moving source: } \frac{f_{\rm observed}}{f_{\rm produced}} = \frac{v_{\rm sound}}{v_{\rm sound} - v_{\rm source}\cos\theta}}$$

What if you are the one who is moving? The answer is *not* just that you should use your speed for $v_{\rm source}$. We actually have to re-think the calculation we did above.

First we should think about what is happening, to figure out what to expect in the answer. If you are moving towards a source, then not only are the pressure peaks moving out towards you; you are moving in at them, so you will cross them more often. Therefore you should hear a higher frequency. Moving away, the pressure pulses are approaching you, but you are running from them, so they reach you less often, and you will hear a lower frequency.

Now let's do the calculation, considering first the case where you are moving straight at the sound producing source. This time, the sound source will not be moving, but you will. Suppose again that the source makes a pressure pulse at time 0 and another at time Δt. If you are at a distance x when the first pressure pulse occurs, then it will reach you at a time, when its distance from the source and your distance from the source are the same. Call that time t_1, and let us try to find it. The sound wave's distance from the source is $v_{\rm sound} t_1$. Your distance from the sound is $x - vt_1$. These equal at

$$v_{\rm sound} t_1 = x - vt_1 \quad \rightarrow \quad (v_{\rm sound} + v)t_1 = x \quad \rightarrow \quad t_1 = \frac{x}{v_{\rm sound} + v}$$

Now we have to find the time when the second pressure pulse meets you. Call that time t_2. The distance from the source of the second pulse is $v_{\rm sound}(t_2 - \Delta t)$, and your position is $x - vt_2$; they equal at time,

$$v_{\rm sound}(t_2 - \Delta t) = x - vt_2 \quad \rightarrow \quad (v_{\rm sound} + v)t_2 = x + v_{\rm sound}\Delta t\,,$$

or,

$$t_2 = \frac{x}{v_{\rm sound} + v} + \frac{v_{\rm sound}\Delta t}{v_{\rm sound} + v}\,.$$

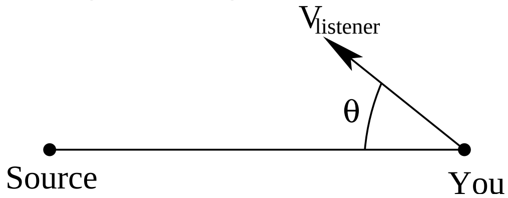
The period of the sound, as perceived by you, is the time difference between pressure peaks:

$$\begin{aligned} T_{\rm observed} &= t_2 - t_1 = \frac{x}{v_{\rm sound} + v} + \frac{v_{\rm sound}\Delta t}{v_{\rm sound} + v} - \frac{x}{v_{\rm sound} + v} \\ T_{\rm observed} &= \frac{v_{\rm sound}}{v_{\rm sound} + v}\Delta t \end{aligned}$$

from which it follows that,

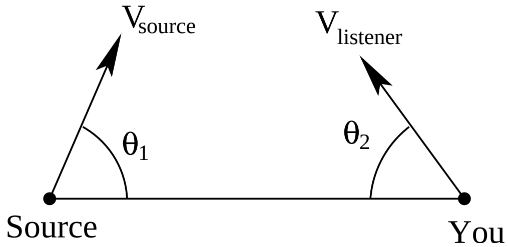
$$\frac{T_{\rm observed}}{T_{\rm produced}} = \frac{v_{\rm sound}}{v_{\rm sound} + v}\,, \qquad \text{or} \qquad \frac{f_{\rm observed}}{f_{\rm produced}} = \frac{v_{\rm sound} + v_{\rm listener}}{v_{\rm sound}}\,.$$

The way that v_{listener} entered here, is that it is how fast you were moving towards the sound fronts. If you are moving straight away, you should flip its sign. If you are moving at some general angle θ, then the result is multiplied by the cosine of that angle:



$$\boxed{\text{Moving listener: } \frac{f_{\text{observed}}}{f_{\text{produced}}} = \frac{v_{\text{sound}} + v_{\text{listener}}\cos\theta}{v_{\text{sound}}}}$$

What if both are moving? You should not be too surprised by the answer:



$$\boxed{\text{Both Move: } \frac{f_{\text{observed}}}{f_{\text{produced}}} = \frac{v_{\text{sound}} + v_{\text{listener}}\cos\theta_2}{v_{\text{sound}} - v_{\text{source}}\cos\theta_1}}$$

There are two cases in which these answers do not seem to make sense. The first is that if $v_{\text{source}} \geq v_{\text{sound}}$ and the source is moving at you, it looks like the frequency can become *infinite*. This is actually **correct**! For a sound source moving at you, as it approaches the speed of sound, the frequency you hear becomes extremely sharp. What is happening is that the sound fronts, pictured in figure 18.1, are piling up and actually touching. All the pressure peaks are happening at once. If you try to go faster than the speed of sound, you run into the pressure pulses you created before you got up to that speed. For an airplane, that means that a huge overpressure in front of the plane is encountered just as it reaches the speed of sound. This is called the **sound barrier**, and it is why commercial airflights all (now that the Concorde is retired) fly slower than the speed of sound, and why it took so long to develop airplanes which could handle this and exceed the speed of sound. Surprisingly, it is much easier to fly an airplane at twice the speed of sound (Mach 2) than right at the speed of sound (Mach 1), for this reason.

The other case which looks strange is what happens if you are moving away from a sound source, $\cos\theta_2 = -1$, and $v_{\text{listener}} \geq v_{\text{sound}}$. Then the numerator becomes zero, and I find a *vanishing frequency*. Can that be right? This is **also correct!** This is the case where you are moving away from a sound source, at or above the speed of sound. Each pressure pulse is moving towards you, but you are moving back away from it at least as fast as it approaches you. That means that it never reaches you; *you can run away from a sound* and it never catches up (if you can go faster than v_{sound}).

Chapter 19

Reflection and Impedance

Let us look again at the question of what happens when a sound wave in the air runs into the surface of a body of water (or a wall, or glass, or whatever). How much of the wave is reflected, and how much becomes a sound wave inside the water?

Recall first that the velocity of the air due to the wave, and the pressure of the air due to the wave, were related via

$$P - P_{\rm atmos} = (\rho_{\rm air} v_{\rm sound,air})(v_{\rm air} - v_{\text{``wind''}})$$

From now on I will write ΔP as the pressure difference from atmospheric and will ignore the possibility of wind speed, so $v_{\rm air}$ is the air motion because of the sound wave.

The relation between pressure change and velocity change in a sound wave is true of *all* media, it is not special to air. The key is that you have to use the speed of sound and density of the medium you are thinking about. Because it comes up so often, the product $\rho v_{\rm sound}$ (a characteristic of a material) is called the **mechanical impedance** Z:

$$\textbf{Mechanical impedance} \quad Z \equiv \rho v_{\rm sound} \, .$$

If you compare Z between air and most other materials, you find out *both* that $\rho_{\rm air}$ is small (typically 500–5000 times smaller than in another material), *and* that $v_{\rm sound}$ in the air is small (typically 3–10 times smaller than in another material). Therefore the mechanical impedance Z of a solid or liquid will be far larger than for air.

Now consider a sound wave in the air hitting the surface of a lake or other body of water. The sound wave approaching the water contributes a pressure which I will call $\Delta P_{\rightarrow}$ and an air velocity $v_{\rightarrow}$, which satisfy

$$\Delta P_{\rightarrow} = Z_{\rm air} v_{\rightarrow} \, .$$

Note that I am measuring v in terms of the direction the sound wave is moving, that is, from the air into the water. There will be a sound wave in the water, moving away from the air.

It will have a pressure and water velocity

$$\Delta P_{H_2O} = Z_{H_2O} v_{H_2O} \,.$$

To relate these, I have to think about what happens right at the surface. The air ends exactly where the water begins. Therefore, however fast the air is moving right at the surface, the water must be moving at the same speed. Also, the air pushes on the water exactly as hard as the water pushes on the air. Therefore,

$$v_{\text{air}} = v_{H_2O} \qquad \text{and} \qquad \Delta P_{\text{air}} = \Delta P_{H_2O} \,.$$

If you try to put these together with the results for the pressures and velocities, you will find that the only way to get *both* things to be true, is either to have the Z's be equal, or to have the ΔP's both be zero. The reason is, of course, that I tricked you by leaving out something important. There will always be a reflected wave going off into the air. It satisfies,

$$\Delta P_{\leftarrow} = -Z_{\text{air}} v_{\leftarrow}$$

because the wave is going in the opposite direction. The total air motion and pressure are the sum of those from the two waves:

$$\Delta P_{\text{air}} = \Delta P_{\rightarrow} + \Delta P_{\leftarrow} \,, \qquad v_{\text{air}} = v_{\leftarrow} + v_{\rightarrow} \,.$$

The next part is derivation and you can skip to the end if you don't want to see it.

[At this point I have 5 equations (3 relations between ΔP and v for each wave, and the two equalities at the water surface) in 6 unknowns (three pressures and three velocities). I can use the relations between pressures to express the v's in terms of ΔP's and write both equalities in terms of pressures:

$$\begin{aligned} \Delta P_{\rightarrow} + \Delta P_{\leftarrow} &= \Delta P_{H_2O} \\ \frac{\Delta P_{\rightarrow}}{Z_{\text{air}}} - \frac{\Delta P_{\leftarrow}}{Z_{\text{air}}} &= \frac{\Delta P_{H_2O}}{Z_{H_2O}} \end{aligned}$$

In the last expression, multiply through by Z_{air}. Then add the top expression:

$$2\Delta P_{\rightarrow} = \frac{Z_{\text{air}} + Z_{H_2O}}{Z_{H_2O}} \Delta P_{H_2O} \,, \tag{19.1}$$

$$\Delta P_{H_2O} = \frac{2Z_{H_2O}}{Z_{H_2O} + Z_{\text{air}}} \Delta P_{\rightarrow} \,, \tag{19.2}$$

$$\text{and also,} \qquad \Delta P_{\leftarrow} = \frac{Z_{H_2O} - Z_{\text{air}}}{Z_{H_2O} + Z_{\text{air}}} \Delta P_{\rightarrow} \tag{19.3}$$

This relates the pressure of the incoming sound wave to the pressure of the sound wave in the water. You see that the sound wave in the water has about twice the pressure of the one in the air, since Z_{H_2O} is much larger than $Z_{\rm air}$; so the relation is almost $\Delta P_{H_2O} = 2\Delta P_{\rightarrow}$. That makes it sound like the wave in the water is not so small after all. But you have to remember the relation between intensity and pressure, and that what really counts is the intensity. The intensity of the sound wave in the water is,

$$I_{H_2O} = \frac{\Delta P^2_{H_2O}}{Z_{H_2O}} = \frac{4Z_{H_2O}}{(Z_{\rm air} + Z_{H_2O})^2}\Delta P^2_{\rm air} = \frac{4Z_{\rm air}Z_{H_2O}}{(Z_{\rm air} + Z_{H_2O})^2} I_{\rm air}\,.$$

Similarly, you can take equation (19.3) and square both sides to find that

$$I_{\rm reflected} = \frac{(Z_{H_2O} - Z_{\rm air})^2}{(Z_{\rm H_2O} + Z_{\rm air})^2} I_{\rm air}\,.$$

That ends the derivation and you can start reading again.]

Now rewrite that for an incident (incoming) sound wave going from a medium of impedance Z_1 to a medium of impedance Z_2. There is a transmitted wave (the one which makes it into the new medium) and a reflected wave, of intensities,

$$\boxed{\frac{I_{\rm transmitted}}{I_{\rm incident}} = \frac{4Z_1Z_2}{(Z_1{+}Z_2)^2} \quad \text{and} \quad \frac{I_{\rm reflected}}{I_{\rm incident}} = \frac{(Z_1{-}Z_2)^2}{(Z_1{+}Z_2)^2}}$$

When I_1 and I_2 are very different (as for air and water), these expressions mean that almost all the energy is reflected and not transmitted: $I_{\rm transmitted}/I_{\rm incident} \simeq 4Z_1/Z_2$ if Z_1 is smaller. Note that the result here is for right-angle incidence. It is different at a different angle, but I will not present that just so as not to confuse matters further.

For air and water, since

$$\begin{aligned}
Z_{\rm air} = \rho_{\rm air} v_{\rm sound,air} = 1.2\ {\rm kg/m^3} \times 344\ {\rm m/s} &= 400\frac{\rm kg}{\rm m^2\,s} \\
Z_{\rm water} = \rho_{H_2O} v_{{\rm sound},H_2O} = 1000\ {\rm kg/m^3} \times 1400\ {\rm m/s} &= 1400000\frac{\rm kg}{\rm m^2\,s} \\
\frac{Z_{\rm air}}{Z_{\rm water}} &\simeq \frac{1}{3500}
\end{aligned}$$

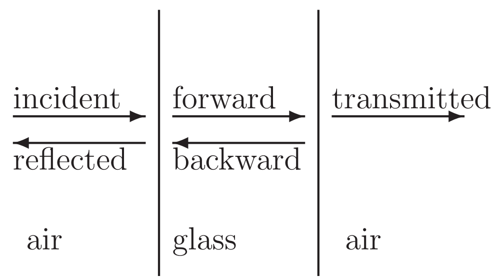
which is indeed tiny. This means, using the formulas above, that when sound reflects off water, only 4/3500 of the intensity makes it through. That is a reduction of almost 30 dB in the intensity of the sound. Note that sound, moving through the water, reflects just as efficiently to stay in the water and let only a tiny amount into the air.

One other thing about the reflection problem just considered. The pressure of the incident wave and of the transmitted wave were of the same sign. What about the reflected wave? This turns out to depend on whether $Z_1 < Z_2$ (as for air to water), or $Z_1 > Z_2$ (as for water to air). For $Z_1 < Z_2$, the reflected wave has the same sign of ΔP and opposite sign of v. For the other case, $Z_1 > Z_2$, the reflected wave has the same sign of v and opposite sign of ΔP.

You can see this in equation (3) above. This proves important in many problems, and so it is worth thinking about a little more.

The way to think about it is, that when Z is small, the medium is easy to move around (like the air–you don't have to push on it hard to make it move). If the other medium is hard to move around, then the velocities of the incoming and outgoing waves will need to cancel, which happens (since they are waves going in opposite directions) when the pressures are the same. When it is the high Z medium (water) which the sound wave is coming from, then the medium is hard to move. It only takes a tiny pressure to get the other medium to move, so the incoming and outgoing waves must have opposite pressure from each other.

Now think about a pane of glass. Instead of three waves–incoming, transmitted, and reflected–there are five:



This problem is more complicated. The problem is that the time when the incoming wave has its peak pressure need not be the time when the forward and backward waves have their peak pressures. I will not drag you through a derivation of what happens in this case, but I will quote two limits for the answer.

First, think about a very thin panel of glass. The forward wave has the same sign of pressure as the incident wave. The backward wave, though, has the opposite sign of pressure, because of the remarks about reflected waves we made above. That means that the forward wave can have a bigger pressure, and still satisfy the matching condition where the air and glass meet. That means that much more sound will get through. The only reason the sound does not *all* go through, is that the short time the wave is moving in the glass gets the pressure peak of the forward and backward moving waves to occur at slightly different times. This depends on the wavelength of the sound and thickness of the glass. The answer, after much work, turns out to be,

$$\frac{I_{\rm transmitted}}{I_{\rm incident}} \simeq \left[\frac{Z_{\rm air}}{Z_{\rm glass}}\right]^2 \left[\frac{v_{\rm sound,glass}}{\pi f d_{\rm glass}}\right]^2 , \tag{19.4}$$

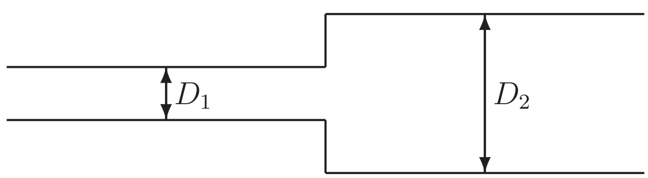
where $d_{\rm glass}$ is the thickness of the glass. This formula shows that glass lets through more low frequency sound than high frequency sound, something you have probably all experienced or can easily check at home. The sounds from outside when you close the window are muffled;

not just quieter, but different in timbre, specifically, losing 6 dB of loudness per octave frequency compared to the original sound.

Now, what about a very thick piece of stuff? In this case, essentially you can forget about the backward wave in considering the transmission from air into the glass. Then you just get the product of the amount of transmission from air to glass, times the transmission from glass back to air.

If we put typical numbers for the walls of a musical instrument into equation (19.4), we find something like 30 to 40 dB of attenuation. That is, the walls of the instrument hold in essentially all of the sound bouncing around inside of the instrument. Therefore, we can think of the sound inside an instrument as being perfectly contained by the walls of the instrument. This leads us to think about sound traveling in a tube.

Consider sound traveling in a tube, where the walls of the tube hold the sound in perfectly. What happens if the tube widens abruptly, say, from diameter D_1 to diameter D_2, both much smaller than the wavelength of the sound?



Consider the sound wave traveling down the narrow part of the tube, reaching this widening. Instead of talking about the air pressure and the air velocity, talk about the air pressure and the air flow rate. Surely, the rate the air is flowing has to be the same just before the widening as just after, since air cannot go anywhere else. The pressure must also be the same. But the air flow rate is just,

$$\text{flow} = Av_{\text{air}} = \frac{\pi D^2}{4} v_{\text{air}}$$

We have to have $A_1 v_1 = A_2 v_2$ and $\Delta P_1 = \Delta P_2$. If we simply define a quantity called **Acoustic Impedance**,

$$Z_A \equiv \frac{\rho v_{\text{sound}}}{A}$$

then the relation between pressure and air flow is,

$$\Delta P = Z_A \, \text{flow} \, .$$

In computing how much air reflects from the juncture of the two pipes and how much is transmitted, we can re-use all our previous work, just using Z_A instead of Z. The transmitted and reflected powers are,

$$\boxed{\frac{\text{Power}_{\text{transmitted}}}{\text{Power}_{\text{incident}}} = \frac{4 Z_{A1} Z_{A2}}{(Z_{A1} + Z_{A2})^2} \quad \text{and} \quad \frac{\text{Power}_{\text{reflected}}}{\text{Power}_{\text{incident}}} = \frac{(Z_{A1} - Z_{A2})^2}{(Z_{A1} + Z_{A2})^2}}$$

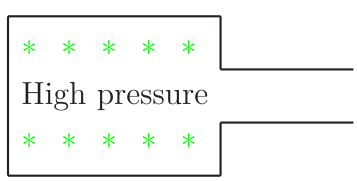
What about the sign of the pressure for the reflected wave? A wide pipe has a low Z_A (it is easy for the air to move), a narrow pipe has a high Z_A (the air moves very little). Therefore, we can repeat the arguments we made about high and low Z material. When a narrow pipe opens out, the large pipe will take an airflow easily. The reflected wave therefore has the same sign of airflow and opposite sign of pressure. When a wide pipe narrows, the pressure of the reflected wave is the same as the incident wave.

Chapter 20

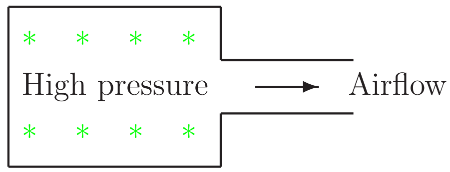
Resonance and Cavities

A key idea–perhaps the key idea–in musical instruments is the idea of resonance. Resonance is what you call it when some system can store energy two different ways, and energy goes back and forth from one form to the other, generally in a sine wave pattern. The example you are probably most familiar with is a mass suspended from a spring. You might be even more familiar, from your childhood, with someone swinging on a swing. In this case, the two ways energy can be stored are, motion of the person on the swing, and the person being raised up into the air (energy stored in gravitational potential). Starting when the person is at the top of the swinging motion, gravity pulls them down and therefore forward. As they gather speed, the energy is moving from gravitational potential into their motion. At the bottom of the swinging motion, all the energy is in motion. But once in motion, you remain so until something stops you, so you keep right on going forward and upward. The energy is then getting taken out of motion into height (gravitational potential) again. At the top of the swing pattern, the whole thing repeats but with backward motion. If it were not for air resistance, losses in the swing chains, and so on, you would swing back and forth forever.

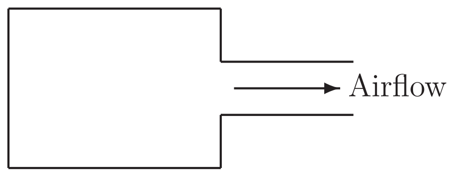
The same idea of resonance comes up in virtually every musical instrument. Since we have been studying sound, air pressure, and air motion, let us look at a nice example involving these ideas. Consider a wine bottle, or any other bottle with a big wide part and a long thin neck. [In physics we call something like this a Helmholtz resonator, after the physicist who first studied one carefully.] Suppose for some reason that the air inside the bottle is initially compressed. What happens to the air in the neck of the bottle?



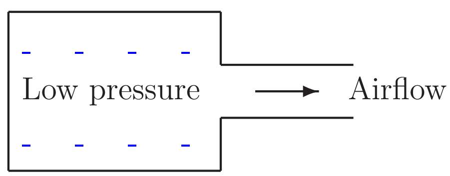
The air in the neck is pushed down by the pressure of the air outside. It is pushed up by the pressure of the air inside. Since that air is compressed, it pushes harder than the air outside. Therefore the air in the neck is pushed upwards. It accelerates upwards. If initially at rest, it will start to move, gathering speed with time. Therefore, an instant later, the situation will be like this:



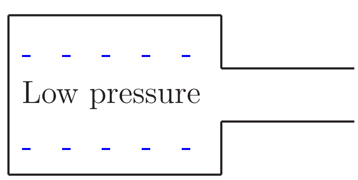
When the air in the neck is rising like this, there is a net flow of air, from the body of the bottle into the air outside the bottle. As air flows out of the body of the the bottle, the pressure inside will drop. After a moment, the pressure inside the bottle has fallen to be the same as the air pressure outside, atmospheric pressure:



However, once in motion, the air in the neck will remain in motion. Nothing needs to push it. Something has to push it the other way to get it to stop. This means that the air in the neck will keep flowing upwards, emptying out air from the bottle and bringing its pressure *below* atmospheric:

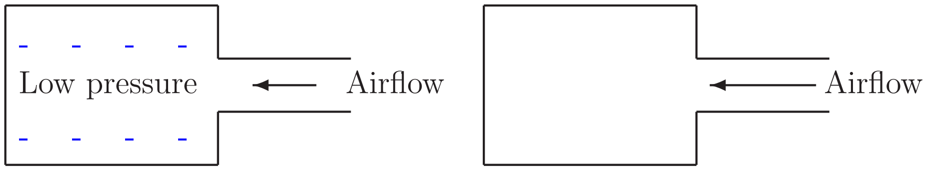


At this point, the air outside the bottle is pushing the neck air downwards more strongly than the air in the bottle is pushing it up. The air in the neck therefore slows down. After another moment, it has come to a stop:

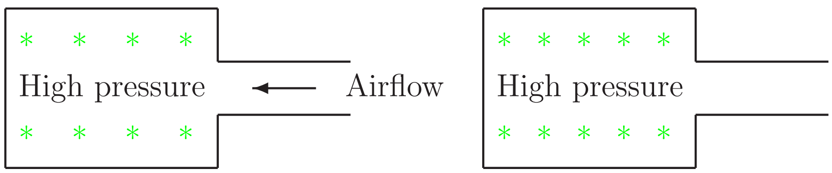


At this point the situation is exactly the opposite of what it was at the beginning.

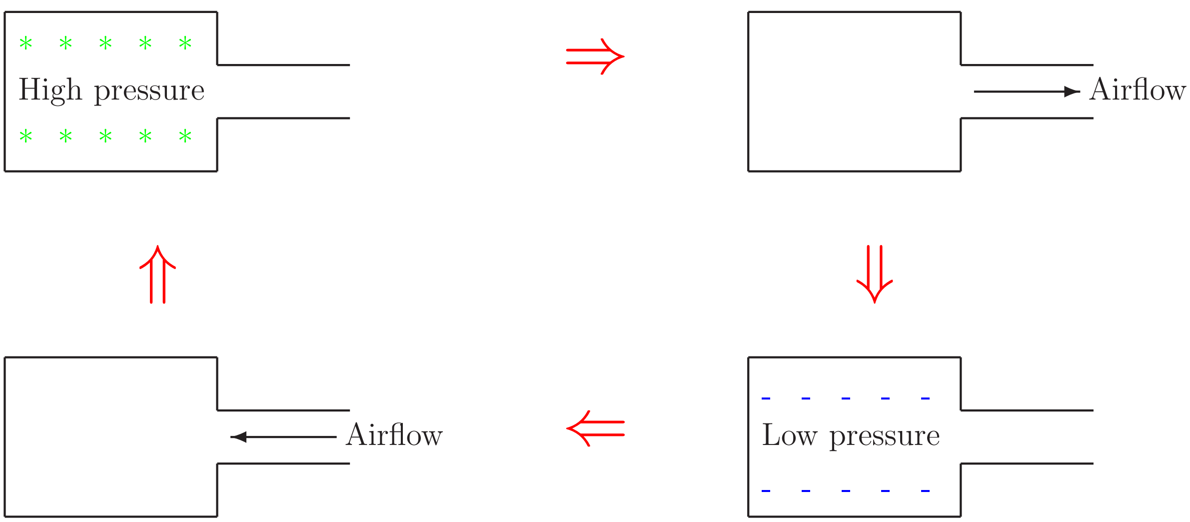
Now the air in the neck will continue to be pushed downwards, and will start flowing into the bottle. After two moments, it will have put the pressure in the bottle back up to normal:



Again the air keeps moving into the bottle until something actually stops it. Therefore it will re-pressurize the air inside the bottle. As the air pressure inside the bottle becomes more than outside, the air in the neck is again being pushed upwards, and will slow down and stop:



At this point, the situation is exactly as it was at the beginning. Therefore, the whole process will now repeat. This produces *periodic* changes in the pressure and air motion. Outside the bottle, the alternate outward and inward flows of air from the bottle will become a sound wave, with a period equal to how long it takes the process to repeat. In other words, the bottle goes round and round through the 4 stages of the resonance:



Let us try to figure out how long it takes for this resonance to go through one full cycle. This is going to depend on three parameters describing the bottle:

- V, the volume of the bottle;
- A, the cross-sectional area of the tube;
- l, the length of the tube.

Further, I will define ΔP to be the peak overpressure in the bottle, and v to be the peak air velocity in the neck. The period of the resonance won't depend on these, but they are useful in writing down the equations. Also, as usual I have to know the density of air, $\rho_{\rm air}$, and atmospheric pressure, $P_{\rm atmos}$.

I will *not* do a true derivation, but I will just make an estimate. A real derivation would not be much harder *if* I were willing to use calculus. Let us start by estimating how long it takes, to go from the bottle being pressurized without air flow, to it being all airflow. We begin by seeing how fast the airflow picks up. The force on the air in the neck is

$$F = P_{\rm bottle}A - P_{\rm outside}A = \Delta P\ A$$

which causes an acceleration:

$$F = ma \quad \rightarrow \quad a = \frac{F}{m}, \quad \text{and} \quad m = \rho A l \quad \rightarrow \quad a = \frac{\Delta P}{\rho l}.$$

Note that the area A canceled out. Now, *if* the pressure on the air stayed the same at all times, the velocity of the air would become $v = \Delta P\, t/\rho l$. The pressure inside the bottle falls because air is leaving the bottle. The volume of air which leaves the bottle is the width of the neck times how far the air in the neck moves. How far the air moves, is its velocity times time. This gives, *roughly*,

$$\Delta V = Avt = \frac{A\Delta P}{\rho l} t^2 .$$

Now, how much air needs to leave the bottle for the pressure to fall down to atmospheric? That will tell me what t has to be, for the pressure to go from maximum to atmospheric. The pressure loss inside the bottle is atmospheric pressure times the fraction of the air I take out. That is, if I take out 1/10 of the air in the bottle, the pressure will fall by 1/10 of an atmosphere, since it is the number of air molecules in the bottle which determines how high the pressure is. Therefore, to get rid of the overpressure, I need,

$$\frac{\Delta V}{V} = \frac{\Delta P}{P_{\rm atmos}},$$

which gives,

$$\frac{A\Delta P\, t^2}{\rho l V} = \frac{\Delta P}{P_{\rm atmos}} \quad \Rightarrow \quad t^2 = \frac{\rho l V}{A \Delta P} \frac{\Delta P}{P_{\rm atmos}} = \frac{\rho l V}{A P_{\rm atmos}}.$$

Recall from long ago, that $P_{\rm atmos}/\rho = v_{\rm sound}^2$, the square of the speed of sound.[1] Therefore, I can re-write this all as,

$$t^2 = \frac{lV}{A v_{\rm sound}^2}.$$

[1] Actually this is not quite right; there was a tiny correction $\sqrt{c_p/c_v}$. However, the exact same correction actually should have come up here. For the physicists, what is important in both cases is actually dP/dV the change in pressure with volume. Since this is the right quantity in both cases, the derivation I made above is actually wrong until the moment I make this substitution, and then it becomes correct!

This is the time for one quarter of the resonance phenomenon to go by. Therefore, my rough estimate is that the whole resonance phenomenon will have a period which is 4 times this long,

$$T = 4\sqrt{\frac{lV}{Av_{\rm sound}^2}} \, .$$

However, this is an under-estimate. The reason is, that I assumed that the full pressure was acting on the air in the neck the whole time; really, once the air starts to move, the pressure is falling. Also, I assumed that the whole velocity was present from the start; really, at first the velocity is zero, and it only builds up with time. Therefore, the real answer will be a little longer than this. You should not be surprised to learn that the right answer is actually,

$$\boxed{T = \frac{2\pi}{v_{\rm sound}}\sqrt{\frac{lV}{A}} \qquad \text{or} \qquad f = \frac{v_{\rm sound}}{2\pi}\sqrt{\frac{A}{lV}} \, .}$$

Now think about this logically. Does each term in the expression belong there?

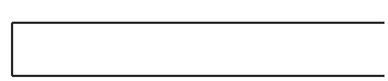
- Dependence on A: if the neck is wider, the air can empty out of the bottle faster. That means the pressure falls faster, and the whole process does not take as long. Larger A gives smaller T, or larger f. That is correct.

- Dependence on l: if the neck is longer, the air in the bottle is pressing on a bigger mass of air. It takes longer for this larger amount of air to get moving. That means the whole thing occurs more slowly. Larger l then gives larger T and smaller f. That is correct.

- Dependence on V: if the volume is bigger, more air has to leave the bottle before the pressure falls. That takes longer, meaning a larger T or smaller f. That is correct.

- Dependence on $v_{\rm sound}$: faster sound speed means air responds faster to its environment, meaning smaller T and larger f. Again, that is the behavior we found in our "derivation."

The bottle we considered is one of the simplest systems one can consider, to find a resonant frequency. For something more complicated, such as the air cavity in the mouth or hands when you whistle or hand-whistle, it is not so easy to do the calculation. There are a few other cavities simple enough that the resonant frequency can be found in closed form, and it turns out that the standard wind musical instruments are each quite close in shape to one of those other cavitites. For instance, a clarinet's bore (the hollow space inside the wood of the instrument) is almost of constant diameter and is almost completely closed at the reed and open at the opening: so it is pretty similar to,



a cylindrical tube. We can make a feeble attempt to use the same formula we found above to describe this tube. The problem is that the whole volume is inside the neck. Still, throwing caution to the winds and admitting that we do not expect the right answer, I can identify the volume of the bottle with the volume of the clarinet and the length of the bottle neck with the length of the clarinet; so $V = Al$. In this case, the formula is $f = (v_{\text{sound}}/2\pi)\sqrt{A/l \times Al}$, or $f = v_{\text{sound}}/2\pi\, l$. The right answer turns out to be, that you should replace 2π in this expression with 4:

$$f_{\text{cylinder}} = \frac{v_{\text{sound}}}{4\,l}\,.$$

This explains why your 2.5 cm meatus enhances sounds with a frequency of 3500 Hertz: $f = v_{\text{sound}}/4\,l \simeq 3400$ Hz.

Non-science people can definitely skip reading the following.

[There is actually a general procedure for finding the resonant frequencies of any cavity with a narrow opening, but it requires differential equations. One assumes that the pressure is changing sinusoidally. The relations between pressure and air velocity are,

$$\begin{aligned} \frac{d\vec{v}}{dt} &= \frac{1}{\rho}\vec{\nabla}P\,, \\ \frac{dP}{dt} &= \frac{c_p P_{\text{atmos}}}{c_v}\vec{\nabla}\cdot\vec{v}\,, \end{aligned}$$

from which it follows that,

$$\frac{d^2P}{dt^2} = \frac{c_p P_{\text{atmos}}}{c_v \rho}\nabla^2 P = v_{\text{sound}}^2 \nabla^2 P\,.$$

Look for solutions which are sinusoidal,

$$\nabla^2 P = -\frac{\omega^2}{v_{\text{sound}}^2}P\,,$$

with boundary conditions of $P = 0$ at the openings. Typically solutions will exist for several discrete choices of ω, which give the resonant frequencies. The reason for the boundary conditions is, that air can escape efficiently from the opening, so it only requires a tiny pressure to allow an airflow at this point. Beyond the narrow opening approximation, life becomes difficult because the resonances are damped and do not have precisely defined frequencies.]

Chapter 21

Pipes, resonances, standing waves

Very few instruments resemble the bottle we considered last time. (Your mouth when you whistle and your hands when you hand whistle, and those simple clay whistles with several finger holes and police whistles are examples of a few which do.) Most wind instruments which appear in orchestras come fairly close to being described as:

- A cylinder with one open and one closed end: the clarinet, some organ pipes and pan pipes (traditional instruments played, for instance, by Peruvians), the tubes underneath a xylophone, etc.

- A cylinder, open on both ends: the flute, some organ pipes, etc.

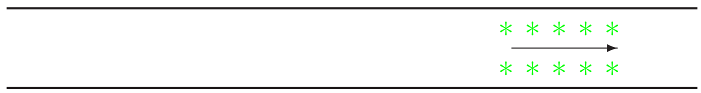
- A cone: oboe, bassoon, saxophone, some organ pipes, etc.

The brass instruments are more complicated; part of the tube is cylindrical, but part opens out, but not like a cone. We will return to that when we discuss brass instruments.

We need to think about how to describe the resonances in these instruments. It is also very important, not *just* to determine the lowest resonant frequency, since the tone quality and useful range of these instruments comes about because of the resonances higher than the fundamental.

To start out, think about a tube, open on both ends. Imagine that for some reason, a sound wave is moving down the tube, close to one end:

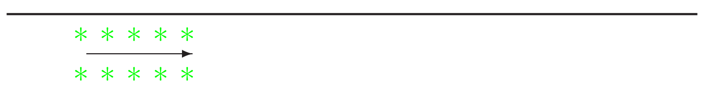
Obviously it moves to the far end of the tube,



where it reflects. Because the tube is opening out (easy for air to flow, but you cannot put up a net pressure), the reflected wave has the same air motion but opposite pressure:

This wave is moving back the other way, and goes towards the beginning of the tube:

Again it will reflect; again the pressure will switch sign but the air motion will not:



We are right back where we started. It is also simple to compute the time it takes for this process to occur: it is the time it takes for the air to go the length of the tube, reflect, and come back. That is the time it takes sound to go *twice* the length of the tube:

$$\boxed{\text{Open-open tube: } T = \frac{2L}{v_{\rm sound}}\,, \qquad f = \frac{v_{\rm sound}}{2L}}$$

for the wave to come back to what it was at the beginning.

We can ask the same thing about a tube, closed on both ends. The difference is that, on each reflection, the pressure will be the same but the air velocity will switch (remember, no air can flow out the end of a closed tube!). Again, the sound wave will look the same after making a full transit of the tube, and

$$\boxed{\text{Closed-closed tube: } T = \frac{2L}{v_{\rm sound}}\,, \qquad f = \frac{v_{\rm sound}}{2L}}$$

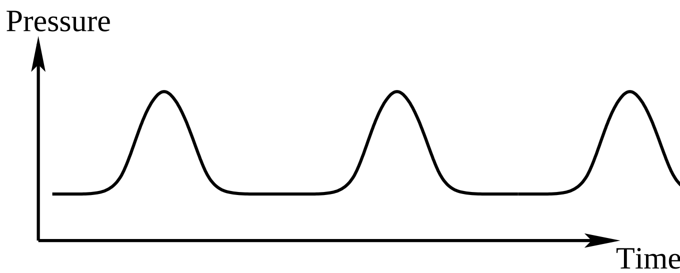
What about a tube, open on one end but closed on the other? That is more subtle, because the pressure flips sign on one reflection (from the closed end), but not on the other. Therefore, the pictures in that case will look like,

- positive pressure wave goes from beginning to end, reflects off open end
- positive pressure wave goes from end to beginning, reflects off closed end
- negative pressure wave goes from beginning to end, reflects off open end
- negative pressure wave goes from end to beginning, reflects off closed end

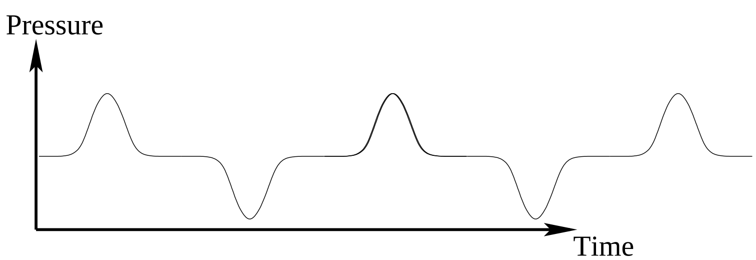
which means that it takes 4 transits of the tube for the situation to return to what it was:

$$\boxed{\text{Open-closed tube: } T = \frac{4L}{v_{\rm sound}}\,, \qquad f = \frac{v_{\rm sound}}{4L}}$$

What will be the appearance of the sound wave emitted from the right hand end of the tube? There will be a pressure pulse each time the sound wave bounces off that end. For the open-open tube, it is always positive pressure approaching, and we get,



while for the open-closed tube, the sound wave alternately approaches with positive and negative pressure (relative to atmospheric), and so the sound wave looks like,



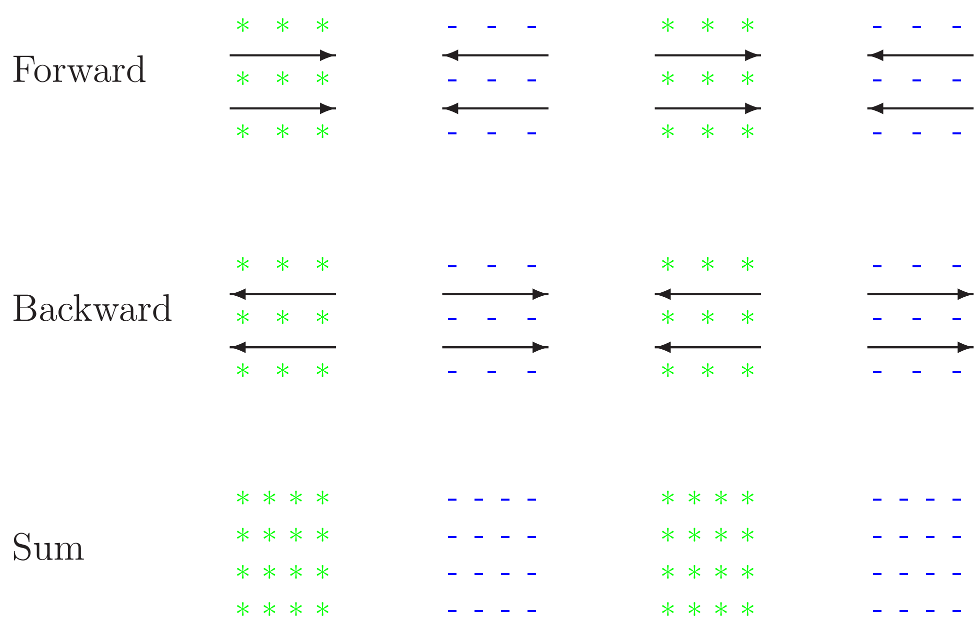
Neither of these is a sine wave, so what I have been talking about is actually having a bunch of sine waves in the tube at once. However, each is periodic, which shows that the sine waves in the tube will have frequencies which are *integer multiples of* (harmonics of) a fundamental frequency which is what I have given above. This is a neat property of such tubes, and is the reason that they are used musically.

Now we have to try harder to figure out exactly what harmonics are possible and what pattern they make inside the tube. This is important to the way musical instruments will function, so we will go through it in a little detail.

A sound wave of a given frequency will be a sine wave. Compared to what we were just considering, that means there will be a wave traveling in each direction along the tube at all times. To understand what will happen, we should think about what it looks like when two sine wave sound waves are going in opposite directions to each other. This can be inside of a tube or in the open air, it doesn't matter too much–the point is to understand what the pattern of pressure and air flow looks like when there are overlapping waves moving in opposite directions. When there are two waves moving in opposite directions and of equal strength, this is called a **standing wave**.

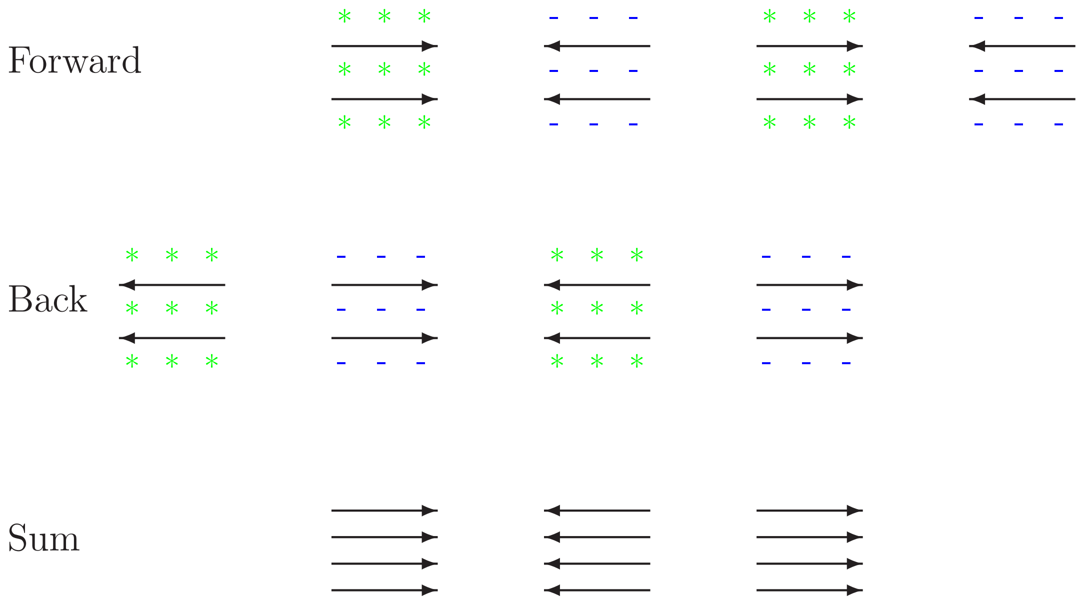
Consider two waves of equal wavelength and equal intensity, going in opposite directions (a standing wave). The two waves coincide in space. The total pressure and velocity of the air is the sum of the contributions from the two waves. To draw what happens, it is convenient to draw the pressure and velocity pattern of each wave, one above the other, and then draw what will be the sum of those two.

At the moment when the pressure peaks are at the same points in space, this looks like,



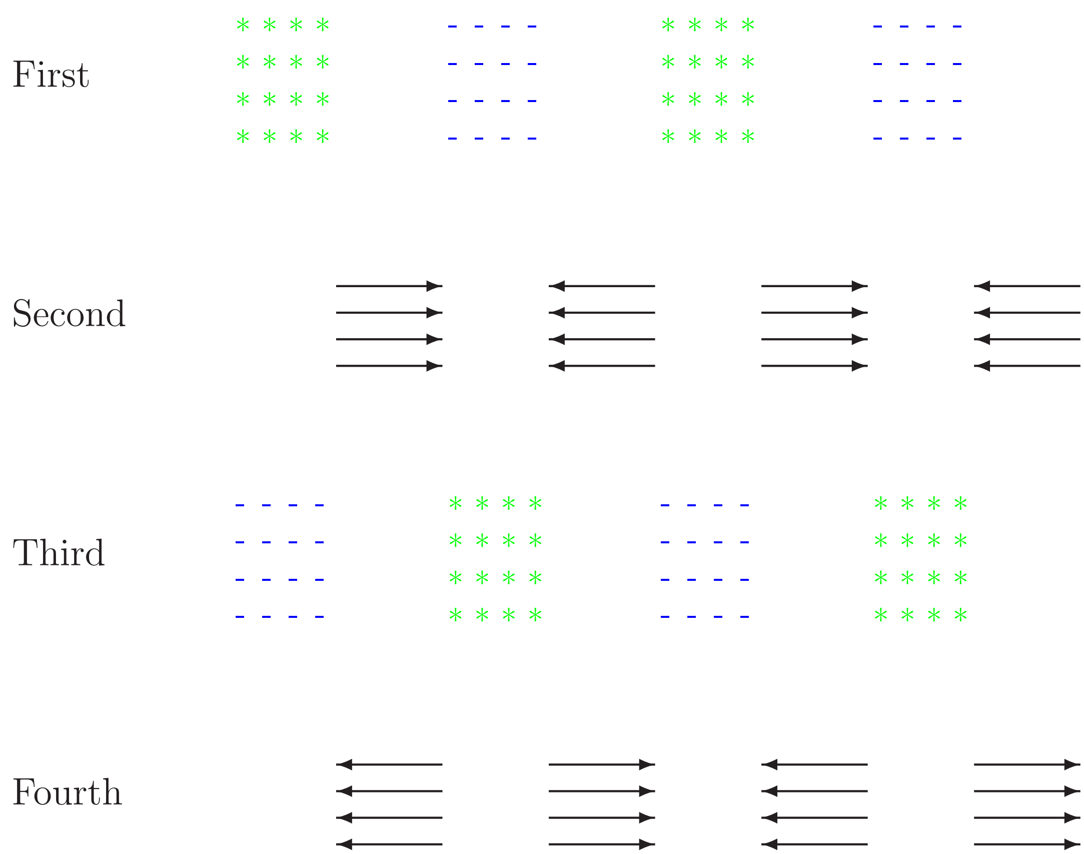
The air motions cancel, but the pressures add. There is a pattern of alternating high and low pressure.

A moment later, when the forward wave has gone forward a quarter wave length and the backward wave has gone backward a quarter wave length, it looks like this:



Two things have changed. First, now the pressures cancel and the air motions add. Second, the location where this happens is different: the spot where the air speeds are maximum is in between the spots where the pressure peaks occurred. The air speed maxima occur 1/4 of a wavelength off from where the pressure peaks occur.

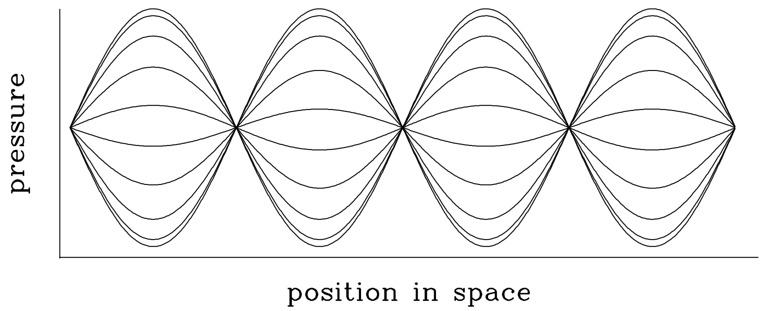
The time series of how a standing wave works, therefore looks like this:



Why? Let's go from "First" to "Second": the area with a high pressure behind it and low pressure in front of it is getting pushed forward. Therefore the air here starts to move forward. The area with low pressure behind and high pressure in front gets pushed backward. The reason that it is the spots *between* pressure peak and trough which start to move, is that those are the places where the force in front and behind are different from each other.

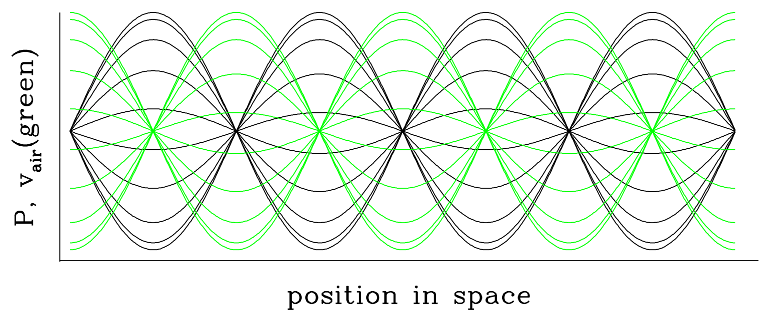
Now go from "Second" to "Third." The spot where the air is going forward from behind and backward from in front, is getting compressed (since air is rushing into that spot). It will therefore become a high pressure region. The spot where air is moving backwards behind and forward in front will become rarefied (less air), and therefore low pressure. The pressure pattern is the reverse of the starting one, so the last two steps are just the reverse of the first two.

If we make a plot of the pressure against position, and superimpose the curves for many different times, it looks like,



The spots where the pressure stays constant are called **pressure nodes**, and the places where it changes the most are called **pressure antinodes**.

Drawing the air motion in green, and putting the two plots on top of each other, the behavior of pressure and of air motion looks like,



Note that, as explained before, the **velocity nodes** (where air velocity stays constant) are at the same place as the *pressure antinodes*, while the **velocity antinodes** (where velocity changes the most) are at the same places as the *pressure nodes.*

Inside a cylindrical tube, a sound wave of definite frequency will look like a standing wave of definite wave length. Note that

the wave length is twice the distance between pressure nodes.

All we have to do is figure out what should happen at the ends of the cylinder.

Air cannot go out through a closed end of a tube, but the pressure can be whatever it needs to be there. Therefore,

Closed ends are velocity nodes and pressure antinodes.

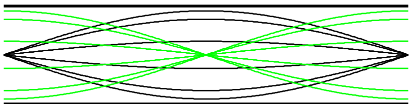
At the open end of a tube, air is free to move in and out. However, the pressure cannot differ from atmospheric (think about what happens when a sound wave in a narrow tube reflects from a widening, and think of the end of the tube as the tube becoming extremely wide). Therefore,

Open ends are velocity antinodes and pressure nodes.

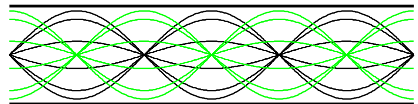
Now we are ready to think about the set of modes in each kind of cylinder. The lowest frequency standing wave that can occur in a cavity (such as a cylinder) is called the **fun-**

damental resonant frequency of the cavity, and the higher ones are called **overtones** of the cavity.

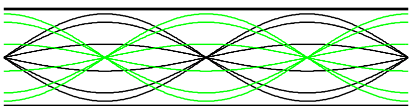
Consider first a cylinder open on both ends. The ends have to be pressure nodes. Therefore, the first few resonant patterns possible in the cylinder are,



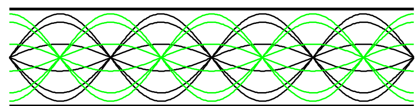
Fundamental



Second overtone: 3×



First overtone: 2×

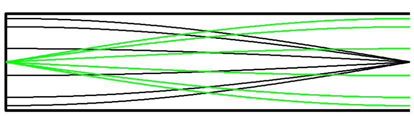


Third overtone: 4×

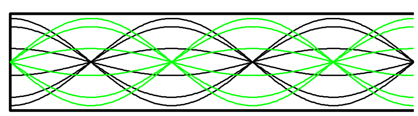
which are respectively, fitting 1/2 wave length, 1 wave length, 3/2 wavelengths, and 2 wavelengths of sound into the tube. Therefore they have wavelengths of $2L$, $2L/2$, $2L/3$, $2L/4$, *etc.* and frequencies,

$$\boxed{\text{Open-open tube, length } L: \quad f = \frac{v_{\text{sound}}}{2L} \times \Big(1\,,\,2\,,\,3\,,\,4\,,\,5\,,\,\ldots\Big)}$$

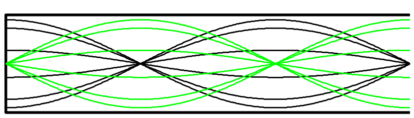
If instead I consider a tube which is closed on one end and open on the other, there should be a velocity node on the closed end (and therefore, a pressure antinode), and a pressure node (velocity antinode) on the other. The first few resonant patterns fitting the bill are,



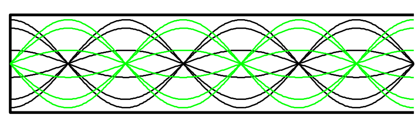
Fundamental



Second overtone: 5×



First overtone: 3×



Third overtone: 7×

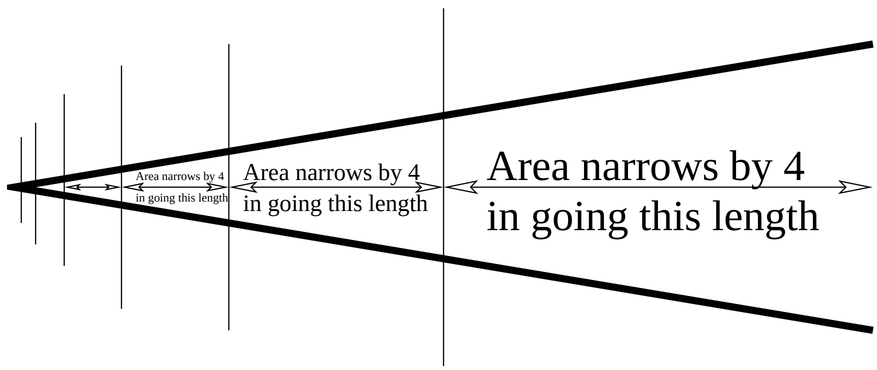
which are 1/4 of a wavelength, 3/4 of a wavelength, 5/4 of a wavelength, and 7/4 of a wavelength fitting inside the tube. Therefore the resonances have wavelengths of $4L$, $4L/3$, $4L/5$, $4L/7$, *etc.* and frequencies,

$$\boxed{\text{Closed-open tube, length } L: \quad f = \frac{v_{\text{sound}}}{4L} \times \Big(1\,,\,3\,,\,5\,,\,7\,,\,\ldots\Big)}$$

This has half the frequency as a fundamental, and is missing the even multiples of the fundamental.

What about a conical tube? This is an important problem in music, because several instruments are shaped approximately conically, including the double reed instruments and the saxophone.

The problem with analyzing the conical tube is that we can no longer just say that a sound wave will propagate along the tube without any change. Recall that, if the width of a tube is changing slowly, a sound wave propagates without reflection. If the area of the tube changes by a large ratio in much less than the wavelength of the sound, though, then most of the sound power is reflected. Now think about a cone:



Any wave will reflect at some point. The longer the wavelength, the further out in the cone it will reflect. Therefore, the cone is closed-open, but it acts like it is shorter for the low modes, and closer to its full length for the high modes. The resonant frequencies turn out to be exactly those for an open-open tube:

$$\boxed{\text{Conical tube, length } L: \quad f = \frac{v_{\text{sound}}}{2L} \times \Big(1\,,\,2\,,\,3\,,\,4\,,\,5\,,\,\ldots\Big)}$$

but the pressure pattern is different; in stead of the pressure coming to zero at the tip, it has a node there.

[For those with advanced mathematical background, the explanation is that finding the pressure pattern inside the cone is solving the wave equation in *spherical* coordinates. We want pressure to satisfy,

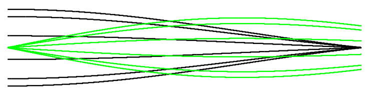
$$\nabla^2 P = \frac{-\omega^2}{v_{\text{sound}}^2} P$$

with $P = 0$ at the opening and $dP/dr = 0$ at the tip. Work in spherical coordinates with the tip of the cone as the center of the sphere. The pressure will depend on r only. The solutions are spherical Bessel functions. In particular, the pressure is given by

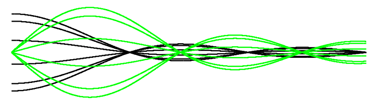
$$P(r) = P_0 \frac{\sin(kr)}{r}\,, \qquad k = \frac{2\pi}{\lambda}$$

which has zeros exactly where sin has them, except with the zero at the origin removed. The zeros are at the same places as for $\sin(kr)$, the function we need for the open-open pipe.]

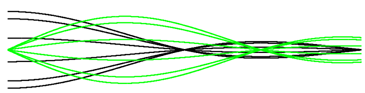
The patterns of pressure and air speed in the conical pipe for the lowest modes, not bothering to draw the cone this time, look like,



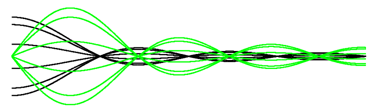
Fundamental



Second overtone: 3×



First overtone: 2×



Third overtone: 4×

These figures are slightly deceptive in terms of showing where the energy in the wave is stored. The pressure and velocity tell what the intensity is, and it is clear that the wave is most intense near the tip of the cone. However, the most *power* is further out, because the power is the intensity times the cross-sectional area of the cone, which becomes small near the tip. Therefore, most of the sound energy does not go back into the tip of the cone.

Chapter 22

Decay of Resonances

Recall that a resonance involves energy bouncing back and forth between two means of storage. In air resonances in cavities, energy is going back and forth between being stored in air compression, and being stored as air motion. However, in any resonant process in the real world, a little of the energy will go into other things each time it goes back and forth between the two main storage mechanisms. For instance,

- when you swing on a swing, energy is going back and forth between gravitational potential (being high up at the top of the swing) and energy of motion. However, there is energy being lost to friction against the air, energy going into the bar or branch holding up the swing, energy lost as friction in the joint or pivot on which the rope or chain swings, and so forth. Therefore, unless you keep pumping the swing, it will slowly go less and less high.

- When sound is resonating inside of a wind instrument, some sound energy is escaping from the instrument into the world through the fingerholes, opening, or walls of the instrument. Some (typically much more) is being lost to friction as the air rubs against the walls of the instrument. Some may be lost to inelasticity in the reed, lips, and so forth.

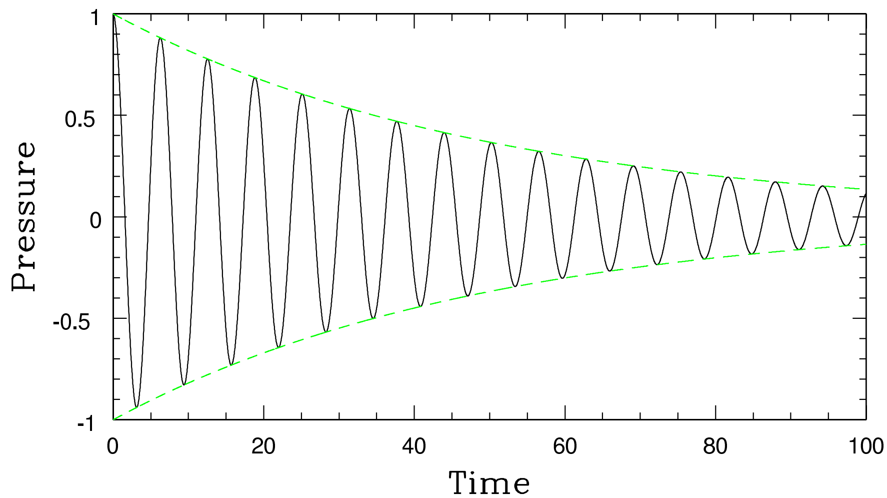
- In a string instrument or piano, the energy in the vibration of the string is lost through the bridge of the instrument into the instrument's body, and against the soft pad of the finger.

- In percussion instruments, energy is lost to radiated sound, but also to inelasticity of elements (especially wood components, drumheads, and other soft pieces) and is carried out through the support structure. It is also removed by deliberate damping devices.

22.1 Resonance decay and quality Q

Just how much of the energy is lost per cycle of the resonance is important in understanding how an instrument built around resonances will behave. Therefore we will try to quantify resonant energy loss in this lecture. We will also try to understand what happens when you **drive** a resonance, which means, when you "push on" something which can behave resonantly, at a frequency which is close to the resonant frequency. This explains how an instrument built around a resonating device can produce a steady tone which does not die away with time.

First, some terminology. The behavior of one variable in a resonance which is dying away typically looks like this:



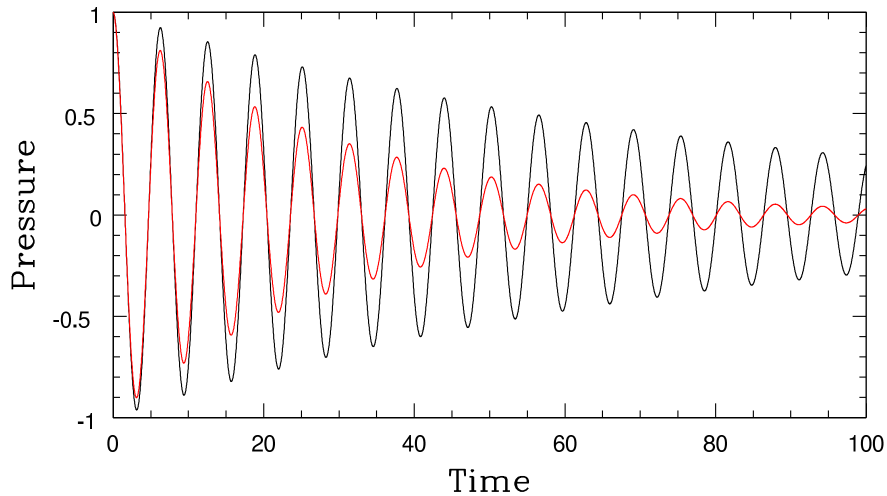
I took the vertical axis to be pressure for the sake of argument; it could be air velocity, your forward speed on a swing, or whatever. The green dotted lines are called the **envelope** of the decaying oscillations.

The curve in the figure is characterized by three things:[1]

- The period, or equivalently, the frequency, of the oscillations. The signal is no longer sinusoidal, but we can define the period as the time between peaks in the pressure.
- The starting size of the pressure oscillations.
- The rate at which the oscillations damp down.

The first two of these we are familiar with. We need some language to describe the last one. For instance, the following two curves have the same period and starting pressure:

[1]There is really a fourth, the starting phase of the oscillations, but it is not really interesting.



How do we describe that the black curve takes a long time to die away, and the red curve dies away faster?

Resonance is a transfer of energy between two things: air pressure to air motion to air pressure to air motion to air pressure .... The relevant question is, what fraction of the energy is lost in each transfer from one form to the other? Since we like resonances which take a long time to die away, we define the **resonant quality** Q as, roughly,

$$Q \sim \frac{\text{Energy in resonance}}{\text{Energy loss per transfer}}$$

where by transfer I mean transfer of the energy from one form of storage to the other. It turns out to be more convenient to define Q as precisely,

$$Q \equiv 2\pi \, \frac{\text{Energy in resonance}}{\text{Energy loss in 1 period}} \, .$$

The 2π is the usual 2π that comes into anything involving sines and cosines.

A large Q means the resonance repeats for many oscillations without losing strength. A small Q means that it only lasts a few.

An equivalent way of writing Q is,

$$Q = 2\pi f \frac{\text{Energy}}{\text{Power loss rate}}$$

[The reason for the 2π in the definition is that $2\pi f = \omega$ the angular frequency, which turns out to be a more natural quantity physically.] Note that it could sometimes be true that a bigger or smaller fraction of energy is lost from the early oscillations than the late ones. However, for *most* resonant phenomena, this is not true; if there is twice as much energy in the resonance, twice as much will be lost.

So how long does it take for a sound to die away? Very roughly, the answer must be $Q/2\pi$ oscillations. However, each oscillation starts with less energy, and so each oscillation loses less energy than the one before it. Therefore, the energy loss is actually exponential:

$$\frac{\text{Energy}}{\text{starting Energy}} = e^{-2\pi\times(\text{number of oscillations})/Q}\,.$$

[Here $e = 2.71828$ is the base of the natural logarithm. If you don't know how that got in here, don't worry about it, just take this as a result.]

How long it takes for a resonance to "die away" depends on what your standards for "die away" are. Technically, the resonance will go on forever, just smaller and smaller. However, at some point it is so small that it has disappeared for practical purposes. Therefore, we will define a **decay time** τ, as the time for the amount of energy to fall by a factor of $1/e$. This is,

$$\frac{1}{e} = e^{-2\pi(\text{number of oscill})/Q}$$

or,

$$1 = \frac{2\pi(\text{number of oscillations})}{Q}$$

or,

$$\text{number of oscillations} = \frac{Q}{2\pi}\,,$$

but the time is the number of oscillations over the period, which is the number of oscillations over the frequency. Therefore,

$$\tau = \frac{Q}{2\pi f}\,,$$

is the decay time.

The ear considers a change in intensity by a factor of 10 to be roughly a factor of 2 change in loudness, as we have seen previously. The time for a factor of 10 loss in intensity is

$$\tau \log_e(10)$$

and for a factor of 100 loss is $\tau \log_e(100)$, and so forth. Roughly,

$$\log_e(10) = 2.3\,.$$

For example, listening to the diedown in the ringing made when I "pluck" the wine bottle with my thumb, I estimate it takes 1/3 of a second for the sound to die away. That gives me $\tau \sim 0.3$ s$/2.3 = 0.13$ seconds. The frequency is 112 Hertz, so the Q is $Q = 2\pi f\tau = 6.28 \times 112 \times .13 = 90$. This is considered a reasonable Q but not super high.

I can try to do a better job by actually looking at the pressure pattern of a recording of "plucking" the bottle. I count how many pressure peaks for the peaks to get smaller by a factor of 2. Since intensity goes as pressure squared, this is a change in intensity of a factor of 4, requiring $\log_e(4) = 1.4$ times the decay time. I count about 16 oscillations for the amplitude to fall by this factor of 2. Therefore the τ is $16/(1.4 \times 112)$ seconds, which is .10 seconds, about the same as my "by-ear" estimate. The Q can be found directly from the number of oscillations for the energy to fall by some factor:

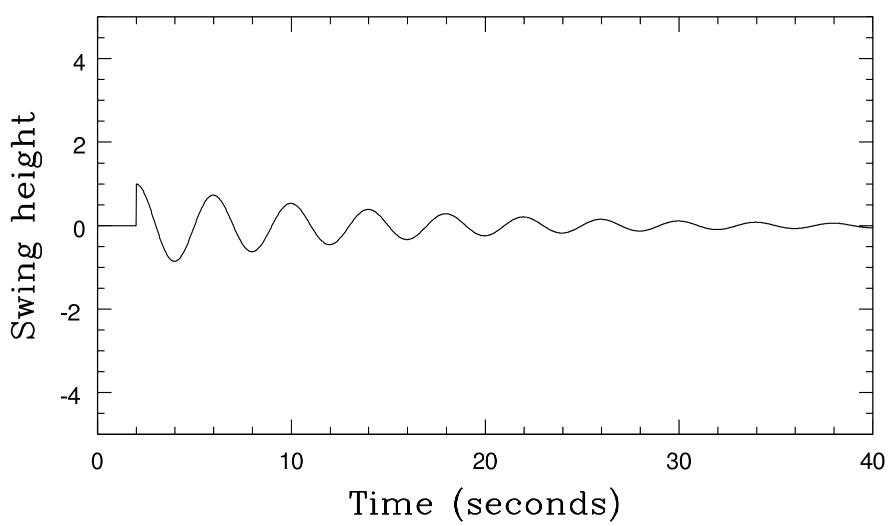
$$Q = \frac{2\pi \times \text{ number of oscillations}}{\log_e(\text{start/final energy})}$$

which gives $2\pi \times 16/\log_e(4) = 72$ for the Q, about the same as the other estimate. Doing the same thing with the open D string on the cello, I find about 60 oscillations for the pressure peak to fall by a factor of 2, so $Q \sim 270$. The Q of instruments are often in this range of 100 to hundreds. Some metal percussion instruments have much higher Q.

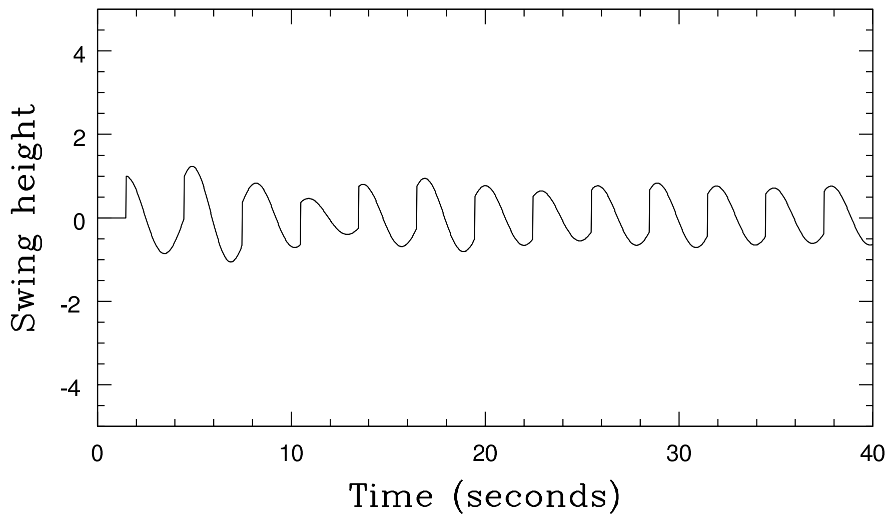
22.2 Driving a resonance

For an instrument to make a steady tone, it has to have energy go in as fast as it is going out from the resonance. I have to **drive** the resonance somehow. Resonances are also useful if something else happens to be producing periodic sound at a frequency close to the resonant one. The size of the resonant oscillations then depends not only on how loud the other sound is, but on how close its frequency matches the resonant frequency. This will explain some things we have already discussed, about the ear and about other things, so let us see it in a little detail.

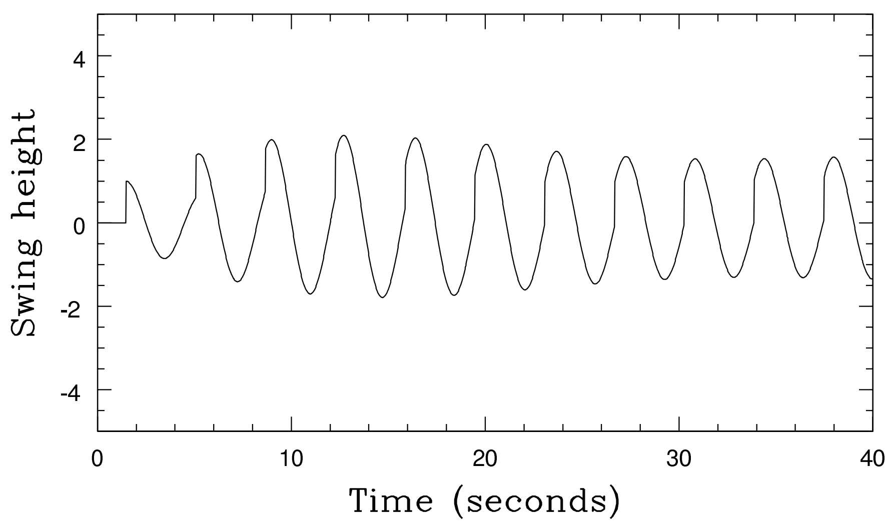
Think about pushing your little brother or sister on a swing. Suppose the swing goes back and forth in 4 seconds. If you push it just once, when it is at the bottom, the swing will do this:



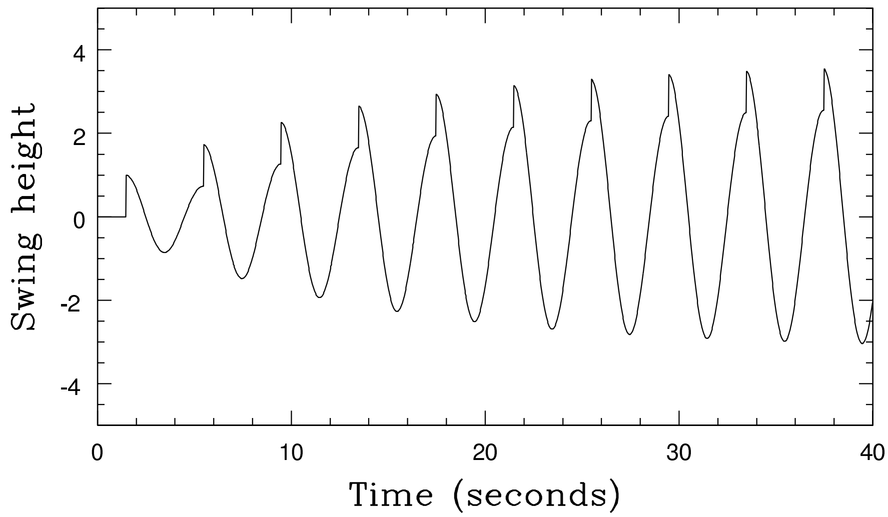
Suppose you push the swing every 3 seconds, which is too often, as you know. It means you will be pushing the swing before it reaches the bottom from the last swinging. The swing will do this:



Because you are not pushing the swing when it is going forward, you don't really get a larger amplitude swinging than the result of each push. What if you push closer to the right frequency, say, every 3.6 seconds?



That worked much better. However, it does not work perfectly, because you are pushing too soon; the swing has not come back to where the push will do the most good. The way to think about it is that the last push you made was .4 seconds too soon; the push before was .8 seconds too soon; going back 5 pushes, it was 2 seconds too soon. That means that the swinging because of that push, is backwards compared to what you are doing now. Only the last 5 pushes are actually adding up. (The reason it is 5 pushes is, that your are pushing 0.4 seconds too early, and it takes 5 pushes for that to add up to half a swing.) If you push every 4 seconds, the ideal frequency, you would get this:



Now each push is coming exactly when it is needed. This leads to the maximum height for the swing. The only reason the swinging does not grow bigger and bigger is that the swing is damped.

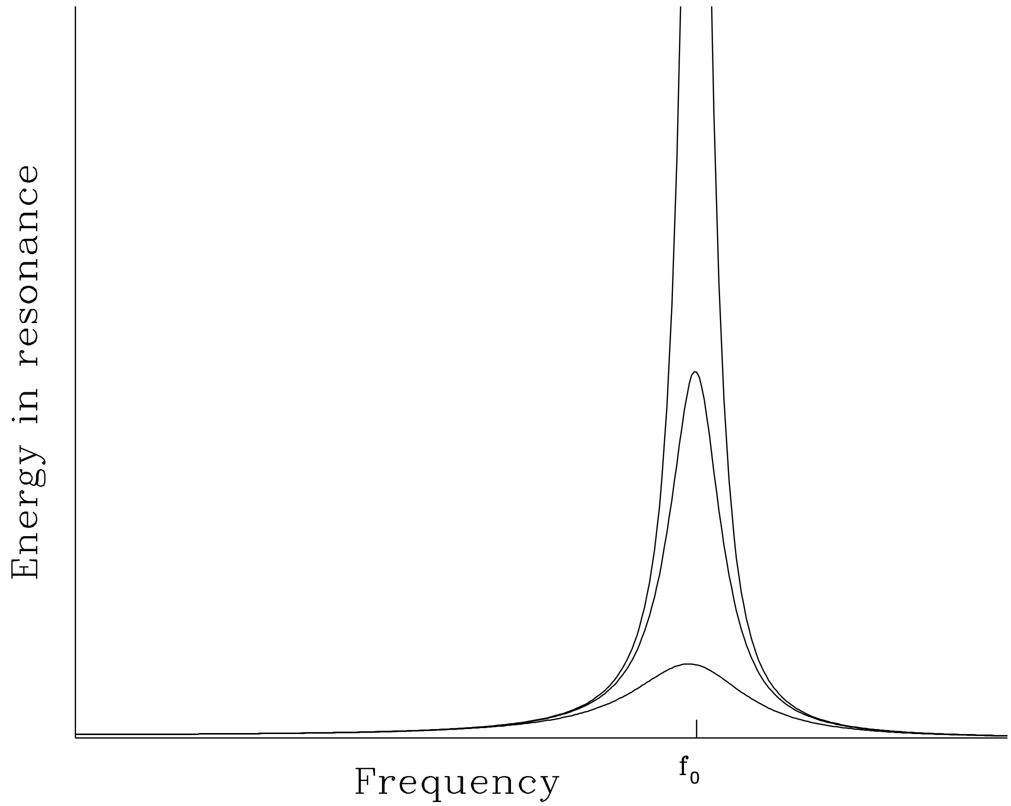
How big can the amplitude get, and how close to the right frequency must the pushes come? I will answer for the case where you push with a sine wave, instead of discrete, separate shoves. Call the ideal frequency f_0, and the frequency you really push with, f. If $|f - f_0| > f_0 Q/2$, then the height is limited by the fact you get out of sync with the result of previous pushes. The amplitude then grows with the accuracy that you match your pushes:

$$\text{Amplidude} \propto \frac{f_0^2}{(f_0{+}f)|f_0{-}f|}$$

(Here || is absoluted values; if the thing inside is negative, choose minus that thing. $\propto$ means "is proportional to.") If the frequencies are better matched than this, the Q is the thing that limits how big the oscillation gets, and

$$\text{Amplitude} \propto \frac{1}{Q}\,.$$

The energy in the resonance goes as the square of the amplitude. Therefore, a forced resonance, with the forcing at the right frequency, has its energy grow as $1/Q^2$. The energy in the resonance, as a function of frequency for a few Q values, is shown below:



The moral is, that the way to get a big amplitude vibration is to have a high Q and drive on resonance.

For a musical instrument, the resonance is typically at hundreds of Hertz. There is no way a person can "push the swing" at just the right time, at hundreds of Hertz. Therefore there has to be an *automatic* way that energy gets fed in at just the right spot in the resonant cycle. This is one of the three *key ideas* of designing a steady-tone musical instrument:

Key ideas in steady tone instruments:

- Have a high Q resonance.
- Have a way (nonlinearity) that energy is automatically fed in at the right point in the resonant cycle.
- Have some way to convert the energy in the resonance efficiently into sound waves.

We will see how various wind and string instruments solve these problems in the coming lectures.

Chapter 23

Reed instruments

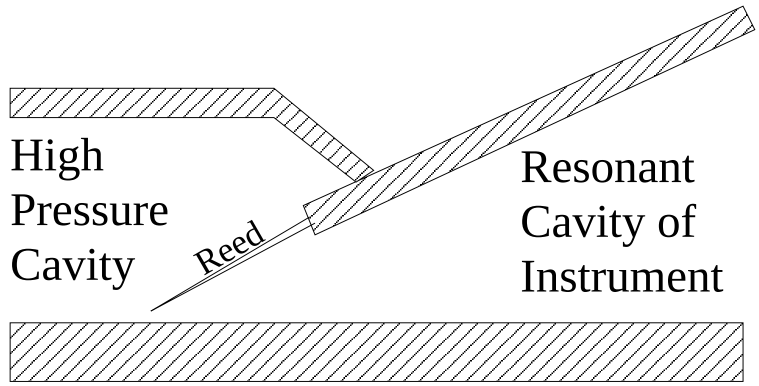
Musical instruments which produce a sustained periodic tone invariably do so by using a resonance to set the frequency of that tone. There remain three problems:

- How is energy to be fed into the resonance?
- How is the resonant frequency to be adjusted so the instrument can play a wide range of notes?
- How is the energy in the resonance to be converted efficiently into sound?

One very nice way to solve the last problem is to use directly a resonance in an air cavity. That way the energy is in sound to begin with. This leaves the problem of tuning the resonant frequency of the cavity, and designing some way for the resonance to be "driven," with energy added every resonant cycle so the resonant phenomenon does not die down.

A reed is a device for putting energy into a resonance, which automatically adds high pressure when the pressure is high and not when it is low. This solves the problem discussed in the last lecture, of making sure "always to push when the swing is going forward." Let us see how it works.

A reed is a thin, stiff piece of material, fixed at one end and free to move back and forth at the other. Since it is thin, it can bend back and forth. Since it is stiff, if it bends, it elastically tries to spring back to its natural position. It is placed at a narrow spot between a source of high pressure air (such as the mouth, the bag of a bagpipe, the compressed air source of an organ, *etc.*) and a resonant cavity or the open air. The design must be such that, if the reed bends one way, it narrows or closes the connection between the high pressure and outside, and if it bends the other way, it opens the connection wider. The cross-section of a reed system might look like this:

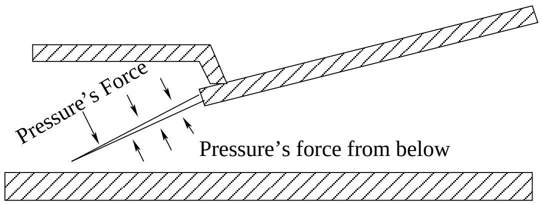


At this point I have to explain something about reeds and distinguish two types. The reed is made out of a stiff, flexible piece of material, sort of like half of a ruler when you clamp the other half against a table with your hand. Because it is flexible, the reed can be bent; because it is stiff and it is attached firmly on one side, it is springy and tries to spring back to its natural position. If you bend the reed and release it, it will vibrate back and forth at a frequency determined by its size, thickness, stiffness, and so forth. This is the resonant frequency of the reed.

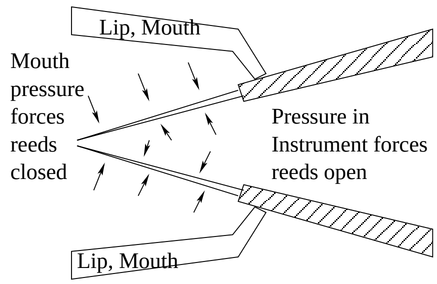
There are two sorts of reeds: free reeds, and reeds coupled to a resonant chamber. The free reeds connect a pressure reservoir and the outside world, and they vibrate at the resonant frequency of the reed. The reeds which make the sound in harmonicas (mouth organs) and accordions are of this sort, as are the party noise makers I will hand out in class. Understanding how the vibration of the reed is amplified by the flow of air through the reed is a little complicated; it involves Bernoulli's principle, and I will avoid talking about it in this lecture. Most musical instruments (such as oboes, clarinets, bagpipes, bassoons, saxophones, and so forth) use the other sort of reed, where the frequency is set by the resonant frequency in a cavity.

Let us understand how this "normal" sort of reed instrument works. The reed sits in between a resonant cavity and a source of pressure (such as your mouth). Specifically, it is at a pressure antinode and velocity node of the resonant cavity, where the pressure variation within the cavity will be maximal when the resonance is excited. The reed almost closes the connection between the mouth and the cavity. The cavity resonant frequency should be lower than the natural frequency of the reed, which means that the reed will respond quickly to the pressures it feels.

The air pressure pushing on one side of the reed is the high pressure in the air source. The pressure pushing on the other side is the pressure in the resonant cavity. The reed must be mounted such that the pressure in the mouth (or high pressure source) is trying to push the reed closed (that is, to close the path for air to go between cavities). The pressure in the resonant cavity is trying to push the reed open. Here is what it looks like for a single reed (clarinet) or double reed (oboe):



Single Reed

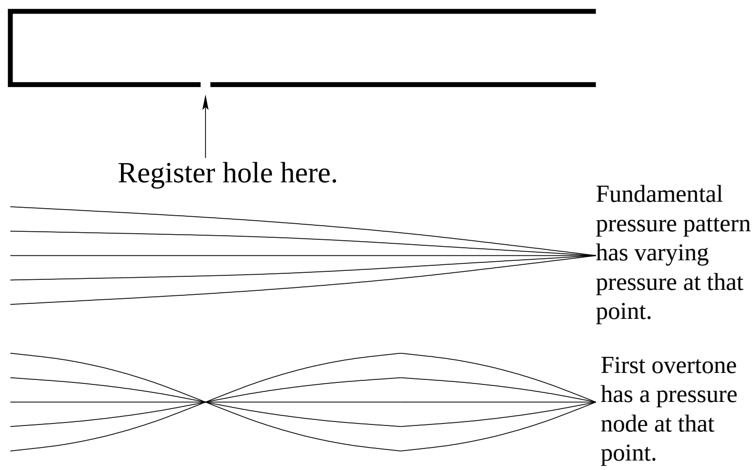


Double Reed

Suppose there is already some sound in the resonant cavity. Then the pressure inside, on the back side of the reed, rises and falls. Since the reed has a high resonant frequency, it moves back and forth as dictated by the pressures on the two sides. When the air pressure in the resonant cavity is high, the reed is forced wider open. That lets through high pressure air from the high pressure source–at just the right time for it to increase the high pressure in the instrument. When the pressure is low in the instrument cavity, that lets the reed fall further shut–at just the right time to keep from spoiling the low pressure in the cavity by adding high pressure from the source. Therefore, the reed *automatically* lets high pressure join the high pressure part of the sound wave, but not the low pressure part, increasing the strength of the resonance in the cavity. If this adds more power than is lost to the finite Q of the cavity, then the size of the sound wave will get bigger.

A few points are in order. First, the resonant frequency of the reed has to be higher than of the cavity. Otherwise, the reed will respond too slowly to the pressures on it, to open and close at the right stage of the oscillation. That is why, when the oboe plays the reed when it is taken off the instrument (in which case it is a free reed, playing at its natural frequency), it makes a high pitched squeak, much higher than the note played by the oboe. To reach the top of the range of a reed instrument, the player often bites down on the reed, which effectively shortens the part of the reed which is free to vibrate, raising its resonant frequency (like moving where the ruler is clamped between your hand and the table, so a shorter stub of the ruler is sticking off the table). Second, the size of the resonance in the instrument body will of course not grow forever. In particular, the peak pressure inside the instrument will never get larger than the pressure supplied by the mouth, since the mouth was only increasing the pressure in the cavity because the mouth pressure was higher. Third, the louder you play, the further the reed moves back and forth. At some loudness, the reed falls all the way shut in the low pressure part of the pattern–it "claps." This makes the pressure pattern very different from a sine wave. Because of the way the reed functions, a reed instrument typically makes something quite far from a sine wave pressure pattern, so reed instruments typically have lots of harmonics in their spectrum. This is especially true if the resonant cavity is designed or chosen such that the harmonics of the note played are also resonant frequencies of the cavity. This is the case for all orchestral reed instruments.

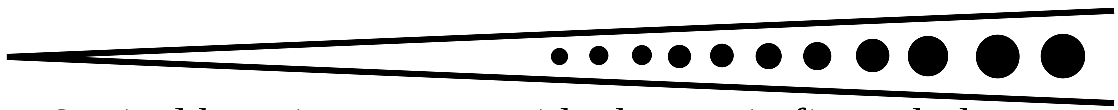
How is a reed instrument to play a wide range of frequencies? Part of the answer is, by playing not just the fundamental frequency of the resonant chamber, but the resonant frequencies of its overtone series as well. This requires some way for the performer to favor the buildup of sound in an overtone of the cavity without building up in the fundamental resonant pattern of the instrument. One way to do this is with a **register hole.** This is a small hole which can be opened in the instrument at the spot where the fundamental resonant pattern has a high pressure, but some higher overtone has a pressure node. For instance, on an open-closed cylinder (the clarinet), a hole one-third of the way along the instrument,



is at a spot where the resonance of frequency $3v_{\rm sound}/4L$ has a pressure node, but the fundamental $v_{\rm sound}/4L$ does not. When the fundamental plays, the high pressure at that spot will force air to leak out of this little hole. That means that energy is lost from the resonance; the fundamental has a low Q. However, since the overtone has no pressure variation at that spot, no air is forced out, and the Q is unchanged. A low Q resonance requires a large strengthening of the resonance pattern each period, while a high Q resonance will build up even if only a little energy is added each period. Therefore the high Q resonance will win out. Most reed instruments use such holes to help the player force the instrument to play on a particular overtone (which is called playing in a particular register).

This is obviously not good enough, since it only lets the instrumentalist play the overtone series of the instrument. To get all the notes inbetween, it has to be possible to change the resonant frequency of the resonance. Since most instruments are built around tubes, this means changing the effective length of the tube. This is done by opening larger holes, called **finger holes**, placed along the tube of the instrument.

If the finger hole were as big around as the diameter of the instrument, then opening a hole would effectively end the instrument's tube there. Since the air is free to leak out at the finger hole, it has to be a pressure node; the condition which determines the frequency played by the instrument is the distance from the closed end of the instrument (at the reed) to the pressure node at the end of the instrument, which has now moved to the finger hole. To play a chromatic scale on a conical bore instrument, one would need 11 such holes, spaced further apart at the end of the instrument and closer together at the middle:



Conical bore instrument with chromatic finger holes

In practice it is a better idea to use holes which are quite a bit smaller than the diameter of the instrument. (For one thing, it is easier to cover them with your fingers!) Since such a hole lets out less air than a big hole, the pressure does not have to be quite at a node at a smaller hole; the resonance pattern extends a little ways further up the bore of the instrument, which means that you can change the frequency a little by opening or closing the holes further down than the first opened hole. This reduces the number of holes needed to play a chromatic scale. Since some instruments are quite long and many have holes too big for fingers to cover accurately, an elaborate system of levers, wires, and spring mounted hole covers is required. The fingering technique is complicated, varies between instruments, and beyond the scope of this lecture.

Two final remarks are in order. First, the timbre of the sound inside the instrument is not the same as the timbre of the sound emerging from the instrument. This is because of the way that the sound radiates out from the holes in the instrument. We saw in lecture that, when a tube opens from area A_1 to a much larger area A_2, only a fraction $4A_1/A_2$ of the sound intensity escapes. What should we use for A_2 to understand sound escaping into the world? The answer turns out to be, roughly $A = (\lambda/\pi)^2$ with λ the wave length of the sound. The higher the frequency, the smaller λ is, and the bigger $A_{\rm tube}/\lambda^2$ is. Therefore, the high frequencies more efficiently escape the instrument, and the sound you hear is richer in harmonics than the sound inside the instrument. Since it involves the square of the wavelength, the effect is a factor of 4, or 6 dB, per octave.

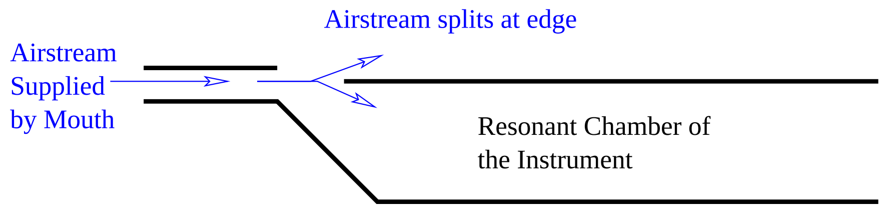
Second, the way the sound radiates from a series of open finger holes, each much smaller in diameter than the tube, and the way sound radiates from the end of the tube, are quite different. That means that if a reed instrument really used a strictly conical or cylindrical bore, the notes with all finger holes closed would have a very different timbre from the notes played with several open holes. The bell (a flaring opening at the end of the instrument) was designed by trial and error as a way of changing the radiation pattern of the notes with all closed fingerholes so it more closely resembles the timbre of a note radiated out from the finger holes.

Chapter 24

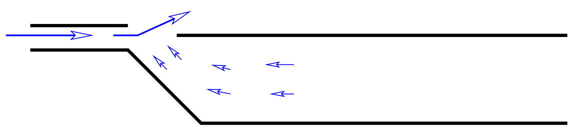
Flutes and whistles

We saw last lecture that the reed represented one way of solving the problem that the energy in an air resonance damps down with time. The reed presents a technique for pumping energy into the resonance by always adding pressure when the pressure is high and never when the pressure is low. Flutes and whistles operate by the same principle, except instead of playing with the pressure at a velocity node and a pressure antinode, they play with the air flow at a velocity antinode and pressure node, that is, an opening in the instrument.

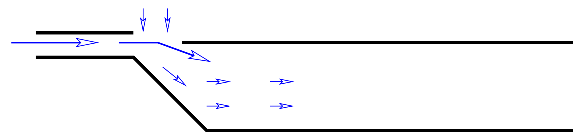
Consider a recorder, which is a kind of whistle. The player blows into a slot or tube, providing a stream of air. (Opposite of the reed instrument, what is important here is that the player is providing air flow, rather than air pressure). The stream of air crosses an opening and is split by an edge. On one side of the edge is the outside, on the other is a resonant chamber:



The air either goes up, into the world, or down, into the resonant chamber. Now suppose there is a resonance going on inside the chamber. The air is crossing an opening, which is one of the places air can get in or out of the resonant chamber. A resonance has a velocity antinode (pressure node) here, meaning it is a spot where there is a maximal airflow during the resonance. When air is flowing out, it will deflect the air stream to go out with it, and when the air is going in, it will deflect the air stream to go in with it:



Behavior when resonance makes air flow out



Behavior when resonance makes air flow in

The airstream you blow therefore joins the air flow, in or out of the resonant chamber, and makes it stronger. That strengthens the resonance, pumping energy into the resonance and making it louder. This is analogous to what a reed does, only it is done by playing with the airflow, rather than with the pressure.

A flute works by exactly the same idea; you blow across a hole (called an **embouchure hole** on a flute) which is an opening in a resonant chamber, and your airstream is split on the other side of the hole, and either goes into or out of the chamber. A resonance in the chamber causes air to flow in or out of the embouchure hole, and that forces your air to accompany it. For the purposes of this class, we will take the distinction between a flute and a whistle or recorder to be, that in a flute, the mouth forms the airstream, while in a whistle or recorder, a length of tube or a slit forms and directs the airstream. Each technique has its advantages, but over time the flute has won out as the more widely used orchestral instrument (though the recorder is enjoying growing popularity).

The problems in designing a flute or recorder are similar to the problems in designing a reed instrument. To play chromatic scales, one must be able to change the resonant frequency of the resonant chamber, which is generally done by making it a cylindrical or nearly cylindrical tube with finger holes. Finger holes means that the far end of the tube must be open, so flutes are almost always open-open instruments. An exception is the pan pipe, where the instrument is made out of a large number of open-closed tubes, often tied together into two rows like the white and black keys on the piano. Each tube plays one note, and the player must move the mouth from place to place to reach different notes.

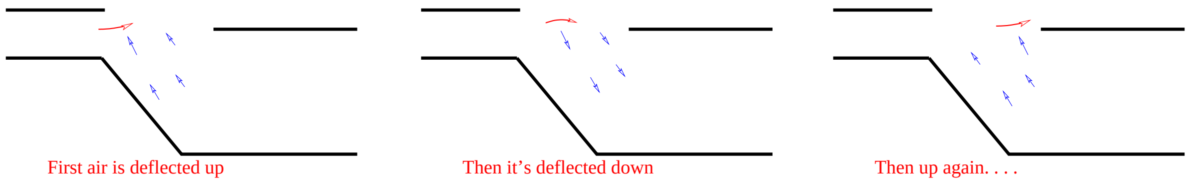
How does a flute or recorder player force the instrument to play overtones of the resonant chamber? A recorder often has a register hole (your thumb covers it on the cheap grade school recorders), but register holes are actually not necessary for flutes and recorders to hit overtones. It can be done instead with the speed of the airflow across the embouchure hole (and in the flute, the placement of the mouth too).

The explanation given above, of how the resonance and the air flow interact, would be true of any resonant frequency of the chamber. However, not all resonances are enhanced by the same amount. It turns out that the airstream crossing the embouchure hole has a natural frequency which it enhances the most efficiently–the frequency where the time for the airstream to cross the embouchure hole is about half of the period of the resonance:

$$f_{\text{most enhanced}} \sim \frac{2v_{\text{airstream}}}{d}$$

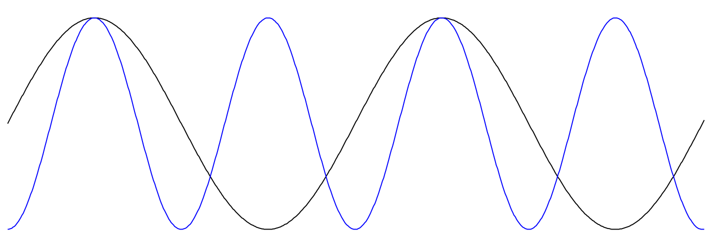
with $v_{\text{airstream}}$ the velocity of the airstream blowing over the opening, and d the distance for the airstream to cross the opening. Let us see why.

First, consider a resonant frequency in the body of the instrument which has a shorter period than the time it takes for the air to cross the opening. In this case, the airstream is first pushed in, then pushed out, then pushed in, . . . on its way across the opening. To see this, imagine following one bit of air from the airstream, on its way across the gap where it gets deflected:



The different deflections do not add up, and do not help make the airstream more or less likely to deflect the right way at the edge on the other side of the embouchure hole. That means that high frequency oscillations do not self-reinforce. For this reason, flute and whistle instruments tend not to be very rich in harmonics (though they have some, if the resonant cavity has harmonically related overtones). We saw this during instrument days when we sampled flutes and/or recorders.

What about a resonance at a frequency lower than the time for the air to cross the opening? This does push in the right direction consistently as the air crosses the opening. However, if more than one resonance has a frequency below $f_{\text{most enhanced}}$, only one of them is typically amplified. To see why, imagine that both the fundamental and a first overtone at twice the frequency were oscillating inside the body of the instrument. They alternate between pushing in the same direction, and in opposite directions, because the two waves we are adding look like this:



When they push the same direction, they will make the airstream move with them, which enhances *both* waves. When they push in opposite directions, the airstream will move with whichever is larger. That enhances the resonance with the larger amplitude, and diminishes the resonance with the smaller amplitude. If the resonant chamber happens to start with mostly the low frequency resonance present, it will be strengthened and will remain. If neither is present, or if the high frequency one is present with a larger amplitude, the high frequency one tends to build up and the low frequency one does not. That means that, once you get the instrument to start making one of the available frequencies, it tends to stick with that frequency. However, it is easier to get the higher frequency started. Tricks flutists know, like using the tongue to briefly stop the airflow, can help the instrument "find" the higher frequency. Also, depending on how fast you blow, the higher frequency harmonic can take over even if the lower frequency one was present first.

In other words, when the flutist controls which overtone of the resonant cavity they play, mostly by varying the way they form the air stream. To play the fundamental register, the musician pulls the lips back so that the whole embouchure hole is exposed, blows slowly, and shapes the airstream with the lips so that it will not be narrow and fast. This leads to a slow airstream crossing a large opening, which can drive the fundamental but not the higher overtones. To skip up an octave, the flutist narrows the lips, speeds the airflow, and moves the lips forward somewhat, shortening the distance the air travels to cross the embouchure hole. This means that it takes less time for air to cross the hole, so higher frequencies can be enhanced. There are also tricks with fingerings which can help the resonance in the chamber find a higher overtone.

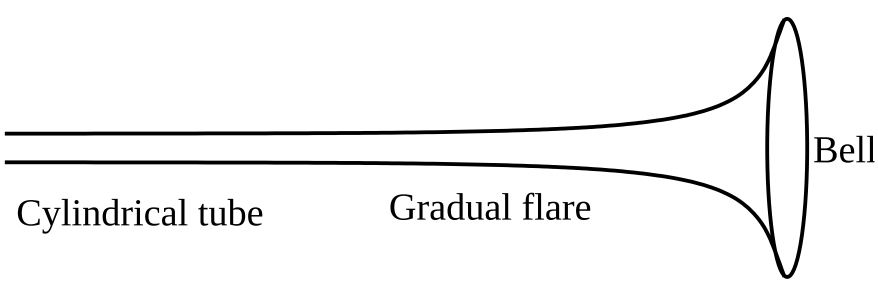
Chapter 25

Brass instruments

There are two key ideas behind brass instruments. The first is to use the lips as a reed. It is quite easy (and annoying) to play your lips as a free reed. You tighten them and bring them almost together, and then blow. By varying how tight you make your lip muscles, you can change the pitch your lips produce. Tightening the muscles makes the lips both thinner and stiffer, and raises the frequency.

As reeds, your lips have two properties which distinguish them from the single and double reeds used by orchestral reed instruments. First, they are heavier. That makes it more difficult for them to move at frequencies far from their resonant frequency. Therefore, the brass instruments will generally play notes quite close to the frequency the lips would make if you moved the brass instrument away. Second, the resonant frequency of the lips is much easier to control, since they are part of you. This makes it practical to have a reed with a highly and rapidly variable vibration frequency.

The other key idea of a brass instrument is that the sound emerges from a wide opening, called the **bell** of the instrument. The bore of the instrument is cylindrical for much of its length, but then it flares out wider, gradually at first and more rapidly at the end:



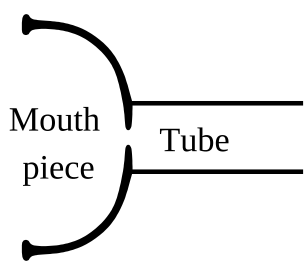
The gradual flare is an impedance matcher which lets the sound move from the narrow to the wide part of the tube with little reflection. By widening the bore before the sound reaches the outside world, the efficiency of the radiation of the sound is improved. Recall that, when a sound goes from being in a pipe of cross-section A_1 to a pipe of cross-section

A_2, the fraction which is transmitted is

$$\frac{I_{\rm transmitted}}{I_{\rm incident}} = \frac{4A_1A_2}{(A_1+A_2)^2} \, .$$

To understand the flaring of the instrument, remember that this expression is for a *sudden* change in the diameter. If the tube changes gradually, one should roughly use this formula repeatedly for the amount of flaring which happens every 1/4 wavelength of the sound in question. When A_2 is close to A_1, almost all sound is transmitted; for instance, even for $A_2 = 2A_1$, 8/9 of the sound is transmitted. Therefore, if the flaring of the instrument is by only a factor of 2 every 1/4 wavelength, the sound intensity continues without reflection. At the opening of the instrument, we should use this expression again, with A_1 the area of the opening and $A_2 = \lambda^2/\pi$ as the "effective area of opening into the outside world."[1]

Now think about a high frequency overtone. It has a short wavelength, and will make it almost unreflected through the flaring of the instrument. It therefore radiates into the world using the full area of the instrument's bell, making for efficient radiation of the sound in the instrument out into the world. For a low frequency tone, on the other hand, the sound makes it unreflected out to where the bore starts to change diameter rapidly. Then most of it is reflected back in, and only a fraction makes it to the bell, which again gives an inefficient radiation because the area is much less than λ^2. Therefore the brass instruments are especially efficient at radiating high frequencies, compared to low frequencies. This is both what makes them loud and what makes them "brassy" (which is the timbre of something with a lot of power in overtones).



On the other end of the instrument, the lips to not attach to the cylindrical tube of the instrument directly. It is easier to play the instrument if there is a **mouthpiece**, a roughly hemispherical piece of metal which provides a larger ring for the mouth to touch and a narrow opening into the tube of the instrument, as pictured to the left. The mouthpiece is a separate piece of metal and can be taken on and off.

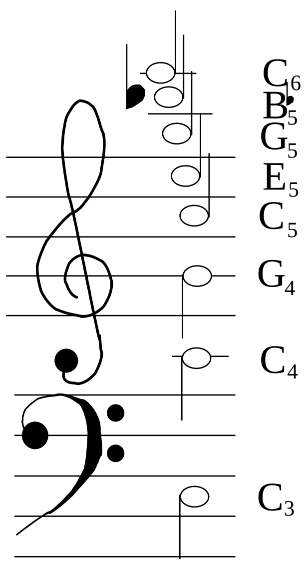
In designing an instrument, one must choose either to have the sound radiated from finger holes, or from a bell. Modern orchestral instruments always use finger holes for reed, flute, and whistle instruments, and bells for lipped instruments. There is no reason in principle that an instrument with a reed and a bell cannot exist, or an instrument played with lips and finger holes. In fact, brass and wind musicians often try this before and after rehearsals, putting their mouthpieces on fingered instruments or reed bocals onto brass instruments in place of the mouthpiece. It works, and PDQ bach wrote a piece for the "tromboon," a trombone played with a bassoon bocal. However, the history of instrument design has

[1] This is the result of a tricky calculation which you don't want to see.

dismissed this idea and preferred brass instruments with a bell played with the lips, and fingered instruments played with a reed or as a flute or whistle.

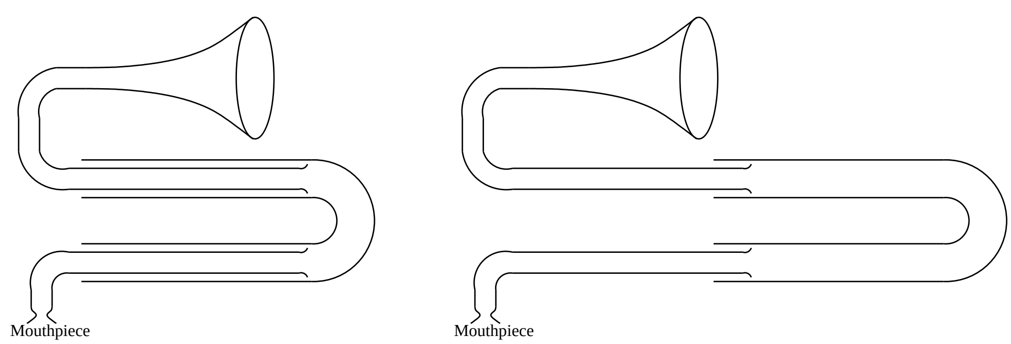
There was a good reason for finger holes, though; it allowed the length of the bore of the instrument to be varied. How is the brass player to play any note without finger holes? One solution, used for the bugle, is that you don't; only the overtone series of the instrument can be played. This is good enough for the army, but not for orchestral use.

To see the problem, let us look at the notes which a trumpet in C can play, without varying the length of the instrument.



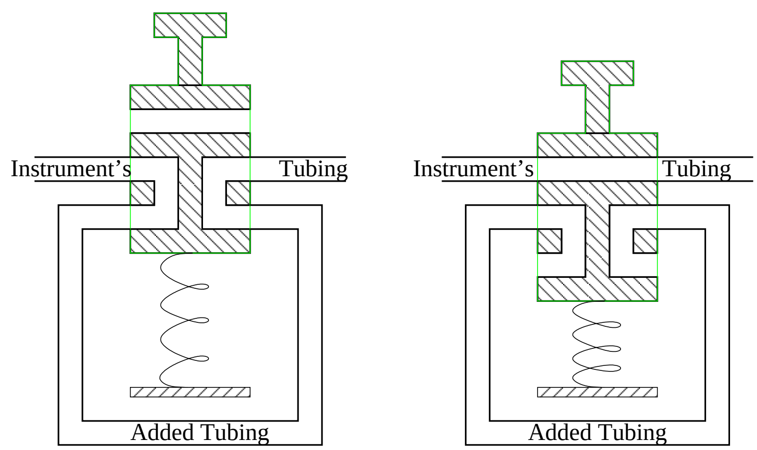
The fundamental is C_3. As we will see below, this is hard to play and rarely used. The harmonics are in the $1, 2, 3, 4, \ldots$ progression, so the next is C_4, followed by G_4, C_5, E_5 (14 cents flat), G_5, $B_5^\flat$ (31 cents flat), C_6, and so forth. Abandoning playing C_3, the largest interval is from G_4 to C_4, 7 half-steps. One needs to be able to play the 6 notes between these ($F_4^\#$ to $C_4^\#$) in order to play any chromatic note. To do so, it must be possible to make the instrument longer by amounts varying from a factor of 1.06 (adding 6% of the instrument's length) to 1.41 (adding 41% of the instrument's length). That way, the instrument can play down from G_4 to $C_4^\#$. This also allows the instrument to play below C_4, down to $F_3^\#$. (Most trumpets are in $B^\flat$, which means that all notes are a whole step lower than the notes I just named.)

There are two ways to vary the length of tubing making up the instrument: with a slide, and with valves. The idea of a **slide** is to have two long, straight cylindrical pieces of pipe at some point in the instrument's tubing. Another pair of pipes, of slightly larger diameter, fit around these pipes. The inner pipe comes to an end, and the air flows into the outer pipe; it has a 180° bend, taking the air to the other straight section of tubing. The outer tube can be pulled back and forth, like so:



which varies the total length of the instrument. It is difficult to design an instrument where the slide makes up more than about half the length of tubing, but it does not have to; it only needs to make up about 40% of the total length of tubing, as we saw. The advantage of this method is that it is possible to tune the instrument to any note in the range allowed by the slide. It also allows certain effects; a glissando (a note sliding in pitch smoothly rather than jumping from one note to another) is easy to achieve using a slide. The trombone is the most common brass instrument using a slide as the main tuning device.

The other way of varying the length of tubing in a brass instrument is by using valves. The general meaning of "valve" is a device which can open or shut a pipe, like the handle you use to turn on or off water at a sink. For a brass instrument, a valve is something which can reroute the air in the instrument so that it either goes through a very short tube (usually, just across the valve) or it goes through a longer section of tube. A cartoon of how this could work is,



One Valve Position: Tubing Added

Other Valve Position: No Tubing Added

The thing in the green box is the valve; the top is a button for a finger to press, on the bottom is a spring. When you push down (on the right in the figure), the tubing of the instrument is routed through the short piece of tubing across the valve. When you let it up (left in the figure), the tubing of the instrument is routed to attach to an extra piece of tubing, thereby increasing the effective length of the instrument's tube. In this cartoon, pushing down the valve button makes the tube shorter, but in practice it is generally done the other way, so pushing down the valve button adds the extra piece of tubing and makes the instrument's tube longer. By having 3 or 4 valves, each attaching a different length of tubing, enough combinations are possible to achieve any desired length of tubing to reach the notes in a 6 half-step range. On a trumpet, this is achieved by having a valve which adds enough tubing to go down a half-step, one valve with enough tubing to go down 2 half-steps, and a valve which adds enough tubing to go down 3 half-steps. This does not work perfectly;

each added half-step requires more tubing than the half-step before it (since pitch goes as the log of the period and therefore the log of the instrument's length). To some extent this is solved by the performer "pulling" the pitch a little by varying the lip tension and shape, called "lipping a pitch up or down." On some instruments (depending on the manufacturer), a very small slide is added to adjust the note by a fraction of a half-step. On the deeper brass instruments with longer tubes, such as the tuba, a fourth valve is added, to allow more possible combinations of length, which makes it easier to design the instrument so that some combination is close to the right tuning for each note.

Different instruments use different kinds of valves, *eg*, piston versus rotary valves. The distinction is important to the musicians and manufacturers, but their role in the instrument is the same.

The key problem in the design of a brass instrument is getting the overtones of the instrument's cavity to lie in the right harmonic series. The problem is, that the instrument is not an open-open cylinder, or an open-closed cylinder, or tapered in a cone. The instrument should be closed or nearly closed on one end (the end with the mouthpiece), so it will have a pressure antinode where it is being played as a reed. It must have a long cylindrical section, and it is chosen to have a flaring section at the end. The tricky thing is that the long cylindrical section will *vary in length* as the slide is moved in and out or as valves are depressed or lifted.

Yet somehow, the overtones need to be in an integer progression ($1, 2, 3, 4, \ldots$ times a fundamental). This is important to make it easy to play the instrument in tune, since a major way the player goes from note to note is by going from overtone to overtone. It is also important to the timbre of the instrument; if the musician is playing C_4, and the overtones of the cavity are $C_5^{\#}$ and A_5 instead of C_5 and G_5, then the harmonics of the C_4 note will not resonate. That makes the tone too nearly sinusoidal, which sounds hollow, and it turns out to make it harder to hold the note in tune.

If there were no flare in the instrument, and the mouth end were really closed, the overtones would be in the progression,

$$\frac{v_{\rm sound}}{2L} \times \left(\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \ldots \right)$$

rather than the desired progression,

$$\frac{v_{\rm sound}}{2L} \times \left(1, 2, 3, 4, \ldots \right)$$

Each note is too low, and the early ones are too low by more. What the flaring of the tube does, is to make the low frequency (long wavelength) modes reflect before reaching the bell, which makes the instrument effectively shorter for them. This lifts the frequency of the low overtones more than for the high ones, which is just what we want. A gradual widening reflects only long wavelength modes. A rapid widening reflects shorter wavelength (higher frequency) modes. To make the instrument shorter for the longer wavelength modes, it should flare gradually at first and more rapidly at the end. The mouthpiece also modifies the resonant frequencies somewhat.

The best way to make the instrument's bore flare is a hard design problem, which had to be solved with the most powerful design technique known: trial and error. Over the years, for each instrument, instrument makers have learned how to make the flare of the instrument just right so that the overtone series is very nearly in harmonic progression. The exception is the fundamental; there is no way to design the instrument so it is in tune. In practice it is well flat of the ideal tone, typically around 0.8 times the desired pitch (which is a major third). This makes it very hard to play, and generally it is not played. The note at the missing fundamental (half the frequency of the first overtone, C_3 for the C_4 trumpet discussed above) *can* be played, even though there is no resonance in the instrument there! The presence of resonances at all the overtones makes this possible. This note is called the **pedal tone**. It is difficult to play and is separated by a gap from the rest of the instrument's register (the trumpet in C discussed above can play $F_3^{\#}$ and a pedal tone at C_3, but nothing between these notes), so it is only used infrequently.

We end with two amusing remarks about wind instruments.

The first is about the pressure required to play. The reed and brass (lip) instruments require a high pressure and a low airflow. The flutes and whistles take a fast airflow but not much pressure. Flute players may get out of breath; reed and brass players get red faces.

Why red faces? The arteries and veins which serve the head pass very close to the pharynx in the throat. The pressure of the air in the pharynx (between the back of the mouth and the larynx) pushes against the arteries and veins. Typical blood pressure is "120/80" in *torr*, or millimeters of mercury. In real units, this is about 16 000 Pascal and 10 500 Pascal, or 16% and 10.5% of an atmosphere. The higher number is the pressure of the blood being pumped *into* the body from the heart, the lower pressure is the pressure of the blood *returning* to the heart. The higher the pressure a reed player puts up, the higher the pressure inside the instrument can be; so you must push hard to get a loud sound. Reed and brass players, playing at maximum dynamic, can use air pressures in excess of 10% of an atmosphere. This is enough pressure to close the veins draining the head. As blood fills into the head without leaving, the face turns red. This is more amusing than harmful.

Very skilled trumpet players can reach even higher pressures, sometimes in excess of 16% of an atmosphere, in very loud playing. This is enough pressure to close off the *arteries* bringing blood *into* the head. The blood in the head then becomes anoxic and turns blue, giving a bluish white complexion. *This is bad for you* and will catch up with you eventually. This gives some trumpet players have health problems late in life.

The second remark is about circular breathing. Skilled reed players and extremely skilled flute and recorder players can sustain a note *indefinitely*. The trick is to use the mouth as an air reservoir. Just before running out of breath, the musician fills up the mouth and cheeks with air. Then he/she closes the mouth off from the throat, and expels the air in the mouth to keep the instrument playing, while breathing in through the nose as fast as they can. This requires coordination and is easier on an instrument which requires a low airflow (reed and brass) than on an instrument which needs a fast airflow (flute and recorder). A musician who is adept in doing this can sustain a note without interruption until they get bored and decide there is something else they could be doing instead.

Chapter 26

String instruments

A string under tension will vibrate back and forth when it is plucked or bowed. A single string under tension has many resonant frequencies, which are in harmonic relation if the string is thin and uniform. This makes strings a particularly simple way to design an instrument which produces periodic tones which are rich in harmonic structure. This lecture will explain why a string under tension has resonances, why their frequencies are harmonically related, and how string instruments overcome the two other problems of musical instrument design, efficient sound production and the production of sustained tones, rather than tones which die off (though only the bowed instruments solve the latter problem.)

A string is under **tension** if something on each end of the string is pulling on it. A string on an instrument is usually not accelerating very fast, which means that the forces have to balance out on the string; so something on each end of the string must be pulling on it with about the same force. The tension is defined to be this force.

It is not easy to see that there is a large force pulling on each end of a string on a violin or guitar because the force is being maintained by a peg which is not moving. The force is maintained by friction on the peg, keeping it from turning. You only really realize how big the tension is when you try turning the peg. On a guitar this is done through a screw mechanism which gives a large mechanical advantage (leverage); the force on the string is much larger than the force you exert on the peg. To clarify, Figure 1 contains a very poor cartoon of the parts of a string instrument. The tension on the string is maintained by the peg pulling the string away from the tailpiece and the tailpiece pulling the string the other way. Of course, this means the string puts force on the peg and the tailpiece, pulling them towards the middle of the instrument. The body of the instrument is under compression from these two forces, and must be fairly mechanically strong.

An "ideal" string has no rigidity; its shape and its motion are completely determined by the tension and the mass per length of the string. The tension on the string tries to keep the string straight. When the string is not straight, it will "snap back" towards being straight.

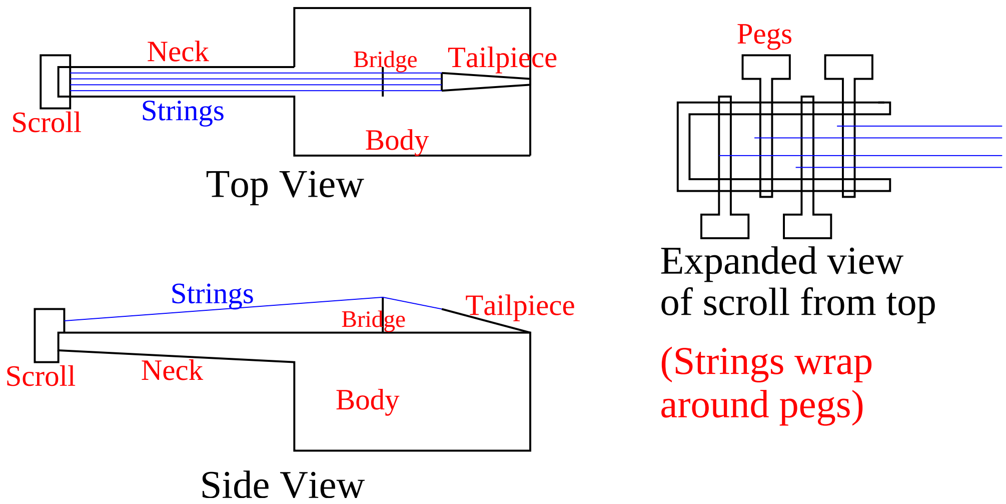
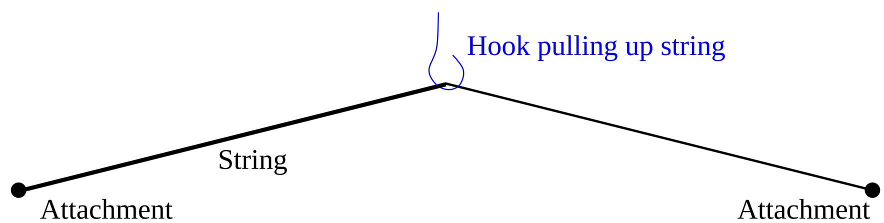
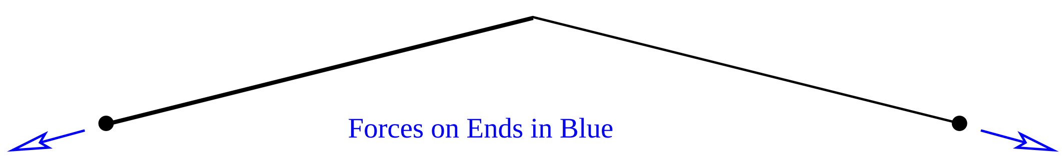


Figure 26.1: Parts of a string instrument. How the peg works is shown in a blow-up of the scroll of the instrument.

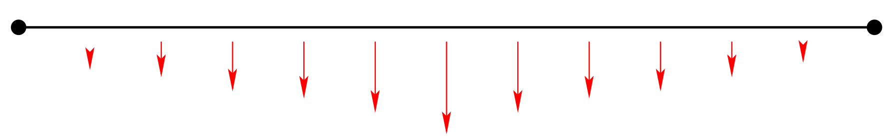
The mass per length of the string limits how fast this "snapping back" actually happens, and means that the string keeps moving past where it becomes straight. In pictures, think of it like this; imagine I attach a string on its two ends, under tension. Now, grab it somehow and pull it upward:



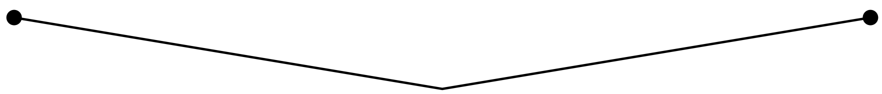
and release it. The two anchor points of the string are no longer pulling it left and right; they are pulling left and a little down, and right and a little down:



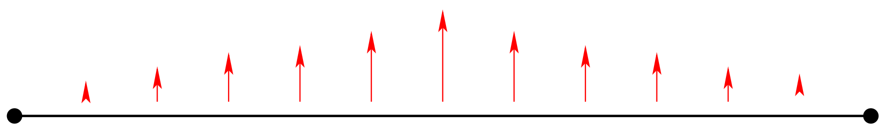
and if I add up the forces, there is a downwards force on the string. That will start the string moving. The bigger the tension, the faster the string will move. The heavier the string, the slower it will move. A moment later, the string will be straight, but moving:

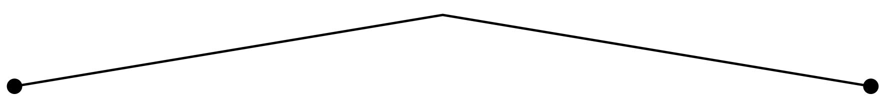


Now there is no *net* force on the string; it is pulled to the left from one side and the right from the other, and these forces cancel. However, since it is moving, it will continue to move until something pulls on it to stop:



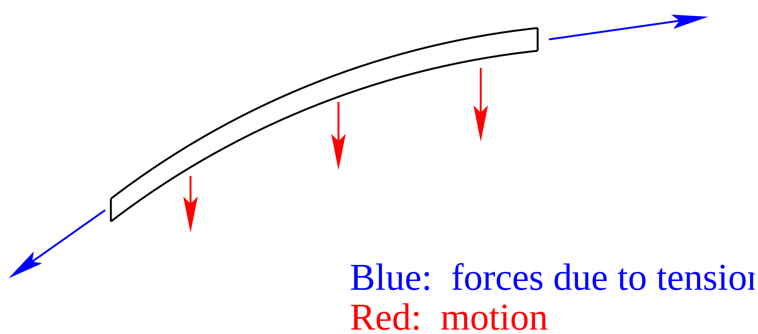
at which point the forces on the end are each pulling it somewhat upward. The motion of the string will reverse:





and we see the string is undergoing periodic, resonant oscillations.

To calculate the possible patterns and frequencies of oscillation that a string can make, we have to think about the string "little bit by little bit." Think of a short bit of string. Call the tension T (unfortunately the same letter as the period, but for this lecture T will be tension), the mass per unit length μ (measured in kg/m, meaning how many kg weight a meter of string would weigh; usually μ is much less than a kg per meter), and imagine that the little bit of string may be bent, and may be moving:



The forces will *not* balance to zero if the string bit is curved. The string bit can be moving, and if the motion is not uniform, the curvature will change.

Recall that, for the air, the thing which made the air move was a *difference* of pressure across a bit of air. Here, for the string, it is a *difference* of slope for the string between the two ends. For the air, the thing which made the pressure change was a *difference* of the air speed, between the front and back of a bit of air. For the string, what makes the slope change is a *difference* of the motion between the front and back. Therefore, there is an analogy between the case of a string and of air:

Air	String
Pressure P	Tension T
density ρ	mass per length μ
Compression	slope of string
Difference in compression	difference in slope
Air motion $v_{\rm air}$	String motion $v_{\rm string}$
Difference in $v_{\rm air}$	Difference in $v_{\rm string}$

The analogy means that bending and motion of the string propagates as a wave along the string, just as compression and motion in the air propagates as a sound wave. The speed of the "sound wave" on a string is,

$$v_{\rm sound,string} = \sqrt{\frac{T}{\mu}}$$

just as the speed of sound in air is $\sqrt{P/\rho}$ (up to that correction $\sqrt{c_p/c_v}$, which is absent for the string).

What are the normal modes of vibration for a string? Invariably, the ends of a string are fixed, which means that the string cannot move there. "Cannot move" is like "air cannot move" which is like what happens at the closed end of a pipe. Therefore, the normal modes of vibration of a string are those of a closed-closed pipe, and the frequencies are

$$f = \frac{v_{\rm sound,string}}{2L} \;\times\; 1, 2, 3, 4, 5, \ldots$$

which means that strings automatically have a harmonic overtone series. The vibration patterns have nodes: the pattern with $f = 2f_{\rm fundamental}$ has a node at the midpoint (1/2) of the string, the pattern with $f = 3f_{\rm fundamental}$ has nodes at the 1/3 and 2/3 points, the pattern with $f = 4f_{\rm fundamental}$ has nodes at 1/4, 2/4, and 3/4 of the way along the string, and so forth. If you gently touch a finger to a point on the string, it is analogous to opening a register hole at the same point on a tube. Any pattern which has motion at that point will be damped, any pattern which does not will continue to resonate. Therefore, putting a finger gently at the middle of the string will "kill" the modes with frequencies of 1,3,5,7,... of the fundamental and leave those with frequencies of 2,4,6,... of the fundamental–that is, it will double the frequency the string plays. Putting the finger at the 1/3 point triples the frequency. Putting it 1/5 of the way along the string makes the frequency go up 5 fold (2 octaves and a major 3'rd). (So does putting at finger at the 2/5, 3/5, or 4/5 point.) This way of playing harmonics of the strings is widely used on the harp and used somewhat on the other string instruments (bowed instruments, guitar, etc). However, it does not give enough tuning options to be sufficient by itself.

There are 3 remaining design issues in making a string instrument. They are,

- How is the instrument designed to be able to produce many different frequencies (preferably a chromatic scale with a range of at least 3 octaves)?
- How is sound to be produced efficiently?
- How is a sustained sound, rather than a sound which dies off, to be produced?

First, let us address the question of producing many frequencies. One method is to have many strings, one for each pitch needed. This is the solution used by the harp. The added advantage is that the instrument can play several notes at once. To some extent this method is used by other string instruments too; most have several strings (violin family instruments have 4, most guitars have 6, some guitar family instruments have more).

Another method is to have a way to change the length of a string. This is done on the violin and guitar family instruments by having the string run, for most of its length, slightly above a **fingerboard**. By using a finger to clamp the string to the fingerboard, the effective length of the string is changed from ending at the nut (a bar or bit of wood which the string runs over just before going into the peg box in the scroll) to ending where the finger presses the string. On guitars and some other instruments, little metal ridges called **frets** are placed along the fingerboard. If the string is pushed down between two frets, it automatically ends at the fret, rather than the finger, which ensures accurate intonation. (This is not done in the violin family of instruments to leave open more freedom in intonation, such as slides and vibrato, which is possible but harder on a guitar.)

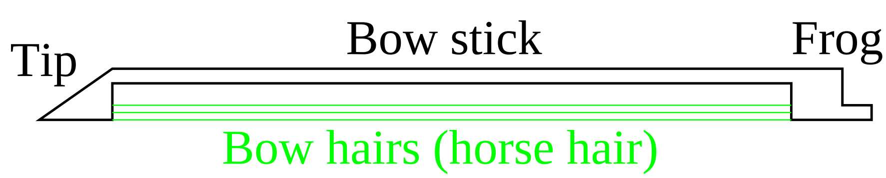
What about efficient sound production? A string is so thin that its motion through the air barely makes any air move, and is very inefficient in making sound. (Turn off the amp on an electric guitar to see this.) What makes sound efficiently is a large surface area moving back and forth by a substantial amount. To move by a substantial amount, it typically needs to be thin. This is accomplished by having the string go over a **bridge**, a piece of wood standing up vertically from the body of the instrument. The string bends by an angle as it crosses the bridge, as shown in figure 26.1, which means that the string exerts a downward force on the bridge. The bridge acts as the other end of the vibrating section of the string (opposite end to the nut, finger, or fret). The bridge absorbs some of the vibration energy of the string and transfers it into the body of the instrument. On some instruments (piano, harp) there is a large flat board called a **sound board**, which acts as the large vibrating surface which carries the sound into the air. On violin and guitar family instruments, this role is played by the front plate of the instrument, and to a lesser extent by the back and sides as well. The presence of an air cavity in these instruments is also helpful.

The sound board, instrument body, and instrument cavity typically have a number of resonances of their own, typically of fairly low Q. These tend to enhance any note or harmonic with a frequency lying close to a resonant frequency of the instrument. One of the design goals or features of string instruments is to make the resonances be widely scattered in frequency. In particular, a violin family instrument's apparent symmetry is much like your body's symmetry; once you look inside at the "guts" you find out it is nowhere near symmetric. The top plate's thickness is nonuniform and differs from left to right side; there is a reinforcing wood bar running most of the length of one side of the body, and a dowel (the soundpost) running between front and back of the instrument, offset on the side of the instrument without the reinforcement. Guitar reinforcement is also quite asymmetric. This ensures that the resonances reinforce many frequencies, rather than giving large reinforcements to a few frequencies.

The final question is how the instrument produces a steady sound. On all plucked, picked, strummed, or struck (hammer) string instruments, the answer is, that it does not. The string is set in vibration by some sudden process. The sound features an abrupt attack and then a decaying ringing. To make the sustain musically interesting, these instruments are usually constructed so the string's vibration will have quite a large Q (which is not too hard to do with a string instrument). In particular, the angle the string bends in going over the bridge is not too large, so the coupling between the string and the bridge is not too large.

For the bowed string instruments (violin family instruments and a number of other "traditional" instruments), the string *is* driven in a way which produces a steady tone. The way the bow acts on the string increases the vibration of the string in a way slightly analogous to how a reed increases the vibration in an air resonance.

Here is the idea. The bow is a wooden stick, supporting a mass of hairs (literally horse tail hairs) held under tension by the stick.



The hand holds the frog of the bowstick, which contains a mechanism for adjusting the tension on the bowhairs. Bowing the string means pushing the hairs against the string and dragging them across the string. This is generally done quite close to the bridge, not in the middle of the string.

The string vibrates back and forth underneath the bowhair. The vibration consists of the string moving with the bowhair, then moving against the bowhair, then moving with the bowhair, then against it, ... The friction of the bowhair against the string means that the hair is always pulling the string to move along with it. Pushing a car in the direction it is already moving adds energy, and pushing a car against the direction it is moving takes energy away. Similarly, when the bow is moving with the string, it is adding energy to the string's motion, but when the string moves back the other way, the bow actually takes energy out from the string.

The key to the operation of the bow is that it is a stick-slip action. While the bow moves with the string, the bowhair and string are actually sticking together–the string and the bowhair just above it are moving at the exact same speed. When the string moves back, it is slipping against the bow hair. The hairs are pulling on the string by friction, and friction turns out to be larger between two things which are sticking together than two things which are sliding against each other. This is why, when you are walking on ice, as long as you put your foot straight down, you are fine, until the moment the foot starts to slip–then there is no stopping it, because its friction against the ice suddenly reduces. Once you start to slip, you slip more easily. This is also the idea behind anti-lock brakes on cars. As soon as your tires start to skid, their friction against the road is reduced, and your car does not stop as well. Actually letting up on the brakes, until the road forces the wheel to turn again, increases your friction against the road and lets you brake faster.

The reason that bows are made with horse hair is that it turns out that hairs have an especially large difference between static (things sticking together) and dynamic (things slipping against each other) friction. This is probably because hairs are covered with tiny microscopic scales, which stick out and catch on the things touching them. (These scales are behind most of the properties of wool, such as its scratchiness and the way wool clothes shrink when they are washed.) To make the friction even larger when the bow sticks, the hairs are covered with tiny grains of rosen (dried pine tree sap).

Because the energy in the string is being continuously replenished by the bow action, the violin family instruments can afford to have a high bridge; the strings bend by a larger angle going over the bridge than in a guitar family instrument, which means that the vibration of the string is more quickly drained into the instrument body, and hence into sound. This (along with a more highly developed instrument body, violin makers would claim) allows the bowed instruments to play substantially louder than guitar family instruments (without amplification). The bridge is also arched, rather than straight, so the strings do not form a straight line across it, making it possible for the bow to play one string at a time. The large coupling to the bridge means that, when violin family instruments are plucked, the sound dies away faster than in a guitar instrument.

Besides the difficulty of good intonation on an instrument without frets, the main reason that beginners on violin family instruments sound so bad is that they have not learned to control the slip-stick action of the bowhair on the string. If the hairs are forced down with too much pressure or not enough bow speed, they tend to stick while the string is moving against the hairs, leading to a “crunching” sound. If the pressure is too small or the bow speed too fast, the hairs tend to slide across the string rather than sticking, leading to a wispy sound. The right range of pressure to get a proper stick-slip action varies depending on where on the spot on the string being played, the pressure, and the bow speed. These control parameters (bow placement and pressure) are also what determines loudness, and how fast the bow moves (which must be controlled depending on the length of the note).

Chapter 27

Percussion

This lecture will just try to outline the issues involved in percussion instruments and illustrate with a few examples. No attempt will be made to be comprehensive. A much better and more complete discussion can be found in the course book.

A percussion instrument is any instrument where you hit something to put some vibrational energy in, and then wait as the instrument converts the vibrational energy into sound. A simple example is a thin plate of metal. The metal can bend in a number of different ways (normal modes). When you hit it with a hammer, the plate starts bending in each of several of these modes at once. Depending on the place you hit and the properties of the hammer, different modes are excited different amounts. (Metal hammers excite a lot of very high frequency modes, which sounds bad, so people tend to go for mallets made with rubber, wood, leather, or other somewhat softer materials, which are in contact with the plate for longer and so don't excite high frequency modes as much.)

Percussion instruments invariably involve resonances in the instrument. Therefore the instrument has to be made in a way that it has some (bending or vibrational) resonances. There are several other design problems to overcome:

- How is sound production to be made efficient–or is it?
- How are overtones to be tuned to be in harmonic relation–or are they?
- How can the relative loudness of different overtones be controlled?
- How long does the resonance last, and how can that be controlled?

Consider a metal plate struck by a hammer, for instance. The sound production comes from the metal moving against the air. The larger the plate, the more air it pushes against. The thinner the plate, the further it moves with the same strength of hammer blow. Therefore, a large, thin plate makes more sound. This leads in the direction of the cymbal, which is a large thin plate of metal. Tuning the overtones in a metal plate is difficult. For a random shaped plate, they are not in tune, and it just makes a noise. A bell can be thought of as a plate which is in a very strange shape (no longer flat), and it turns out that the shape and thickness is cleverly designed to make several harmonics be approximately harmonic. The relative loudness of overtones can be controlled by the material of the hammer and the position of the hammer strike, at least to some extent. If I hold a metal plate with my hand, the resonance is quickly damped away into my hand. Suspending it by drilling a small hole through the plate and hanging it with a string eliminates this damping and lets the resonance last much longer.

Now we briefly survey some of the percussion instruments. One family of percussion instruments involves long thin bars, struck by a mallet or hammer. A uniform bar, much longer than it is wide and much wider than it is thick, and suspended so the ends are free, has resonant frequencies at about

$$f = \frac{f_{\rm fund}}{9} \times \Big(9 \, , \, 25 \, , \, 49 \, , \, 81 \, , \, \ldots \Big)$$

where the numbers are 3^2, 5^2, 7^2, and so forth. This is not a harmonic series.

In the glockenspiel, one makes a chromatic scale of bars with frequencies in a range around 800 to 4000 Hertz. The suspension and mallets are chosen so that only the fundamental is audible. There is a sharp attack which contains other harmonics, but they quickly die away, leaving a pure sine wave sustain.

The xylophone is similar, but with two modifications. First, to enhance the loudness, tubes are placed under the bars, with resonant frequencies matched to the resonant frequencies of the bars. The tubes support air resonances at the same frequency, strongly enhancing the conversion of sound from bar vibration into sound in the air. This allows the instrument to make sound efficiently at much lower frequencies, so the range can extend much lower than in the glockenspiel. Further, the bars (typically made of wood) are carved to be thinner in the middle than on the ends. This lowers the fundamental relative to the overtones. The amount of thinning is chosen so the first overtone has 3 times the frequency of the fundamental, which also happens to be the first overtone of the (open-closed) resonator tube. Because the instrument possesses a harmonic as well as the fundamental, it has a more interesting tone color.

Chimes are also based on long bars. They use that three of the frequencies, $81 = 9^2$, $121 = 11^2$, and $169 = 13^2$, are pretty close to $2 : 3 : 4$ relation. By putting a weight on the top of the bar, the upper overtones are pulled flat, making the relation more nearly $2 : 3 : 4$. The ear ignores the low tones and fills in the missing fundamental to hear a tone with harmonics.

Cymbals *revel* in the randomness of the overtones. They are thick near the center and thin at the edges to make sound production efficient and vibration retention longer lasting.

Because the bending motion on the cymbal can be large, vibrations can transfer from overtone to overtone (something which does not usually happen; vibrations are usually linear). Therefore the cymbal can store vibration in low frequency forms and transfer it to high frequency (more easily heard and more efficiently converted to sound) some time after the cymbal is struck.

Bells are plates which have been bent so much that it is no longer useful to think of them as plates. By carefully designing the shape and the way the thickness varies, the bell is given a series of overtones of frequencies

$$f = f_{\rm main} \times \left(\frac{1}{2} \, , \, 1 \, , \, 1.2 \, , \, \frac{3}{2} \, , \, 2 \, , \, \frac{5}{2} \, , \, 3 \, , \, 4 \, , \ldots \right) \, .$$

Depending on the strike point, different harmonics are louder. Several harmonics roughly form a tone and harmonics.

The simplest drums consists of a membrane (traditionally made from an animal hide) stretched over a circular frame. The mathematical problem of finding the vibrational modes of an ideal membrane stretched over a perfectly stiff frame have been solved long ago, and unfortunately, the resonant frequencies are in a nearly random succession, getting closer and closer together in frequency as the frequency goes up.

[To find the resonant frequencies for this ideal situation, one should solve the wave equation

$$\nabla^2 z = \frac{1}{c^2_{\rm sound,membrane}} \frac{d^2 z}{dt^2}$$

in cylindrical coordinates. Here z is the height of the membrane. To find periodic (resonant) solutions, one should use $z = z(r) e^{-i\omega t}$. The boundary conditions are that z is regular at $r = 0$ and vanishes at $r = R$ the radius of the drumhead. The problem can be solved by separation of variables, writing $z = z(r)e^{in\theta}$, with n an integer; the equation for $z(r)$ becomes,

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \frac{n^2}{r^2} + \frac{\omega^2}{c^2} \right) z(r) = 0$$

which is called the **Bessel equation**. The solutions are some nasty functions, $J_n(r)$, called Bessel functions of the n'th degree. A great deal is known about these functions but there is no simple statement about where they have zeros (which one must know to determine the condition on ω^2/c^2 for $J_n(R) = 0$ to hold).]

For the ideal drumhead, relative to the lowest frequency, the vibration frequencies are,

$$f = f_{\rm lowest} \times (\mathbf{1}, 1.59, 2.14, \mathbf{2.30}, 2.65, 2.92, 3.16, 3.50, \mathbf{3.60}, 3.65, 4.06, 4.15, \ldots) \, .$$

If you see a pattern, you know something no-one else knows. The modes in bold are the ones where the drumhead is moving up and down at the middle; for the other modes, the drumhead moves up and down on either side of the middle but the exact middle is at rest. If you strike a drum with these resonant frequencies, it will sound like noise which is not periodic and not of a definite pitch (which is true of a lot of drums). The different resonances appear as different patterns on the drumhead; striking the drumhead at a particular point excites the patterns which vibrate a lot at that point more than the patterns which vibrate at other points. This is why the drum sounds very different when struck in different places.

Many drums have the underside of the drumhead closed off rather than open to the air. That means that, when the drumhead moves up and down, it compresses and decompresses the air inside. This acts as an extra "spring," stiffening the drumhead–but only for those modes where the drumhead as a whole moves up and down, which are the ones in bold in the previous list. These frequencies are raised, especially the first one. They also damp down the fastest, especially the first one. By clever design, for a kettle drum the resonant frequencies become (naming the next to lowest frequency the "prime" frequency),

$$f = f_{\rm prime} \times (0.85, 1, 1.51, 1.68, 1.99, 2.09, 2.44, 2.89, \ldots)$$

The sequence $1, 1.51, 1.99, 2.44, 2.89$ is almost $1, 1.5, 2, 2.5, 3$ which is the harmonic series of a missing fundamental. Your ear reconstructs this as "almost" a definite pitch.

There are more sophisticated drums which make more nearly periodic sounds. Perhaps the most sophisticated are the tin drums used in the Carribian, which are discussed extensively in the text.

In conclusion, the challenges of designing a percussion instrument are to make it loud, to make it sustain, to make it produce a periodic sound, and to make it possible to control the relative loudness of different overtones. Not all instruments solve all of these problems, nor is it always the intention to. Percussion is really a grab-bag of a large number of very distinct instruments.

Chapter 28

Piano, a Percussion String Instrument

One of the problems–often the most severe problem–in the design of a percussion instrument was, that it is difficult to get the resonant overtones to fall in a harmonic series. Well, a string fixed on both ends automatically has an overtone series which is harmonic (frequencies in ratio 1:2:3:4:5:. . .). Why not have a percussion instrument where the things being struck are long, thin strings under tension?

This is an old idea; hammer dulcimers existed in antiquity. However, the piano takes another idea, also very old, of having a keyboard, and combines them. In a hammer dulcimer, the player holds the hammers in the hands and directly strikes the strings with them. In a keyboard instrument, there is a row of keys which each actuate some sound producing device, such as a hammer or pick. This has a very substantial advantage for the performer; it makes it easy to hit the right sound producer and to hit many at a time. A keyboard is such a good idea that it is used in a very wide variety of instruments. For instance, in a harpsichord, each key operates something to pluck a string; in a celesta, each key operates a hammer which strikes a metal rod (as in a glockenspiel); in an organ, each key operates a valve which lets compressed air flow into a pipe. In the piano, each key operates a hammer which strikes a string (or more than one string tuned to the same frequency).

There are two big design complications associated with this idea:

- The action of the hammer must be quite complicated. The harder you push the key, the harder the hammer must be thrown at the string. The hammer must be thrown at the string, so that it is not left resting against the string (which would damp the vibration away again). It is generally not desired that the string should ring until the vibration dies away; the player should be able to control how long the string vibrates. There must therefore be dampers, one for each key, and the way the key is pressed and released must allow the player to control when the damper will and will not press against the string and absorb the vibration.

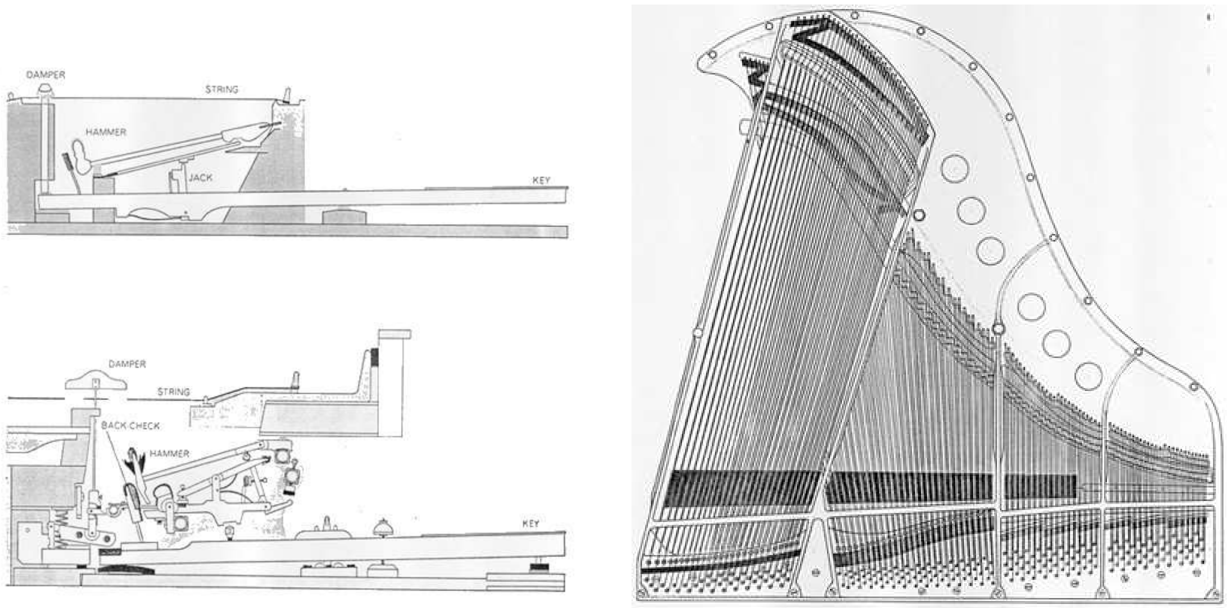


Figure 28.1: Left: illustration of a piano action, intended just to give the idea of how complicated it is. Right: illustration of how the strings are arranged to fit more compactly inside the body of the piano.

- The instrument is best if it has an enormous frequency range. Pianos conventionally have a range of more than 7 octaves, from A_0 to C_8. This (you can check) is about a factor of 154 in frequency. If all the strings are of the same thickness and are under the same tension, then if the shortest string is a few centimeters long, the largest must be a few *meters* long.

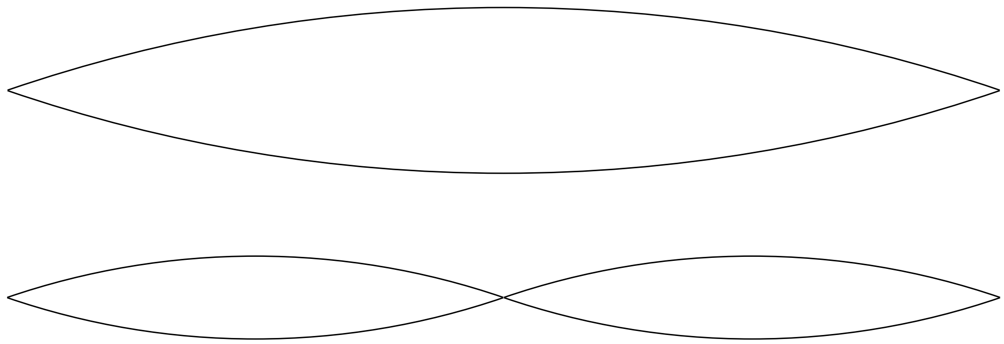
The first problem is solved by having a tremendously complicated "action" actuated by pressing each key. As the key is depressed, the force is transferred into throwing a hammer at the strings. Simultaneously, as the key is depressed, the damper is drawn back from the strings. The harder the key is pushed, the harder the hammer is thrown (hence "piano-forte," or "soft-loud")–this is the key advantage over the older harpsichord. The damper remains up as long as the key remains depressed. The key is counterweighted so that it will rise as soon as the finger is lifted off, and the damper will move up and stop the vibration of the string. Therefore, how long you hold down the key controls the "sustain." To make life more interesting and the design of the action more difficult, there are pedals, which modify what the action does. The exact function of each pedal varies between different piano designs, but usually includes a "sustain" pedal which removes all the dampers from the strings. The action must be designed so that pushing the pedal moves something which changes the way the action behaves. Piano actions are terribly complicated, as illustrated in figure 28.1.

The second problem requires that, for compactness, some of the strings go at an angle above others, as also shown in the figure. Even so, it is impossible to make the strings as long as would be desirable. The speed of sound on the strings for the lowest notes must be made lower than on the higher notes, so they will not be as long as otherwise required. Either they must be made under low tension (which would lead to a wimpy sound), or they must be made thicker (the solution actually chosen). The lowest part of the piano register is therefore made of thick, wound strings which have a larger mass per length μ, and therefore a lower speed of sound. How many strings are wound, and how fat they are, depends on how big the piano is. The main advantage of a grand piano over an upright is that the lowest strings can be longer, and therefore do not need to be as heavily wound.

How is the vibration to be converted into sound? We already discussed that a string moving through air makes almost no sound. A large surface has to move back and forth against the air for efficient sound production. This is done in the piano, as in other string instruments, by having the piano strings pass over a bridge on a soundboard, a large flat piece of wood (typically spruce, a preferred material for other string instruments as well) with reinforcing braces. It is the vibration of the soundboard which produces the sound.

The problem of tuning a piano turns out to be quite complex. first, consider tuning each individual note. On the piano, most notes are played not by a single string, but by three. (This is not true of the low pitched, wound strings.) If there were only one string, or if the three were exactly in tune with each other, then their vibration energy would quickly be absorbed by the bridge into the soundboard. This would lead to a good attack but no sustain–a dead "plunk" sound. Instead, there are three strings which are deliberately slightly mis-tuned from each other. Right after they are struck, they pull up or down on the bridge in unison. Energy rapidly goes into sound production, giving a good attack to the note. But after half a beat frequency of the deliberate mis-tuning, one string is "up" when another is "down." That means that the 3 strings are no longer all pulling on the bridge at the same time. Their forces on the bridge mostly cancel out, and they no longer lose their energy so efficiently into the soundboard. This allows a better "sustain".

The other complication of piano tuning is that the strings do not actually have perfectly harmonic overtones. The very highest strings are very short, and the very lowest strings are very fat. In both cases, the strings do not actually act as perfectly "floppy." The rigidity of the strings is important. Compare the first and second vibration modes of a string:



For a string under tension, it is the slope of the string which tells how much energy is stored. Recall that, in the analogy between sound on a string and in the air, slope was analogous to pressure in the air. However, for a string with its own stiffness, it is how much bending of the string there is, that is, the curvature is what costs energy. For a given amount of slope, the curvature is higher for higher overtones. While, for an ideal string fixed at both ends, the overtone series is,

$$\text{ideal string:} \qquad f = f_0 \times \Big(1, 2, 3, 4, 5, \ldots\Big)$$

for an ideal bar (the extreme case of the string's stiffness deciding what happens), fixed on both ends, the series is,

$$\text{ideal bar:} \qquad f = f_0 \times \Big(1, 4, 9, 16, 25, \ldots\Big)$$

For a string which has some stiffness, the series goes like,

$$\text{Real string:} \qquad f = \Big(f_T + f_b\,,\ 2f_T + 4f_b\,,\ 3f_T + 9f_b\,,\ 4f_T + 16f_b\,,\ \ldots\Big)$$

where f_T is the frequency of the fundamental because of the string tension and f_b is the frequency because of the stiffness of the string. Even in the top and bottom strings on the piano, the stiffness is only a small correction to the frequencies of the notes. Therefore, we can treat f_b as a correction in the above. Defining $f_0 = f_T + f_b$, and $A = f_b/f_T$ (the anharmonicity), the frequency series for a real string is,

$$\text{Real string:} \qquad f = f_0 \times \Big(1\,,\ 2{+}2A\,,\ 3{+}6A\,,\ 4{+}12A\,,\ \ldots\Big)\,.$$

We see that this correction will make the overtones' frequencies sharp compared to the fundamental. That is, for a string where you cannot neglect the stiffness, the overtones get sharper as you go up the overtone series. Remember that, even if this is a small effect, your ear is very sensitive to accuracy in frequencies, so it is still important.

What is most important to the listener is that there are no beats. When playing a note and its octave, this means that the octave must be in tune with the first overtone of the lower note–rather than being at exactly twice the frequency of the lower note. In the middle of the piano's range, the strings are long and thin, so A is almost zero. The overtones are harmonic, and tuning should be done in the normal way. However, the highest strings are short, so their stiffness is more important. The lowest strings are fat, so their stiffness is also more important. Therefore, at the very top and the very bottom of the range, the overtones are sharp and the intervals must be "stretched." That means that the top notes on the piano must be deliberately tuned sharp. For instance, G_7 must be more than twice the frequency of G_6, because the overtone of G_6 is sharp and G_7 needs to coincide with that overtone. C_1, on the other hand, must be tuned flat, so that its too-sharp overtone will line up with C_2. (Recall that, in tuning a piano, one starts in the middle and tunes towards the edges of the range.) This "stretch tuning" is particularly severe for the bottom range of upright pianos, where the strings are much shorter than would be ideal and must be very heavily wound. This is why the bottom octave or so of an upright piano sounds so bad, no matter how well tuned and how high quality the upright is.

Chapter 29

Voice

29.1 Anatomy of the Voice

The voice is the oldest and in many ways most complicated of the musical instruments. We will discuss it in a series of lectures which will all be covered in this set of notes. They are broken up into a discussion of the anatomy of the vocal system, a discussion of vowel production, and a discussion of consonant production.

The vocal system is the upper part of the respiratory system. It has many parts, some completely familiar to everyone and some a little less so; we will spend more time describing the less familiar but important components.

The vocal tract is shown in figure 29.1. It begins at two openings to the outside, the nose and the lips. The nose leads into the **nasal cavity,** the mouth into the **oral cavity**. These meet at the back of the mouth and join into a single tube or cavity called the **pharynx**, which in turn divides into two tubes, the **esophagus**, which goes to the stomach, and the **trachea**, which goes to the lungs. The top of the trachea is a formation called the **larynx** or voice-box.

It should be emphasized that all of these anatomical features serve several roles, some of them more important than speech. As you know, the most important roles for the vocal tract are in breathing and in eating. The fact that the same features which are needed for these jobs (such as lips, teeth, and tongue, all essential in eating) can also be used so elegantly in sound production. Keep this in mind when we look at the form and function of each piece.

The nasal cavity runs as far back into the head as the oral cavity, and is quite complicated. It contains a series of bony plates with soft tissue on them, the olfactory organ, responsible for the sense of smell. It also connects to a number of air cavities further up into the head, called sinuses, which regulate the pressure in the brain cavity. Its role in speech is minor but not irrelevant. (You can see how big by pinching your nose closed and then talking. Most of the sounds are unchanged; a few, particularly “n,” “m,” and “ng,” are ruined.) The cavity

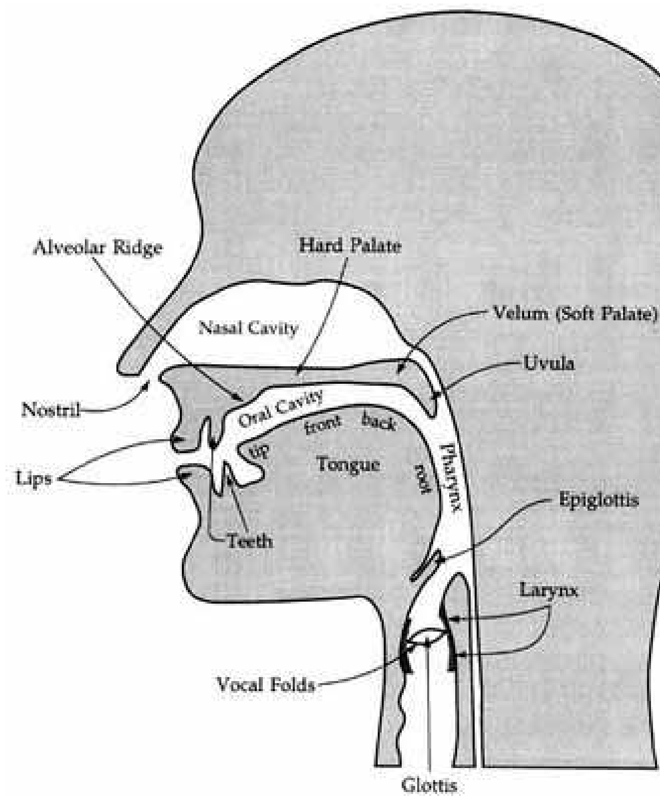


Figure 29.1: Main parts of the vocal system.

has no flexibility so there is no way to control its shape, which means that its role in speech is passive.

The oral cavity has a number of features everyone is familiar with, and some people are less familiar with. At the opening are the lips, then the teeth, which are held in by bones: the mandible below and the maxillary bones above. The maxillary bones also form the *front half* of the roof of the mouth, the **hard palate**, which is inflexible. The underside of the oral cavity is mostly taken up by the tongue, which (as you know) is a very flexible organ composed mostly of muscles. Motion of the tongue and lower jaw allows tremendous variability in the size and shape of the air cavity in the mouth; motion of the jaw and lips allow great control of the size and shape of its opening. The muscles of the jaw go up the sides of the head and attach to the temporal bones of the skull. Most of the muscles of the tongue are anchored on a bone in the throat, called the **hyoid bone**. This horseshoe shaped bone is at the front top of the throat just above the voicebox. It is held in place by tendons attaching it to muscles and by cartilages; it is the only bone in the body which does not attach directly to any other bone.

The back of the roof of the mouth is the **soft palate** or velum. If you touch it with your finger or tongue, you will feel that it is soft and flexible (and will notice that it is the site of your gag reflex). It can move up and down a little. When it moves up, its tip, the uvula, closes off the nasal cavity from the pharynx (important in swallowing, breathing through the mouth, swimming, and much of speech). The oral cavity can be opened and closed by the base of the tongue, the tip of the tongue, or the lips (or any combination).

The pharynx is a slightly flexible tube, connecting at the top to the nasal and oral cavities and at the bottom to the trachea and esophagus. Everything you breathe and swallow goes into this one tube, which then splits again. Just as the soft palate and base of the tongue control whether the two cavities are open or closed into the pharynx, there *must* be a way to control which of the two tubes is open for things to go down into. The **epiglottis** is a flap at the top of the larynx. Depending on whether it is raised or lowered, it either opens or closes the opening of the larynx. When you swallow you reflexively lower the epiglottis to close the larynx and trachea. This occasionally fails, usually when you are suddenly distracted. Your choking reflex is to use a burst of air to clear out anything which accidentally gets past the epiglottis.

What about the larynx, or "voice box"? It is a cartilage-cased "box" made up of cartilages, ligaments and tendons, muscles, and membranes. The airway goes through the larynx as a roughly tube shaped passage through its middle, but the walls of the tube have a number of pieces and processes which make it far from an even, straight tube. Only two will be of real interest to us. The first is the epiglottis, which we already described. The second are the vocal folds. These are two pieces of tissue, roughly semi-circular, one on either side of the trachea. They extend from the walls towards each other, mostly closing off the airway, see figure 29.2. They are made of membrane, tendon, and muscle. At the back, they are attached to two cartilages called the **arytenoids**; at the front they attach to a cartilage called the **thyroid cartilage**. In men the larynx is larger and the thyroid cartilage sticks out in the front of the throat, forming the "Adam's apple." A number of muscles attach to their edges and to the arytenoids, and a few muscles actually pass through the vocal folds themselves.

The vocal folds can be opened or closed by using the muscles attached to the arytenoids. One muscular action brings the arytenoids together; the vocal folds then meet in the middle and close off the airway entirely. Relaxing these muscles or using other muscles draws the arytenoids apart, so a space opens between the vocal folds. This gap, through which you breathe, is called the **glottis** (epiglottis means "above the glottis").

The vocal folds are *not* cords. They extend from the walls of the air passage almost or completely to the middle (depending on how you pull the muscles).

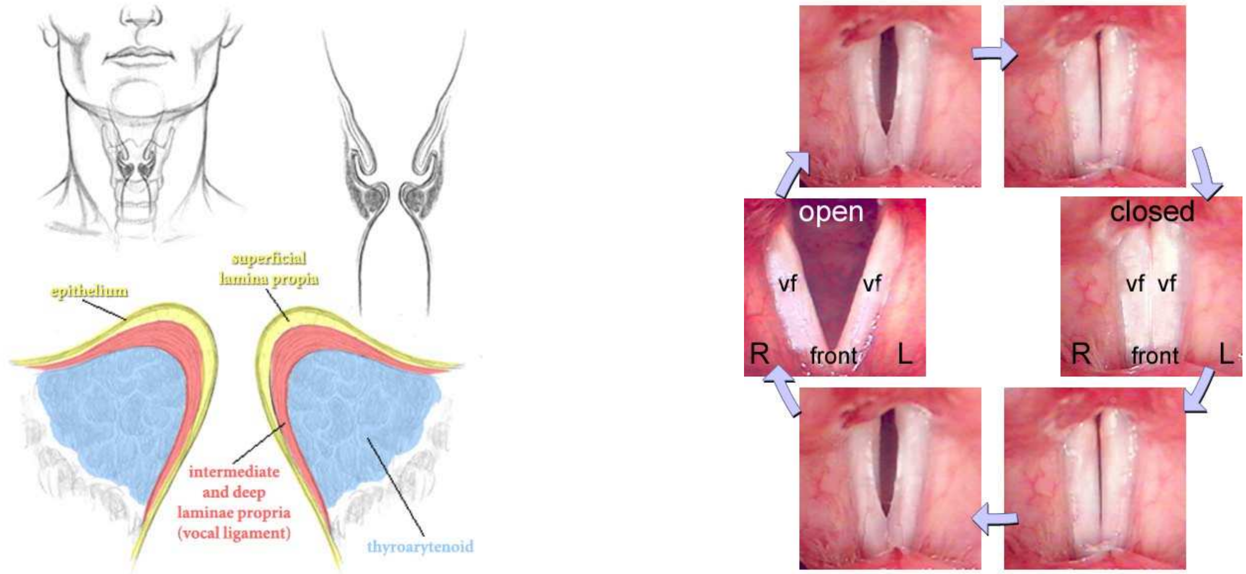


Figure 29.2: Vocal folds. Left: front view. Right: top view (actual photographs) of how they open and close.

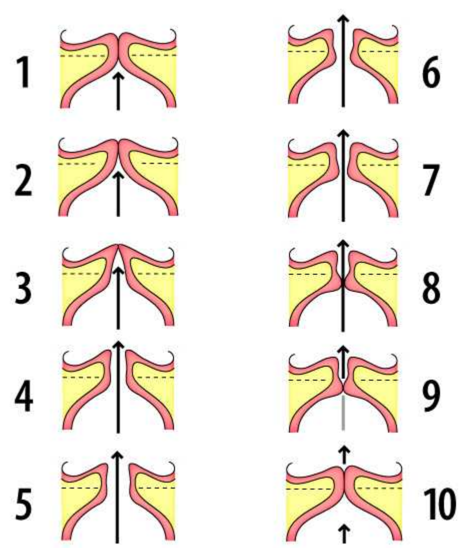


Figure 29.3: Time series of how the vocal folds move as air is forced through them during sound production.

29.2 Sound production and Vowels

When you close the vocal folds and then force air up through them, they make sound. Let us see why. The muscles of the chest compress the lungs and raise the pressure of the air in the lungs and larynx up to the focal folds. This causes a situation rather like blowing out through your pursed lips. The pressure of the air underneath the vocal folds forces them up and open. A burst of air escapes. The vocal folds are not soft and flabby, though; they are elastic membranes, which spring back towards their natural position. They have a resonant frequency, set by their size, thickness, and stiffness (tension), and they will resonate at about this frequency, emitting a burst of air each period of the resonance. See figure 29.3. Each burst of air temporarily lowers the pressure just below the folds, so there is less upward force as they fall shut than as they open. Therefore the act of blowing through them increases the energy in the resonance, which is what keeps it going. Turning the air stream into a series of puffs generates sound, because it means the pressure and air speed above the vocal folds will vary with time. The sound will be periodic, with a frequency set by the resonant frequency of the vocal folds.

Several properties of the sound can be controlled at the vocal folds. First, since there are muscles attached to the folds and within the folds, their shape, size, and tension can be adjusted. By relaxing all the muscles, the vocal folds are made loose, which gives a low frequency of vibration. By tensing the muscles within the vocal folds, they become thicker, but also much stiffer (more elastic). A thicker fold is heavier and slower to move, but a stiff fold springs back faster; the latter effect is larger, so the frequency goes up. By tensing the muscles attached to the edges of the fold, they can be made thinner and tighter, raising the frequency. An untrained voice (most of us) can change the frequency of vibration of the vocal folds by about a factor of 4 (two octaves). Vocal training not only increases the proficiency in controlling these muscles, but increases their strength and conditioning, and can increase the range of the voice. Since most of the muscular actions increase the frequency, it is easier to train someone to increase the upper end of their singing range than to enlarge the lower end of the singing range.

In typical speech, men use frequencies around 100 to 150 Hertz, women use frequencies around 200 Hertz, and children (whose whole apparatus is smaller and therefore produces higher frequencies) are more typically around 300 Hertz. The top range of male singing depends on the voice, and might be in the 350 to 450 Hertz range. Womens' singing voices can extend much higher, towards 1000 Hertz, though this also depends on the woman. Mens' voices are classed as **bass, baritone,** and **tenor**, in order from low to high. Women are **alto, mezzo-soprano,** and **soprano** in order from low to high. Alto is higher than tenor; womens' voices are almost invariably higher than mens'. A few men sing counter-tenor (same range as alto) and a few women sing contralto (same range as tenor).

The exact way the muscles are used to stiffen the vocal folds also has an influence on the timbre of the resulting sound. If the folds are not pulled tight all the way, then the folds do not fall all the way shut. This gives a "breathy" tone. If the arytenoids are pulled together tighter, the folds are shut for more of the sound producing cycle, giving more harmonics and a gruffer tone. The two strategies for raising the frequency of the voice also give different sounding voice. When the muscles within *and* around the vocal folds are tightened, the folds become thick and tense. They shut more tightly, producing more harmonics. This is called the "chest voice." When the muscles within the folds are left loose but the muscles around the folds are tightened, so the folds are thin but tight, they do not close completely and fewer harmonics are produced. This singing style is called the "head voice" or "falsetto." This sound making strategy can usually achieve higher pitches than the chest voice, often much higher pitches, but some people find that it sounds silly.

This explains how the frequency of your voice can be adjusted, and a little about the tone color. What makes the difference between different vowel sounds, like "ah," "eh," "oo," and so forth? Making these sounds will convince you that it is the mouth. The key is that the sound is being produced at the bottom of a tube, running from the vocal folds to the lips, the nose, or both (depending on the position of the soft palate). The tube has resonances and the frequencies produced by the vocal folds which lie close to a resonance are enhanced.

The mouth is about 9 cm from front to back and the pharynx is about 8 cm from the back of the mouth to the top of the vocal folds. These numbers are approximate; they vary between individuals and they change slightly because the shape and length of the pharynx can be slightly adjusted by muscular action. *If* the mouth and pharynx formed a tube of uniform width, then since it is (almost) closed at the vocal folds and open at the mouth, the resonant frequencies would be at approximately,

$$f_{\rm res} = \frac{v_{\rm sound}}{4L} \times 1,3,5 = 500\,,\ 1500\,,\ 2500\,,\ \ldots\ {\rm Hz}$$

Note that all of these frequencies are substantially higher than the normal speaking range of frequencies. Therefore, unlike most musical instruments, *the human voice is not played at a resonant frequency of the air cavity*[1] . The frequency is determined by the tension and thickness of the vocal folds, which are heavy enough that their resonation is almost unaffected by the resonances in the air cavity above them. The role of the resonances of the vocal tract is entirely in strengthening certain harmonics of the voice and not others; namely, the harmonics which come close to coinciding with a resonant frequency of the vocal tract are enhanced. Because of this role in "forming" the timbre of the voice, the resonant frequencies of the vocal tract are called **formants**. They are numbered, from the lowest to

[1]There is one notable exception: sopranos singing high in their register are singing as high as the lowest resonant frequency. Well trained operatic sopranos actually adjust their throat and mouth with the note they sing to shift the resonance to match the note they are singing, greatly enhancing loudness. This is why operatic sopranos are so famous (infamous) for their loudness.

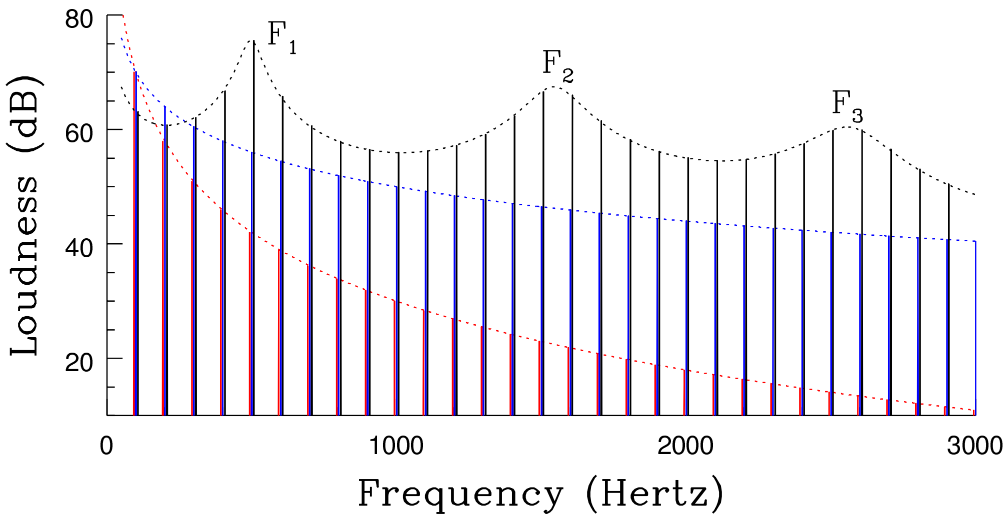


Figure 29.4: The effect of radiation and formants on sound. Since the voice is periodic, the frequency spectrum is a series of lines at the harmonic frequencies of the vocal frequency. The curve which shows how high these lines go is called the envelope. If the vocal folds were open directly to the air, the sound production would be like the red lines and envelope. Adding in the 6 dB per octave of enhancement for high frequencies to radiate out from the mouth opening gives the blue. The resonant enhancements near the formants gives the black.

higher frequencies, that is, they are called the **first formant** F_1, the **second formant** F_2, the **third formant** F_3, and so forth.

The resonant frequencies of the vocal system are not particularly high Q. This is mostly because the vocal folds are not completely closed for the full cycle. When they are partly open, they reflect most of the sound wave but let a part through, down the trachea to the lungs. The lungs are soft spongy tissues which almost perfectly absorb sound. This energy loss mechanism limits the Q. Recall that a resonance enhances any frequency which is close to the resonant frequency. Therefore the formants modify the tone color whether or not a harmonic of the sung frequency happens to coincide exactly with a formant frequency. The timbre of the voice is also affected by the efficiency of radiation from the mouth opening; like the wind instruments we have already discussed, a larger fraction of high frequencies escape an opening than for low frequencies, enhancing the harmonics by about 6 dB per octave. The influence of the formants and radiation efficiency on the spectrum of the voice is illustrated in figure 29.4.

The key to producing all the different vowel sounds is, that the shape of the vocal tract can be modified very widely and with great precision by adjusting the tongue, lips, and jaw (and to a lesser extent, the throat and soft palate). What this does is to change the resonant frequencies, modifying which harmonics of the voice are enhanced and which are not. For instance, in the letter "long e," (and even more in the German "ü"), the mouth and lips are narrowed. The pharynx is a wider tube and the mouth a narrower tube. This is a little like the wine bottle (Helmholtz resonator) we encountered in an earlier lecture. The bottom frequency F_1 is lowered, and the two next frequencies F_2 and F_3 move towards each other at around 2000 Hertz. This leads to an enhancement of low frequencies and of a range of frequencies around 2000 Hertz, but nothing in between. (Remember that the appearance of "ee" was close to a sine wave or sine plus sine with twice the frequency, with prominent, much higher frequency ripple. The low frequencies were from F_1, the prominent wiggles were from F_2 and F_3 being at close to the same frequency.) On the other hand, if you open the mouth and lips wide, "aah," then the tube is a narrower tube (pharynx) opening to a wider tube (mouth) opening to the world. The resonances are roughly the resonance of each tube, $f = v_{\rm sound}/4(L/2)$; so F_1 and F_2 move together around 1000 Hertz, F_3 moves up in frequency around 3000 Hertz. All of the other vowels are other ways of holding the mouth which cause the formants to move other places or change in Q.

Actually, in English many of the vowel sounds are actually **diphthongs**, that is, two sounds run together so tightly and so regularly that we think of them as a single sound. For instance, long i, as in "like", is really two sounds: "aah" and "yy". Long o, as in "own," is really "ooo" and "www". English uses diphthongs more than most languages.

29.3 Consonants

This does not begin to exhaust the list of tricks the vocal tract has for making sounds. The other tricks are grouped together as **consonants**. Most consonants involve the narrowing or closing of the airway to produce or modify a sound. However, the definition of what is a consonant and what is a vowel is made partly on usage (linguistics) rather than on the method of sound production. Some consonants involve the vocal folds; others do not. Some can be sustained; others have to be short.

The consonants in English are grouped into 5 types according to how the sound is produced:

- **Fricatives**: The fricatives are hissing sounds made then air is forced through a very narrow passage. This sound production is similar to how sound is produced by a flute, only without the flute. It relies on the physics of turbulence, which we will not try to understand in this course. The thing to know is, that when air passes rapidly through a short, narrow opening, it oscillates erratically as it emerges. If you are blowing through a horizontal slit, the air will alternately go somewhat upwards and downwards as it emerges. The time scale for this alternation is about twice the time it takes a bit of air to rush through the narrow spot (just as the ideal rate to blow across a flute is so the air takes about half a period to cross). For instance, if the air is going 50 meters/second and the slit is 5 mm across, the air needs $.005/50 = .0001$ seconds to get through, and the airstream will oscillate at around 5000 Hertz. It is *not* a periodic oscillation. It is noise made of frequencies in a range around this central value.

The fricatives are the consonants made by forming a narrow spot in the airstream and forcing air through it. If the vocal folds are also vibrated, the fricative is **voiced**; if they are not, it is **unvoiced**. The location of the narrowing and resulting consonants are,

- lip and teeth: "f" (unvoiced) and "v" (voiced)
- tongue tip and upper teeth: "s" (unvoiced) and "z" (voiced)
- tongue and both sets of teeth: "th" as in "thistle" (unvoiced) and "th" as in "this" (voiced)
- tongue and forward hard palate: "sh" (unvoiced) and "zh" (voiced)
- vocal folds themselves: "h"

Other languages have additional fricatives; for instance, in German the tongue and the back of the hard palate are used to produce "ch" as in "chanukah."

The fricatives sound different partly because the range of frequencies produced are different, and partly because the fricatives occurring further back in the mouth can be modified by the resonant structure still in front of them.

- **Plosives**: Plosives are consonants made by closing the airpath completely and then letting it open suddenly. There are three ways to do this: with airflow (aspirated); voiced; and both voiced and aspirated, a combination not used in English. (Neither voiced nor aspirated would not produce any sound.)

  The plosive sound depends on where the vocal tract is closed:

  - lips: "p" (aspirated), "b" (voiced), "bh" (both)
  - tongue and teeth: "t" (aspirated), "d" (voiced), "dh" (both)
  - tongue and hard palate: "k" (aspirated), "g" (voiced), "gh" (both)

  There are a number of other possibilities which are not used in English; the tongue and soft palate, the glottis (glottal stop), and so forth.

  Plosives cannot be "sustained," unlike vowels and the other consonants; they are always short.

- **Nasals**: Nasals are voiced sounds where the mouth is completely closed but the soft palate is moved to open the nose, so the sound emerges from there. The resonances of the pharynx-nasal cavity-oral cavity system vary depending on where the mouth is closed off, so different sounds are produced by closing the mouth at different locations:
  - mouth closed at the lips: "m"
  - mouth closed by tongue and hard palate: "n"
  - mouth closed by tongue and soft palate: "ng"
- **Liquids**: Liquids are voiced sounds where the airway is constricted almost as much as in a fricative, but not enough airflow is produced for a fricative sound. There are two in English, "r" and "l".
- **Semi-vowels**: Semi-vowels are vowel sounds which involve a narrowing of the airway which is more than in most vowel sounds. The two semi-vowels in English are "w" and "y". The distinction between a semivowel and a vowel is made mostly on its linguistic use rather than on the sound production mechanism; both semi-vowels are used as parts of diphthong vowels in English, as well as their use as consonants.

To clarify, the above discussion does not exhaust the list of sounds the vocal system can make, it only lists the consonants commonly used in English (and a few which are not). The full set of sounds the vocal system can make is vast, and no language uses all of them. Different languages use different ones, and make finer distinctions between some sounds which other languages consider equivalent. This is all learned and cultural. Your vocal system is capable of making any sound used in any language. The difficulty is that your brain has not trained to use the sounds you do not need in the language(s) you speak, and your ear has not trained to recognize the important differences between sounds needed for some other languages. The vocal training needed is really nontrivial, because in normal speech you make 10 to 20 different vowel and consonant sounds per second, which means that the muscular motions have to be memorized. Training the vocal system becomes more difficult with age, which is why people who learn languages after the age of about 12 typically retain an accent.

The other aspect of speech is **prosodic** features. This is the rising and falling of the voice frequency, changes in loudness (accents and emphasis), and changes in speed or silent gaps. These convey additional meaning, such as urgency, sarcasm, emphasis, the distinction between a question and an answer, and so forth. In Indo-European languages their use is analogous to punctuation in written language. In *tonal languages,* they are varied syllable by syllable, and depending on how the pitch rises or falls, the syllable has a different meaning (is a different word). Mandarin, Cantonese, and Thai are examples of tonal languages.

29.4 Singing versus speech

The main thing which makes singing singing rather than speech is the way the pitch and timing of the voice are modified. Typically, each syllable is made the length of a note and is pitched to be steady on a note of the scale, rather than hopping up and down to convey prosodic information. This is the first reason that it is harder to understand song than speech. The other reason is that there are typically some modifications of the way the vocal tract is used, to increase the power of the voice at the expense of intelligibility.

Singers, especially trained choral or operatic singers, typically widen the jaw opening and change the mouth and lip configuration to increase sound radiation and power (think of your choral director yelling over and over to open your mouths wider). There is also a lowering of the larynx and a widening of the pharynx (increasing the size of the pharynx). This often leads to a resonance of the pharynx cavity, called the **singer's formant**, at approximately 3000 Hertz. This happens to be the most sensitive range of human hearing, and is also above the dominant frequency region of most musical instruments. This is what allows a tenor to compete with an orchestra and still be heard.

What allows a soprano to compete with an orchestra and still be heard, especially in their upper register, is **formant tuning**. They are singing at a high enough frequency that a vocal formant *can* lie within the range of the fundamental of their voice. By adjusting the shape of the vocal tract, they can match a formant frequency of the vocal tract to the note they are singing. This causes a dramatic resonant enhancement of the loudness of the fundamental frequency. It is hard to do, because the vocal tract must be held differently for each note. This is particularly challenging on very fast passages. Therefore it is only highly trained operatic sopranos who can apply this technique. It also means that the note cannot be formed into any specific vowel sound. However, in this range of frequency (500 to 1000 Hertz), you could not tell apart different vowels anyway, because the harmonics of the voice are so far apart that they are not "sampling" the locations of the formants enough for you to tell where the formants lie. Therefore operatic writers only put words intended to be intelligible in the lower part of a soprano's range. It is the combination of sound power and the very large fraction of the sound power in the fundamental which allows opera singers to shatter wine glasses (if they are made of high quality crystal, so the vibration of the glass has a very high Q, and the singer tunes their sung frequency to coincide exactly with the vibration frequency of the glass, and the glass is held right in front of the soprano's mouth to get the maximum sound intensity, then the vibration of the glass grows so large that it shatters).