Sound - An Interactive eBook

Kyle Forinash and Wolfgang Christian

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 04/15/2025

Introduction to Sound

Sound: An Interactive eBook

Sound: An Interactive eBook consists of 33 interactive simulations which require the reader to click buttons, move sliders, etc. in order to answer questions about the behavior of waves and sound in particular. There are also dozens of links to YouTube videos and other online resources that pertain to the topics being covered as well as suggestions for laboratory exercises and sound clips for understanding the fascinating subject of sound and music. The goal was to create an engaging text that integrates the strengths of printed, static textbooks and the interactive dynamics possible with simulations to engage the student in actively learning the physics of sound.

This book began in the way most textbooks do, as notes put together for a new course. The physics of sound, however, lends itself particularly well to examples, demonstrations and student participation in experiments. There are thousands of YouTube videos of interesting sound phenomena and dozens of simulations related to the physics of sound and music. This book was created from trying to provide access to these resources in a single source, first from a web page, then as interactive simulations on web pages and finally as this interactive textbook.

Interactivity

Gutenberg's invention of movable type around 1450 did not revolutionize the content or the format of the information being provided. It did have the important consequence of speeding up and broadening access to information. In a similar way, much of our modern technology has accelerated and expanded access to the world's knowledge base. Instructors today routinely provide a course syllabus, course information, instructor notes, assignments, sample tests, supplementary reading, and web links to other material, all online using a course management system or simple web pages. Many university students now receive access to a PDF version of the course textbook when they register for a course.

These uses, however, are *not* interactive. Much like an enhanced printing press, this technology serves to accelerate the one way transfer of material from the instructor to the student. In this regard it is not much different from what was already being done 560 years ago by Gutenberg; the information flow is unidirectional, albeit much faster. While, in hindsight, Gutenberg's creation was seminal to mass education, the communications revolution of the past century has yet to produce comparable improvements in human learning. However, today's technology has the capacity to function much more interactively of which this book is an example.

> **Note**
>
> Not all platforms allow JavaScript access to the sound hardware. Reset the simulation and try the Sound button again if the sound fails to play.

The above simulation (found in Chapter 10) is an example of the interactive nature of the simulations in this book. Play the simulation (you may have to click the reset button for the sound to turn on). What do you hear? What is the mathematical relationship between the beats you hear and the two frequencies? Use the slider to gradually change the difference in frequencies. At what point do you no longer hear beats? This unpleasant sound is called *dissonance*. At what difference in frequencies does this occur? If you continue to separate the two frequencies you will eventually hear two separate tones. When does this occur?

Acknowledgements

There are a great many people and institutions that have contributed to our efforts, and we take great pleasure in acknowledging their support and their interest. The Open Source Physics project has benefited tremendously from collaborations with U.S. and non-U.S. universities. In particular, we give special thanks and recognition to the Easy JavaScript Simulations developers Francisco Esquembre and Felix Garcia at the University Murcia, Spain. We would also like to thank the AAPT- ComPADRE team headed by Bruce Mason, Lyle Barbato, and Matt Riggsbee.

W.C. would like to thank his Davidson College colleague and collaborator Mario Belloni for his support and many contributions to the Open Source Physics project. He would also like to thank the numerous students who have worked with him over the years developing programs for use in undergraduate physics education. In particular, he would like to single out Drew Glassman for his work on the JavaScript adaptations of the original Java-based Physlet material.

K.F. would like to thank Indiana University Southeast for research reassigned time which was used for the preparation of the initial version of this tutorial and to the many students over multiple semesters who proof read the material and offered suggestions for improvement.

TABLE OF CONTENTS

5: Wave Speed

6: Wave Behavior

7: Pitch, Loudness and Timbre

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

CHAPTER OVERVIEW

1: Basics

In this chapter various physics concepts and definitions needed for the study of sound, acoustics and musical instruments are presented. If these terms are already familiar to you, you may wish to skim this chapter and skip to the next one on simple harmonic motion.

In science words have specific narrow definitions that sometimes don't correspond to their use in everyday language. These definitions are relatively easy to memorize but surprisingly hard to apply correctly. The only way to really understand these concepts and how to apply them is to work the examples (group work, homework, simulation exercises, etc.) that are part of this book.

It is also important to realize that science is in the business of measuring things. This means that, unlike a math class, any number you see is a measurement of something and has a unit attached to it. So we never have 9.5 but we can have 9.5 centimeters or a frequency of 9.5 hertz (Hz). This book uses the metric system but other sets of units are mentioned. Here is a useful site that converts *metric units* to other systems: Conversion Calculator.

Key Terms:

Position, displacement, time, velocity, speed, acceleration, mass, weight, density, linear density, force, tension, Newton's three laws, pressure, the ideal gas law, Bernoulli's principle, energy, conservation of energy, power, the first and second laws of thermodynamics.

1.1: Basics- Work and Energy

Various physics concepts and definitions needed for the study of sound, acoustics and musical instruments are presented.

1.1.1: Motion and Force

Page ID 26128

The location of an object is called the ***position*** and is generally measured in *meters* (or centimeters or inches or feet, etc.) starting from an agreed upon starting point (called the origin). Sometimes it is convenient to set up a coordinate system with and x-axis, y-axis and z-axis in order to locate something precisely in three dimensional space. We may want to also know how far something is from where it started or from its normal, at rest position. This is the displacement or ***displacement from equilibrium*** and is also measured in meters. For example, when a violin string is plucked or bowed, different points on the string are displaced from their equilibrium positions by different amounts initially.

Speed, v, is the rate of change of position and is measured in *meters per second*, m/s (or ft/sec or miles per hour, mph, etc.). If there is no acceleration, speed is related to distance traveled as rate times time equal distance; $vt = d$. ***Velocity*** is speed with additional information about which direction something is moving. So if you are traveling at 50 mph (no direction given) you have a speed but if you are traveling 50 mph and headed north-east we call that a velocity.

Acceleration is a word that has a specific definition in science but gets confused with speed and velocity in everyday use. Acceleration is a change in velocity over a given time and is measured in *meters per second squared*, m/s^2 (or ft/s^2, etc.). A car going down the road in the same direction might have a constant speed of 50 mph. If neither the speed or direction changes then the acceleration is zero. If the car speeds up, slows down or changes direction then the velocity has changed and there is acceleration. The gas pedal in your car is both an acceleration pedal (if you push it to the floor or ease off the pedal you change your speed and get faster or slower) and also a speed pedal (if you hold it constant you keep a constant speed with no acceleration). Your brake pedal is also an acceleration pedal; it slows you down. Your steering wheel is also an accelerator control because it changes your direction.

For objects near the surface of the earth there is a special acceleration, called the ***acceleration of gravity***. For something falling in the case where air resistance doesn't slow it down too much, the speed changes by 9.8 m/s every second so we say it has an acceleration of 9.8 m/s per 1s or 9.8 m/s^2 (or 32 ft/s^2 in the English system). So if a pot gets pushed off an upper story window ledge it starts from rest (zero speed) and will be falling at a speed of 9.8 m/s after one second, $9.8 \text{ m/s} + 9.8 \text{ m/s} = 19.6 \text{ m/s}$ after two seconds, $9.8 \text{ m/s} + 9.8 \text{ m/s} + 9.8 \text{ m/s} = 29.4 \text{ m/s}$ after three seconds, etc. This acceleration of gravity works the same for any object, regardless of size or weight, as long as air resistance is small and you are near the earth's surface. This is what Galileo figured out by dropping different sized objects off the leaning tower of Pisa; objects accelerate (get faster) at the same rate no matter what their size (again, if you can ignore air resistance).

The amount of 'stuff' an object has regardless of where it is (in space, on earth, on the moon), its 'heft' you might say, is called ***mass*** and is measured in *kilograms*, kg (or sometimes grams where $1000 \text{ g} = 1 \text{ kg}$). The English units of mass are the slug and the stone but we seldom use them.

A ***force*** is a push or shove and is measured in *newtons* , N, which is a kilogram-meter per second squared, kg m/s^2. The tension in a guitar string is the force pulling at either end of the string (the force pulling each end is the same and balances out so the string doesn't go anywhere). The English unit of force is the pound which is also the unit used to measure weight. This can lead to some confusion because the pound is also sometimes incorrectly used for mass.

Weight is the force that gravity causes an object to exert on the earth. Weight is measured in *newtons* in the metric system or pounds in the English system. This is how mass and weight get confused; in everyday use we use weight and mass interchangeably which, strictly speaking is incorrect. For example in the grocery store you may see a bag of sugar that says 5 lbs or 2.3 kg which is really comparing different things; lbs measures weight but kg measures mass which is different. What the label should say is "this item weights 5 lbs or 22.3 newtons" (because 2.3 kg times 9.8 m/s^2 is 22.3 newtons which is a unit of weight). Or it could say "this item has a mass of 2.3 kg or 0.16 slug" (5 lbs divided by 32 ft/s^2 is 0.16 slugs and slugs and kg measure mass). So if you see something that claims to have a weight of 2.3 kg (or 5 lbs) remember what that really means is its mass is 2.3 kg.

Questions on Motion:

1. Why are units (such as meters or seconds or miles) important in science?
2. What is different about numbers in a science course as compared to numbers in a math course?
3. What is the difference between speed and velocity?
4. What is the difference between velocity and acceleration?
5. Does a car speedometer measure speed or velocity? Explain.

6. If you go around a curve at constant speed, do you have an acceleration? Explain.
7. Can a rapidly moving object have the same acceleration as a slowly moving one? Explain.
8. Can an object have an instantaneous velocity of zero and have a non-zero acceleration? Give an example.
9. At the end of its arc, the velocity of a pendulum is zero. Is its acceleration also zero at this point? Why or why not?
10. Assuming air resistance can be ignored, which gets to the ground first, a bowling ball or a tennis ball if they are dropped from the same height at the same time? Explain.
11. What is the difference in saying something is moving at a constant velocity and saying it is moving with zero acceleration?
12. Explain the difference between mass and weight.
13. Consult a table for the speed of sound in various substances (found in chapter six). If you have one ear in the water and one ear out while swimming in a pool and a bell is rung that is half way in the water, which ear hears the sound first?
14. At 20° C the speed of sound is 344 m/s. How far does sound travel in 1 s? How far does sound travel in 60 s?
15. Compare the last two answers with the distance traveled by light which has a speed of $3.0 \times 10^8 \text{ m/s}$. Why do you see something happen before you hear it?
16. The speed of sound in water is 1482 m/s. How far does sound travel under water in 1 s? How far does sound travel under water in 60 s?
17. What would an orchestra sound like to someone in the audience if different instruments produced sounds that traveled at different speeds?
18. For the previous question, would it make a difference if you sat further away from the orchestra?

1.1.2: Newton's Three Laws

Newton came up with several important laws of physics, three of which we may find useful for the physics of sound.

First Law: Objects continue at rest or in straight-line motion with constant velocity unless an unbalanced force acts on them. You might think that if no forces act the object is stationary. But this is only one case; the first law also says that if all the forces add to zero the object could still be moving at constant velocity. So why do you need to run the motor in your car if you are going down the road in the same direction at the same speed? The motor supplies enough force to overcome friction. Not more force than friction but a force *exactly equal* to friction so that the total force is *exactly zero* and you continue moving at a constant velocity. Similarly, a performer playing a wind instrument must continue to supply force on the air in the instrument to overcome friction as the air passes through the instrument. ***Inertia*** is the term used for the property of an object that causes it continue in straight line motion unless acted on by an unbalanced force.

Second Law: Forces cause accelerations (forces do not cause velocity but *changes* in velocity). The second law is also known as $F = ma$ where F is the total force acting on mass m, and a is the acceleration of the mass. Note that as was stated in the first law, a net total force is not needed to maintain a constant velocity; the first law says once something is moving it will keep moving unless a force (for example friction) acts to stop it. When you accelerate your car there is a net force on it. But when you reach cruising speed the net force goes to zero (the motor force exactly cancels resistance forces). In this case (when the net force adds to zero) the object (your car) obeys Newton's 1st law and you continue moving in a straight line at constant velocity.

Third Law: Anytime one object exerts a force on a second object, the second object exerts an equal force (but in the opposite direction) back on the first object. Forces always act in pairs on two different objects; an action force acting on one object and a reaction force acting on the second object. Another way to say this is you cannot touch something without it touching you back, and just as hard as you touched it. This seems simple if you are pushing on the wall; the wall obviously pushes back with the same force. But what about the force on a baseball as you throw it (before release)? Is the force pushing back by the baseball the same as the force you apply? YES! The laws of physics are always true with no exceptions. So how can you throw the baseball if the two forces are equal in magnitude but opposite in direction? The key is to realize the action and reaction forces act on different objects. Your force on the ball makes it go forward until it leaves your hand. The equal reaction force back from the ball acts on your hand (not the ball) so you can feel the ball as you throw it.

Video/audio examples:

- Newton's first law: egg drop.
- Newton's first law (car crash testing) one, two, three. Watch the movement of the dummy's body and head when the car suddenly stops. Because there is no force acting on the head, it keeps moving, according to Newton's 1st law.
- Newton's first and second law: sky diving. If you understand this example you have a good idea of how Newton's laws work.
- Newton's first, second and third law applied to a bicycle.
- Newton's third law: examples on the space station.

Questions on Newton's First Law:

1. If you are sitting at a stop sign and get hit from the rear, your head seems to fly back and hit the headrest. Explain, using Newton's law of inertia, what really happens.
2. If you are standing in a bus and it suddenly stops, you feel like you are being thrown forward. Using Newton's law of inertia, explain what is really happening.
3. Using Newton's law of inertia, explain why a headrest prevents whiplash.
4. Using Newton's law of inertia, explain why using a seatbelt in a car, plane, rollercoaster, etc. is a good idea.
5. Once a satellite is in orbit it doesn't need to fire any rockets to stay there. Why not? What keeps it going?
6. What keeps the earth going around the sun?
7. A magic trick you can preform at home is to pull the tablecloth out from under some dishes on a table by giving the tablecloth a very quick horizontal jerk. Explain why the dishes don't move.
8. Bowling balls slow down slightly as they roll down the lane. Explain why this does not violate Newton's law of inertia?
9. If you quit pushing a shopping cart it stops. Explain how this does not violate Newton's law of inertia.
10. The earth is rotating such that objects on the surface are traveling at close to 1000 km per hour (this is slightly different depending on latitude). Using Newton's law, explain why you don't get slammed by the wall if you jump straight up into the air.
11. When you are traveling in an airplane at cruising altitude, why does an object that is dropped not fly to the back of the plane?

12. A person drops a wrench from the top of the mast of a sailboat that is moving forward at constant velocity. Where does the wrench land relative to the mast if the boat has a speed of 10 m/s and the mast is 20 m high?
13. A driver heading towards a left curve encounters some ice on the road. Describe the motion of the car (drawing a picture will help) assuming the ice prevents any friction force from acting on the car. Which of Newton's laws tells you what will happen?
14. A small ball rolls in a frictionless tube that is flat on a table shown from above in the drawing. Draw the trajectory of the ball when it leaves the tube and justify your answer.

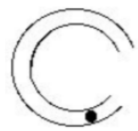


Figure 1.1.2.1

Questions on Newton's Second Law:

1. Does a book at rest on the table have no forces acting on it? List and explain the forces that act on the book.
2. A car traveling North down the road at constant velocity has zero acceleration. The net force has to be zero since $F = ma$. Why then do you need to keep the engine running and the gas pedal pushed down?
3. Can you have an object traveling forward with a net force acting on it in the opposite direction? Explain and give an example.
4. You exert 800 N of force to push a box across the floor at a constant speed (zero acceleration). Is the friction force larger, smaller or exactly equal to 800 N? How do you know?
5. Can you make an object go around a curve without applying any force? Explain.
6. You jump out of an airplane and open your parachute. With the parachute open you travel at constant speed. How does the upward force of the parachute compare with the downward force of gravity?
7. A load of lumber in the back of a pickup truck accelerates at the same rate the truck does. What applies the force to make this happen? What happens if this force isn’t large enough?
8. You throw a ball upward. Once it leaves your hand, what force acts on it on the way up? On the way down? What is the effect of this force on the way up? What is the effect of the force on the way down?
9. Explain why Newton’s first law is really a special case of Newton’s second law.
10. A car is traveling at a constant 55 mph in a straight line. What is the net force acting on the car?
11. An astronaut is in a spaceship far from the effects of gravity. She pushes with the same force on a baseball and a bowling ball. Indicate which of the following is true and explain why:
    a. they both accelerate with the same speed because they are weightless;
    b. they accelerate differently since their mass is different but they end up with the same terminal velocity;
    c. they have different accelerations.
12. The maximum tension on a guitar string is about 900 N (202 pounds of force). Suppose the peg holding the string weighs 0.002 kg and comes loose so that the 900 N causes the peg to accelerate. What will be the acceleration of the peg (in m/s^2)?
13. Is the acceleration in the previous problem dangerous? Hint: If the force acts over half the length of the guitar, say 0.20 m the final velocity will be $v = \sqrt{2ax}$ where a is the acceleration and x is the distance traveled.
14. Redo the previous two questions for the case of a piano string with 700 N of tension and a distance of 50 cm traveled. Assume the peg has the same mass.

Questions on Newton's Third Law:

1. Can you push on your left hand with a larger force using your right hand? Explain.
2. You tie a rope to a box in order to pull it across the floor. According to Newton’s second law, the box pulls back on the rope with the same force that you pull on the rope. Explain how you can move the box if these forces are exactly equal and in opposite directions.
3. Explain the action and reaction forces when you push against the ground with your foot in order to take a step forward.
4. What are the action and reaction forces in these cases:
    a. a tennis racket hits a tennis ball;
    b. while walking, your foot pushes off from a curb;
    c. you push down on the pedal of a bicycle;
    d. during the windup of a baseball pitcher, up until he releases the ball?

5. For each case in the previous question, state which force is the larger force.
6. A force F pushes towards the left on a box. A friction force, f, between the floor and the box resists the movement of the box. These are the only forces acting in the horizontal direction. For the following three cases state which is bigger (or the same size), F or f and why.
   a. The box does not move.
   b. The box moves to the left with constant velocity.
   c. The box moves to the left and accelerates.
   d. The box moves to the left and decelerates.
7. A bowling ball collides with a tennis ball. Which object has the larger impact force on the other? Which has the greater acceleration? Explain.
8. Before space travel some people thought rockets would not work in space because there was no atmosphere for the rocket exhaust to push against. Explain the error in this thinking using Newton’s third law.
9. You are in a railroad car but the tracks are very smooth and the windows closed so you cannot tell if you are moving or not. You drop a tennis ball and it falls straight down and lands directly below your hand. What can you conclude about the motion of the car from this observation?
10. When you hit a xylophone bar with a mallet the mallet will bounce back. Which of Newton's three laws explains where the force comes from which causes the mallet to bounce back into the air?
11. A guitar string pulls on the peg mechanism with a tension of $600\ \mathrm{N}$ ($134\ \mathrm{lbs}$ of force). What minimum force must the mechanism be able to pull back with to avoid having the string change tension (which also changes the pitch)?
12. What force must a piano frame be able to withstand if the tension in the tightest string is $900\ \mathrm{N}$?

1.1.3: Density and Pressure

The ***density*** of something is its mass divided by its volume (m/V) and is measured in *kilograms per cubic meter*, $\mathrm{kg/m^3}$ (or sometimes grams per cubic centimeter; $\mathrm{g/cm^3}$). So a kilogram of feathers and a kilogram of iron have the same mass and weigh the same but since the kilogram of feathers takes up more space (larger volume) it is less dense. One other version of density we will use is called the ***linear density*** which is the mass per length in $\mathrm{kg/m}$. Bass strings on guitars and in pianos have a larger linear density than the strings used for the higher notes. We will see why later on.

Pressure is defined to be a force acting over an area; $P = F/A$. There are several units of pressure; we will use the *pascal*, Pa, which is a newton per meter squared, $\mathrm{N/m^2}$. Other units are the bar; *atmospheres*; millimeters of mercury, mmHg; inches of water; torr, etc. A larger force over the same area increases pressure but the same force over a larger area decreases pressure. A dull knife does not apply the same pressure as a sharp knife because the area of contact of the dull blade is larger than the area of contact of the sharp blade. Being stepped on by the heel of a high heeled shoe hurts a lot more than if the same person steps on you with a low heel because the same force (the persons weight) applied with a high heel acts over a smaller area so the pressure is much higher. As we will see, the loudness of a sound wave is related to pressure; high volume sound exerts more pressure on average, and therefore more force on the surface of your eardrum.

For gasses in a closed container, pressure and volume are inversely proportional (pressure increases as volume decreases). Pressure and volume are also each directly proportional to temperature (either pressure or volume or both will increase if the temperature increases). These properties are sometimes summarized as the ***ideal gas law*** which can be written as $PV = nRT$. Here P is pressure, V is volume, T is temperature in kelvin, K. The variable n indicates how much gas there is (in moles where a mole is 6.0×10^{23} atoms or molecules) and R is a constant equal to $3.14\ \mathrm{J/mol\ K}$

Pressure in a liquid or gas is the weight of the liquid pushing downward on an area at a given depth (and is measured in the same units as pressure) or $P = \rho g h$ where h is the depth, ρ (Greek letter rho) is the density in kilograms per meter cubed, and g is gravitational acceleration. We sit at the bottom of a sea of air that pushes down on us. This pressure is called atmospheric pressure and it varies a little bit from day today because the air above us is moving and also because of changes in temperature and humidity (and so its density changes). When you use a straw you are decreasing the pressure inside the straw and atmospheric pressure outside the straw pushes the liquid up into the straw. This is why a straw would not work in a vacuum. If you are under water, the water above you pushes down on you in addition to the air above the water which pushes down on the water. Since water is much more dense than air, pressure changes a lot faster as you go deeper under water than it does if you change altitude in the air.

Bernoulli's principle says that if the speed of a fluid (liquid or gas) increases, the internal pressure in the fluid decreases. Take a strip of paper one inch wide and 12 inches long. Hold the short end up to your lips and blow. You'll notice that the strip pulls upward to meet the flowing air. This is because the moving air above has a slightly lower pressure than the stationary air below. A similar effect causes baseballs to change direction (curve balls) and airplane wings to have lift. Some wind instruments and the human voice operate in part because of forces due to the Bernoulli effect, as we shall see.

Video/audio examples:

- Pressure and a bed of nails: with a balloon, with a person. What would happen to the pressure if the number of nails is reduced? What would this do to the person lying on them?
- Bernoulli's principle: Hair dryer. Why does the ping pong ball stay suspended? Soccer ball. Why does the ball not follow Newton's first law and travel in a straight line? Several Examples. Why do the balloon's come together?

Questions on Density and Pressure:

Density

1. Does the mass of a car change if it is crushed into a cube? Explain.
2. Does the density of a car change if it is crushed into a cube? Explain.
3. Does a dieting person lose mass? Weight? Density? Explain your answers.
4. How does the density of water change when it freezes into ice?
5. Which is more dense, a kilogram of feathers or a kilogram of iron?
6. Which weighs more on the earth, a kilogram of feathers or a kilogram of iron?
7. What is the difference between density and linear density?

8. Given what you know about Newton's second law ($F = ma$), why would you expect a denser guitar string to vibrate more slowly when plucked with the same force?

Pressure

1. Why does a sharp knife cut better than a dull knife (even when you apply the same force)?
2. An old time magic trick (that originally came from India) was to lie down on a bed of nails (hundreds of nails sticking up through a board) without getting hurt. Using the definition of pressure, explain how this is possible.
3. Does a bathroom scale measure pressure or force? Explain. (Try this at home: Stand with both feet on the scale, look at the reading and then stand on one foot on the scale and check. Are the readings different?)
4. Which exerts more pressure on the ground, the foot of an elephant or a person in high-heeled shoes? State your reasoning.
5. You may notice that an unopened bag of chips is soft while on the ground but is puffs out to be firm when at cruising altitude in an airplane. Explain why.
6. Why would you want the bottom of a dam to be stronger than the top?
7. Why would it be slightly more difficult to suck soda through a straw on top of a high mountain as compared to sea level?
8. You decide you want to use a piece of garden hose with one end above water to go to the bottom of a pool 3 m deep and be able to breath. What is wrong with this plan?
9. A siphon is a tube that transfers liquid from a higher level to a lower level. How does it work?
10. What causes the 'lift' on an airplane wing?
11. Why is it easier on your heart when you are lying down compared to when you are standing?
12. Why does a lightweight shower curtain move in towards you when the shower is running?
13. How does an airplane wing provide lift?
14. The pressure variation in a sound wave from a jet engine is around 200 Pa (Pascal). What is this in N/m^2?
15. Suppose a marimba mallet makes contact with a wooden marimba bar and applies a strike force of 600 N. You measure the contact area of the mallet to be 5 square millimeters ($5.0 \times 10^{-6} \text{ m}^2$). What pressure (in N/m^2) was exerted on the marimba bar? Convert this to atmospheres (1 atmosphere $= 101325 \text{ N/m}^2$). Do you think this could do damage to a marimba bar?
16. For sound waves from a normal conversation the pressure at the listener's ear fluctuates by around 0.2 N/m^2 . How much does the force change on the eardrum if the area is 1 cm^2 ($1.0 \times 10^{-4} \text{ m}^2$)?
17. Suppose the pressure of a sound wave reaching a microphone fluctuates by 0.002 atmospheres. What is the force on the microphone if it has an area of 2 cm^2 ($2.0 \times 10^{-4} \text{ m}^2$)?

1.1.4: Energy and Power

Energy is the capacity to do work and it is measured in *joules*; $\mathrm{J=kg\ m^2/s^2}$. Other common units of energy are the Btu (British thermal unit); the calorie ($1\ \mathrm{cal}=4.18\ \mathrm{J}$); the *food calorie* ($1\ \mathrm{kcal}=1000\ \mathrm{cal}=4186.8\ \mathrm{J}$); the kilowatt-hour ($1\ \mathrm{kWh}=3,600,000\ \mathrm{J}$). Notice that the calories listed for food are actually a thousand scientific calories each.

There are many forms of energy (all measured in joules) and you can convert from one form to another. The capacity of a musical instrument for converting mechanical energy of vibration (a vibrating guitar string for example) into vibrations in the air determines how loud the instrument will sound.

- ***Work*** is one form of energy. Work is defined scientifically as force times the displacement caused by the force; $W=Fd$. We only count the part of the force that acts in the same direction as the displacement; so a force acting perpendicular to a displacement does not do work. Likewise, if something doesn't move, no work is done.
- ***Kinetic energy*** is energy of motion. If we do work on a ball (apply a force over a distance) and then release it the ball will have kinetic energy. Kinetic energy is directly proportional to mass (double the mass of an object at the same speed and you have twice as much energy) and directly proportional to velocity squared (double the speed of an object and you have four times as much kinetic energy); $KE=1/2mv^2$.
- ***Gravitational potential energy*** is the energy (work) you can get out of an object due to letting it fall. Or if you do work in lifting a mass against the pull of gravity you store up energy that you can get back by letting it fall. Gravitational potential energy is directly proportional to mass and how high it is: $GPE=mgh$, where $g=0.8\ \mathrm{m/s^2}$ is the acceleration of gravity.
- You can do work on a spring by either stretching it or compressing it a distance x in which case there is stored ***spring potential energy***. The stiffness of a spring is given by a constant, κ, and the energy stored is $SPE=1/2\kappa x^2$.
- ***Electromagnetic radiation*** is energy carried in the form of electromagnetic waves. Examples of electromagnetic waves are light, radio signals, Wi-Fi signals, blue tooth, cell phone signals, x-rays, gamma rays, microwaves, infrared, ultraviolet, etc. The difference between each kind of electromagnetic wave is the size of the wavelength and the energy it carries. Except for visible light, we cannot detect electromagnetic waves. In general we need some electronic device to detect electromagnetic signals. For example a car radio turns electromagnetic waves from the radio station into audible sound waves.
- A ***chemical reaction*** occurs when two or more atoms interact by re-arranging where their electrons are located (they may share electrons, donate or borrow electrons, or have other complicated interactions). When this happens energy may be emitted or absorbed in the form of heat and/or electromagnetic waves. A burning candle is an example; the molecules making up the candle are interacting with oxygen and giving off heat (increased random molecular energy) and light (electromagnetic energy). The chemical energy stored in a battery is another example; molecules in the battery can combine in a way to give energy to a flow of electrons in a wire.
- As Einstein famously showed, there are certain types of changes in the nucleus of some atoms that give off heat. This ***nuclear energy*** comes from the atom changing a very small amount of mass directly into energy via $E=mc^2$, where c is the speed of light. In other words, in certain special atoms (called radionuclides) something happens to make the atom either randomly split or give off part of its nucleus. If you could weigh the pieces after the reaction you would find a tiny bit of mass was missing. It is this mass that has been turned into energy via Einstein's famous equation. This is the energy used in nuclear reactors and also the energy source of the sun.

In the above examples there is only one mass or object involved. But we know all matter is made of atoms and chemically bound combination of atoms called molecules that are too small to see, even in a microscope. In a solid these atoms are not stationary but vibrate around an equilibrium position. For liquids and gasses they move relative to each other as you saw in the pressure simulation. In both cases the average kinetic energy is proportional to something we call ***temperature***. Temperature is not a type of energy but is proportional to the internal kinetic energy of the molecules that make up a substance and is measured in fahrenheit, $^\circ\mathrm{F}$, *celsius*, $^\circ\mathrm{C}$, or kelvin, K.

In addition to random kinetic energy molecules can bend, vibrate and rotate in both solids, liquids and gasses. If we place an object that has a high temperature (high internal random motion) in contact with an object that has a low temperature (low random internal motion) energy will flow from the high temperature object to the low temperature object (the molecules of each will bump into each other so they eventually have the same average random energy). When this happens we call the energy that moves from the hot object to the cold object ***heat*** which *is* measured in joules. Notice that heat and temperature are not the same thing. Heat is a flow of energy (measured in Joules) and temperature is a number in celsius that is proportional to the internal kinetic energy of the molecules making up a substance.

There is one other term, related to energy, which is how fast energy is being used or delivered. The rate at which energy is used or work done is called ***power*** and it is measured in *watts*, W, and horsepower ($1 \mathrm{~hp}=746 \mathrm{~W}$). In the US we use watts for electrical power but hp for mechanical power. It would make more sense to either measure light bulbs in hp or cars in watts so that everything had the same units. Power is directly proportional to the amount of energy delivered and inversely proportional to the time it takes to deliver the energy; $P=W / t$. So accelerating your car up to a certain speed will require the same amount of energy regardless of whether you do it slowly or rapidly. But in order to accelerate faster (reach the same kinetic energy in a shorter time) you need a motor that is more powerful.

Questions on Energy and Power:

Mechanical Energy (work, kinetic, gravitational potential, spring potential, heat)

1. Does a baseball pitcher do any work on a baseball as they throw a ball (before release)?
2. Does a baseball pitcher do any work on a baseball after they release the ball?
3. Suppose you are hired to stand and push on a sheet of plywood to keep it stationary while other workers paint it. Are you doing any work in the physics sense?
4. Does the force of gravity do any work on a ball rolling across the floor? What about a satellite in orbit around the earth?
5. We know that for every force on an object there is an equal force in the opposite direction (Newton's third law). So if you push a filing cabinet across the floor and there is an equal force pushing back on you, does this mean you do no work on the cabinet? Explain.
6. Bullets leaving a rifle typically are traveling at a much higher velocity (more kinetic energy) than bullets leaving the barrel of a pistol. This has something to do with the length of the barrels. Explain.
7. What is the difference between “conserving energy” (i.e. turning off lights, turning the thermostat down) and conservation of energy (a law of physics)?
8. Suppose a $10 \mathrm{~kg}$ mass is held at a height of one meter so that it has a potential energy of $100 \mathrm{~J}$. Answer the following:
    a. If it is released, how much kinetic energy does it have right before it hits the floor?
    b. How much kinetic energy does it have when it is half way down?
    c. What happens to this energy after the mass comes to rest on the floor?
9. A simple pendulum consists of a mass swinging back and forth at the end of a long string or rope. When is the gravitational potential energy a maximum? When it the kinetic energy a maximum? What is the relationship between the kinetic energy and the gravitational potential energy?
10. Describe the changes in types of energy when you throw a ball up into the air during each of the following steps:
    a. You apply a force over a distance to get the ball started;
    b. You release the ball and it starts upward;
    c. The ball slows at it goes upwards until it reaches its highest point;
    d. The ball turns around and begins increasing speed on the way down;
    e. Just before the ball reaches the ground it has its maximum speed;
    f. The ball hits the ground and comes to rest.
11. A kid reaches the bottom of a slide in the playground with $1200 \mathrm{~J}$ of kinetic energy. Based on the height of the slide he had $1400 \mathrm{~J}$ of potential energy at the top. What happened to the missing $200 \mathrm{~J}$?
12. Years ago the Wham-O company sold a “superball” with the claim that it would bounce higher than the height at which it was dropped. Is this possible? Explain.
13. Is it possible to build a rollercoaster that has peaks that are higher than the starting point without using any motors? Explain.
14. A Ping-Pong ball and a golf ball have the same kinetic energy. Which has the higher speed?
15. Helium molecules are lighter than oxygen molecules. In a mixture of these gasses at the same temperature they have the same kinetic energy. Which type of molecule is moving faster?
16. Suppose you do $25 \mathrm{~J}$ of work on a guitar string by stretching the middle it to some maximum position. Then you let it go.

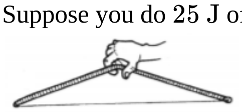


Figure 1.1.4.1

    a. How much kinetic energy does the string have when it passes through its equilibrium (straight) position?

Figure 1.1.4.2

b. How much potential energy does the string have when it reaches the maximum in the other direction?

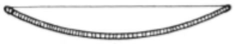


Figure 1.1.4.3

c. What happens to the energy when, after a while, the string comes to rest?

17. Why do the brakes on your car get hot when you stop?
18. Does a car burn more gasoline if the lights are on? What about if the engine is off? Does the gas mileage change in either case? Explain.
19. A gasoline engine only converts about 20% of the energy in the gasoline into useful mechanical energy (making the car go, running the lights, etc.). What happens to the rest of the energy?
20. If a car engine could be 100% efficient (it cannot), would it give off heat? What about sound? What about vibrations?
21. Calories are another unit for energy (a food calorie, like is listed on a soft drink can is actually 1000 calories). We need energy to generate body heat and to move around. What happens over time to someone if input calories (what you eat) is less than the output calories (body heat and movement)? What happens if someone takes in more calories than they burn?
22. Why would you expect the temperature of a jar of water to increase when shaken vigorously?
23. The temperature of the water at the bottom of a waterfall is slightly higher than at the top. Explain why.
24. Suppose there is a sudden loud sound in a closed room. Eventually the sound dies away. What happened to the energy in the sound waves that were produced? What can you say about the temperature of the walls of the room after the sound has died away?

Power

1. If the power company can't provide enough electricity fast enough, is this a power crisis or an energy crisis? Explain.
2. When you reach the top of a hill, have you used more power if you go straight up versus if you take a zigzag path (Hint; it takes more time to walk the zigzag path)?
3. Betty and Bob have the same mass and race up the stairs. Betty gets there first. Who does more work? Who uses more power?
4. Use a calculator to find out how many horsepower a 100 Watt light bulb is capable of putting out. ($1\text{ hp} = 746\text{ Watts}$)
5. Use a calculator to find out how many Watts a 250 hp car motor is capable of putting out.

Thermodynamics

1. An inventor claims to have the following new system. An engine runs by burning hydrogen. The engine turns a generator that makes electricity. The electricity runs through water to make hydrogen. The hydrogen is used in the engine. The inventor claims the system produces more energy than it uses. Should you invest in this new system? Why not (name the law which is broken)?
2. An inventor claims to have invented a motor that is 100% efficient. Would you invest in this device? Why not?
3. Why can't a gasoline car engine be 100% efficient?
4. The actual upper limit of efficiency for a gasoline engine is probably something less than 45%. Where does this limit come from?
5. Why do amplifier circuits for electric guitars generate heat?
6. Many instruments (violins and guitars for example) have to be returned after they have been played for a few minutes. This is because they get warmer. One source of the heat that causes them to warm up is heat from the hands of the performer. What other source of heat is involved?

1.1.5: Two Energy Laws

One very important concept in the physics of sound (and for all of physics) is the ***law of conservation of energy***: Energy cannot be created or destroyed but it can be transformed from one type of energy to another.

Consider the following energy conversions. You do 10 J of work on the spring of a toy gun so that 10 J of potential energy are stored. When you pull the trigger the spring energy is converted to 10 J of kinetic energy and a ball leaves the gun going upward. The kinetic energy gradually turns into gravitational potential energy as the ball goes upward. Halfway up it has 5 J of kinetic energy and 5 J of potential. At the top there is no kinetic left but you have 10 J of gravitational potential energy. On the way down the potential energy is gradually converted to kinetic energy. Just before the ball hits the ground it has 10 J of kinetic energy. Once it hits the ground the 10 J is converted to sound and heat; the ball is just a tiny bit warmer than it was.

You can also trace the original 10 J back in time before you pushed the spring on the toy gun. The energy for moving your arm came from chemical energy stored in your body. This chemical energy came from the plants and/or animals you ate. That energy can eventually be traced back to the nuclear energy released in the form of electromagnetic waves from the sun. Plants absorbed this light energy to form stored chemical energy for you to eat. Energy was never created or destroyed in any of these conversion steps. The law of conservation of energy is also called the ***first law of thermodynamics***.

There is a second energy law that is also very important but we will not need to know the details of how it works. The ***second law of thermodynamics*** says that anytime you change energy from one form to another, some of the energy must (that is the law!) end up in a much less useful form, usually as random thermal energy (heat). When the mechanical energy of bowing a violin is turned into vibrational energy in the strings and then that energy is turned into sound energy in the air, a small amount of the total energy always ends up as heat. The violin and surrounding air will be just a tiny bit warmer. Likewise, amplifiers for radios, stereos and electric guitars give off heat because they are converting electrical energy into sound energy; according to the second law the conversion process has to give off some heat.

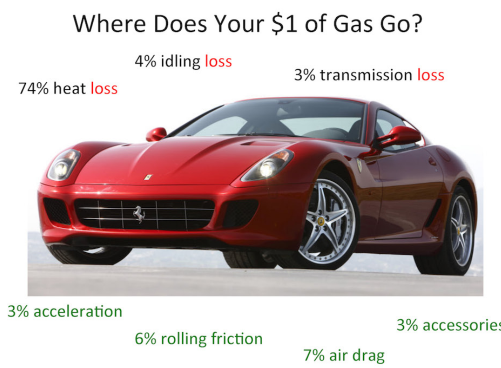


Figure 1.1.5.1

As can be seen in the picture above, as the result of the second law, sometimes the usable energy is only a small fraction of the total energy used. Your car motor is very inefficient, losing about 74% of the energy as heat. New technology may improve things some but as long as you are burning gasoline the efficiency of a car motor will remain well below 50% because of the second law of thermodynamics. Many other energy conversion processes are more efficient than process which burn a fuel source. For example, large electric motors and generators can be more than 90% efficient in converting electrical energy to mechanical energy or vice versa. The second law still applies but the consequence isn't as severe in the case of electrical conversions.

Video/audio examples

- Energy conservation examples: Bowling ball, real roller coaster, animation of roller coaster.

1.1.6: Conservation of Energy Simulation

This simulation shows a test dummy bungee jumping from a tower. It records the kinetic energy (KE), gravitational potential energy (GPE), elastic or spring potential energy (SPE) and total energy (TE) of the jumper/bungee system. There is no friction in this simulation. You can change the length of the bungee cord, L_o. If the bungee cord is too long, the dummy crashes into the ground.

Simulation Questions:

1. What is the longest length of the bungee for a safe jump? (Try several lengths to find out.)
2. For a bungee cord length of one meter, describe what happens to the kinetic energy, the gravitational potential energy and the spring potential energy as the dummy falls (you can use the step button to see how the different energies are changing)?
3. What happens to the total energy during a jump?
4. The total energy shows how much is kinetic, graviational potential and spring potential. Explain what happens to the various components of the total energy during a jump.
5. As you change the length of the bungee, what happens to the total energy? the KE, PEs?
6. For the most exciting (safe) jump, what makes it exciting? How can you describe it in terms of energy transfers?

1.2: Basics- The Molecular Basis of Matter

This simulation shows particles interacting with a slight attraction which will cause them to stay connected with each other to form a a solid at low temperature. But if they have enough thermal energy they will begin to move around each other to act like liquid. Additional thermal energy causes them to act like a gas.

1.2.1: Molecular Dynamics Simulation

The following simulation shows an adjustable number of gas molecules in a container. The molecules interact with a slight attraction which will cause them to stay connected with each other to form a a solid. But if they have enough kinetic energy they will begin to move around each other while still connected and the solid becomes a liquid. With even more kinetic energy they separate from each other and become a gas. Energy can be added or removed from the molecules with the 'cool' and 'heat' buttons to show changes of phase from solid to liquid to gas. The total energy (E), temperature and pressure (P) are given below the simulation.

The units used in the simulation are called 'natural units' where many constants (such as Boltzmann's constant, the mass of the molecules, the width of the molecules, etc.) are set equal to one. For the values to be in the usual units of joules for energy, pascal for pressure and kelvin for temperature they would have to be multiplied by scale factors which will depend on exactly what gas is being modeled. For our purposes here we can ignore these details and just talk about energy, pressure and temperature; see the references below the simulation for further details.

The initial simulation shows the system as a gas. The molecules are yellow if they have a lot of kinetic energy, red if they have a medium amount and blue if they have very little kinetic energy. There are several preset choices which can be selected from the pull down menu.

Credits

The Molecular Dynamics Exploration was developed by Wolfgang Christian at Davidson College using the Easy Java/JavaScript Simulation (EjsS) modeling and authoring tool created by Francisco Esquembre. This EjsS simulation is based on a pure JavaScript + HTML 5 simulation developed by Daniel V. Schroeder, Physics Department, Weber State University. Schroeder's simulation is described in Interactive Molecular Dynamics article (pdf), published in the American Journal of Physics **83** (3), 210-218 (2015), arXiv:1502.06169 [physics.ed-ph]. (Thanks to John Mallinckrodt for inspiring several of the presets.) Many of these presets use features that are only available in the original simulation by Dan Schroeder. We have imported the configuration file generated with Schroeder's simulation into the EjsS model.

Simulation Questions:

1. Start the simulation. What state is the system in when the simulation starts?
2. Do all the molecules have the same kinetic energy? (Does the simulation show them all with the same color?)
3. The average kinetic energy of the molecules in a substance is proportional to the temperature. Use the 'cool' button to decrease the temperature. What do you notice about the average kinetic energy as the temperature is lowered?
4. At about what temperature do the molecules begin to clump together and form a liquid? (Heat it up and cool it off several times to check your initial guess.)
5. Recall that pressure is force divided by area. In this case the force is supplied by the molecules hitting the sides of the container. What happens to the pressure as the temperature goes down? What if the temperature goes up?
6. Cool the system even more until it is a solid. At about what temperature do the molecules stop moving around each other to form a solid?
7. Stop the simulation and use the pull down menu to choose the bouncing ball preset. Run the simulation. What do you notice about the kinetic energy of the molecules (lighter color means higher kinetic energy) after the ball hits the ground?
8. For the bouncing ball, we know energy is conserved. But the ball does not return to its original height after it bounces so it doesn't end up with the same gravitational potential energy it started with. Based on your observations in the previous question, what happened to some of the gravitational potential energy of the ball after it hit the ground?
9. Reset the simulation and run the 'Hot and Cold' preset. The molecules in the upper solid have more kinetic energy than the molecules in the lower solid. What do you think will happen when the two solids touch? Run the simulation to see what happens. Were you correct?
10. Reset the simulation and run the 'Friction' preset. You will notice the molecules of the incline and the sliding object gain kinetic energy as the objects slides down the incline. Where does this energy come from?
11. Reset the simulation and run the 'Plucked String' preset. What part of the string has the most energy when it is plucked?

Chapter Two Summary

Forces cause accelerations, not velocities (objects will keep moving with constant velocity if the net force is zero). For every force there is always a second reaction force of the same amount but acting on a different object and in the opposite direction. Pressure is force distributed over an area. Bernoulli's principle is the result of air flow at different speeds and different pressures and causes baseballs to curve and the lips of a trumpet player to buzz. Energy comes in many forms and the total amount of energy is conserved (we can't create it or destroy it, only convert it from one type to another). When energy is converted from one form to another some of it has to end up in a less useful form (heat that generally isn't useful). Also, as you go through this book remember that: Science words have specific narrow definitions; Any number you see in science is a measurement of something, unlike in math class; and The laws of physics are always true with no exceptions.

CHAPTER OVERVIEW

2: Vibrations

All sound starts with something that vibrates. The reed in a clarinet vibrates, the vocal cords in a singer's throat vibrate, the air flowing over the mouthpiece of a flute oscillates, and the speaker cone on your stereo or in an ear-bud vibrates. In this chapter we investigate a particular kind of vibration called simple harmonic motion. Most of the vibrations in musical instruments and the human voice can be described approximately by ***simple harmonic motion***.

Key Terms:

Periodic, period (cycle), linear restoring force (Hooke's law), non-linear restoring force, amplitude, displacement, phase, frequency, natural frequency, spring constant, simple harmonic motion, damped harmonic motion, damped driven harmonic motion.

2.1: Vibrations

All sound starts with something that vibrates. The reed in a clarinet vibrates, the vocal cords in a singer's throat vibrate, the air flowing over the mouthpiece of a flute oscillates, and the speaker cone on your stereo or in an ear-bud vibrates. In this chapter we investigate a particular kind of vibration called simple harmonic motion.

2.1.1: Simple Harmonic Motion

An object will not vibrate if there is no restoring force causing it to want to return to its equilibrium position. If this force is proportional to the distance from equilibrium it is a ***linear restoring force*** and obeys ***Hooke's law***. Hooke's law says that if we double the displacement from equilibrium, the force acting to return the object to the equilibrium position also doubles. If the displacement is one third as big the force is one third as big and so on. Most springs obey Hooke's law; the more you stretch the spring, the larger the force.

What if the force is not proportional to the displacement? Such a force is called a ***non-linear force*** which does not obey Hooke's law. An example is the modern compound bow used in archery. A system of pulleys causes the force to be the smallest when the displacement is greatest. This makes it easier for the archer to hold the bow at maximum displacement while he or she aims at the target. Non-linear forces can be quite complicated but fortunately most forces involved in sound and musical instruments are close enough to linear that we can ignore non-linear effects. The few times a non-linear force acts will be explicitly mentioned; in all other cases you can assume the forces are linear.

The simplest of all vibrations occurs when there is a Hooke's law force and no friction acts. This type of motion is called ***simple harmonic motion*** and will be the model we will use for vibrations in musical instruments. A free hanging mass on a spring and a pendulum swinging with low amplitude approximately obey simple harmonic motion.

If friction acts the motion will gradually stop. This is called ***damped harmonic motion***. To maintain a constant vibration when there is friction, a periodic force must be applied. Harmonic motion that has damping and an applied periodic force is called ***damped, driven harmonic motion*** and will be discussed further in the next chapter.

Questions on Simple Harmonic Motion:

1. How is sound created?
2. How would you determine the period of a pendulum?
3. How are frequency and period related?
4. Define the units of frequency.
5. What is the difference between ‘amplitude’ and ‘displacement’? In the strict sense, why aren’t these terms interchangeable?
6. If the period of an oscillation doubles, what happens to the frequency?
7. The frequency of middle C sound wave is 262 Hz. What period of oscillation which produces this sound?
8. What is the period of oscillation of a string if the frequency is 200 Hz?
9. What is the frequency of oscillation if the period is 1.2 s?
10. A cork fishing float bobs up and down 15 times per minute. What is period of oscillation in seconds? What is the frequency in Hertz?
11. What is the period of the second hand of a watch for going all the way around once?
12. The frequency of a local radio station is 89.3 MHz ($\text{M} = \text{mega} = 10^6$). What is the period of oscillation of the electromagnetic waves of this signal?
13. If the CPU of a computer from the year 2000 is 200 MHz, what would the period of oscillation be?
14. If you hang a larger mass on the same spring, what happens to its period?
15. The average guitar has six strings each in ascending thickness. How might the thickness affect the frequency of the sound from each string when plucked?
16. Suppose a vibrating guitar string moves a total distance of 1.0 cm from it’s maximum in one direction to the maximum in the other direction. What is the maximum amplitude for this motion?
17. Suppose a clarinet reed vibrates with a maximum amplitude of 0.04 cm. How far does it travel in a complete cycle (all the way back to its starting point)?
18. What does the phase of an oscillation tell you about its motion?
19. A phase of 270 degrees is how many radians?
20. A phase of 200 degrees is how many radians?
21. Define simple harmonic motion.
22. What conditions are required for simple harmonic motion to occur?
23. What is Hooke’s law and why is it important?
24. What is the difference between a linear force and a non-linear force?
25. Which has the larger period, a stiff spring or a soft spring?

26. Which has the larger frequency, a stiff spring or a soft spring?
27. Which has the larger period, a small mass hanging from a spring or a large mass hanging from the same spring?
28. What kind of clarinet reed would more easily play low frequency notes, a stiff reed or a soft reed (assuming the mass is the same)? Explain your thinking.
29. What kind of saxophone reed would more easily play low frequency notes, a thick, heavy reed or a thin, light reed (assuming the stiffness is the same)? Explain your thinking.
30. The mathematical description of SHM is given by $y(t) = A\cos(2\pi ft + \phi)$. Explain what each of the terms (A, cos, π, f, t, ϕ) represent in the motion of a mass on a spring.

2.1.2: Period, Frequency, Amplitude, Restoring Force, Phase

A motion that repeats itself over time is said to be periodic and has a ***period***, T, measured in *seconds*. Simple harmonic motion is an example of periodic motion. The period is defined to be the time it takes for the vibrating object to make one cycle or oscillation and return to its original position. The inverse of period is called the ***frequency***, f where $f = 1/T$. Frequency is measured in *hertz* (Hz) which is a cycle or oscillation per second. So if a vibration repeats every $0.004\ \mathrm{s}$ it has a period of $0.004\ \mathrm{s}$ and a frequency of $1/0.004\ \mathrm{s} = 250\ \mathrm{Hz}$. A $500\ \mathrm{Hz}$ frequency means that whatever is vibrating oscillates with a period of $1/500\ \mathrm{Hz} = 0.002\ \mathrm{s}$.

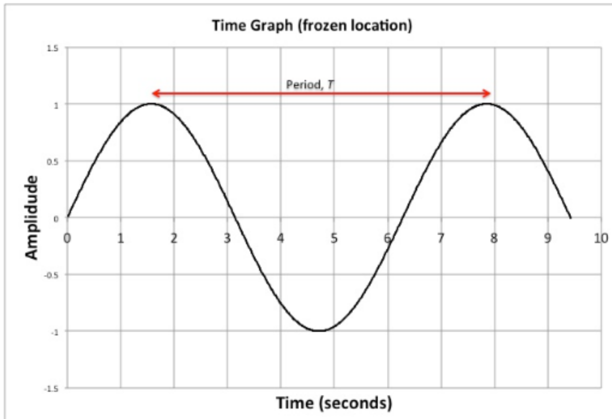


Figure $2.1.2.1$

When a guitar string or clarinet reed is at rest it is in its equilibrium position. When it is vibrating its displacement (Chapter 2), measured in meters or centimeters, is constantly changing. At some point in time a point on the string or reed reaches a maximum displacement and then moves in the opposite direction until it reaches a maximum in the other direction before turning around again. This maximum displacement from equilibrium at a given point is called the ***amplitude***. You may hear the word 'amplitude' used to mean 'displacement' at times other than the maximum but strictly speaking this isn't correct. Amplitude can also describe the maximum of other quantities which vary periodically. For example the speed of a vibrating reed also changes periodically and so has a speed amplitude which is the maximum speed.

Not only does the displacement change over time (from a maximum in one direction to a maximum in the other) but the velocity of the vibrating object is continually changing. As the object reaches its maximum displacement it slows down, stops and turns around (think of a mass attached to the end of a spring). We know from Chapter 1 that a change in velocity is an acceleration and accelerations are caused by forces. So to vibrate, an object has to have a force acting on it. This force always acts to push the object back towards its equilibrium position. This is called a ***restoring force***.

The ***Phase*** tells where we are in a cycle when we start making a time measurement of the motion. Phase is measured as an angle in degrees or *radians* ($2\pi\ \mathrm{radians} = 360\ \mathrm{degrees}$). Imagine a pendulum swinging back and forth and we want to measure its period. We could start our stopwatch when the pendulum is all the way to the right and then stop it when the pendulum returns to its original position. Or we could start when the pendulum is all the way to the left and wait for it to return all the way to the left. Or we could start the stopwatch exactly when the pendulum is straight up and down, moving to the right and stop the measurement when the pendulum is again straight up and down, moving to the right again (it would be straight up and down once in between but moving to the left; the motion hasn't fully repeated so this isn't a full period). In all three cases we would get the same number of seconds for the period, T. The difference is where the pendulum was in its motion when we started our stopwatch. By convention we start with the pendulum straight up and down and moving to the right which is given a phase of zero (degrees or radians). Once it has gone through an entire cycle it has traveled $360\ \mathrm{degrees}$ or $2\pi\ \mathrm{radians}$.

If the pendulum is all the way to the right it is $1/4$ of the way through the period from when we started compared to the straight up and down case. So it has an initial phase of $1/4$ of 360 or $90\ \mathrm{degrees} = \pi/2\ \mathrm{radians}$ of a cycle. With the pendulum all the way to the left it is $3/4$ of the way through the cycle or $270\ \mathrm{degrees} = 3\pi/4\ \mathrm{radians}$ when it starts. We can take data for the period starting at *any* time (not just at a maximum amplitude) and still accurately describe the motion by including the phase (sometimes called the ***phase angle***) in our description. As we will see, many interesting sound phenomena occur when two sound waves arrive at a microphone or your ear which are not in phase with each other.

Note

Both simple harmonic motion and many waves can be described by the trigonometric functions sine and cosine. What is the difference between these two functions? A sine function starts with an amplitude of zero ($\sin(0) = 0$) whereas a cosine function starts with an amplitude of one ($\cos(0) = 1$). If we slide the sine function to the left by 90 degrees $= \pi/2$ radians it exactly matches a cosine curve. In other words the sine and cosine are out of phase by 90 degrees; $\sin(x + \pi/2) = \cos(x)$ for any x. So we can use either function by adjusting the phase appropriately.

Video/audio examples:

- Examples of oscillation (some are simple harmonic, others not).
- Simple harmonic motion: explanation and examples.
- Tuning fork in water in slow motion.
- Brass player's lips vibrating.
- Example of a nonlinear spring. Note that the frequency shifts upward over time, unlike a linear system where the frequency remains constant when the amplitude changes.

2.1.3: Simple Harmonic Motion Simulation

The following is a simulation of a mass on a spring. The graph shows the y (vertical) location of the mass at different times. The force acting on the mass in this case is called a ***Hooke's Law*** force: $F=-\kappa y$ where κ is called the *spring constant*, in $\mathrm{N/m}$ indicating the stiffness of the spring and y is the location of the mass from some equilibrium position.

The points on the graph of the motion of the mass on a spring can be described by the mathematical function $y(t)=A\cos(-2\pi ft+\phi)$ where cos is the cosine function, f is the frequency, A is the amplitude (maximum displacement) and ϕ is the phase in radians. For any time, t, the displacement of the mass can be found by calculating $y(t)=A\cos(-2\pi ft+\phi)$ if the frequency and starting point (phase) are known.

Simulation Questions:

1. Drag the mass to some initial location using the mouse. Click 'play' button (lower left) to see the motion and the graph of the mass location. Determine the period of oscillation from the graph by finding the difference in time between two peaks.
2. Now find the period from the times of two consecutive times when the mass is at the bottom of its oscillation. Is this the same period you got in the first question? Explain.
3. Try a different starting displacement and measure the period. Does changing the amplitude of the oscillation change the period?
4. Clicking on the graph shows the coordinates of the mouse in a yellow box at the lower left. Determine the period and frequency of this motion from the values on the graph (Hint: Frequency in Hz is the inverse of period $f=1/T$.
5. Check the 'V' box to see graphs of both position and velocity. Where is the mass when the velocity is a maximum? Where is the mass when the velocity becomes zero?
6. Try different spring constant values, κ, between $0.5\ \mathrm{N/m}$ to $5.0\ \mathrm{N/m}$, resetting and releasing the mass at the same point each time. What is the relationship between spring constant and frequency?
7. For a given spring the frequency is determined by the mass hanging on it and the stiffness of the spring; $f=(\kappa/m)^{1/2}/(2\pi)$. Measure the frequency for the case of a spring constant, κ equal to 2.0. What must be the mass on the end of this spring? Verify that you get the same mass by measuring the frequency for several different spring constants (this is equivalent to hanging the same mass on several different springs).
8. According to the equation for frequency, what would happen to the period of oscillation of a spring-mass system if the mass is doubled?
9. Why is it more convenient to use a cosine function with a phase of zero for the description of the motion of the mass in this case? (Hint: Does the mass start with an amplitude of zero or a maximum amplitude?)

Advanced Questions:

1. As we will see in Chapter 6, angular frequency is given by $\omega=2\pi f$. What is the angular frequency of the motion of the mass?
2. The points on the graph of the motion of the mass on a spring can be described by the mathematical function $y(t)=A\cos(-\omega t+\phi)$ where cos is the cosine function, ω is the angular frequency and ϕ is the phase in radians. Using this equation and a calculator (in radian mode!), what is the location of the mass when $t=0$ and $\phi=0$? What is the location of the mass when $t=0$ and $\phi=\pi$ (don't forget we are in radian mode)? Explain what the phase angle, ϕ tells you about the initial position ($t=0$) of the mass on the spring.
3. If $y(t)=A\cos(-\omega t+\phi)$ is the location of the mass on the spring and the time derivative (using calculus) of location is velocity, then the velocity of the mass is given by $v(t)=-A\omega\sin(-\omega t+\phi)$. Check the 'V' box and run a simulation with an initial amplitude of $6.0\ \mathrm{m}$ and a spring constant of 2. From the graph find the angular frequency and calculate the speed amplitude $v_{max}=A\omega$. How does this number compare with the maximum value on the velocity graph, are they the same?
4. Find an expression for the acceleration of a mass on a spring, based on the position given by $y(t)=A\cos(-\omega t+\phi)$ and the fact that acceleration is the second derivative of position with respect to time (this requires calculus).

Chapter Three Summary

Vibrating objects can be described by their period (in seconds) or the inverse, their frequency (in Hertz). The phase angle tells where in the cycle a measurement of time or space is starting. Anything that experiences a Hooke's law force (a force proportional to displacement) without being driven and without friction will undergo simple harmonic motion. Simple harmonic motion can be described mathematically by a sine or cosine function.

CHAPTER OVERVIEW

3: Resonance

Resonance is a key concept in the production of sound in instruments and in acoustics. We will come across it many more times in this book. The ***natural frequency*** (f_o measured in *hertz*) is the frequency at which an oscillating system naturally wants to vibrate. For a mass on a spring, this is determined by the size of the mass and the stiffness of the spring; a stiffer spring has a higher natural frequency as we saw in the previous chapter. To keep a system vibrating in the presence of friction we have to keep pushing it with a periodic force. The frequency of this periodic driving force is called the ***driving*** frequency, f which is totally independent of the natural frequency (we can push our mass on a spring at a frequency different from the frequency at which it wants to vibrate).

Key Terms:

Natural frequency, driving frequency, angular frequency, damped harmonic motion, driven harmonic motion, resonance, resonator, Helmholtz resonance, Quality or Q-factor.

3.1: Resonance

Resonance occurs in an oscillating system when the driving frequency happens to equal the natural frequency. For this special case the amplitude of the motion becomes a maximum. An example is trying to push someone on a swing so that the swing gets higher and higher. If the frequency of the push equals the natural frequency of the swing, the motion gets bigger and bigger. Resonance is a key concept in the production of sound in instruments and in acoustics and we will come across it many times.

3.1.1: Resonance Examples

Resonance occurs in an oscillating system when the driving frequency happens to equal the natural frequency. For this special case the amplitude of the motion becomes a maximum. An example is trying to push someone on a swing so that the swing gets higher and higher. If the frequency of the push equals the natural frequency of the swing, the motion gets bigger and bigger.

For many systems we can make a graph of amplitude versus frequency and see where resonance occurs. Suppose we have a mass on a spring and attach a vibrator with a frequency we can choose. We start at low driving frequencies and measure the amplitude of the motion (how far it bounces) at each frequency. We might end up with a graph like the one below. Notice that the amplitude was a maximum at a driving frequency of 2.5 Hz. So the natural frequency of the system without the vibrator was also 2.5 Hz. In other words, we are driving the spring at resonance.

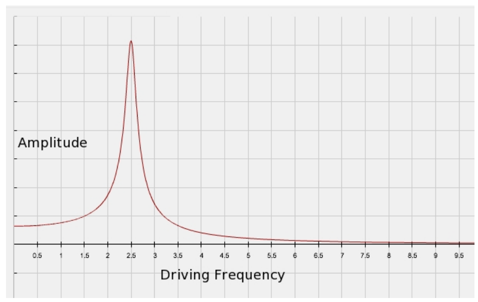


Figure $3.1.1.1$

Video/audio examples:

- Swing Resonance. What is the driving force in this case?
- Air track resonance cart. Why is the amplitude of the cart larger at one particular frequency?
- Resonance cart. Note there are three different natural frequencies in this example. Why is this so?
- Tidal resonance at the Bay of Fundy. The high water mark is reached every 12 hours. Why is it 12 instead of every 24 hrs?

3.1.2: A Few Other Examples of Resonance

Resonance can occur in any system that has a natural frequency. You probably have experienced a rattle or buzz in your car that only occurs at a certain speed. This is an example of resonance; the tires provide a periodic driving force which changes frequency as you change speed. Various parts of the car have different natural frequencies, particularly parts that have come loose. If the driving frequency of the rotating tires matches the natural frequency these parts will vibrate at a larger amplitude, making a buzz or rattle. Here are some other examples.

Video/audio examples:

- Helicopter ground resonance. What is the driving force in this case?
- Bridge resonances: bouncing, twisting, Tacoma Narrows, Russian bridge.
- Breaking glass: 1, 2. What are the driving forces in these two cases? [NOTE: Don't do this at home! The breaking glass can go into your eyes or mouth.]
- More Singing rod. What is the driving force that makes the rod resonate?
- A web page with more examples of resonance.
- A description of building resonance as a result of a dance class.

You probably have blown across the top of a bottle to get a note. Different sized bottles make different notes and partially filling the bottles with water also changes the pitch. These are examples of ***Helmholtz resonators***; a container of gas with a single opening that will resonate at a specific pitch. Blowing across the top causes a driving force on the air inside which has a natural frequency due to the 'springiness' of the air and the size of the container. As we will see, acoustic stringed instruments such as the guitar consist of a hollow body which acts like a Helmholtz resonator. Here are more details on Helmholtz resonance.

If you have ever been annoyed by a low frequency sound while driving down the highway with one window partially open, you were experiencing a Helmholtz resonance where the air blowing past the open window caused the air inside the car to vibrate.

Resonance also occurs in electrical circuits. In fact this is one way a radio receiver can tune to a certain broadcast frequency. The parameters of the circuit are adjusted so that the circuit has a resonance equal to the frequency of the station you are trying to listen to. With those parameters the circuit has much more current flowing for that particular frequency, thus making current flow for that frequency much larger than any other frequency. Here is a circuit simulation that shows three circuits. The top circuit is being driven below resonance so the current flow is small. The bottom circuit has a driving frequency that is too large. The middle circuit is driven at the resonance frequency of 41.1 Hz and has the largest current oscillations.

You may have heard the word resonance applied in the medical world. Magnetic Resonance Imaging (MRI) is used in medicine to get images inside the body without doing harm to the living tissue. Every different molecule in the body has a different natural frequency of oscillation. Each molecule also has electrical charges on them so that they can be driven by an oscillating magnetic and electric field (an electromagnetic wave). When the driving frequency of the oscillating magnetic field equals the natural frequency of the molecule, the molecule absorbs the energy and undergoes larger oscillations. This absorbed energy doesn't pass through the body and so is not detected outside after the magnetic field passes through. By changing the driving frequency (the frequency of the electromagnetic wave) it is possible to map the location of different types of molecules in the body, giving a map of the internal structures. This tutorial on MRI gives more details and shows some pictures generated using magnetic resonance imaging.

Questions on Resonance:

1. What is meant by damped harmonic motion? Give an example.
2. What is meant by driven, damped harmonic motion? Give an example.
3. What is the difference between natural frequency and driving frequency?
4. Define resonance.
5. List as many examples of resonance in every day life as you can think of.
6. In the YouTube of the singer who broke a wine glass using his voice, explain why the wine glass broke.
7. What is a Helmholtz resonator? Give an example.
8. In the simulation with several different masses, explain why different masses resonate at different driving frequencies.
9. The natural frequency of an oscillating object is 10 Hz. At what frequency would you want push it in order to make the oscillations bigger?
10. For the previous question, how often in seconds should you push the object to make the oscillations bigger?

11. State two ways to change the resonance frequency of a mass-spring system.
12. Why does the amplitude of the driving force not matter as much as the frequency of the driving force?
13. In the YouTube video with three different masses attached to rods on the cart, why are there three different resonance frequencies?
14. Suppose you recorded the amplitude of a driven spring system for many different driving frequencies and got the following graph of amplitude versus driving frequency. According to the graph, what is the resonance frequency of the system?

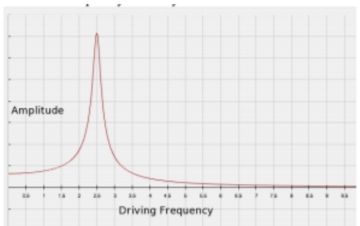


Figure $3.1.2.1$

15. What is the resonance frequency of a kid on a swing if the period of oscillation is 2.4 s?
16. If the period of oscillation of a kid on a swing is 3.0 seconds you could push it with a period of 3.0 s and the oscillations would get larger. What would happen if instead you pushed every 6.0 s?
17. For the 3.0 s oscillation in the previous question, what period other than 6.0 s could you push the swing so as to make the oscillation larger? We will talk about these higher frequencies (called harmonics) later.
18. What would be best for a stringed instrument: a high Q-factor or a low one? Explain.
19. In this chapter, you learned that a radio can be tuned to different frequencies by the principle of resonance. Would a higher Q-factor be beneficial for tuning a radio? Why or why not?
20. Why does changing the driving force, Fo, not change the resonance frequency (Hint: Look at the equation for the resonance amplitude in the simulation exercise.)?

3.1.3: Harmonic Motion and Resonance Simulation

The following simulation shows a driven, damped harmonic oscillator; a 1 kg mass on a spring with spring constant 2 N/m. The displacement of the motion is graphed versus time. The driving frequency, f, can be adjusted so we expect that for one particular frequency we will see the amplitude (maximum displacement) of the motion to be very large; in other words, resonance will occur.

Several other parameters can also be adjusted. b is the amount of friction in Ns/m (this could be air resistance or sliding friction or friction in the spring itself); v_o in m/s is the initial velocity of the mass, and F_o is the magnitude of the driving force in Newtons.

In this (and many other simulations we will use) it is easier to write $\omega = 2\pi f$ where ω is the ***angular frequency*** in radians per second instead of having to write $2\pi f$ everywhere.

Simulation Questions:

1. Start with the default parameters, drag the ball to the maximum starting position (or enter 10.0 m in the box for x_o) and hit play. How does this motion compare with simple harmonic motion (in the last chapter)?
2. Reset the simulation, change the friction parameter, b to 0.5 Ns/m, drag the ball to the maximum starting position (10 m) and hit play. (You can also click the button at the top for under damped motion.) What happens? This motion is called ***damped harmonic motion***.
3. Try different values for b. How is the behavior of the mass for small values of b (less than one) different than for values larger than 2 Ns/m (be sure to use the same starting position each time)?
4. If the mass oscillates at least once before stopping the damping is called ***underdamped*** motion. If the mass never quite gets back to the equilibrium position the motion is called ***overdamped***. The case where there is just enough damping so that an oscillation does not occur (the mass just barely makes it back to equilibrium) is called ***critically damped*** motion. Click the buttons on the top for under, over and critically damped motion. Describe the differences between these three cases.
5. Reset the simulation, set the friction parameter, b to 1.0 Ns/m and the magnitude of the driving force, F_o to 1.0 N. The angular frequency should be 1.0 rad/s also. Drag the ball to the maximum starting position and hit play. In this and the next few questions you will need to wait 10 s or so for the motion to stabilize. Describe the stable motion after the initial oscillations for this case. This motion is called ***driven, damped harmonic motion***.
6. Experiment with different amounts of force, keeping friction and angular frequency equal to one. Also start from the same position each time. What is the effect of larger values of force amplitude, F_o on the final, stable motion of the mass?
7. With damping, b set to 0.2 Ns/m and F_o set to 1.0 N try several values of the angular driving frequency, ω. Start with a value of 1.1 rad/s and increase by 0.1 each time until you get to 1.8 rad/s. In each case, wait until the animation ends and measure the amplitude by clicking on the top of the curve near the end. The second number in the yellow box is the amplitude in m. Write down the amplitude for each driving frequency. Which driving frequency ended up giving the largest amplitude? (This is the resonance frequency of the system.)
8. The ***natural frequency***, written as f_o is given by the stiffness of the spring, κ, and the mass; $f_o = (\kappa/m)^{1/2}/2\pi$ and is measured in Hz. In this simulation the mass is 1 kg and the spring constant is 2 N/m so $f_o = 0.225\text{ Hz}$ and the natural angular frequency, $\omega_o = 2\pi f$ equals 1.41 rad/s. In the previous question you should have seen the maximum amplitude for a driving frequency of 1.41 rad/s. In other words a driving frequency, ω of 1.41 rad/s leads to resonance (maximum amplitude) because it equals the natural frequency, ω_o. In the previous question, did you get a maximum amplitude for a driving frequency of 1.41 rad/s?
9. Use a calculator to find the natural frequency for a spring and mass system with $m = 2.0\text{ kg}$ and $\kappa = 5.0\text{ N/m}$. What do you expect the resonance frequency to be for this case?

Advanced Questions:

1. The mathematical formula that describes damped harmonic motion is $Ae^{-\gamma t}\cos(\omega t + \phi)$ where $\gamma = b/2m$. Notice that this is the same cosine function for simple harmonic motion but the amplitude, A, is multiplied by an exponentially decreasing function of time, $e^{-\gamma t}$. So we expect the oscillation of a damped harmonic oscillator to be an up and down cosine function with an amplitude that decreases over time. Check to see if the formula $Ae^{-\gamma t}\cos(\omega t + \phi)$ really does describe the behavior of damped harmonic motion in the simulation. To do this, use a graphing calculator (or go to meta calculator and chose graphing calculator or use desmos calculator) and plot $y = 10 * exp(-.2 * x) * cos(1. * x)$. This is the case $A = 10\text{ m}$; $b = 0.4\text{ Ns/m}$; $m = 1.0\text{ kg}$; and $\omega = 1.0\text{ rad/s}$. (You can cut and paste the equation into the online calculator).

How does this graph compare with the simulation for these same parameters (note: you are only interested in the positive *x* values)?

2. The mathematical formula that describes driven, damped harmonic motion is $A_o \cos(\omega t + \phi)$ where $A_o = (F_o/m)/((\omega^2 - \omega_o^2)^2 + 4\gamma^2\omega^2)^{1/2}$. In this case the amplitude, A_o does not change over time but it is dependent on the driving frequency, ω. What happens to the amplitude, A_o, when $\omega = \omega_o$ (assume all the other factors are constant numbers)? Is there any other combination of ω and ω_o that gives a larger amplitude?
3. Verify your previous answer by making a plot of $A_o = (F_o/m)/((\omega^2 - \omega_o^2)^2 + 4\gamma^2\omega^2)^{1/2}$ for a range of driving frequency. To do this, use a graphing calculator to graph $y = 1.0/((1.4^2 - x^2)^2 + 4 * .04 * 1.4^2)^{.5}$. Where does this graph have a maximum for positive values of x?
4. Based on the previous two questions, what is the resonance frequency of this system and how does this compare with the natural frequency?

3.2: Resonance Springs

This simulation shows five different masses, each attached to a spring of the same stiffness. The springs are mounted on a mechanical device that shakes the springs and attached masses.

3.2.1: Driven Springs Simulation

The following simulation shows five different masses, each attached to a spring of the same stiffness. The springs are mounted on a mechanical device that shakes the springs and attached masses. You can adjust the driving frequency, f in Hz, of the shaking mechanism, the amplitude of the driving force, F_o, in Newtons, the amount of friction, b in $\mathrm{Ns/m}$, and the stiffness of the springs, κ, measured in $\mathrm{N/m}$.

Simulation Questions:

1. Start the simulation. Do any of the masses have a very large amplitude?
2. Increase the amplitude of the driving force, F_o. Now do any of the masses have a very large amplitude?
3. Reset the simulation and change the driving frequency, f, to $0.5\ \mathrm{Hz}$. Wait a few seconds. What do you see now?
4. Why is the oscillation of mass number five much larger than the other ones now?
5. Reset the simulation and change the driving frequency to $2.0\ \mathrm{Hz}$. What do you notice now?
6. See if you can determine the resonance frequency of the center mass by trial and error. What is its resonance frequency?
7. In the previous simulation we saw that the natural frequency, written as f_o is given by the stiffness of the spring, κ, and the mass; $f_o = (\kappa/m)^{1/2}/(2\pi)$. In this simulation the large mass is $10\ \mathrm{kg}$ and the spring constant is initially $100\ \mathrm{N/m}$ so $f_o = 0.5\ \mathrm{Hz}$. This is why it has a large oscillation when driven at $0.5\ \mathrm{Hz}$; it will resonate with a driving frequency equal to its natural frequency. The center mass (number three) is $2.5\ \mathrm{kg}$ so the natural frequency is $1.0\ \mathrm{Hz}$. Did you find a resonance frequency of $1.0\ \mathrm{Hz}$ for this mass?
8. Mass number two is $1.25\ \mathrm{kg}$. Calculate its natural frequency. Verify your result by trying it out in the simulation.
9. Reset the simulation so that none of the masses are in resonance. Why doesn't the driving amplitude have much of an effect on the oscillation of the masses?
10. Change the spring stiffness, κ to $150\ \mathrm{N/m}$. (This changes the stiffness of all the springs.) Do the masses have the same resonance frequencies? Explain.
11. Use the formula for the natural frequency to calculate the natural frequency of the $10\ \mathrm{kg}$ mass (mass number five) with a spring constant of $150\ \mathrm{N/m}$. Verify your result using the simulation.

3.3: Quality Factor

Many systems, including musical instruments, have a wide range of frequencies at which the system will resonate. We study how this range depends on the damping coefficient.

3.3.1: Quality

So far we have been talking about systems that have a single resonance frequency. The peak in the graph in section 4.2.1 is very sharp which means that driving the system at any frequency other than 2.5 Hz will not produce a very large amplitude vibration. Many systems, including musical instruments, have a wide range of frequencies at which the system will resonate. Several different resonance curves appear in the graph below. Notice that, although there is always a single resonance frequency for any one curve (the frequency of the peak), there are many frequencies near resonance that will cause almost the same amplitude of vibration. The width of the resonance curve is called the ***Quality*** or *Q-factor*. The higher the Q-factor, the narrower the curve and the more selective the system is about what frequencies will excite it into resonance.

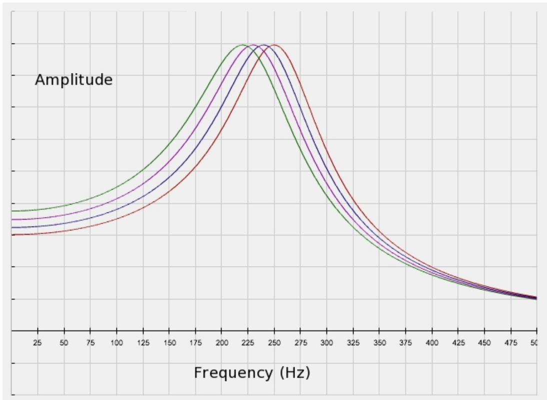


Figure $3.3.1.1$

The mathematical calculation (not given here) of the Q-factor depends on the amount of friction or damping in the system. Less damping means a narrower and higher resonance curve and it also means the frequencies will die away more slowly. An instrument designed with a high Q-factor would resonate loudly and for a long time but only at one frequency. Decreasing the Q-factor by having more damping means the resonance is not as sharp or as loud but it allows the instrument to play many different notes with nearly equal loudness. As we will see, the body of stringed musical instruments demonstrates a trade-off between sharp resonances (high Q) and broad range (low Q).

3.3.2: Q-Factor Simulation

This simulation expands on what was shown in the previous simulation. Now there are fifty-one independent, damped, driven harmonic oscillators instead of five. Each oscillator is independent of the others and has a different mass which means their natural frequencies are all slightly different. The ratio of the driving frequency, f, to the natural frequency, f_o for each oscillator is given on the horizontal axis. The masses are heaviest on the left (blue) and lightest on the right (red). Each mass is driven by the same sinusoidal driver with the same amplitude and the same driving frequency. A red arrow shows the driving force being applied to each of the masses. The amplitude of the driving force (red arrow) is the same for all damping but will appear different because the scale on the left changes to match the amplitude of the center mass.

In the previous simulations we examined the effects of damping, driving amplitude and driving frequency on a damped, driven harmonic oscillator. In this simulation, the only adjustable parameter is the amount of damping, b which is the same for each oscillator. The grey dots show the maximum displacement of each of the masses for a given amount of damping but note that the scale on the left changes depending on amplitude. Keep in mind the masses are independent and do not interact with each other.

Simulation Questions

1. On which side of the center is the driving frequency higher than the natural frequency? On which side is the driving frequency lower than the natural frequency?
2. Damped driven harmonic oscillators settle into a steady state after a certain time as you saw in the previous two simulations. How long does it take for the steady state behavior (when the center mass reaches its maximum shown by the grey circles) to emerge if $b = 0.1 \text{ Ns/m}$? If $b = 0.2 \text{ Ns/m}$?
3. The time to reach steady state is also an indication of how long the oscillations take to relax to zero if the driver is turned off. Based on your previous answer, which amount of damping will let the system 'ring' the longest?
4. How does the damping coefficient affect the maximum displacement of the center mass (look at the scale on the left for different values of damping)?
5. Describe the shape of the maximum amplitudes of all the masses when damping is small compared to when damping is large. In which case are more oscillators moving, low damping or high damping?
6. Based on what you have seen, what is the relationship between damping and Q-factor?

Chapter Four Summary

Oscillators will vibrate at their natural frequency which is determined by the physical properties of the system (mass, stiffness, etc.). Most real oscillators have some damping (friction) so that they will gradually stop unless external forces are acting. If an oscillator is pushed (driven) with a periodic force it may have several different amplitudes of vibration, depending on the frequency of the driving force. The largest amplitude occurs when the driving frequency equals the natural frequency. This condition is called resonance.

Resonances may occur at a broad range of frequencies (low Q) or a very sharp, single frequency (high Q) in which case the vibrations will die away more slowly once the driving force is turned off.

CHAPTER OVERVIEW

4: Wave Types

Transverse waves are the type of wave you usually think of when you imagine a wave. The motion of the material constituting the wave is up and down so that as the wave moves forward the material moves perpendicular (or ***transverse***) to the direction the wave moves. Examples of transverse waves include waves on a string and electromagnetic waves. Water waves can be approximately transverse in some cases.

Key Terms:

Transverse wave, longitudinal wave, compressional wave, torsional wave, wavelength, wave vector, wave period, amplitude, wave frequency, angular frequency, electromagnetic waves, sound waves, water waves, S-waves, P-waves, Rayleigh waves, Love waves, $v = \lambda f$.

4.1: Transverse Waves

Transverse waves are the type of wave you usually think of when you imagine a wave. The motion of the material constituting the wave is up and down so that as the wave moves forward the material moves perpendicular (or transverse) to the direction the wave moves.

4.1.1: Transverse Waves

Waves move over time which makes it hard to draw on a piece of paper. There are two possible representations. Suppose there is a cork floating in the water that is fixed at a certain location and we record the displacement (how high and low it is from equilibrium) at different times. If we plot the displacement versus time we have a frozen position graph:

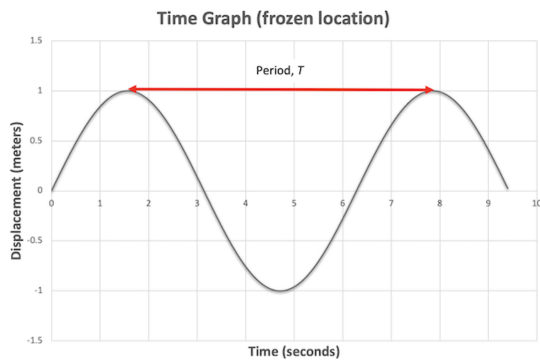


Figure 4.1.1.1

The displacement from the equilibrium position will be measured in meters but the horizontal peak to peak distance is a *time* measurement. This graph looks just like the graph for simple harmonic motion from Chapter 3. In fact the cork is undergoing simple harmonic motion and the horizontal peak to peak distance in the is the ***period***, T of the wave (just like the peak to peak distance of a time graph of harmonic motion was the period). According to the graph the first peak is at 1.5 s when the second is at 7.8 s so the period is $7.8\text{ s} - 1.5\text{ s} = 6.3\text{ s}$. As in the last two chapters, the wave frequency is given by $f = 1/T$ in Hertz so this oscillation has a frequency of $1/6.3\text{ s} = 0.16$ Hz.

We can make a second kind of graph of the wave if we take a snapshot (frozen time) of the wave and then plot the height (***amplitude***) versus position. This gives us the following picture of a wave frozen in time (much like a photo of a wave):

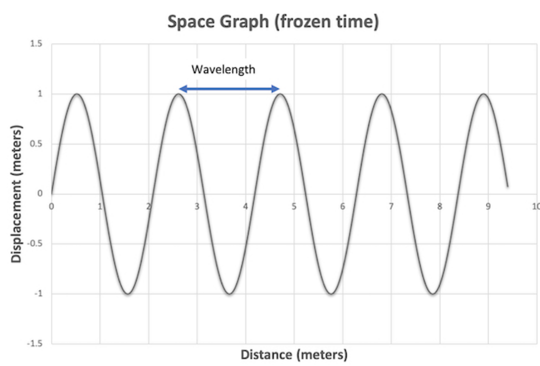


Figure 4.1.1.2

The displacement from the equilibrium position will be measured in meters (or centimeters, etc.) just like in the previous graph. We can also measure the horizontal distance from one peak to the next but this time it will be in meters instead of seconds. This *distance* is called the ***wavelength*** ,λ. In the graph the peak to peak distance is from 2.5 m to 4.6 m which gives a wavelength of $4.6\text{ m} - 2.5\text{ m} = 2.1\text{ m}$ (this is also the bottom to bottom distance or distance from where it is increasing and has a displacement of zero to the next place where it is increasing and has a displacement of zero, etc.).

Now imagine a long line of corks floating on the surface of a lake. As a wave passes by, the closer end of the row of corks starts moving up and down and then the rest of the row. Each cork is undergoing simple harmonic motion (which we studied in Chapter 3) but at a slightly different phase. Each cork has a time graph like the first graph above but taking a picture of the group as a whole gives a frozen time graph like the second graph above. The equations describing a wave will be sine and cosine functions, just as for the simple harmonic motion we saw in Chapter 3 but now the variables will depend on both time and distance (the simulation exercise below explains this in detail).

The speed of the wave, its frequency and its wavelength are related. If two waves are traveling at the same speed but have different wavelengths, a cork floating on each will bob up and down at different rates and so have different frequencies. A shorter wavelength will make the cork bob more often while a longer wavelength will make the cork bob less often. Mathematically this relationship is expressed as $v = \lambda f$ where v is the speed of the wave in meters per second, λ is the wavelength in meters and f is the frequency of the wave in Hertz. This equation applies to all types of waves and we will use it many more times in this book.

Video/audio examples:

- Lecture from Kahn Academy on amplitude, frequency, wavelength, 15 min.
- Transverse and longitudinal wave lecture.
- Transverse and longitudinal wave lecture.

4.1.2: Transverse Wave Simulation

The following simulation shows a (frozen position) graph of the motion of one location, the red circle, on a string which has a transverse wave on it. Notice that, while the wave moves forward along the string, the red circle does not (in fact none of the circles move forward). The speed of the wave is how fast the crest of the wave moves forward, not the up and down speed of the circles.

A *second* velocity associated with a wave is how fast the material of the wave moves up and down at a single location (the vertical speed of the circles in the simulation). This velocity, the ***transverse velocity***, is not a constant but is a function of location and time (different places on the wave move upward or downward at different speeds at different times).

The vertical location of points on the string (represented by the circles) as a function of horizontal location along the x-direction and time is described mathematically by $y(x,t) = A\sin(2\pi x/\lambda - 2\pi f t + \phi)$. If the wavelength, frequency, amplitude and phase are known, the height of the wave at any location and any time can be found by substituting time and location into the equation.

Simulation Questions:

1. Play the animation. From the top (frozen location) graph, what are the amplitude and period of the motion of the red dot?
2. Clicking on the lower panel gives the mouse location (in the yellow box) which in this case are the *x* and *y* location of points on the wave. Use these numbers to determine the wavelength of the wave (this is easiest to do with the animation paused or finished).
3. From the period and wavelength that you just measured, calculate the forward speed of the wave.
4. You can check your answer in the following way. Pause the wave just after it starts, click on the top of the wave to find its x-position (first number in the yellow box) and record the time. Let the wave move forward for a while and pause it again and record the x-position and time. Subtract the two x-locations to find the distance traveled and divide by the time interval between the two recorded distances. Does this match what you calculated? Why or why not?
5. Check the 'V' box to see graphs of both position and transverse speed of the red dot. What is the maximum transverse speed of the red dot? How does this compare with the forward speed of the wave you just calculated, are they the same or different?
6. Carefully state the relationship between position and velocity of the red dot. When the position is zero (equilibrium position of the red dot) what is the velocity? When the position is a maximum, what is the velocity?

Advanced Questions:

The forward speed of the wave is constant but the vertical speed of the material of the wave is not. Since velocity is the rate of change of position, the transverse velocity in the y direction is given by the derivative of the displacement with respect to time: $v(x,t) = \partial y(x,t)/\partial t = -A\omega\cos(kx - \omega t + \phi)$ where $k = 2\pi/\lambda$ is called the wave number and $\omega = 2\pi f$ as before. We use a partial derivative here because $y(x,t)$ is a function of two variables.

Notice that the maximum speed of a section of the wave at location x and time t will be given by $v_{max} = A\omega$.

1. Why is it more convenient to use a sine function for the description of the motion of the red dot in this case rather than the cosine function used for the mass in the previous chapter?
2. What is the wave number for the wave on this string?
3. Click the 'V' box to show speed and then 'play'. The upper graph now gives the speed of the red dot in the y-direction as a function of time. What is the maximum speed (approximately) of the red dot according to the graph? How does this compare with the forward speed of the wave which you found in the last question, are they the same or different?
4. Where is the red dot (relative to the rest position before the wave passes) when the maximum transverse speed occurs? Where is it when the transverse speed is approximately zero?
5. Using the amplitude and v_{max} from the graph and $v_{max} = A\omega$, what is the angular frequency? How does this compare with the value calculated from the period?
6. How does the speed amplitude (maximum speed) from the graph compare to the speed amplitude given by $v_{max} = A\omega$?
7. Since points on the wave change their transverse velocity over time there must also be a vertical or *transverse acceleration*. Since acceleration is the time rate of change of velocity we have $a(x,t) = \partial v(x,t)/\partial t = -A\omega^2\sin(kx - \omega t + \phi)$ where the *maximum acceleration* is $a_{max} = A\omega^2$. Calculate the maximum acceleration of the red dot. What are the units of this acceleration if amplitude is in meters and angular frequency in radians per second squared?
8. Based on the equation for acceleration, where will the red dot be when the acceleration is a maximum? Where will it be when the acceleration is approximately zero? What is the phase difference between the acceleration and the speed?

9. Carefully state the relationship between position, speed and acceleration. When the position is zero (equilibrium position of the red dot) what are the speed and acceleration? When the position is a maximum, what are the speed and acceleration?

4.2: Longitudinal Waves

Longitudinal waves are waves where the motion of the material in the wave is back and forth in the same direction that the wave moves. Longitudinal waves are sometimes called compressional waves. Sound waves (in air and in solids) are examples of longitudinal waves.

4.2.1: Longitudinal Waves

In Chapters 3 and 4 we saw that vertical springs and horizontal springs behave the same way and can be described by the same equation. In this chapter so far we have seen that transverse waves can be described by a line of masses undergoing transverse (up-and-down) harmonic motion. Can we also have a wave where the particles move with harmonic motion in the horizontal direction?

YES! ***Longitudinal waves*** are waves where the motion of the material in the wave is back and forth in the same direction that the wave moves. Longitudinal waves are sometimes called ***compressional waves***. Sound waves (in air and in solids) are examples of longitudinal waves. When a tuning fork or stereo speaker vibrates it moves back and forth creating regions of ***compression*** (where the pressure is slightly higher) and regions in between where the air has a lower pressure (called a ***rarefaction***). These compressions and rarefactions move out away from the tuning fork or speaker at the speed of sound. When they reach your ear they cause your eardrum to vibrate, sending signals through the rest of the ear to the brain.

4.2.2: Longitudinal Wave Simulation

Longitudinal waves can be described mathematically by the same equation as transverse waves: $y(x,t) = A\sin(2\pi x/\lambda - 2\pi ft + \phi)$. Only now, $y(x,t)$ is the *horizontal* displacement at time t and location x of the material in the wave from equilibrium instead of the vertical displacement from equilibrium. As was the case for transverse waves the forward velocity of a longitudinal wave is given by $v = \lambda f$.

The following simulation shows a graph of the longitudinal motion of one row of molecules, the red dots, in a collection of molecules which has a longitudinal wave passing through it, much like sound passing through air. A vertical line marks the equilibrium location of the red circle. Random thermal motions are not shown.

Simulation Questions:

1. Click 'play'. Do any of the dots travel all the way across the simulation to the other side? Explain.
2. Left clicking (single click for Mac users) on the upper panel gives the time and displacement of points on the graph in the yellow box. Determine the amplitude (maximum displacement from equilibrium) and the period of oscillation from the graph.
3. Left clicking on the lower panel gives the x and y locations of points on the wave in a yellow box. Pause and step the animation until the red dots are at their equilibrium position. Find the wavelength of the wave using the mouse by finding the distance between one place where the circles are clumped together to the next location (or from two successive locations where the circles are furthest apart). What is the wavelength?
4. From the period and wavelength find the speed of this wave (Hint: The same equations work for both longitudinal and transverse waves).
5. For sound the wavelength (or frequency) tells us something about the ***pitch*** of the sound. There are other aspects of pitch perception which involve other physical features of the wave but the main component of pitch is the frequency. What is the frequency of the wave in the simulation?
6. Click the 'V' box and then 'play'. How does the top graph with displacement and transverse velocity compare to the graph for the transverse case?
7. Write an equation of the form $y(x,t) = A\sin(2\pi x/\lambda - 2\pi ft + \phi)$, filling in the values of A, f and λ for this wave assuming the phase angle is zero.

Advanced Questions

Notice that the circles in the simulation move back and forth with a variable speed around an equilibrium position while the wave moves only in one direction with a constant speed. The velocity of the individual particles is given as before by the derivative of the displacement: $v(x,t) = \partial y(x,t)/\partial t = -A\omega\cos(kx - \omega t + \phi)$ where k and ω are defined the same way as in the transverse wave case.

1. Click the 'V' box and then 'play'. The upper graph now gives the velocity of the red dots as a function of time. What is the maximum velocity (approximately) of the red dot according to the graph? How does this compare with the velocity of the wave which you found in 6.4? How does it compare with $v_{max} = A\omega$?
2. In your own words, explain the difference between wave speed and particle velocity for a longitudinal wave.
3. Where is the red dot relative to the vertical line when the maximum velocity occurs? Where is it when the velocity is approximately zero? What is the relationship between position and velocity?
4. How do your answers for the previous question compare with your answers for this question in the transverse wave case?
5. Take a derivative of velocity to find an expression for acceleration of particles in the material (the red dot). Show that the maximum acceleration is given by $a_{max} = A\omega^2$.
6. Calculate the maximum acceleration of the red dot using $a_{max} = A\omega^2$. If amplitude is in meters and angular frequency in radians per second, what are the units of this acceleration?

4.3: Other Waves

If you stretch a slinky out between two points and gently twist it at one end, the twist will travel down the slinky as a wave pulse. This is an example of a torsional wave.

4.3.1: Torsional Waves

A third type of wave is a ***torsional*** or twisting wave. If you stretch a slinky out between two points and gently twist it at one end, the twist will travel down the slinky as a wave pulse. This is an example of a torsional wave. We saw two YouTube examples of bridges undergoing torsional motion in the last chapter: twisting bridge, Tacoma Narrows. These were examples of standing torsional waves.

Video/audio examples:

- Transverse, longitudinal and torsional waves lecture.
- Longitudinal and transverse waves on a slinky.
- Animations of different wave types.
- Torsion pendulum lecture demo.
- Wilberforce Pendulum which converts up and down motion to twisting motion and back.

4.3.2: Examples of Waves

Sound waves are an example of a longitudinal wave, the air molecules vibrate back and forth as the sound wave passes through them. We will look at the details of sound waves in the next few chapters. We are mainly interested in sound waves in this book but many other types of waves exist and have the same properties and behavior as sound waves. It is also the case that the same equations apply to many types of waves, although we won't go into great detail about this. If you are interested in the deeper mathematical properties of waves, have a look at Forinash's Wave Tutorial.

Although waves on the surface of a pond or ocean look like examples of transverse waves, they are actually a bit more complicated. A cork on the surface also has a little bit of back and forth motion as it bobs up and down. For a particle suspended in the water at a certain depth the back and forth motion increases with depth and the up and down motion decreases.

An earthquake occurs when there is a sudden movement of two parts of the earth's crust relative to each other. If the movement is inside the earth, two different types of waves, S-waves and P-waves are formed. S- or Shear waves are transverse waves; the earth moves up and down (or back and forth) as the wave moves through. P- or Primary waves are longitudinal (compressional) waves; the earth moves back and forth relative to the direction the wave is traveling as the wave passes. P-waves travel faster than S-waves. If these disturbances cause a wave to travel along the surface of the earth they are called surface or L-waves which generally move more slowly than S- or P- waves. The main types of surface waves are Rayleigh waves, Love waves and Stonely waves. Rayleigh waves are rolling waves, similar to transverse waves but with more bending motion. Love waves are a type of transverse wave where the motion is side to side instead of up and down. Love waves generally cause more damage than Rayleigh waves since buildings can often withstand an up and down jolt but are not typically built to withstand side to side motion. Stonely waves are transverse waves at an intersection or boundary inside the earth. All of these types of waves (and several others) travel at different speeds and the difference in speed can be used to locate the origin of the earthquake.

Video/audio examples:

- Wave pulse in slow motion.
- Sound waves can be visualized using a technique called Schlieren Flow Visualization (Wikipedia). Several examples are given in this video.
- Wikipedia has simulations of both S-waves and P-waves as well as other seismic waves.
- Video examples of earthquakes: mostly P-wave, mostly S-wave.

4.3.3: Water Wave Simulation

Many real, physical waves are combinations of three kinds of wave motion; transverse, longitudinal and torsional (which we have not yet discussed). The following is a more accurate simulation of water waves (but it still does not show breaking wave behavior which will come later). The dots are locations of water molecules or small objects floating in the water.

Simulation Questions:

1. What two types of wave motion are represented in the simulation?
2. Are the wavelengths and periods the same for both types of motion (Hint: Use 'pause' and hold the left mouse button down to make measurements; numbers in the yellow box give the cursor location in meters)?
3. Determine the wavelength, period and speed of the wave (use the 'pause' and 'step' buttons to measure the length of time it take a peak to pass a given location in the simulation).
4. How does the motion at the top of the water compare with the motion at the bottom?
5. Describe the overall motion of one of the red dots; what path does the dot follow?
6. Earthquakes can produce several different kinds of waves, each traveling at a different speed. Search for a reliable source and find a definition for P-waves, S-waves, Rayleigh waves and Love waves. Be sure to include comments on their speeds and which ones are more destructive.

4.4: Electromagnetic Waves

This simulation shows an oscillating electron in a sending antenna on the left. Because electrons have an electric field, an accelerating electron will create a wave in the electric field around it. Magnetic fields are created by moving charges so a magnetic wave is also formed by an accelerating charge. Only the y-component of the change in the electric field is shown (so an oscillation frequency of zero will show nothing, because there is only a constant electric field).

4.4.1: Electromagnetic Waves

Electromagnetic waves are a type of wave that has two transverse components. A wave in the electric field travels perpendicular and in-phase with a wave in the magnetic field moving at the same speed. Light, X-rays, infrared, ultraviolet, gamma rays, radio signals, wifi signals, Bluetooth signals, cell signals, broadcast TV and radio signals are all electromagnetic. They all travel at the same speed, the speed of light, $\mathrm{c} = 3.0 \times 10^8\ \mathrm{m/s}$. We will investigate the electromagnetic waves used in transmitting radio and TV in a later chapter. The following is a chart of the electromagnetic spectrum by Victor Blacus (Wikipedia File).

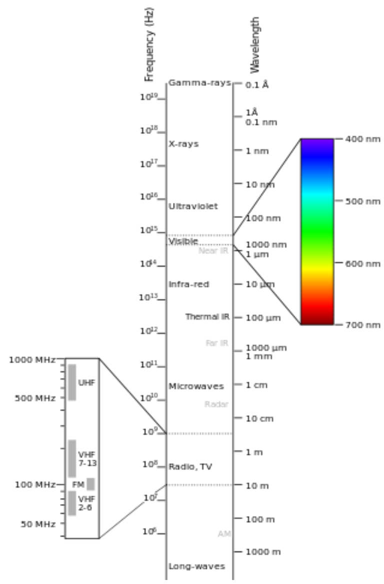


Figure 4.4.1.1

If light and radio are the same thing, why can't we see radio signals? There are two differences between all these types of electromagnetic waves. The first is the frequencies are different for each type (and since $v = \lambda f$, the wavelengths also have to be different). In general you need an antenna close to the size of the wavelength of the wave in order to detect it. Our eyes have special molecules that are just the right size to detect visible light but not X-rays, cell phone signals or other electromagnetic waves. Car antennas are around a meter or so in length because the wavelength of radio signals is in this range. Cell phone antennas (which are built inside the phone) are a few centimeters in size because this is how big the waves making up the cell phone signal are.

The second difference between different types of electromagnetic waves has to do with the energy of the photon. Einstein proved that the energy of an electromagnetic wave arrives in a single pulse called a ***photon*** whose energy is proportional to frequency. The energy of an individual photon for a wave with frequency f is given by $E = hf$ where $h = 6.63 \times 10^{-34}\,\mathrm{J\ s}$ is ***Planck's constant***. This energy has nothing to do with the brightness (intensity) of the wave, only the frequency. Intensity or brightness is related to amplitude of the wave and the number of photons, not the energy of an individual photon. Each X-ray photon carries lots energy (hence they can penetrate the body and also cause cancer). Cell phone signals and visible light are lower frequency electromagnetic waves and carry less energy per photon (as a result they cannot cause cancer, no matter how intense they are). Electromagnetic waves with frequencies higher than the visible part of the spectrum are said to be ***ionizing*** because they have enough energy to ionize (remove an electron) from a molecule. Removing electrons from biological molecules at the wrong time causes them to react differently and this can lead cellular chemistry to become abnormal, sometimes causing cancer.

Video/audio examples:

- NASA YouTube about the electromagnetic spectrum.

4.4.2: Antenna Simulation

In this simulation you see an oscillating electron in a sending antenna on the left. Because electrons have an electric field, an accelerating electron will create a wave in the electric field around it. Magnetic fields are created by moving charges so a magnetic wave is also formed by an accelerating charge. Only the y-component of the *change* in the electric field is shown (so an oscillation frequency of zero will show nothing, because there is only a constant electric field). The magnetic fields are not shown.

The simplest type of receiving antenna can be approximated by free charges (electrons) constrained by a metal wire. For a receiving antenna oriented in the y-direction, an oscillating field traveling in the x-direction will cause the charges in the receiver to oscillate in the y-direction with the same frequency as the wave (the charges cannot move in the x-direction because they are confined to the wire). This oscillating current can then be analyzed by electronic circuitry to extract the transmitted signal. A general rule of thumb is that for strongest reception, the receiving antenna should be roughly the same length as the wavelength of the wave it is trying to receive.

For simplicity this simulation has oscillating positive charges in the sending and receiving antennas. (Electrons feel a force in the opposite direction of the applied field.) Time is measured in microseconds (10^{-6} s).

Simulation Questions:

1. Run the simulation and describe what you see. Does the receiving antenna charge start oscillating immediately? Why or why not?
2. Try different oscillation frequencies. How does the frequency of the sending charge compare with the frequency of oscillation of the charge in the receiving antenna?
3. Use the step button to find the time lapse between when the source charge starts to oscillate and when the receiver starts oscillating. If the receiving antenna is $1.6 \times 10^3\ \mathrm{m}$ away, what is the speed of the wave?
4. Repeat the previous exercise with different oscillation frequencies. Does changing the oscillation frequency change the speed the wave travels in the x-direction? What does change if the oscillation speed changes? (Hint: Recall that $v = \lambda f$ where v is the speed of the wave.)
5. What do you notice about the amplitude of the wave as it travels away from the sending antenna? Explain.

Chapter Five Summary

For transverse waves the motion is up and down while the wave moves forward. Examples include water waves, electromagnetic waves, and earthquake S-waves. In longitudinal (compressional) waves the motion is back and forth in the same direction the wave travels. Sound waves and earthquake P-waves are examples. The same equations and behavior applies to all types of waves.

Questions on Wave Types:

$$f = 1/T \qquad v = f\lambda \qquad v = \omega/k \qquad k = 2\pi/\lambda \qquad \omega = 2\pi f \qquad y(x,t) = A\cos(kx - \omega t + \phi)$$

1. Give a definition of wavelength. In what units is wavelength measured?
2. What is the difference between wavelength and period? How do you measure each?
3. Give a definition and example of a transverse wave, a longitudinal wave and a torsional wave.
4. Suppose you have a slinky stretched between you and another person. Describe what you would do to your end to make
   a. a transverse wave
   b. a longitudinal wave
   c. a torsional wave.
5. In the following graph, determine the period and wavelength of the wave on the string. Each block on the lower graph is one meter, time is in seconds as shown on the upper graph.

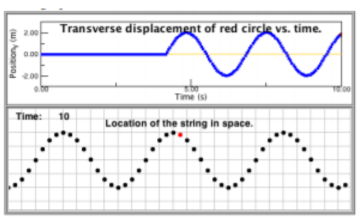


Figure 4.4.2.1

6. In the following graph, determine the period and wavelength of the wave on the string. Each block on the lower graph is one meter, time is in seconds as shown on the upper graph.

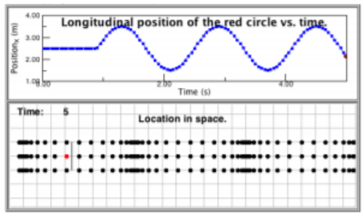


Figure 4.4.2.2

7. Why are longitudinal waves sometimes called compressional waves?
8. What are compressions and rarefactions in sound waves?
9. How is simple harmonic motion (a mass on a spring) connected to wave motion for transverse waves (Hint: Look at the transverse wave simulation.)?
10. How is simple harmonic motion (a mass on a spring) connected to wave motion for longitudinal waves?
11. What are S-waves and P-waves?
12. Describe the major types of surface earthquake waves.
13. Is a water wave longitudinal, transverse or both? Explain.
14. Find the wavelength of a 20 Hz sound wave (about the lowest note humans can hear) and the wavelength of a $20,000\text{ Hz}$ sound wave (about the highest note humans can hear). Assume the speed of sound is 343 m/s.
15. What is the connection between simple harmonic motion and the points on a string when the string has a wave traveling on it?
16. In the following graph, determine the wavelength of the wave. X is measured in centimeters.

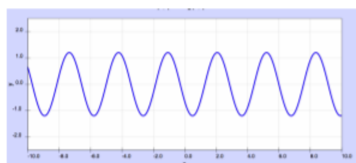


Figure 4.4.2.3

17. In the graph above, what is the amplitude of the wave?
18. In the following graph, determine the period of the wave.

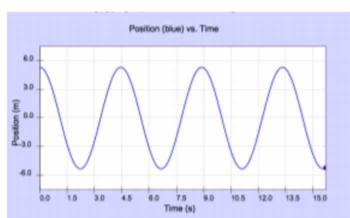


Figure 4.4.2.4

19. For the graph above, what is the frequency of the wave?
20. For the preceding graph, what is the amplitude of the wave?
21. What is the angular frequency, ω, for a 20 Hz sound wave? (Hint: See top of page).
22. What is the wave vector, k, for a 4.0 meter sound wave?
23. What are electromagnetic waves? Give some examples.
24. What is the difference between sound waves and radio signals?
25. Why can we see visible light but not other types of electromagnetic waves?
26. What is the difference between a gamma-ray and visible light?
27. What is the difference between ionizing radiation and non-ionizing radiation?
28. Why can't cell phone signals cause cancer but x-rays can (they both electromagnetic waves)?

29. What is the wavelength of a radio signal of 89.3 MHz, given than $\text{M} =$ Mega $= 106$ and the speed of radio signals are the same as light ($3.0 \times 10^8\text{ m/s}$)?
30. Calculate the frequency of your favorite color. Hint: Nanometers $= \text{ nm} = 10^{-9}\text{ m}$.

CHAPTER OVERVIEW

5: Wave Speed

There are two different speeds involved with describing a wave. In previous chapters we saw that the individual points on a wave oscillate (up and down for transverse waves, back and forth for longitudinal waves) with simple harmonic motion, just like masses on springs. A point on a transverse wave moves fastest as it passes through the equilibrium point, slows down as it reaches the maximum amplitude, stops, turns around and increases speed in the opposite direction. So this speed is changing over time; points on the wave accelerate up and down.

But the up and down speed of a point on a transverse wave doesn't tell us how fast the wave moves from one place to the next. The ***wave speed***, v, is how fast the wave travels and is determined by the properties of the medium in which the wave is moving. If the medium is uniform (does not change) then the wave speed will be constant. The speed of sound in dry air at $20^{\circ}\mathrm{C}$ is $344\ \mathrm{m/s}$ but this speed can change if the temperature changes.

It is also possible to include the direction and define a vector called the ***wave velocity*** which is speed and direction but for now we will just talk about wave speed.

Key Terms:

Speed of a point on the wave, speed of the wave, velocity of the wave, bulk modulus (compressibility), Young's modulus (stiffness), speed of electromagnetic waves, speed of sound in different media, linear waves, $v = \lambda f$, wave number, angular frequency, index of refraction.

5.1: Wave Speed

There are two different speeds involved with describing a wave. In previous chapters we saw that the individual points on a wave oscillate (up and down for transverse waves, back and forth for longitudinal waves) with simple harmonic motion, just like masses on springs. But the up and down speed of a point on a transverse wave doesn't tell us how fast the wave moves from one place to the next. The wave speed, v, is how fast the wave travels and is determined by the properties of the medium.

5.1.1: Speeds of Different Types of Waves

The speed of a wave is fixed by the type of wave and the physical properties of the medium in which it travels. An exception is electromagnetic waves which can travel through a vacuum. For most substances the material will vibrate obeying a Hooke's law force as a wave passes through it and the speed will not depend on frequency. Electromagnetic waves in a vacuum and waves traveling though a linear medium are termed ***linear waves*** and have constant speed. Examples:

- For sound waves in a fluid (for example air or water) the speed is determined by $v=(B/\rho)^{1/2}$ where B is the bulk modulus or compressibility of the fluid in newtons per meter squared and ρ is the density in kilograms per cubic meter.
- For sound waves in a solid the speed is determined by $v=(Y/\rho)^{1/2}$ where Y is Young's modulus or stiffness in Newtons per meter squared and ρ is the density in kilograms per meter cubed.
- For waves on a string the speed is determined by $v=(T/\mu)^{1/2}$ where T is the tension in the string in Newtons and μ is the mass per length in kilograms per meter.
- Although electromagnetic waves do not need a medium to travel (they can travel through a vacuum) their speed in a vacuum, $c=(1/\mu_o\varepsilon_o)^{1/2}=3.0\times 10^8\ \mathrm{m/s}$ is governed by two physical constants, the permeability μ_o and the permittivity, ε_o of free space (vacuum).

Speed of Sound in Various Substances (CRC Handbook)

Glasses ($0^\circ\mathrm{C}$)	**Substance**	**Speed of Sound** ($\mathrm{m/s}$)
	Carbon Dioxide	259
	Hydrogen	1284
	Helium	965
	Nitrogen	334
	Oxygen	316
	Air (21% Oxygen, 78% Nitrogen)	331
	Air ($20^\circ\mathrm{C}$)	344
Liquids ($25^\circ\mathrm{C}$)	Glycerol	1904
	Sea Water (3.5% salinity)	1535
	Water	1493
	Mercury	1450
	Kerosene	1324
	Methyl Alcohol	1103
	Carbon Tetrachloride	926
Solids	Diamond	12000
	Pyrex Glass	5640
	Iron	5960
	Granite	6000
	Aluminum	5100
	Brass	4700
	Copper (annealed)	4760
	Gold	3240
	Lead (annealed)	2160
	Rubber (gum)	1550

Table 5.1.1.1

Here is a more comprehensive list of the speed of sound in various materials.

As we saw in the previous chapter, there is a relationship between the period, wavelength and speed of the wave. The period of a cork floating in the water is affected by how fast the wave passes (wave speed) and the distance between peaks (wavelength). The ***relationship between speed, period and wavelength of a sine wave*** is given by $v = \lambda/T$ where wavelength and period for a sine wave were defined previously. This can also be written as $v = \lambda f$ since frequency is the inverse of period and is true for all linear waves. Notice that, since wave speed is normally a fixed quantity the frequency and wavelength will be inversely proportion; higher frequencies mean shorter wavelengths.

Often it is easier to write $\omega = 2\pi f$ where ω is the ***angular frequency*** in radians per second instead of having to write $2\pi f$ everywhere. Likewise it is easier to write $k = 2\pi/\lambda$ where k is the ***wave number*** in radians per meter rather than having to write $2\pi/\lambda$ a lot. (Note that k is *not* a spring constant here.) Using these new definitions the speed of a wave can also be written as $v = f\lambda = \omega/k$.

If the medium is uniform the speed of a wave is fixed and does not change. There are circumstances where the speed of a particular wave does change, however. Notice that the speed of sound in air depends on the density of the air (mass per volume). But the density of air changes with temperature and humidity. So the speed of sound can be different on different days and in different locations. The temperature dependence of the speed of sound in air is given by $v = 344 + 0.6(T - 20)$ in meters per second where T is the temperature in Celsius (T here is temperature, not period). Notice that at room temperature ($20^{\circ}\mathrm{C}$) sound travels at $344\ \mathrm{m/s}$.

The speed of sound can also be affected by the movement of the medium in which it travels. For example, wind can carry sound waves further (i.e. faster) if the sound is traveling in the same direction or it can slow the sound down if the sound is traveling in a direction opposite to the wind direction.

Electromagnetic waves travel at $\mathrm{c} = 3.0 \times 10^{8}\ \mathrm{m/s}$ in a vacuum but slow down when they pass through a medium (for example light passing from air to glass). This occurs because the material has a different value for the permittivity and/or permeability due to the interaction of the wave with the atoms of the material. The amount the speed changes is given by the ***index of refraction*** $n = c/v$ where c is the speed of light in a vacuum and v is the speed in the medium. The frequency of the wave does not change when it slows down so, since $v = \lambda f$, the wavelength of electromagnetic waves in a medium must be slightly smaller.

Video/audio examples:

- What is the speed of sound in a vacuum? Buzzer in a bell jar. Why is there no sound when the air is removed from the jar?
- Demonstration of speed of sound in different gasses. Why is there no sound when the air is removed from the jar?
- These two videos demonstrate the Allasonic effect. The speed of sound is different in a liquid with air bubbles because the density is different. As the bubbles burst, the speed of sound changes, causing the frequency of sound waves in the liquid column to change, thus changing the pitch. Example: one, two. What do you hear in each case?
- The Zube Tube is a toy that has a spring inside attached to two plastic cups on either end. Vibrations in the spring travel at different speeds so a sound starting at one end (for example a click when you shake the tube and the spring hits the cup) ends up changing pitch at the other end as the various frequencies arrive. In other words this is a nonlinear system. See if you can figure out from the video which frequencies travel faster, high frequencies or low.

Mini-lab on measuring the speed of sound.

Questions on Wave Speed:

$$f = 1/T, \quad v = f\lambda, \quad v = \omega/k, \quad k = 2\pi/\lambda, \quad \omega = 2\pi f, \quad y(x,t) = A\cos(kx - \omega t + \phi), \quad v = \sqrt{B/Q}$$

1. Light travels at $3.0 \times 10^{8}\ \mathrm{m/s}$ but sound waves travel at about $344\ \mathrm{m/s}$. What is the time delay for light and sound to arrive from a source that is $10,000\ \mathrm{m}$ away (this can be used to get an approximate distance to a thunderstorm)?
2. What two mistakes are made in science fiction movies where you see and hear an explosion in space at the same time?
3. Consult the table for the speed of sound in various substances. If you have one ear in the water and one ear out while swimming in a lake and a bell is rung that is half way in the water some distance away, which ear hears the sound first?
4. At $20\mathrm{C}$ the speed of sound is $344\ \mathrm{m/s}$. How far does sound travel in 1 s? How far does sound travel in 60 s?
5. Compare the last two answers with the distance traveled by light which has a speed of $3.0 \times 10^{8}\ \mathrm{m/s}$. Why do you see something happen before you hear it?

6. The speed of sound in water is 1482 m/s. How far does sound travel under water in 1 s? How far does sound travel under water in 60 s?
7. What happens to the speed of sound in air as temperature increases?
8. Using the equation for the speed of sound at different temperatures, what is the speed of sound on a hot day when the temperature is 30°C? Hint: $v = 344 \text{ m/s} + 0.6(T - 20)$ where T is the temperature in Celsius.
9. Using the speed of sound at 30°C from the last question, recalculate the distance traveled for the cases in question four.
10. Suppose on a cold day the temperature is -10°C (14°F). You are playing in the marching band outside. How long does it take the sound from the band to reach the spectators if they are 100 m away?
11. What is the difference in the speed of sound in air on a hot day (40°C) and a cold day (0°C)?
12. What would an orchestra sound like if different instruments produced sounds that traveled at different speeds?
13. The speed of a wave is fixed by the medium it travels in so, for a given situation, is usually constant. What happens to the frequency of a wave if the wavelength is doubled?
14. What happens to the wavelength of a wave if the frequency is doubled and has the same speed?
15. Suppose a sound wave has a frequency 200 Hz. If the speed of sound is 343 m/s, what wavelength is this wave?
16. What factors determine the speed of sound in air?
17. Why do sound waves travel faster through liquids than air?
18. Why do sound waves travel faster through solids than liquids?
19. The speed of sound in a fluid is given by $v = \sqrt{B/Q}$ where B is the Bulk Modulus (compressibility) and Q is the density. What happens to the speed if the density of the fluid increases?
20. What must be true about the compressibility, B, of water versus air, given that sound travels faster in water and water is denser than air?
21. The speed of sound in a fluid is given by $v = \sqrt{B/Q}$ where B is the Bulk Modulus (compressibility) and Q is the density. Can you think of a clever way to measure the Bulk Modulus of a fluid if you had an easy way to measure the speed of sound in a fluid? Explain.
22. The speed of sound on a string is given by $v = \sqrt{T/\mu}$ where T is the tension in Newtons and μ is the linear density (thickness) in kg/m. You also know that $v = f\lambda$. Give two ways of changing the frequency of vibration of a guitar string based on the knowledge of these two equations.
23. For the previous question, increasing the tension does what to the frequency? What does using a denser string do to the frequency?
24. The following graph is of a wave, frozen in time at $t = 0$. The equation describing the wave is $y(x,t) = A\cos(kx - \omega t + \phi)$. Sketch the effect of doubling the amplitude, A.

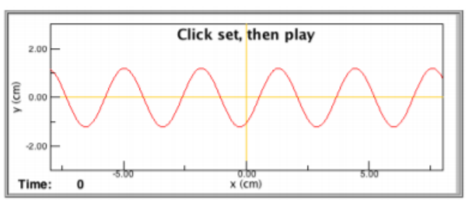


Figure 5.1.1.1

25. For the following graph of a wave, sketch the effect of doubling the wavelength.

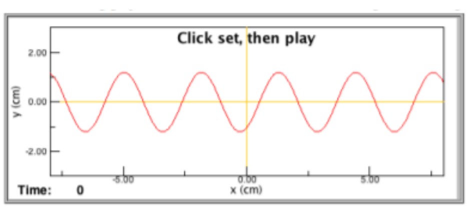


Figure 5.1.1.2

26. The mathematical description of a sine wave is given by $y(x,t) = A\cos(kx - \omega t + \phi)$. Explain what each of the terms (A, k, ω, ϕ) represent.

5.1.2: Speed of a Wave Simulation

The ***velocity*** of the wave, v, is a constant determined by the properties of the medium in which the wave is moving as we saw above. The velocity is a vector which gives the forward speed of the wave and the direction the wave is traveling. For now we will not worry about direction since the waves being discussed will all be assumed to travel along the x-axis. The speed of a sine wave is given by $v = \lambda/T$ where wavelength and period for a sine wave were defined in the previous exercise. Note that this is *NOT* the up and down speed of a point on the wave. Here we are talking about the forward speed of the wave crests.

In this simulation the original wave will remain in the window so that as you make changes to $f(x,t)$ you can see how the moving wave (in red) compares to the original stationary wave ($g(x,t)$, in blue).

This simulation is *misleading* in one important way. In the simulation you can set any combination of angular frequency and wavenumber you choose and so have any speed you want for the wave. But as we saw above, for mechanical and acoustic waves the speed is determined by the medium in which the wave travels. Since $v = \omega/k$ it is the case that the angular frequency and wavenumber are inversely proportional with a constant v.

Simulation Questions:

1. Determine the speed of the wave in the simulation using $v = \lambda/T$ where wavelength and period are determined from the simulation (wavelength is the peak to peak distance and period is the time for one peak to travel to the location of the next peak). What is the forward speed of this wave?
2. The speed of this wave is also given mathematically by $v = \omega/k$ since $\omega = 2\pi f = 2\pi/T$ and $k = 2\pi/\lambda$. What is the speed of this wave based on the values of ω and k in the equation? Does this match the speed you got from the simulation?
3. Reset the initial conditions and experiment with values of the wavenumber both smaller and larger than $2.0\ \mathrm{rad/m}$ keeping the angular frequency fixed. How does the wavenumber change the speed of the wave?
4. Reset the initial conditions and experiment with values of the angular frequency both smaller and larger than $0.8\ \mathrm{rad/s}$ keeping the wavenumber fixed. How does the angular frequency change the speed of the wave?
5. Reset the initial conditions. For a wavenumber of $4.0\ \mathrm{rad/m}$ experiment to find the correct angular frequency which gives the original speed of the wave you found in questions one and two (you should be able to see from the simulation when the new wave is traveling at the same speed as the original).
6. Calculate the wavenumber which gives the speed of the original wave for angular frequencies of $0.4, 0.6, 1.0$,and $1.2\ \mathrm{rad/s}$ using the relationship in question two. Check your answers with the simulation if you are in doubt.

Chapter Six Summary

The speed of a wave is generally fixed by the properties of the physical material through which the wave is traveling, the exception being electromagnetic waves which require no medium (they can travel through a vacuum). In all cases of linear waves the wavelength and frequency are inversely proportional and given by $v = f\lambda = \omega/k$.

CHAPTER OVERVIEW

6: Wave Behavior

In this chapter (the longest in the book!) we investigate the behavior of waves in various circumstances. Some of our examples will be for light but remember that all waves have similar behavior; sound waves will also obey the laws of reflection and refraction, scattering, diffraction, interference, etc.

Key Terms:

Reflection, specular and diffuse reflection, refraction, constructive and destructive interference, path difference, diffraction, dispersion, Doppler shift, standing waves, scattering, beats, ultrasound.

6.1: Doppler Shift

If the wave source or receiver is moving, the waves will appear to have a different frequency. For example if you are moving towards a sound source you catch up with the next peak in the wave sooner than you would expect because you are moving towards it. This effect is called the Doppler Shift and occurs for both light and sound.

6.1.1: Doppler Shift

The movement of a source of waves does *not* change the speed of the waves. Neither does the movement of a receiver of waves. However, if the source or receiver is moving, the waves will appear to have a different *frequency*. For example if you are moving towards a sound source you catch up with the next peak in the wave sooner than you would expect because you are moving towards it. This effect is called the ***Doppler Shift*** and occurs for both light and sound. For sound, if the object moves towards you the frequency seems higher. If the object is moving away the frequency seems lower.

> **Note**
>
> It isn't the loudness we are talking about here, which also increases as the source gets closer and decreases as the source moves away; the Doppler shift is about frequency, not volume.

The Doppler shift for light changes the frequency or color of the light. Objects moving at high speed towards you will have their color shifted a little towards the blue end of the spectrum while objects moving away have colors shifted towards the red end of the spectrum. This cannot be seen with the eye but can be measured and is used in astronomy to determine the speed and direction of stars and galaxies. The Doppler shift tells us that most galaxies are moving away from ours which tells us the universe is expanding.

Police radar and weather radar both use the Doppler Effect to measure the speed of something. For a stationary object (a car, a raincloud) the reflected radar beam returns with the same frequency it originally had. Measuring how long it takes for the beam to get back gives us the distance to the object but not its speed. If the object (car, raincloud) is moving the Doppler Effect will cause a shift in the returning frequency. A computer compares the outgoing and reflected beam to determine not only how far away the object is (from the time it takes to return), but also how fast it is moving (from the Doppler shift in frequency). For weather radar the amount of reflected radar also tells us something about the water content in the cloud. Hail and snow have different reflection patterns than rain so radar can also tell us what is in the cloud in addition to distance and speed.

Video/audio examples:

- Doppler shift of a fire engine. Notice that the loudness is NOT what is important; the change in *pitch* is the Doppler effect.
- For a source traveling faster than the speed of sound the waves pile up into a shock wave which produces a sonic boom.
- Once piece of evidence (there are others) that the universe is expanding is the Doppler shift of light from stars.
- Sound waves with frequencies above 20,000 Hz are called *ultrasound* sound waves. You may be familiar with pictures of unborn children made using ultrasound. The Doppler shift can also be used with ultrasound to see the movement of blood in the heart. This is known as echo cardiology.
- Directions for building your own radar detector and a lecture on how it works.

6.1.2: The Doppler Effect Simulation

This simulation looks at the Doppler effect for sound; the black circle is the source and the red circle is the receiver. The time is measured in centiseconds (10^{-2} s), distances are in meters. A similar effect occurs for light but in that case the source and receiver cannot travel faster than the wave speed (the speed of light).

Simulation Questions:

1. Click the play button to see a stationary source and receiver (Animation 1). Reset and use the pause and step buttons to verify that the period at the receiver (time elapsed from when one wave reaches the receiver until the next one reaches it) is 5.0×10^{-3} s. What is the frequency of this wave?
2. After there are several waves in the simulation pause it and use the mouse to find the wavelength (distance between two successive crests). What are the wavelength and speed of the wave (wavelength/period)?
3. Now look at Animation 2, where the receiver is moving. Use the step button above to find the period (time between crests) *as measured by the moving receiver* when it is on the right of the source (moving towards the source). What is the frequency at the receiver if it is moving towards the source?
4. When the receiver gets to the left of the source (moving away from the source) pause the simulation and measure the period. What is the frequency at the receiver if it is moving away from the source?
5. Now look at Animation 3 which shows the source moving but the receiver stationary. Again find the frequency while the source is on the left, moving towards the receiver and the frequency when it is on the right moving away.

Advanced Questions:

Animation 4 shows the effects of a moving source and a moving observer at the same time. The equation for the Doppler shift with both a moving source and observer is given by $f' = f(v \pm v_o)/(v \mp v_s)$ where f' is the received frequency, f is the original frequency, v is the speed of the wave, v_o is the speed of the observer and v_s is the speed of the source. The upper signs in the equation are used if either the observer or source is moving towards each other and the lower signs are used if the either object is moving away from the other (so if the observer is moving towards the source but the source is moving away from the observer the equation to use is $f' = f(v + v_o)/(v + v_s)$).

1. For the case of the moving receiver and stationary source ($v_s = 0$) use the original frequency you found in question one, the shifted frequency (f') you found in question three and the speed of sound you found in two to find the speed of the observer.
2. Animation 5 shows a source moving faster than the speed of the sound wave. In this case all of the wave crests arrive together forming a shock wave or "sonic boom". Why can this not happen in the case of light from a moving light source?
3. Electromagnetic waves will also undergo a Doppler shift except that the relative velocity between the source and observer can never be larger than the speed of light and the formula for calculating the shift is slightly different. For electromagnetic waves we have $f' = f((c+v)/(c-v))^{1/2}$ where v is the relative speed between the observer and source (positive if they are approaching and negative if they are moving away from each other) and c is the speed of light.
4. If the speed of the wave is known and the original and received frequencies are know the speed of the source or observer can be found. Explain how you could determine the speed of a car or thunderstorm by bouncing radio or microwaves off of them. (Police radar and thunderstorm tracking both use the Doppler Effect.)
5. If a car goes past with its radio blaring we easily hear the Doppler shift for sound as the car passes (the sound appears to shift from a pitch which is too high to one which is too low). (Note: We are talking about the change in pitch, *not* the change in volume.) But if a car goes past with its lights on we do not notice the Doppler shift for light (the color does not seem to shift towards the red frequencies). Explain why this is so. (Hint: Try plugging in some numbers for a car speed in the equation for the Doppler shift for light).
6. If an astronomer notices that the spectrum of colors coming from a star are all shifted towards the red end of the spectrum (the frequencies are lower than they should be) what can she conclude about the motion of the star relative to the earth? (This is one of the pieces of evidence that the universe is expanding; nearly all the stars and galaxies around us are moving away from us.)

Chapter Seven Summary

Waves (of all types) will: reflect (diffuse or specular) at the same angle when they hit a solid surface; bend if their speed changes (refraction); bend if they run into an object or opening that is near the same size as the wavelength (diffraction); change frequency (NOT loudness!) if the source or receiver are moving (Doppler shift). Two waves can add constructively or destructively to give a wave with larger amplitude (louder sound or brighter light) or sound cancellation, standing waves, beats, interference (colors on soap bubbles, moth wings, bird feathers).

Questions on Wave Behavior:

1. Give a definition of the law of reflection.
2. Do sound waves obey the law of reflection? Explain.
3. A police officer has a new sound gun that stuns the criminal. They can see their suspect who is around the corner, reflected in a mirror. Should she simply aim at the image of the target in the mirror? Explain.
4. Trucks sometimes have a sign on the back that says "If you can't see me in my mirror, I can't see you." Explain the physics here.
5. Why is the image in a mirror reversed left to right but not top to bottom? (A drawing might help your explanation.)
6. Some storefront windows are angled so the bottom is further in and the top comes out towards the street. Explain how this would help reduce glare on a bright day. (A drawing might help your explanation.)
7. What is the difference between diffuse reflection and specular reflection?
8. Why are matte finishes for photos and books generally better than glossy finishes?
9. Why do sound studios often have the walls covered with egg carton shaped foam?
10. Why is it harder to see the road at night in the rain?
11. Give a definition of the law of refraction, explaining the difference between reflection and refraction.
12. How does refraction depend on the speed of a wave?
13. Give some examples of common, everyday objects that use refraction to operate.
14. Why are images blurry underwater if you don't have goggles?
15. Why are images not blurry underwater if you are wearing goggles?
16. If a fish wore goggles to come above the surface, why would it want to have goggles filled with water?
17. If you place a glass test tube in water you can still see it but if you place it in soybean oil it disappears. What does this tell you about the speed of light in glass, water and soybean oil?
18. If you want to spear a fish under water from the shore should you aim below it, at it or above it? Explain.
19. If you want to zap a fish with a laser, should you aim below it, at it or above it? Explain.
20. Why does a pond or lake with very clear water look shallower than it really is? Explain using a diagram.
21. Suppose color X bends more when passing through glass than color Y. Which moves slower in the glass?
22. What is total internal reflection? When does it occur?
23. What modern devices depends on total internal reflection?
24. On a windy day, why can you hear someone clearly if they are downwind but can't hear them as well if they are upwind?
25. Why is it difficult to hear someone on the other side of a lake during the day when the air above the lake is cool but very easy to hear voices at night when the air above the lake is warmer than air higher up? (Hint: Sound travels faster in warmer air.)
26. The term 'heat lightning' is sometimes used to describe lightning that we can see off at a distance but not hear. Why don't we hear it like ordinary lightning? (Hint: Think about the previous question.)
27. What is dispersion and what causes it? Give an example.
28. What is constructive and destructive interference? When does each occur?
29. Explain how does path difference cause constructive and destructive interference?
30. Explain how the phenomena of beats occurs.
31. Why do some bird feathers appear to be iridescent, changing color when viewed from different angles?
32. What causes the different colors on a CD disk?
33. What causes the different colors on a soap bubble?
34. Suppose you are standing directly in front of a pair of stereo speakers. Why would you expect the sound not to be quite as loud if you move slightly to the left or right?
35. What is diffraction? Give some examples.
36. Under what circumstances do you expect to see the effects of diffraction?
37. Why can you hear sound from the other room, even when you cannot see into the room?
38. Why does light not bend when it passes through a doorway but sound does?
39. What is scattering?
40. Why is the sky blue?
41. What is the Doppler effect?

42. A friend hears an ambulance go by and says this is an example of the Doppler effect because the sound got louder and then softer. Correct your friend's mistaken definition of the Doppler effect.
43. Give an example of the Doppler effect for light and one for sound.
44. Does the Doppler shift depend on whether the source or the receiver is moving?
45. Does the speed of the wave change when there is a Doppler shift? Explain.
46. We can easily hear the Doppler shift of a car passing by but we do not notice the Doppler shift of light from its headlamps. Why is that?
47. Originally radar was just used to find the distance to a plane or thundercloud by measuring how long it took for the signal to return. What additional information does measuring the Doppler effect provide?
48. What is one piece of information that tells us the universe is expanding?

6.2: Mirrors and Reflection

In many cases waves of all types will travel in a straight line, reflecting off of objects and surfaces at the same angle that they strike the surface. This is called the law of reflection and is true for sound waves as well as light as long as the surface is smooth relative to the wavelength.

6.2.1: Reflection

In many cases waves of all types will travel in a straight line, reflecting off of objects and surfaces at the same angle that they strike the surface. This is called the ***law of reflection*** and is true for sound waves as well as light as long as the surface is smooth relative to the wavelength. In the pictures below the arrows show the direction the light or sound is traveling. On the left for reflection off a smooth flat surface the angle of incidence, θ_i equals the angle of reflection, θ_r.

The law is still obeyed for a smooth curved surface but with a different result as shown in the figure on the right. In this case rays starting from one special point called the ***focal point*** still obey the law of reflection but they end up traveling outward parallel to each other. If we reverse the rays they will arrive parallel but all reflect through the focal point. Flashlights have curved mirrors behind the bulb with the bulb at the focal point so that the light coming from the flashlight is approximately a parallel beam. Outdoor concert shells are often built with a curved shape with the performer located at the focal point. This is so the sound is directed outward and spread evenly in the direction of the audience.

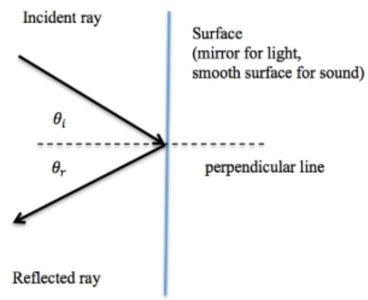


Figure 6.2.1.1

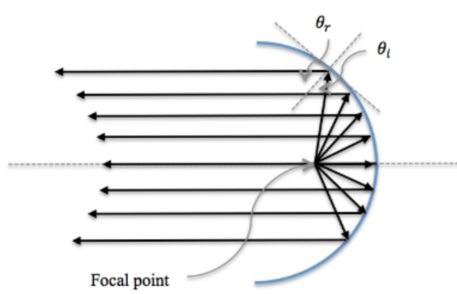


Figure 6.2.1.2

In the two pictures below the blue lines give the direction the wave is traveling and the small gray lines represent the crests of the waves (so the distance between the gray lines is the wavelength). For ***specular reflection*** (picture on the left) the surface is smooth relative to the size of the wave and the entire wave obeys the law of reflection. For ***diffuse reflection*** (picture on the right) the surface is rough relative to the size of the wave and, while individual components of the wave obeys the law of reflection, the wave as a whole is broken up by the reflection. To avoid strong echoes and other unwanted acoustic effects, surfaces in concert halls are designed to absorb sound and also to reflect sound diffusely, as we will see later.

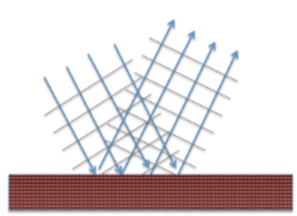


Figure 6.2.1.3

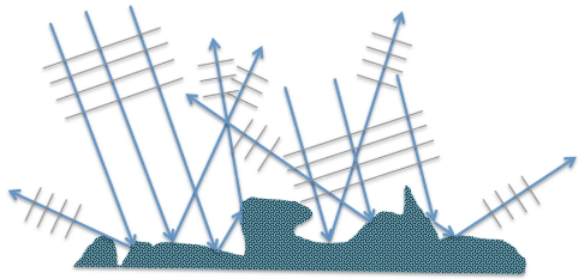


Figure 6.2.1.4

Video/audio examples:

- Reflection of light.
- Ripple tank reflection of water waves.
- This Ripple Tank simulation by Paul Falstad lets you look at wave reflection from a parabolic mirror and inside a circular or elliptic enclosure. Directions: Choose Setup: Parabolic Mirror 1, Mirror 2, Circle or Ellipse. Describe what you see in each case. For the ellipse, what happens to sound that starts at the focus of an elliptical chamber? This is the explanation behind the strange acoustics in some large building such as the nation's capital.
- Animation of what images look like in a convex mirror.

6.2.2: Reflection Simulation

The simulation below allows for a brief exploration of specular reflection. In this case we are simulating mirrors but keep in mind that, if the surface is smooth relative to the wavelength, **any** wave will obey the same rules. The mirror in each case in the center and everything to the right is behind the mirror (virtual). The units of height, h, and distance, d, are arbitrary (cm, inches, etc.).

Simulation Questions:

1. The simulation starts with a flat mirror, an object on the left (a candle) and the candle's image (shown on the right). Two real rays (white arrows) are shown leaving the object and reflecting off the mirror. Carefully describe each ray. What does the parallel ray do when it reflects off the mirror? What does the ray that goes to the center of the mirror do when it reflects?
2. The small protractor can be dragged to different locations and the arrow tip can be moved to change the angle. Use it to measure the incident and reflected angles (angles are measured from the horizontal line which is perpendicular to the surface of the mirror, not from the surface of the mirror). What are the incident and reflected angles for the ray going from the object to the center of the mirror? Is the law of reflection obeyed?
3. Our eyes and brain do not perceive the rays as reflecting. Instead, our brains perceive the reflected rays as coming from behind the mirror (the magenta arrows) from an imaginary or ***virtual*** image behind the mirror. The object can be moved using the mouse. Does the size of the image change if the object is moved back and forth?
4. Now choose the convex mirror button. Move the object back and forth. Describe what happens to the image as the object moves back and forth. What happens to a parallel ray when it reflects? What happens to a ray going to the center of the mirror?
5. The red dot on the right is the focal point of the mirror and is $1/2$ the radius of curvature of the mirror. The turquoise dot locates the center of curvature and the radius of curvature is from the center of curvature to the mirror. What happens to a ray that starts on the left and heads towards the focal point on the other side; how does it reflect?
6. Use the protractor to measure the incident and reflected angles for the ray striking the center of the mirror for the convex case. (Note: The curvature of the mirrors is exaggerated; place the protractor on the turquoise line rather than the mirror itself.) Are the incident and reflected angles equal?
7. The other rays also obey the law of reflection but the mirror surface at those locations is not perpendicular to the x-axis. For a very short object, place the protractor at the location (on the turquoise line) where the top ray strikes the mirror and rotate the arrow to point directly away from the center of curvature (turquoise dot). This is the direction of a perpendicular to the mirror and should fall precisely between the incoming and outgoing rays. What is the angle? (Note: These measurements only work for the very center of the mirror; in other words for a very short object relative to the height of the mirror.)
8. For the flat and convex mirror, does the size of the object change the laws of reflection? Describe what you see.
9. Now click on the concave mirror button. Slide the object back and forth, moving it from very far away from the mirror to very close. Describe where the image is, its size relative to the object and orientation (Hint: There are three different cases; Closer than the focus; further away than the focus but closer than twice the focal length; further away than twice the focal length.)
10. Images appearing behind the mirror (to the right in the simulation) are ***virtual images***; we can only see them by looking into the mirror. Images appearing on the left side, in front of the mirror are ***real images***. These can be seen in the mirror and can also be projected onto a screen. For the flat mirror were the images real or virtual? Are the images for the convex mirror real or virtual? How do you know? In which cases were the images real and virtual in the concave mirror?
11. Describe the three reflected rays for the case of a virtual image in the concave mirror case. Do they obey the law of reflection? Explain.
12. Describe the three reflected rays for the case of a real image in the concave mirror case. Do they obey the law of reflection? Do the laws of reflection change depending on the size of the object in the concave case? Explain.
13. Go back to the flat mirror and use the mouse to find the distance from the mirror to the object and from the mirror to the image. (Use the absolute values of the numbers in the yellow box. You may also have to click the mouse down and slide over to the object to get a value without moving the object.) How is the distance to the mirror from the object related to the distance to the image for the flat mirror?
14. For a flat mirror the image height and image distance are the same as the object's. For a curved mirror the relation between the distance to the object, s and to the image s' are related to the focal length, f which is equal to one half the radius of curvature of the mirror. The relationship is $1/f = 1/s + 1/s'$ where the focal length is positive for a concave mirror and negative for a convex mirror. The distance to the image will be positive if the image is real and negative if virtual. Use the mouse to find the distance to the object and to the image for several different distances in the convex mirror case. Use these values to find the

focal length, f (use the absolute values of the distances given in the yellow box and the rule that distances to the image will be positive if the image is real and negative if virtual).

6.3: Refraction

A wave that changes speed as it crosses the boundary of between two materials will also change direction if it crosses the boundary at an angle other than perpendicular. This is because the part of the wavefront that gets to the boundary first, slows down first. The bending of a wave due to changes in speed as it crosses a boundary is called refraction.

6.3.1: Refraction

A wave that changes speed as it crosses the boundary of between two materials will also change direction if it crosses the boundary at an angle other than perpendicular. This is because the part of the wavefront that gets to the boundary first, slows down first. The bending of a wave due to changes in speed as it crosses a boundary is called ***refraction***. As mentioned in the last chapter, light in air or a vacuum travels at $c = 3.0 \times 10^8 \text{ m/s}$ but slows down when passing through glass. As shown in the diagram below, this will cause light to change direction a little. For a piece of glass with flat surfaces this isn't very noticeable unless the glass is very thick. But for a curved surface the light ends up leaving the glass going in a different direction and this is how lenses for glasses, telescopes, microscopes, binoculars, etc. are made.

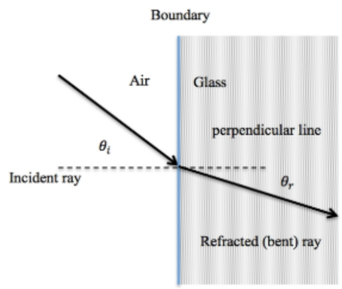


Figure $6.3.1.1$

What about sound? Sound also undergoes refraction. Recall from the last chapter that wind can change the speed of air. In the following picture notice that Jill can hear Jack because the wind speeds up the upper edges of the sound, bending it back towards the ground. Jill can't hear Dana because the wind bends the sound upward.

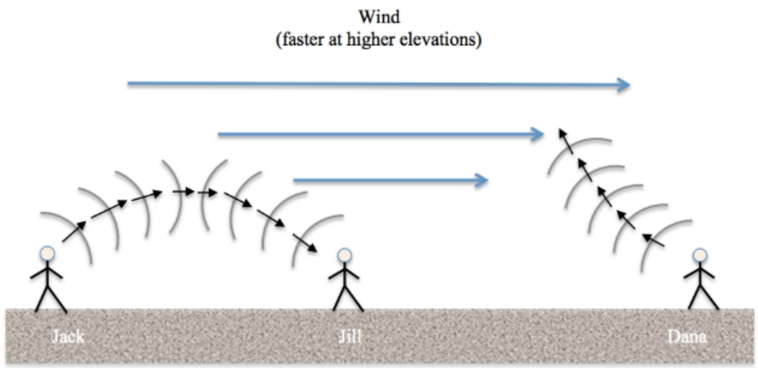


Figure $6.3.1.2$

Likewise we know that the speed of sound depends on density which changes with temperature and humidity. In the following picture notice that Jill can hear Jack because the warmer temperature speeds up the upper edges of the sound, bending it back towards the ground. In the second picture there is a temperature inversion with warmer air trapped underneath cooler air. Jill sees but does not hear the lightning (this is sometimes called *heat lightning*, as shown in the second figure below).

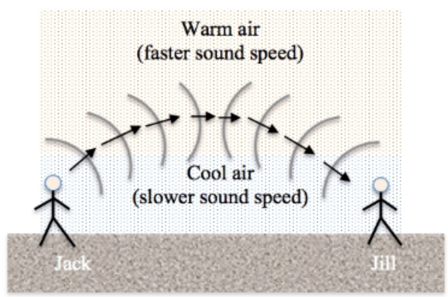


Figure $6.3.1.3$

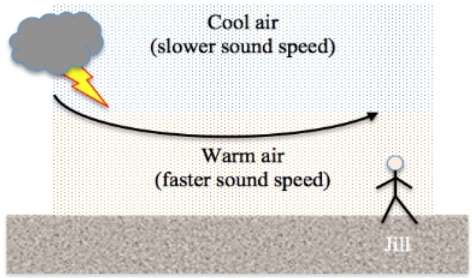


Figure 6.3.1.4

Video/audio examples:

- The broken straw illusion (due to refraction of light).
- Sound refraction example by Paul Hewitt.
- Explanation of ripple tank including Snell's law.
- Example of refraction in a swimming pool.
- Optical illusions due to the refraction of light.
- Snell's law tells you how much a light wave will bend when going from air to glass or vice versa.
- Refraction of sound in a balloon filled with Carbon Dioxide.
- Light going into glass ends up with a refracted angle that is smaller than the incident angle. Going the other way (glass to air) the light ends up with a larger refracted angle than the incident angle. In this case, what happens if the refracted angle tries to exceed 90 degrees? It reflects back into the glass, rather than passing into air. This is known as total internal reflection and is a consequence of Snell's law.
- An example of total internal reflection in a water stream. The same thing happens in a fiber optic cable; light stays inside the cable because of total internal reflection.
- Glass disappearing due to the index of refraction.
- Mini Lab on Ray Tracing.

6.3.2: Refraction Simulation

The ratio of the speed of light in a material to the speed in a vacuum ($c = 3.0 \times 10^8\ \mathrm{m/s}$) is called the ***index of refraction***; $n = c/v$ where v is the speed of light in the medium. In this simulation we will investigate the effects of a change in the speed of a wave as it moves from one material to another. Although our example is for light, the same behavior can be demonstrated with other waves. For example, much of what we know about the interior of the earth, the sun and other planets comes from tracking earthquake waves when they refract as they pass through layers of material that have different densities.

The relationship between the index of refraction and the change in the direction angle of the a ray as it goes across a medium is given by ***Snell's law***: $n_1 \sin\theta_1 = n_2 \sin\theta_2$ where n_1 and θ_1 are values measured on side of the boundary and the subscript *2* is for the material on the other side of the boundary.

Simulation Questions:

1. The simulation shows a single ray of light passing through a medium surrounded by air on both sides. You can move the protractor around and use it to measure angles by dragging the tip of the arrow. The source can also be moved. What can you say about light striking the medium perpendicular to its surface ($\theta = 0$)? What if the source is inside the medium; what happens to a perpendicular beam leaving the medium? Are these observations still true if the index of refraction is something other than 1.00?
2. Set the index of refraction to 1.40. Change the angle of the rays from the source. The ***angle of incidence*** is the angle the ray makes with a perpendicular to the surface (not the angle between the ray and the surface). Use the protractor to measure the angle of incidence (put the bottom left edge of the protractor where the ray enters the glass and line up the arrow with the ray). Move the protractor to the boundary and line up the arrow with the rays inside the material to get the ***refracted angle*** (the angle inside the medium).
3. The index of refraction of air is approximately one (light travels about as fast in air as it does in a vacuum). Use Snell's law to calculate the index of refraction in the medium. ($n_1 = 1;\ \theta_1 =$ the incident angle; θ_2 is the refracted angle; you are solving for n_2). NOTE: It is very difficult to get the protractor arrow to line up exactly with the rays so your answer may be off a little.
4. For the same angle you used in the previous two questions, move the protractor once again to the right edge of the material and measure the angle the ray leaves the material. How does this compare with the incident angle on the left?
5. Choose a different angle and use Snell's law to find the index of refraction outside the medium on the right. This time the incident angle will be inside the material at the right boundary, and n_1 will be the index inside the material (your answer to question three).
6. Drag the source inside the medium and try different angles. What do you notice?
7. When light goes from a material where the index is higher to a lower index there are angles for which the rays cannot leave the material. The angle at which a ray starts to reflect off the boundary instead of passing through is called the ***angle of total internal reflection***. You can see this if you look at the surface of the water in a pool from under water; at a certain angle you see a reflection of the bottom of the pool instead of objects above the water. Total internal reflection is also why light remains in a fiber optic cable instead of being absorbed by the coating. Find the angle for total internal reflection two ways. First experiment with the angle of incidence (source inside the medium) for a fixed index of refraction of 1.40. At what incident angle does the ray reflect? You can also find the angle using $n_1 \sin\theta_1 = n_2 \sin\theta_2$ where n_1 is the index inside the medium and $\theta_2 = 90^\circ$. Do your two values match?
8. Verify the reflected beam obeys the law of reflection using the protractor (Note: You may have to slightly move the source to get the program to update the reflected ray position).

6.4: Lenses

A wave that passes all the way through a piece of material with parallel sides leaves the material at the same angle that it entered. The wave un-bends when it exits the material by the same amount that it bent when entering but this is only true if the sides of the material are parallel. Convex and concave lenses have sides that are not parallel (except near the center). In this case parallel rays of light end up exiting in different directions.

6.4.1: Lens Simulation

You may have noticed in the last simulation, a wave that passes all the way through a piece of material with parallel sides (for instance light through a flat slab of glass) leaves the material at the same angle that it entered. The wave un-bends when it exits the material by the same amount that it bent when entering. This is only true if the sides of the material are parallel, however. In the simulation below we have convex and concave lenses where the sides of the glass are *not* parallel (except near the center). In this case parallel rays of light end up exiting in different directions. This is the basis for any optical device that uses lenses, for example cameras, binoculars, microscopes, glasses, eyes of various animals, etc.

At each surface the waves obey the law of refraction (Snell's law) but the result is that parallel rays that enter are not parallel when they exit. Although our example is for light it should be kept in mind that the same behavior occurs for other types of waves when they enter a medium where their speed is different (as in the previous refraction of sound in a balloon filled with Carbon Dioxide).

For this simulation we use the thin lens approximation which assumes the lens thickness is small compared to the curvature of the lens. This allows us to approximate the bending as if it occurs all at once at the middle line of the lens (instead of some bending at each surface which in fact is what happens).

The object (a candle) in the simulation can be moved using the mouse. White arrows show where the light rays actually travel. The purple lines are imaginary extensions of the real light rays. Our brain/vision system assumes light always travels in a straight line and does not bend. For light that does bend due to refraction, our brain interprets the light as following the purple paths and constructs and image based on this information. The units of height, h, and distance, d, are arbitrary (cm, inches, etc.)

Simulation Questions:

1. The definition of the focal length of a converging lens is the distance to the point where rays initially parallel to the axis meet after passing through the lens. The point is marked by a red circle called the focal point. Why is there a focal point on each side of the lens? Does it make any difference which way light travels through a thin lens?
2. Drag the object back and forth. Describe what you see. What two things are different about the image if the object is closer than the focal length, as compared to when it is further away from the focal length?
3. Use the slider to change the height of the object. How does the height of the image compare to the object height? Does the height of the object change any of your conclusions from the previous question? Explain.
4. For all cases a one ray goes straight through the center of the lens. Why is that? (Hint: Read the introduction.)
5. Carefully describe the other two rays. What happens to a ray that enters the lens parallel to the horizontal axis? What happens to a ray that goes through the focus (if the object is further away from the focus)? What happens to a ray that appears to come from the focus (if the object is closer than the focus)?
6. The previous two questions are about the rules for drawing light rays for a converging lens: 1. Rays parallel to the axis bend and go through the focus on the other side of the lens; 2. Rays going through the focus (or coming from the focus if the object is closer to the focus) bend to exit the lens parallel to the axis; and 3. Rays through the center go straight through without bending. Using these three rules, it is possible to determine where the image will be and how big it will be for any converging lens. Go back and verify these rules. Are they true?
7. Now choose the diverging lens case and experiment. How is it different from the converging case? How does the image size compare with the object size? Is there any case where the image is bigger than the object?
8. One of the rules for drawing rays for a diverging lens is the same as for a converging lens. Which one?
9. Carefully state what happens to a ray that is parallel to the axis when it exits a diverging lens. Also describe what happens to a ray that starts from the object and heads towards the focus on the opposite side. How are these rules different from the converging lens case?
10. As in the case of mirrors, some images from lenses are real (can be projected onto a screen) while others are virtual (are only seen by looking through the lens). For lenses, real images appear inverted and on the other side of the lens. Which cases above had real images and which had virtual images?
11. Your eye has a single lens which projects a real image onto your retina. The retina turns the image into nerve impulses which go to the brain to be interpreted. What is the orientation of this image? Is this surprising?

6.5: Dispersion

The speed of a wave can depend on the frequency of the wave, a phenomenon known as dispersion. Although this effect is often small. it is easy to observe with a prism.

6.5.1: Dispersion Simulation

So far in this chapter and the previous one we have assumed that the speed of a wave does not depend on its frequency or wavelength. This is generally true; for example all the sounds of the instruments in an orchestra reach your ear at the same time, no matter what frequencies they are playing. However, it is the case that under some circumstances speed can depend on the frequency of the wave, a phenomenon known as ***dispersion***. For electrical signals in a cable this means the signal gradually deteriorates in quality because high frequency components travel at a different speed compared to lower frequency components. Different colors of light travel at slightly different speeds through glass (in other words the index of refraction depends on frequency) which is how a prism separates out the different frequencies of white light.

In the previous simulations on refraction we assumed that all wavelengths bend by the same amount. But with dispersion if we start with several different wavelengths (different colors) we expect there may be some situations in which the colors will separate. If the sides of the medium are parallel, each color unbends by the same amount that it bent going into the medium so all the colors are again going in the same direction. However if the sides are not parallel, such as a prism or lens, there will be a separation of color. This is in fact how a prism and water droplets separate colors and why good camera lenses (which compensate for this effect by using compound lenses) are expensive.

Dispersion occurs for all types of waves. For example, longer wavelength surface waves on the ocean travel faster than shorter wavelength waves. There is not much dispersion for sound waves in air but acoustic waves in solids do experience significant dispersion. The simulation below is for visible light passing through a prism. You can choose the color and see what the index is for that wavelength.

Simulation Questions:

1. Use the slider at the bottom of the simulation to try different wavelengths. Which visible wavelength is bent the most? Which the least? Note that the wavelength is given in nanometers (nm).
2. What would you see on the right if the source were a white light composed of all wavelengths?
3. The speed of light is $c = 3 \times 10^8\ \mathrm{m/s}$ and the index of refraction is $n = c/v$ where v is the speed in the medium. Using the index given in the simulation for the chosen wavelength, what are the maximum and minimum speeds for colors in the visible spectrum?
4. For one of the wavelengths use the protractor to measure the incident and refracted angles (as you did in the mirror simulation) for the exiting ray on the right. Calculate the index of the prism for that color using Snell's law. Don't forget that the angles are measured from a perpendicular to the surface (you will have to correct for the fact that the prism sides are slanted at 60 degrees). What is your answer? Do you get the index of refraction shown in the simulation for that color?

6.6: Adding Wave Pulses

When two waves of the same type come together it is usually the case that their amplitudes add. So two overlapping water waves have an amplitude that is twice as high as the amplitude of the individual waves. This is called constructive interference and it can occur for sinusoidal waves as well as pulses.

6.6.1: Adding Waves

When two waves of the same type come together it is usually the case that their amplitudes add. So two overlapping water waves have an amplitude that is twice as high as the amplitude of the individual waves. In the picture below the waves arrive in phase or with a ***phase difference*** of zero (the peaks arrive at the same time). This is called ***constructive interference***.

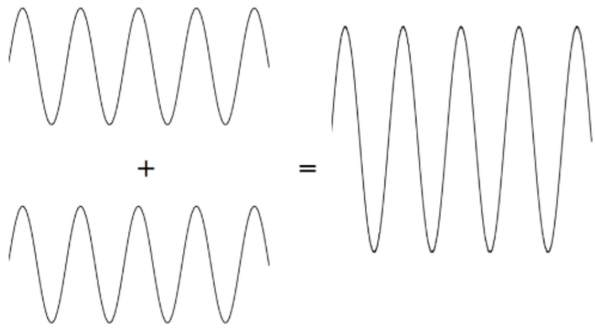


Figure $6.6.1.1$

In the picture below the waves are exactly 180 degrees out of phase so that the peak of one wave lies in the trough of a second wave. This is called ***destructive interference*** and the waves cancel. Noise canceling technology in cars and stereo headphones use this idea. Incoming noise is sampled and some electronics creates an identical sound wave except with a phase difference of 180 degrees so the two waves cancel. Destructive interference can also occur for stereo speakers if they are not wired properly. If the wires to one speaker are accidentally twisted the sounds from the speakers will be out of phase and tend to sound softer (the destructive interference isn't complete because of echoes and different distances to the listener).

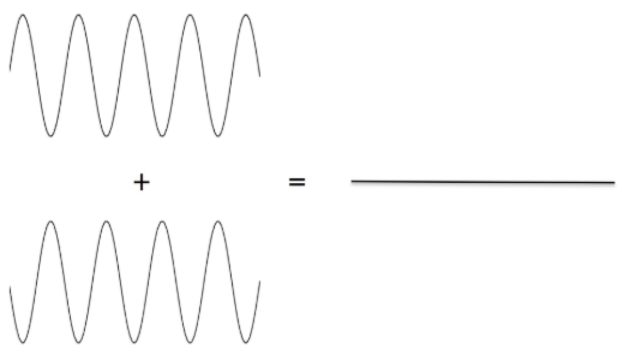


Figure $6.6.1.2$

Two other places where two waves combine to give interesting effects are standing waves and beats. When a guitar string is plucked two waves move out from the contact point in opposite directions. These waves hit the ends where the strings are tied and reflect. The result is a ***standing wave*** where the motion is up and down (you no longer see a wave moving to the left or right on the string). As we will see, standing waves determine the fundamental frequency of stringed instruments.

If two waves don't have exactly the same wavelength they may start out in phase (constructive) but gradually end up out of phase (destructive). This repeats over time and is known as the phenomena of ***beats***. Two sound waves with slightly different frequencies will have slightly different wavelengths (remember frequency times wavelength equals speed which is fixed so sound waves of different frequencies also have different wavelengths but travel at the same speed). So at times they will reinforce each other and at other times tend to cancel. What you will hear is a single note that gets louder and softer. The beat frequency (the frequency of the loudness and softness) is the difference between the two original frequencies. So a 500 Hz sound together with a 502 Hz sound will sound like a single note of 501 Hz but with a varying loudness that changes every 2 Hz. Standing waves and beats are best demonstrated in videos and simulations, below.

Video/audio examples:

- Colliding pulses I, Colliding pulses II. There are many others involving slinkys.
- Short discussion on microphone phase.

- Standing waves in water. Notice that the waves start from the far end of the tank, reflect off the near end and eventually make a set of standing waves.
- Audio demonstration of beats.
- Sound cancellation in a car muffler.
- Interesting water waves in a Japanese laboratory.
- A few other interesting applications of sound waves.

6.6.2: Adding Two Wave Pulses (Superposition)

In this simulation you see two wave pulses (red and green) and then the sum of the two pulses (blue). You can choose to have the two pulses both have an initial positive amplitude (Animation 1) or have the second pulse start with a negative pulse (Animation 2). When two waves are linear (the forces involved are Hooke's law forces) they can be added point by point at each instance of time. This is called ***superposition*** and occurs for any shape of linear wave.

Simulation Questions:

1. Step through Animation 1 until both pulses are on top of each other (a single blue pulse). What is the amplitude of the combination compared to the amplitude of the two individual pulses?
2. Now step through Animation 2 until both pulses are on top of each other (a single blue pulse). What is the amplitude of the combination compared to the amplitude of the two individual pulses? Why is the combination amplitude zero when the pulses collide?

6.7: Adding Sinusoidal Waves

Sinusoidal waves have the property, called superposition, that their amplitudes add linearly if they arrive at the same point at the same time. This gives rise to several interesting phenomena in nature which we will investigate in this and the next few simulations.

6.7.1: Adding Two Linear Waves Simulation

The waves we have been discussing so far and the ones that are most often seen in everyday life, such as light and sound, are for the most part ***linear waves***. Linear waves have the property, called ***superposition***, that their amplitudes add linearly if they arrive at the same point at the same time. This gives rise to several interesting phenomena in nature which we will investigate in this and the next few simulations.

The simulation shows the function $f(x,t)$ in red, $g(x,t)$ in the blue and $u(x,t) = f(x,t) + g(x,t)$ in grey. The check boxes on the lower right determine which functions are visible. You can enter your own functions for $f(x,t)$ and $g(x,t)$ using the same notations used for spreadsheets and calculators.

Simulation Questions:

1. Uncheck the $g(x,t)$ function so that just $f(x,t)$ is showing. What is the amplitude (half the height from highest point to lowest point) of $f(x,t)$? Now measure the amplitude of $g(x,t)$ (they should be the same). Now find the amplitude of the sum, $u(x,t)$. How does the amplitude of the sum compare with the amplitude of $f(x,t)$ or $g(x,t)$? This is an example of ***constructive interference***; the two waves add to give a wave with an amplitude which is the sum of the two amplitudes.
2. Start the simulation (button on lower left). How does the wavelength, frequency and speed of $f(x,t)$ or $g(x,t)$ compare with the wavelength, frequency and speed of $u(x,t)$?
3. Change $f(x,t)$ to have a phase of π (the simulation reads pi as π; type or cut and paste $2.0*\sin(\text{x - t} + \text{pi})$ for $f(x,t)$ and press Return or Enter). Run the simulation. What happens to the amplitude of the sum of the two waves, $u(x,t)$? This is an example of destructive interference. Write a definition for destructive interference in your own words.
4. Experiment with cases in between total destructive and total constructive interference by changing the phase of $f(x,t)$ to be $\pi/2,\ \pi/3$,and $\pi/4$. Stop the simulation each time and record the amplitude of the sum compared to the amplitude of $f(x,t)$ or $g(x,t)$.
5. Click the reset button (fourth button on lower left) and then change the amplitude of $f(x,t)$ from 2.0 to 3.0 and the amplitude of $g(x,t)$ from 2.0 to 1.0 (Hit Return or Enter to update the values). What is the amplitude of $f(x,t)+g(x,t)$ in this case? How does this amplitude compare to the original case?
6. Go back to the original functions but change one of the minus signs to a plus sign (so now $f(x,t) = 2.0*\sin(x+t)$ and $g(x,t) = 2.0*\sin(x-t)$). The sum $u(x,t)$ is called a ***standing wave*** in this case (an example would be the waves on a guitar string as we will see later). Describe the behavior of $u(x,t)$. How does the period and wavelength of the combined wave compare to the period and wavelength of two components? How is the maximum amplitude of the sum related to the amplitudes of the two components? What can you say about the speed of the sum?
7. For standing waves on a string a ***node*** is a location where there is no motion and an ***anti-node*** is a location where there is maximum motion. For the standing wave in the previous exercise, how many nodes are there? How many anti-nodes?

Advanced Questions:

1. Use trigonometric identities to show that the sum of $f(x,t) = A\sin(kx+\omega t)$ and $g(x,t) = A\sin(kx-\omega t)$ equals $2A\cos(\omega t)\sin(kx)$. We can interpret this as a time dependent amplitude, $2A\cos(\omega t)$, multiplying a sine wave which is fixed in space. What happens to the amplitude as time increases? What fixes the location of the maximums and minimums of the standing wave (Hint: $k = 2\pi/\lambda$)?
2. Notice that the standing wave has zero amplitude on both ends in the simulation. This means that only certain wavelengths will "fit" on a given length. See if you can adjust x min and x max so that you have a wave with more nodes and anti nodes that fits on a longer string with the amplitude still zero on the ends. (Hint: $6.28 = 2\pi$.)
3. Now enter the following functions: $f(x,t) = 2.0*\sin(x-t)$ and $g(x,t) = 2.0*\sin(1.1*x - 1.1*t)$ (you can cut and paste instead of typing). Watch the just the sum $u(x,t) = f(x,t) + g(x,t)$ for a while and describe what happens (it changes slowly). Two waves with slightly different frequencies added together give rise to the phenomena of ***beats***. Now turn $f(x,t)$ and $g(x,t)$ on. Are these waves still traveling at the same speed as $u(x,t)$? Find the beat frequency the following way: Stop the simulation when the two source waves exactly cancel ($f(x,t)+g(x,t)$ is a straight line) and record the time (use the step buttons if you overshoot). Start the simulation and stop it again the next time the waves cancel. Record the new time and subtract to get the elapsed time. The beat frequency is 1/(time elapsed). How does this compare to the frequency of $f(x,t)$ subtracted from the frequency of $g(x,t)$?

6.8: Path Difference

If two sources of waves are in phase to start with, when they reach a distant location they may be in-phase (leading to constructive interference) or out-of-phase (leading to destructive interference) depending on slight differences in the distance traveled. This path difference gives rise to many interesting phenomena such as interference patterns (in the case of light) and dead spots in auditoriums (in the case of sound).

6.8.1: Interference

If two sources of waves are in phase to start with, when they reach a distant location they may be in-phase (leading to constructive interference) or out-of-phase (leading to destructive interference) depending on slight differences in the distance traveled. This ***path difference*** gives rise to many interesting phenomena such as interference patterns (in the case of light) and dead spots in auditoriums (in the case of sound). The colors on a soap bubble and an oil slick are caused by path differences.

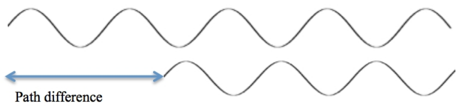


Figure 6.8.1.1

Video/audio examples:

- Two source ripple tank. You are looking down onto a tray of water where two circular waves are being formed. On a line directly in front of the waves (perpendicular to a line connecting the two sources) there is constructive interference. At angles on either side are lines of destructive interference. This is because the waves travel the same distance to reach the points straight out in front but waves arriving at the destructive points have traveled different distances.
- Simulation of double slit interference from Wolfram (you may need to download their plug in to play with this demonstration).
- CD and DVD disks also have colors that depend on the angle at which they are viewed. This is because the data is recorded as tiny pits in the disk (as we will discuss later). Light reflected from the bottom of the pits travels a different distance than light from the surface leading to constructive or destructive interference depending on the wavelength. CD diffraction. [Note: This is sometimes called diffraction which we discuss next but it is really a path difference effect.]
- DVD and Blu-Ray diffraction.
- Many bird feathers and moth wings have colors that change with the angle. So what color are they really? When you look at them under a microscope you see they actually don't have a color. Instead there are layers of feathers (or scales in the case of moths or butterflies). Light reflecting off each different layer travels a different distance back to your eye so some colors will undergo destructive interference while others will have constructive interference. The path difference is slightly different at different angles, hence the changing colors. Here is a video of some peacocks which get their colors from interference rather than pigment in the feather.
- An example of a butterfly changing colors due to angle. Why does the color depend on the angle?
- Why do soap bubbles have colors? Light reflects off the inside and outside surfaces of the soap film that makes up the bubble. The light off the inside surface has to travel slightly further to get to your eyes than light reflecting off the outside surface. Different colors of light have different wavelengths. The path difference can cause destructive interference for some wavelengths while causing constructive interference for other colors. Here is a video of soap bubbles.
- Mini Lab on Interference.

6.8.2: Interference Due to Path Difference Simulation

In this simulation the top two waves are identical but may be set to start at different locations. The bottom graph shows the sum of the two waves. Depending on the path difference the two waves may end up exactly in phase (leading to constructive interference), exactly out of phase (destructive interference) or something in between. The step thickness, D, controlled by the slider on the lower left, determines the distance between the starting locations of the two waves.

Simulation Questions:

1. Start the simulation with the step thickness, D, equal to zero. Are the waves in phase? Slowly increase D until the waves are exactly out of phase. What step thickness causes this?
2. Reset the simulation and slowly increase D to find the first three thicknesses that cause destructive interference. Verify that destructive interference occurs at step thicknesses given by $1/2\lambda,\ 3/2\lambda$,and $5/2\lambda$ (find the wavelength by subtracting the x-location of two successive peaks of the wave).
3. Reset the simulation and enter 1.57 for D (this is the case of half a wavelength path difference so the waves cancel). Increase the wave number, $k=2\pi/\lambda$, until you find the next wavelength that experiences destructive interference (don't change D). What is the wave vector and wavelength of this wave?
4. For light, changing the wavelength changes the color. Can the same step thickness cause destructive interference for all colors? Explain.
5. The general formula for ***destructive interference due to a path difference*** is given by $\delta=(m+1/2)\lambda/n$ where n is the index of refraction of the medium in which the wave is traveling, λ is the wavelength, δ is the path difference and $m=0,1,2,3,\ldots$ What can you say about the various choices of m in this equation; what physical cases do they represent (assume $n=1$ for now)?
6. This simulation starts the two waves at different locations but path differences can also occur due to reflection from a surface that has multiple layers. Imagine what would happen if this simulation represented ***monochromatic*** (single wavelength) light reflecting off a surface with two levels (waves come in from the right and reflect back to the right). In this case the path difference would be *twice* the depth of the step so $\delta=2\times D$. How would this change the results? In which case would there be no reflection from the surface? In which case would there be constructive interference?
7. For a *reflected wave* $\delta=2\times D$ is the actual path difference; a wave reflected off the surface in the top panel must travel an extra distance equal to twice the step thickness to catch up with a wave reflected off the surface in the second panel. Reset the simulation, change the step thickness, D, to find a case of destructive interference. Now click the check box to simulate the case of reflected waves (instead of two waves starting from the left). What do you notice about the combined waves in the case of reflection?
8. A music CD has information stored on it in the form of tiny divots blasted into the surface with a laser. Suppose you see constructive interference for red light (wavelength of $650\ \mathrm{nm}$). What is the minimum ($m=0$) depth of the divots? (Hint: The path difference is twice the divot depth.)
9. Explain why you only see one particular color when looking at a small region of a CD at a fixed angle. What happens to the other colors?
10. If you look at the colors being reflected from a CD you will notice that the color changes depending on the angle. How does the path difference change as you look at the divots at different angles? (Hint: Imagine the waves in the simulation coming in at different angles instead of horizontally. Now the path difference is the hypotenuse of a triangle, one side of which is the Step height.)
11. Suppose you wanted to make a "stealth" jet plane which was non-reflective to a particular wavelength of radar. Describe one way you might try to do this by modifying the surface of the plane.
12. Some insect wings and the feathers of some birds (for example peacocks) exhibit a feature known as ***iridescence***. From a fixed angle only one color of reflected light can be seen. Explain this phenomena given the fact that insect wings and feathers consist of overlapping layers causing the surface to be multi-layered.
13. Soap bubbles show different colors at different places on the bubble. So do oil slicks on water. In both cases light reflects off the upper and lower surfaces of the layer of soap or oil. Explain the different colors in terms of path difference (Hint: draw a picture where the wall of the soap bubble is nearly the same thickness as one wavelength and explain why the path difference is twice the thickness of the soap).
14. There are two other details needed to explain the soap bubble and oil slick color phenomena completely. While the light is inside the soap or oil it travels at a different speed so the wavelength is different. This is why the index of refraction, n, is

included in the formula $\delta=(m+1/2)\lambda/n$. For the case of $m=1$ in the formula, what would happen to a wavelength which reflected from a thickness with constructive interference for an index equal to one if instead the index was equal to 1.5?

15. The second detail for light reflecting off a soap bubble or oil slick is light reflecting from the top surface is going from a "soft" medium (air) to a "stiff" medium (soap) but the light reflecting from the bottom layer of the soap is going from a "stiff" medium (soap) to a "soft" medium (air inside the bubble). This causes a phase change of 180° at the top surface but not at the bottom. If the path difference for a particular thickness of soap film was just right for destructive interference but there was a 180° phase change for the top reflected wave but not the bottom reflected wave, what would happen to that color?

6.9: Interference

This simulation shows a top view of waves interfering on the surface of a tank of water (imagine tapping the surface of a pond with the end of a stick at regular intervals). The white circles coming from the spot represents the wave crests with troughs in between. Two sources can be seen at the same time and the separation between them and the wavelength of both can be adjusted.

6.9.1: Interference- Ripple Tank Simulation

This simulation shows a top view of a source making waves on the surface of a tank of water (imagine tapping the surface of a pond with the end of a stick at regular intervals). The white circles coming from the spot represents the wave crests with troughs in between. Two sources can be seen at the same time and the separation between them and the wavelength of both can be adjusted. The simulation can also be imagined to represent two sound sources (for example two speakers), each producing exactly the same wave. The wavelength, λ, and distance between sources, d, are in the same arbitrary distance units (meters, cm, μm, etc.). Speed is a unitless parameter that controls the rate at which the simulation is refreshed.

Simulation Questions:

1. After looking at the left and right waves to verify they are the same, click on 'Both' to see them together. Waves from each source will cancel in some places (destructive interference) but add in other places (constructive interference). How many lines of constructive interference do you see?
2. With two sources change the wavelength of the sources. How does the number of destructive interference lines change with wavelength? Write a statement about the relationship between wavelength and the number of interference lines.
3. With two sources and a wavelength of 2.0, change the separation between the sources. How does the number of destructive interference lines change with separation? Write a statement about the relationship between source separation and the number of interference lines.
4. Reset the simulation to the original values and click 'Both'. Suppose instead of a ripple tank this simulation represented two light sources (which have the same wavelength and start off in phase- for example laser light from a single source shining through two small openings). The light starts at the middle of the simulation and reaches a screen at the top edge of the simulation. How many bright spots would be seen on the screen in the simulation for the this case?
5. If the simulation represented a double slit light source, changing the wavelength would be equivalent to changing the color. Describe the difference in the location of the bright spots for the color represented by wavelength equal to 1 compared to the location of the spots for the color represented by wavelength equal to 4. Do they occur in the same location on the screen (at the top edge of the simulation)?
6. What would be the result on the screen of shining light which was a mixture of two colors through a double slit?

Advanced Questions:

The formula for double slit interference is given by $d\sin\theta_{bright} = m\lambda$ where $m = 0, \pm 1, \pm 2, \pm 3 \ldots$ for the case of constructive interference. For destructive interference $d\sin\theta_{dark} = (m + 1/2)\lambda$ where $m = 0, \pm 1, \pm 2, \pm 3 \ldots$ In both cases d is the distance between the center of the openings (the separation of the sources), the angle θ is the angle from the central maximum out to a minimum or maximum and m numbers the maximums (or minimums) starting from the center ($m = 0$).

Note

In order to actually measure this effect for light the slits must be similar to the wavelength of light; in other words, very small and close together with the screen quite a distance away. The light must also be coherent (have the same phase as is the case for laser light).

1. For 600 nm light and a separation of 0.01 mm, what is the angle (in radians) to the first maximum? What is the color of this light?
2. For the previous question, how far away would the screen have to be in order to have a 2 mm separation between the first and second maximum?

6.10: Diffraction

Sometimes waves don't travel in a straight line, even if their speed does not change (as in the case of refraction). For example, you can hear the conversation in the next room even though you cannot see the source. This is because sound waves undergo diffraction, bending as they go through the doorway between the two rooms.

6.10.1: Diffraction

Sometimes waves don't travel in a straight line, even if their speed does not change (as in the case of refraction). For example, you can hear the conversation in the next room even though you cannot see the source. This is because sound waves undergo ***diffraction***, bending and spreading as they go through the doorway between the two rooms. Diffraction only occurs when the wavelength is close to the size of the opening or object. In the picture below Jack can hear Jill talking although they cannot see each other.

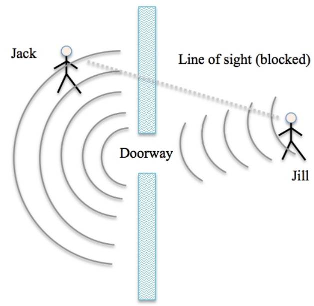


Figure 6.10.1.1

Why does sound diffract in the picture above but light does not? Sound wavelengths are typically between a few meters down to a few centimeters so they are close to the size of the doorway. Light waves have much smaller wavelengths, on the order of a few hundred nanometers (10^{-9} m). We only notice diffraction when the opening or object is close to the size of the wavelength, so to see diffraction of light it needs to pass through a much smaller opening than a doorway.

Scattering is a similar phenomenon that occurs when a wave interacts with an object that has a resonance frequency the same as the wave frequency. The wave is first absorbed and then re-emitted in all directions (or sometimes perpendicular to the incident direction). The sky is blue because clusters of nitrogen and oxygen molecules (which make up most of the atmosphere) have resonances at the same frequency of violet light. Violet and a little blue light is scattered but since our eyes are not as sensitive to violet we see the blue. The other colors pass through. The sun looks a little more yellow than it really is because the violet/blue part of the spectrum has been removed (scattered out in other directions). Likewise sunsets are orange because when the sun is on the horizon the path the light travels to reach us passes through more atmosphere and even more violet/blue is removed.

Video/audio examples:

- Ripple tank diffraction. Here water waves travel through an opening about the same size as the wavelength and change their direction.
- Ripple tank diffraction. You are looking down onto the surface of a tray of water. Notice that the plane waves on the right spread out into a circle on the left after passing through the small opening.
- Red laser diffraction. A red laser beam is shone through several different small openings. The first is a square opening, the second a hexagonal opening. Then the laser is shone through single openings of different sizes. Finally the laser is shone through a series of double slits. Why is the light pattern complicated instead of a simple spot? What is the difference in the light pattern between the single slits and the double slits?
- A diffraction grating is a piece of glass or plastic with a series of very small grooves, each of which acts like a slit. Since different colors diffract by different amounts, white light seen through a diffraction grating will spread out into its component colors as shown in this YouTube of incandescent and florescent diffraction.
- Mini Lab on Diffraction.

6.10.2: Diffraction Simulation

The simulation shows what happens to a planewave light source (below the simulation, not shown) as it passes through an opening. The wavelength (color for light, pitch for sound) of the waves and the size of the opening, a, are in the same arbitrary units (meters, cm, μm, etc.) and can be adjusted. The waves in the simulation represent light, sound or any other type of linear wave.

Simulation Questions:

1. Leave the opening width fixed and experiment with the wavelength of the waves. Describe what you see.
2. You should have noticed that for a long wavelength the opening basically becomes a point source of waves. The waves on the other side of the opening move outward in all directions. But once the wavelength is much smaller than the opening the waves do not spread as much and appear to be more like plane wave all headed in the same direction. Light, with it's very small wavelength, passes through a doorway without bending because the door is much larger than the wavelength. Sound, however, is a wave with wavelengths close to the size of the opening of a door. Explain why we can hear noise through a doorway to another room even though the source (a person, radio, TV. etc.) is not in our direct line of sight.
3. All optical instruments (telescopes, microscopes, even radio telescopes which look at radio waves instead of light waves) have openings to allow light in. This means diffraction will be a problem for that instrument for some sizes of waves. If you want to reduce the effects of diffraction for a particular instrument, would you want to try to use longer or shorter wavelengths? (Hint: Electron microscopes can provide much higher magnification because electron waves can be much smaller than light.)
4. Reset the simulation and leave the wavelength fixed while changing the size of the opening. Describe what you see. How does the opening size affect the diffraction pattern?

Advanced Questions:

Diffraction can also be explained as a type of interference resulting from a path difference from multiple sources. Recall in the ripple tank simulation of two sources waves from the source on the left must travel a longer path to get to a point at the top right of the simulation than waves from the source on the right. This path difference changes depending on how far to the right we look resulting in spots of destructive and constructive interference along the top. For a single opening instead of two separate sources we can imagine a row of many sources filling up the single opening. Again there will be a path difference from the different sources but the pattern will look different because there are now many sources lined up next to each other.

1. The formula for the location of destructive interference in the case of single slit diffraction is given by $\sin\theta_{dark} = m\lambda/a$ where a is the opening size and $m = 0, \pm1, \pm2, \pm3\ldots$ where a is the width of the opening and θ is the angle to each successive dark spot, labeled with the number m.
2. For 600 nm light and an opening of 0.01 mm, what is the angle (in radians) to the first minimum?
3. For the previous question, how far away would the screen have to be in order to have a 2 mm separation between the central maximum ($\theta = 0$) and the first minimum?
4. Red light has a longer wavelength than green light. Which color bends the least when going through a small opening, red or green?
5. What would be the result of shining white light through a small opening?

CHAPTER OVERVIEW

7: Pitch, Loudness and Timbre

The mechanism of human hearing does not operate as a perfect scientific instrument. In this chapter we relate a few *subjective* measurements of sound (things people report after hearing a sound) to *objective*, scientific measurements (measurements made in a laboratory using scientific instruments). The three subjective quantities of pitch, loudness and timbre are related to laboratory measurements of a sound wave's fundamental frequency, amplitude and waveform, respectively.

Key Terms:

Pitch, fundamental frequency, $v = f\lambda$, loudness, sound intensity (in $\mathrm{W/m^2}$), sound intensity level (SIL in dB), decibels (dB), inverse square law, just noticeable difference (loudness and frequency), timbre.

7.1: Pitch, Loudness and Timbre

The mechanism of human hearing does not operate as a perfect scientific instrument. In this chapter we relate a few subjective measurements of sound (things people report after hearing a sound) to objective, scientific measurements (measurements made in a laboratory using scientific instruments). The three subjective quantities of pitch, loudness and timbre are related to laboratory measurements of a sound wave's fundamental frequency, amplitude and waveform, respectively.

7.1.1: Pitch

The main component that gives us the perception of the pitch of a musical note is the ***fundamental frequency***, measured in hertz. In the modern musical scales used today the piano note middle C has a frequency of $261.63\ \mathrm{Hz}$ (we will look at how scales are constructed a bit later). This chart has the frequencies of all musical notes, based on a frequency of $440\ \mathrm{Hz}$ for the note labeled A_4 (A above middle C on the piano). As we will see in the next chapter, musical sounds are usually composed of many frequencies but the fundamental frequency gives us the basic quality we perceive as pitch.

Recall that for both sound and light, frequency times wavelength equals speed ($v = f\lambda$) but the speeds of light and sound are very different ($v = 3 \times 10^8\ \mathrm{m/s}$ for light $v = 344\ \mathrm{m/s}$ for sound). We can use the equation to find that a $440\ \mathrm{Hz}$ sound wave has a wavelength of $0.78\ \mathrm{m}$. In the case of light, frequency tells us the color of light. Green light for example lies in the frequency range $525\ \mathrm{THz}$ to $575\ \mathrm{THz}$ (T is tera or 10^{12}). A $525\ \mathrm{THz}$ electromagnetic signal has a wavelength of $571\ \mathrm{nm}$ (n is nano $= 10^{-9}$). The previous chart also has the wavelengths of notes on the musical scale in centimeters.

A person whose hearing is not damaged can hear frequencies as low as $20\ \mathrm{Hz}$ and as high as $20,000\ \mathrm{Hz}$. But very few people today can hear this range of frequencies. Exposure to normal sounds in everyday modern life tends to do at least some damage to most peoples hearing at an early age. Hearing is also affected by normal aging.

Video/audio examples:

- Article by Peter L. Tyack in *Physics Today* about Human-generated sound and marine mammals.
- Christian Huygens in 1693 noticed that the fountain at Chantilly, France produced an audible pitch. He determined this was the result of the echo of sound from the fountain being reflected off of a set of nearby steps. The steps were about half a meter in depth, causing sound to return from each subsequent step with a time delay (period) of 1m divided by $340\ \mathrm{m/s}$ ($vt = d$). A set of pulses with this period will have a frequency of $340\ \mathrm{Hz}$. This is sometimes called a ***repetition pitch***. He was one of the first scientists to connect sound with the frequency of a wave.
- Sounds reflected from the steps of the Mayan ruins at Chichenitza also produce a pitch, similar to Huygens' fountain but the pitch changes over time, due to the height of the steps. Sound from the bottom of the stairs has a repetition pitch depending on just the depth of the first step. Sound reflecting off the upper steps has a different repetition pitch because of the angle (the sound travels along a hypotenuse connecting the edge of one step to the next, rather than the shorter distance from the edge directly to the back of the step). This lower repetition pitch also takes longer to return because of the further distance to the higher steps. The result turns a hand-clap into a chirp: Sounds.

7.1.2: Loudness

Generally the loudness of a sound is related to the amplitude of the sound wave; a wave with bigger variations in pressure generally sounds louder. For any type of wave the energy carried by the wave is proportional to amplitude squared. This means doubling the amplitude increases the power by a factor of four (two squared). But the amount of energy reaching your ear also depends on the frequency since a wave with more oscillations per second (higher frequency) will mean the same amplitude hits your eardrum more often. ***Sound intensity*** is defined to be the energy per second (power in watts) reaching a given area (measured in square meters). Normal conversation has an intensity of about $10^{-6}\ \mathrm{W/m^2}$.

Unobstructed sound from a small sound source spreads out in all directions in an expanding spherical shape. Because the energy is spread over a larger and larger area as time goes on, the intensity decreases as you move further from the source. The area of a sphere is given by $4\pi r^2$ so the decrease in intensity is proportional to r^2 where r is the distance from the source. This is known as an ***inverse square law*** and many other laws in physics follow this same law (diagram of the inverse square law). For example the gravitational field of the earth (or any other object) decreases as you move away from it proportional to an inverse square law. So does the electric field around an electron or proton. For practical purposes, what this means is doubling the distance decreases the strength by $1/2^2 = 1/4$. If you move three times as far away the strength is $1/3^2$ or $1/9\mathrm{th}$ as much. Shortening the distance by half means the intensity will be four times as much.

7.1.3: The Decibel Scale

The human ear is an amazing instrument that can detect intensities as low as $10^{-12}\ \mathrm{W/m^2}$ and can hear intensities as high as $10^3\ \mathrm{W/m^2}$ (although this is loud enough to cause damage to the ear). To make this huge range easier to write down, a second scale of loudness was created called the ***sound intensity level***, measured in *decibels*. The relationship between sound intensity, I measured in watts per meter squared and sound intensity level (SIL) measured in decibels (dB), is given by $SIL = 10\log(I/I_o)$. Here log is the logarithm and $I_o = 10^{-12}\ \mathrm{W/m^2}$ is a reference sound intensity at about the threshold of human hearing.

We know sound waves are longitudinal fluctuations in the pressure of the air (or other medium) in which the sound is traveling. This is a fluctuation above and below normal atmospheric pressure and is measured in pascals (Pa), the standard unit of pressure. If you tried to report an average value of the fluctuation you would get zero since the pressure varies from above to below normal atmospheric pressure. We could just give the (maximum) amplitude of the fluctuation but what is usually done is to report the ***rms*** value which is the square root of the average of the squares of the amplitude. By first squaring the amplitude and then taking an average and then a square root you have a kind of average value that is not zero. In the chart below the root mean square (rms) variation in pressure from normal atmospheric pressure (in pascals Pa), the sound intensity level (SIL in dB) and the intensity (I in $\mathrm{W/m^2}$) are given for several sounds.

Source	**Pressure** rms (Pa)	**Sound Intensity Level** SIL (dB)	**Intensity** ($\mathrm{W/m^2}$)
Jet engine at 10 m		150	10^3
Jet engine	200	140	100
Jack hammer	60	130	10
Car horn	20	120 (pain threshold)	1
Rock band	6	110	0.1
Machine shop	2	100	0.01
Train	0.6	90	10^{-3}
Vacuum cleaner	0.2	80	10^{-4}
TV	0.06	70	10^{-5}
Conversation	0.02	60	10^{-6}
Office	0.006	50	10^{-7}
Library	0.002	40	10^{-8}
Hospital	0.0006	30	10^{-9}
Broadcast studio	0.0002	20	10^{-10}
Rustle of leaves	0.00006	10	10^{-11}
Threshold of hearing	0.00002	0	10^{-12}

Table 7.1.3.1

Here are a few examples and rules of thumb for converting intensity ($\mathrm{W/m^2}$) into intensity levels (in dB):

- A 10 fold increase in intensity equals an addition of 10 dB. So going from a car horn to a jackhammer *multiplies* the intensity by 10 (1 $\mathrm{W/m^2}$ to 10 $\mathrm{W/m^2}$) but *adds* 10 dB to the intensity level (120 dB to 130 dB).
- A two fold increase in intensity (twice as loud in $\mathrm{W/m^2}$) equals and addition of 3 dB to the SIL. Suppose one trombone produces a sound level of 40 dB. How loud are four trombones? Doubling the number of trombones to two adds 3 dB, doubling again to four adds 3 dB more so the new sound level is 46 dB.

- Suppose the sound intensity is $100\ \mathrm{W/m^2}$. What is the sound level? $\mathrm{SIL} = 10\log(100/10^{-12}) = 10\log(10^{14}) = 10*14 = 140\ \mathrm{dB}$.
- Suppose the sound level is $110\ \mathrm{dB}$. What is the sound intensity? $110\ \mathrm{dB} = 10\log(I/10^{-12})$. Divide both sides by 10 to get $11 = \log(I/10^{-12})$. Now take inverse $\log\ 11$ (same as 10^{11}) to get $10^{11} = I/10^{-12}$. Multiply both sides by 10^{-12} to get $0.1\ \mathrm{W/m^2} = I$.
- A Sound Conversion web site that converts between sound level, sound pressure and sound intensity.

Both sound intensity ($\mathrm{W/m^2}$) and sound intensity level (SIL) are numbers that can be measured precisely in the laboratory (objective). The human ear, however, is an imperfect measuring instrument. We hear better at a mid-range of frequencies than we do at very low or very high frequencies. The ***Loudness Level*** (L_L) scale is a *subjective* measurement of loudness. This scale is arrived at by asking real humans to compare the loudness of different notes and an average is taken for many people (subjective). The units of the Loudness Level are the *phon*.

The diagram below (modified from an MIT OpenCourseWare graph) relates sound intensity level (SIL, measured in dB with laboratory instruments), intensity (measured in $\mathrm{W/m^2}$ with laboratory instruments) and Loudness Level (measured in phons). A SIL of $110\ \mathrm{dB}$ is considered painful while a SIL of 0 is at the threshold of hearing. If our ears were the same as laboratory instruments the lines would go straight across. The L_L scale and the SIL scale do give approximately the same number only for frequencies around $1000\ \mathrm{Hz}$. In other words our subjective perception of loudness (in phons) and the laboratory measurement (in dB) agree but only for sounds with a frequency of $1000\ \mathrm{Hz}$.

Notice there is a dip in all the curves between $1000\ \mathrm{Hz}$ and $5000\ \mathrm{Hz}$ indicating we are more sensitive to these frequencies and this is true for all loudness readings. For example suppose we perceive a sound at $4000\ \mathrm{Hz}$ to be 45 phons (labeled by a blue X in the diagram). The chart shows that at this loudness and frequency the dB reading in the laboratory is actually around $36\ \mathrm{dB}$ (dotted line to the SIL axis). So we perceive a sound of $36\ \mathrm{dB}$ (measured in the lab) as being much louder (45 phons) if it occurs at $4000\ \mathrm{Hz}$. Most other animals have a similar curve allowing them to hear better in a certain frequency range although the dip is usually much more narrow and does not go as low as for humans. Many animals can hear frequencies above and below what humans can hear, but we can hear much softer sounds in this range of frequencies than almost any other animal, owls being an exception. Your cat or dog can hear higher frequencies than you can but you can hear softer sounds in the $1000\ \mathrm{Hz}$ to $5000\ \mathrm{Hz}$ range.

This improved ability to hear softer sounds in the $1000\ \mathrm{Hz}$ to $5000\ \mathrm{Hz}$ range is not surprising once you realize these are important frequencies for human speech; our hearing mechanism is built to hear human voices better than sound with much higher or much lower frequencies. This greater sensitivity around $3500\ \mathrm{Hz}$ is due to the tube resonance of the auditory canal (see chapter 12 for tube resonance and chapter 10 for a picture of the auditory canal).

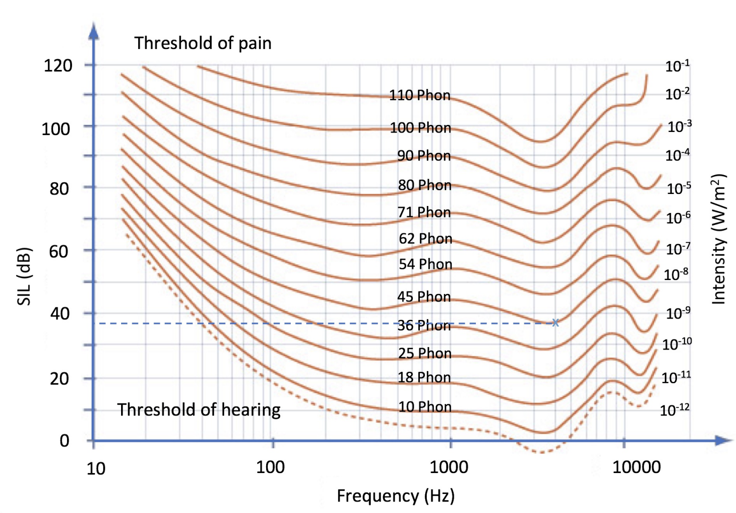


Figure 7.1.3.1

It is also the case that intensity has an effect on perceived frequency; the same laboratory frequency will appear to be a slightly different frequency if the intensity is different. High frequencies are perceived to be a slightly higher pitch than normal if they are very loud. Low frequencies are perceived to be slightly lower than expected if they are very loud. Medium loudness doesn't change the perceived pitch very much.

The above curves are very much like the ***frequency response curves*** of microphones and speakers. No microphone has the same sensitivity to all frequencies and no speaker reproduces all frequencies equally well, as we will see in Chapter 18 on electronics. Likewise our hearing does not have the same sensitivity at all frequencies.

7.1.4: Just Noticeable Difference

Two other differences in human hearing as compared to laboratory measurements are ***Just Noticeable Difference in frequency*** ($\mathrm{JND\ Hz}$) and the ***Just Noticeable Difference in loudness*** ($\mathrm{JND\ dB}$). If a group of people are asked to decide if two frequencies are the same or slightly different most people can tell if the frequency is different by $1\ \mathrm{Hz}$ for low frequency sounds. So the $\mathrm{JND}\ (\mathrm{Hz})$ for a $500\ \mathrm{Hz}$ sound is about $1\ \mathrm{Hz}$; most of us can tell the difference between $500\ \mathrm{Hz}$ and $501\ \mathrm{Hz}$. At frequencies above $2000\ \mathrm{Hz}$ however, most people start having more difficulty telling two frequencies apart. For example at $4000\ \mathrm{Hz}$ the $\mathrm{JND}\ (\mathrm{Hz})$ is about $8\ \mathrm{Hz}$ meaning that the two frequencies must be about $8\ \mathrm{Hz}$ apart before the notes sound different. We can't distinguish a $4000\ \mathrm{Hz}$ pitch from a $4001\ \mathrm{Hz}$ or even a $4007\ \mathrm{Hz}$ pitch. Probably for this reason no musical instrument produces fundamental frequencies above $5000\ \mathrm{Hz}$; we wouldn't be able to tell if the instrument was in tune.

If a group of people are asked to decide if two tones are or are not the same loudness, it turns out that the majority of them will make different decisions depending on the frequency of the note and the initial loudness. It is easier to tell if two sounds are the same loudness when they are both very loud. For example most people can tell if the SIL level changes by $0.5\ \mathrm{dB}$ when the sound is at $80\ \mathrm{dB}$ but need a change of $1.5\ \mathrm{dB}$ to detect a difference if the sound is at $40\ \mathrm{dB}$ to start with. There is also a slight difference in the perception of loudness differences at different frequencies, which is not surprising given the difference in perception at different frequencies (the phon scale, above).

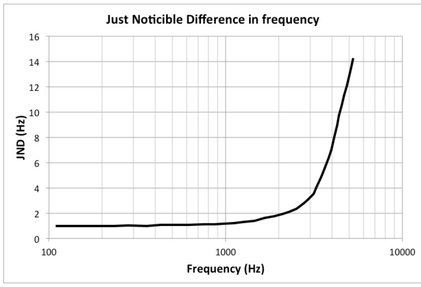


Figure 7.1.4.1

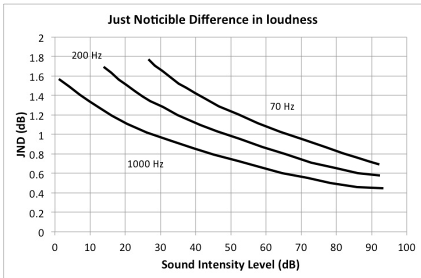


Figure 7.1.4.2

Video/audio examples:

- An online test for JND in frequency. Take the test. Record your answers (we will compare everyone's response in class). What did you find out about your own Just Noticeable Difference in frequency?

7.1.5: Timbre (the first time)

If a trumpet and a clarinet play the same note we can still tell the difference between the two instruments. Likewise, different voices sound different even when singing the same note. Why? We now know that if they are playing or singing the same pitch the fundamental frequency is the same for both so it is not the pitch that enables us to tell the difference. These differences in the quality of the pitch are called ***timbre*** and depend on the actual shape of the wave which in turn depends on the other frequencies present and their phases. Pure tones such as from a tuning fork have a pure sine wave shape and a single frequency. However the notes from musical instruments and voices are more complex and normally contain many frequencies, as we will see in the next chapter. We will also come back to other aspects of the human perception of sounds in Chapter 10 on Perception. For now the main point is that the subjective perception of pitch, loudness and timbre are each related to more than one quantity that can be measured in the laboratory. The following diagram shows some of the connections between objective (laboratory) measurements and subjective perception.

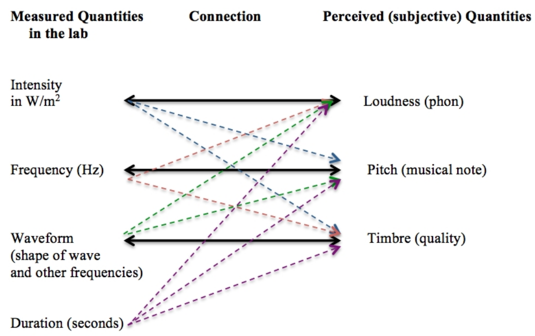


Figure $7.1.5.1$

Notice that our perception of loudness is mainly determined by the intensity of the sound (energy per second per square meter) but also is influenced by frequency and waveform of the sound. Likewise our perception of pitch is mainly determined by the fundamental frequency but also influenced by intensity and waveform. Finally, timbre is determined by waveform (which is determined by the other frequencies present and their phases) with influences from intensity and the fundamental frequency. As we will see later (in a demo in class) the duration of a sound also affects how we perceive its pitch, loudness and timbre.

Summary

Pitch is primarily determined by the fundamental frequency of a note. Perceived loudness is related to the intensity or energy per time per area arriving at the ear. Timbre is the quality of a musical note and is related to the other frequencies present. Laboratory instruments measure the fundamental frequency in Hz and sound intensity in $\mathrm{W/m^2}$ of a sound wave as independent properties. As we will see we can also measure the other frequencies present which determines the waveform. Our hearing mechanisms, on the other hand, perceive the subjective qualities of timbre, pitch and loudness of a musical note. The objectively measured quantities are related to the subjective perceptions but the relationship is not precise. For example we perceive loudness differently for different frequencies. Our ears are better at distinguishing differences in frequency ($\mathrm{JND\ Hz}$) at low frequencies than high. And we distinguish loudness differences better for loud sounds ($\mathrm{JND\ dB}$). As we will see there are several other interesting features of our hearing system that make the perception of sound different from measurements made in the lab.

Questions on Pitch, Loudness, Timbre:

1. What does the fundamental frequency of an instrument determine?
2. What is the difference between pitch and fundamental frequency?
3. What subjective quantity corresponds to the scientific measurement of the fundamental frequency?
4. What is the typical range of frequencies that a human can hear if they have perfect hearing?
5. Although human hearing has a theoretical range, this isn't always the case in practice. Why is that?
6. How are the frequency ranges of animals different from the range humans can hear? (Look at the chart in Chapter 10.)
7. What is ultrasound?
8. How do we measure sound intensity in the laboratory?
9. What subjective quantity corresponds to the scientific measurement of sound intensity?

10. What is the difference between sound intensity and sound intensity level? What units are used for each?
11. The energy per second per area in $\mathrm{W/m^2}$ is one way to measure sound intensity. What other scale is used?
12. What is the threshold for pain in terms of sound intensity?
13. What is the difference between the phon measurement and the decibel measurement?
14. Suppose one clarinet has a measured loudness of $30\ \mathrm{dB}$. How loud will five identical clarinets together be?
15. What is the sound intensity level in dB for a sound with intensity of $1\ \mathrm{W/m^2}$?
16. A vacuum cleaner is about 100 times as loud as ordinary conversation as measured in $\mathrm{W/m^2}$. How much of a difference is this in dB?
17. How loud are rustling leaves in dB? In $\mathrm{W/m^2}$?
18. In the graph of phons compared to SIL and sound intensity at different frequencies, why does the phon curve take a dip between $3000\ \mathrm{Hz}$ and $5000\ \mathrm{Hz}$?
19. What is just noticeable difference in frequency?
20. According to the chart in this chapter, what is the normal JND (in Hz) at $1000\ \mathrm{Hz}$? At $4000\ \mathrm{Hz}$?
21. What is just noticeable difference in loudness?
22. For a $1000\ \mathrm{Hz}$ sound, what is the normal JND (in dB) at $20\ \mathrm{dB}$? At $50\ \mathrm{dB}$?
23. What is timbre and what causes it?
24. Pitch is mainly related to the fundamental frequency. But what other factors affect perceived pitch (what else affects the pitch that we think we hear)?
25. Loudness is mainly related to the sound intensity. But what other factors affect perceived loudness (what else affects the loudness that we think we hear)?
26. Explain the difference between subjective measures of pitch, loudness and timbre as compared to the objective measurements of fundamental frequency, sound intensity and waveform.

CHAPTER OVERVIEW

8: Fourier Series

In the previous chapter we hinted that timbre is determined by the waveform or shape of the wave. So far we have only looked at waves that can be described by the mathematical functions of sine and cosine. How are differently shaped waves related to simple sine or cosine waves? What gives different waveforms different shapes?

Key Terms:

Fourier's theorem, Fourier analysis, Fourier synthesis, synthesizer, Fourier series, Fourier spectrum, fundamental frequency, uncertainty principle, harmonic, overtone.

8.1: Sound Texture

This simulation explores the aural texture of four basic periodic waveforms: sine, triangle, square, and sawtooth. The sine waveform has a single frequency and is the building block of other periodic waves by summing harmonics in a Fourier Series as we will see in the next section. The richness of the sound is called the timbre (defined in the previous chapter) and is determined by the amplitude of the harmonics in the Fourier sum.

8.1.1: Wave Shape

An ***oscilloscope*** is an electronic device that allows us to capture the shape of a sound wave coming from a microphone. The microphone turns the sound waves into an analog electric current which varies in voltage the same way the sound wave does in pressure. The oscilloscope shows the voltage on a screen. Although the sound wave is longitudinal (back and forth) the voltage is plotted vertically (transverse). The variations in voltage are actually too fast to see if they were plotted in real time. However, if the variations are periodic (repeat over time) the oscilloscope can make these variations visible by repeatedly plotting the same variation over and over on the screen. Some oscilloscopes can also capture a short, rapidly changing sound by taking a time snapshot of the sound. Note that oscilloscope pictures are *time graphs* of the sound wave; the oscilloscope does not show the space picture of the wave.

Video/audio examples:

- Oscilloscope used to capture sound waves from a whistle.
- Oscilloscope showing sound waves changing in amplitude and then frequency.
- Oscilloscope showing different frequencies of a quitar.
- There are also smart phone apps that will show the shape of a sound wave in real time (for example SignalScope or oScope for the iPhone or SpecScope for Android).
- There are other, more artistic ways to visualize sound as shown in this sound waves art exhibit. In this exhibit sound waves drive the motion of tiny styrofoam balls. Why do they form into the shapes that you see?

8.1.2: Sound Waveforms Simulation

This simulation explores the aural texture of four basic periodic waveforms: sine, triangle, square, and sawtooth. The sine waveform has a single frequency and is the building block of other periodic waves by summing harmonics in a Fourier Series as we will see in the next section. The richness of the sound is called the timbre (defined in the previous chapter) and is determined by the amplitude of the harmonics in the Fourier sum.

Click on the waveform image to hear the difference types of sounds these waveforms produce and drag up and down within the image to change the waveform frequency. Drag left and right to adjust the volume. The sine waveform produces the smoothest sound because it consists of a single fundamental frequency F_o. The triangle waveform has a richer and higher timbre because the dominant frequency F_o is joined by the odd harmonics $3F_o, 5F_o, 7F_o$, etc. The square waveform also has only odd harmonics but it sounds higher than the triangle waveform because the amplitude of these harmonics is greater than for a triangle waveform. In other words, the triangle waveform more closely matches a sine waveform than the square waveform. The sawtooth waveform has the most complex timbre because every harmonic is present.

The Sound Waveforms JavaScript Model uses the HTML 5 Web Audio API. This API is still under development and may not be supported on all platforms. Press the Reset button to reinitialize the simulation if the sound does not play when the simulations is first loaded.

Note

Press Reset if sound does not play when the simulation first loads.

Simulation Questions:

When two waves have the same fundamental frequency they have the same pitch, according to our ears. If there are other frequencies present (overtones) the pitch stays the same but the timbre is different. Note: Computer speakers do not play very low or very high frequencies accurately; this simulation works better if you use better speakers or headphones.

1. Listen to each wave by clicking on the appropriate picture without changing the fundamental frequency. Describe how each one sounds (what is the timbre of the sound for each one?).
2. Can you distinguish each waveform from the others even when the fundamental frequency is the same?
3. Now drag the mouse cursor up and down on one picture to change the fundamental frequency (or change the number in the box). What do the other waves sound like at this frequency? Is the pitch the same? Can you still tell them apart?

8.2: Fourier Series

The French Mathematician Jean Baptiste Joseph Fourier showed any periodic function can be formed from an infinite sum of sines and cosines. This is very convenient because it means that everything we know about sines and cosines applies to a periodic function of any shape. Although the sum is infinite in theory, in many cases using just a few terms may be close enough to provide a good approximation.

8.2.1: Fourier Series

From examples of real sound waves it might seem that trying to use a cosine or sine function to describe a sound wave is pointless because real sound waves ave very complex waveforms. However the French Mathematician Jean Baptiste Joseph ***Fourier*** showed *any periodic function* can be formed from an infinite sum of sines and cosines. This is very convenient because it means that everything we know about sines and cosines applies to a periodic function of any shape. Although the sum is infinite in theory, in many cases using just a few terms may be close enough to provide a good approximation.

Suppose you wanted to make a square wave but only had sine waves to work with. The pictures below were made with the graphing calculator we used earlier. A series of seven sine waves are plotted on the left and the sum of those waves is plotted on the right, below. Although the square wave is not perfect, the more waves that are added, the closer it becomes to a square wave. (You can try this; here is the sum that made the graph on the right: $1.0 * \sin(t) + .3333 * \sin(3 * t) + .2 * \sin(5 * t) + .1428 * \sin(7 * t) + .1111 * \sin(9 * t) + .0909 * \sin(11 * t) + 0.0769 * \sin(13 * t)$. Try plotting just the first term, then the first two terms, then the first three, etc.)

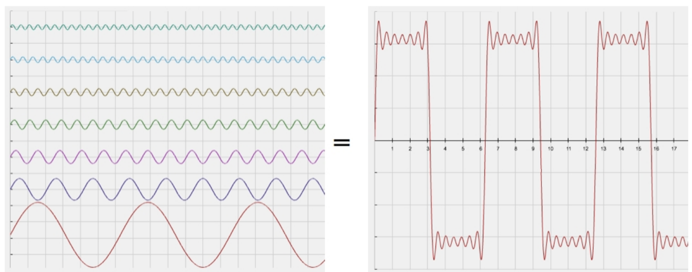


Figure 8.2.1.1

As you can see from the example of the square wave, any periodic shape can be formed by adding sine waves. This is the concept behind constructing an electronic instrument called a ***synthesizer***. Electric pianos also use this principle. By combining the right frequencies and amplitudes of sine and cosine waves the synthesizer can duplicate the sound wave of any other instrument. The exact electronics the do this are somewhat complicated but the principle is simple; the right combination of sine waves can create the sound of any musical instrument.

> **Note**
>
> Modern synthesizers also often use digitally recorded samples of instruments or other sounds which are electronically modified for output, in addition to pure sine and cosine waves.

8.2.2: Fourier Analysis

Suppose we took the components of our square wave (above) and instead of plotting the sine function we just make a bar graph of the amplitude and the frequency number. The result is called a ***Fourier Spectrum*** and is a kind of short hand for the waveform. Instead of drawing the complicated square wave we can simply list the amplitude and frequency number for each component as shown below:

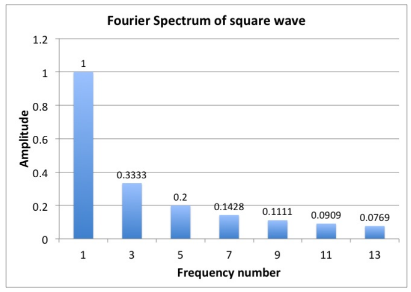


Figure 8.2.2.1

The process of breaking down a waveform into its component sine and cosine waves (the reverse process from Fourier synthesis) is called ***Fourier Analysis***. Fourier analysis and synthesis can be done for any type of wave, not just sound waves. Generally you don't get a nice bar graph like the one above but you get something similar; a graph that tells you how much of each component sine wave is present for each frequency in the sound sample. There are various programs that will do this (for example Audacity is a free sound analysis program we will use to analyze sound waves) and even some smart phone apps (for example SignalScope or oScope for the iPhone) which will show the Fourier spectrum of a sound wave. The Fourier spectrum shows which frequencies are present and the relative strength of each frequency. The lowest frequency present is called the ***fundamental frequency***. The following examples show the waveform (what you would see on an oscilloscope) along with the Fourier spectrum for several different instruments:

Video/audio examples:

- Demonstration of the iPhone app oScope which shows a waveform and Fourier Analysis of it. The waveform is in green, then the FFT is in red.
- YouTube explanation of Fourier synthesis. A bit boring but he eventually makes a nice connection between the interference that causes beats and the combination of sine waves that give rise to interference that we perceive as separate notes.
- Mini-lab on sound analysis using Audacity. [NOTE: Please bring a musical instrument to class!].

As mentioned above, if you use Audacity or a smartphone app (as in the above YouTube example) to measure the Fourier spectrum of a complex sound wave the graph is a bit messier than the bar chart shown above for the square wave. This is because the Fourier analysis process is limited in accuracy by the number of samples taken during the sampling process. We will not go into the details of Fourier analysis here but below are the Fourier spectrums, measured by Audacity, of the four sounds in the previous simulation:

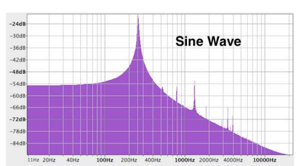


Figure 8.2.2.2

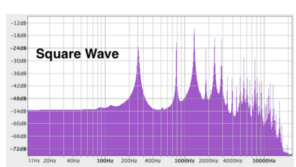


Figure 8.2.2.3

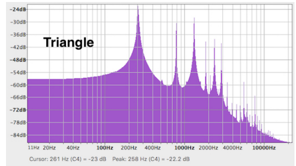


Figure 8.2.2.4

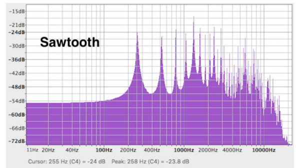


Figure 8.2.2.5

If we go back to the components of the square wave in the bar graph in this section, you will notice that we can reproduce the square wave with just a few numbers from the graph (we really only need the amplitudes since the frequency numbers are just a sequence of whole numbers- odd whole numbers for the square wave). If we want to store a picture of the wave we would need a lot of disk space on our computer. If instead we store the numbers (amplitude and frequency number) and reconstruct the wave from these components when we need it we can keep the same information in much less disk space. As we will discuss later, the various file formats used to compress pictures, sound and movies (JPG, MPEG, etc.) are based on this idea and are in fact versions of the Fourier analysis.

You will also notice from the simulations and exercises that for the square wave, the more terms that are included in the Fourier series the more the wave looks like a square wave. A pure sine wave is smooth and has no specific time location, no start or end. Because of this smoothness and unspecific time only one term is required in the Fourier Series to describe it. The square wave, however, has a very sharp edge at a specific time. This requires many frequencies in the Fourier Series (technically an infinite number to make a perfectly sharp edge). When short pulses of sound such as clicks, cymbal crashes and percussion strikes are Fourier analyzed they show a large number of frequencies and the shorter the sound pulse, the more frequencies are present. This is an example of the ***uncertainty principle*** which says there is a trade off in what you know about when a sound occurs and what you know about the frequencies that are present in the sound. For the sine wave you can know a lot about frequency (it has exactly one number representing frequency) but not much about when it occurs. For a cymbal crash you can know more about when it occurs but at the cost of having to worry about many frequencies. The time/frequency uncertainty principle is only one of several found in physics but is the only one we need for understanding sound.

8.2.3: Fourier Series Simulation

This simulation shows the sum of up to eight harmonics (five on mobile devices) of a sine wave. Initially the speed is set to zero to make the visualization simpler. A ***harmonic*** of a sine wave is a sine wave that has a frequency which is a whole number multiple of the frequency of the original wave. The first harmonic (set with the slider $A1$) is called the ***fundamental***. So if the fundamental is $f_1 = 200 \text{ Hz}$ the second harmonic is $f_2 = 400 \text{ Hz}$, the third harmonic is $f_3 = 600 \text{ Hz}$, etc. We know $v = \lambda f$ so for a fixed speed, doubling the frequency means the wavelength of the second harmonic is half that of the fundamental. Conversely, holding your finger down in the middle of a guitar string and plucking one side cuts the wavelength in half which doubles the frequency being played. For standing waves on a string fixed at each end (such as a guitar string) each harmonic is also called a ***normal mode*** of vibration.

The graph on the upper right in the simulation is the Fourier spectrum and is a short hand way of showing how much of each harmonic is present in the graph on the left. Fourier series usually include sine and cosine functions and can represent periodic functions in time or space or both. In this simulation we only have combinations of sine waves. The Fourier series for the wave function showing in the left graph is given by $y(t) = \sum_{n=1} A_n \sin(n2\pi x/\lambda - n2\pi f_1 t)$. Here t is time, n is the number of the harmonic or mode ($n = 1$ for the fundamental, 2 for the second harmonic etc.), A_n is the amplitude of harmonic or mode number n and f_1 is the fundamental frequency ($f = 1/T$). Amplitude is in arbitrary units, scaled between 1 and -1.

Note

The Fourier Series and Sound JavaScript Model uses the HTML 5 Web Audio API. This API is still under development and may not be supported on all platforms. Press the Reset button to reinitialize the simulation if the sound does not play when the simulations is first loaded.

Simulation Questions:

1. Try adjusting the slider A_1 to different values. What does this slider do? If you have speakers, turn the sound on and listen. This is a 200 Hz sine wave.
2. You may have noticed the amplitude shows up in the graph on the right. You can also see the magnitude of the amplitude by holding the mouse the mouse button down and moving the mouse to the top of one of the peaks on the graph on the left. Does it match the value of $A1$ set by the slider? Use the mouse to find the wavelength (distance between peaks on the left graph), what is the wavelength of this wave?
3. Use the 'reset' button and move slider A_2 (the second harmonic). What is the wavelength of the second harmonic? How does this wavelength compare to the wavelength of the fundamental? This is a 400 Hz sine wave. What does it sound like?
4. The pitch of a sound wave is determined by the fundamental frequency. Turn on the velocity (344 m/s for sound at room temperature) by clicking the simulation run button. With only amplitude A_1 showing, run the simulation. Find the period by measuring the time (in 10^{-3} sec) between when one peak passes the origin and when the next peak passes (use the step button to get an accurate time measurement). What is the frequency of the fundamental?
5. Now find the period of a wave with several harmonics. What is the period of the combination (the time between successive highest peaks)? Although the wave looks more complicated it has the same fundamental frequency and therefore the same pitch. The additional, smaller peaks are due to the harmonics and give a sound its *timbre*. A trumpet and trombone playing the same note have the same fundamental frequency but sound different because of the number and amount of harmonics present.
6. To get the exact shape of an arbitrary periodic function we would need an infinite number of terms in the Fourier series but in this simulation we can only add a maximum of 8 terms. Try the following combination of harmonics (you can type the amplitudes into the boxes next the sliders to get exact values):
   $A_1 = 1.0,\ A_2 = 0,\ A_3 = 0.333(=1/3),\ A_4 = 0,\ A_5 = 0.20(=1/5),\ A_6 = 0,\ A_7 = 0.143(=1/7),\ A_8 = 0$. What is the approximate shape of this wave? If you have speakers, turn the sound on and listen.
7. Reset the simulation and try the following combination of harmonics (you can type the amplitudes into the boxes next the sliders to get exact values):
   $A_1 = 1.0,\ A_2 = -0.5,\ A_3 = 0.333,\ A_4 = -0.25,\ A_5 = 0.20,\ A_6 = -0.166,\ A_7 = 0.143,\ A_8 = -0.125$. What is the approximate shape of this wave? If you have speakers, turn the sound on and listen.
8. Suppose a clarinet and a trumpet both play the same note (have the same fundamental frequency). Why is it that you can still tell them apart, even though they are playing the same note?
9. Find a definition of timbre and write it in your own words. Based on your answers to the above questions, what causes timbre?

10. Suppose you wanted to build an electronic instrument which added waves together to imitate other instruments (this is how some musical synthesizers work). What would you need to know about the sound a trumpet makes in order to reconstruct that sound? (Hint: think about the information contained in the graph at the top right.)

8.2.4: Timbre (again!)

If you did the mini lab or played with one of the Fourier analysis apps you should have detected a different Fourier spectrum for instruments or voices that were different, even when they were playing the same note. According to Fourier, complex waveforms can be constructed from combinations of sine waves. It is these additional frequencies that are the main property that give a musical tone its *timbre*. As shown in the graphs below, we can tell a trumpet from a trombone, even when they play the same note because there are different frequencies present. These variations in frequency change the waveform (top two graphs which can be seen with an oscilloscope) and the Fourier spectrum (lower two graphs which can be determined from a Fourier analysis).

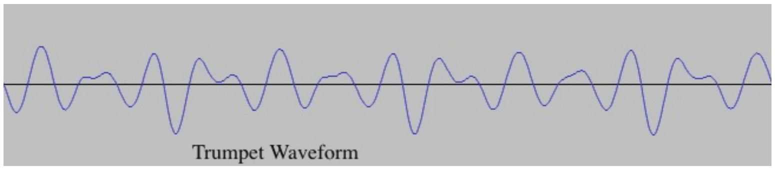


Figure 8.2.4.1

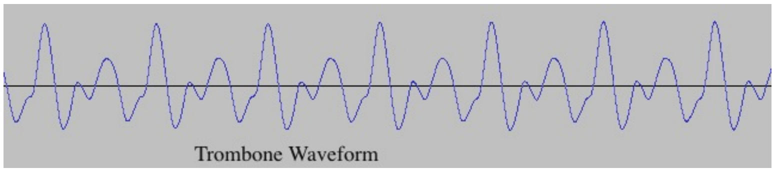


Figure 8.2.4.2

Below the Fourier analysis of each of the above waveforms using the program Audacity. Notice that the fundamental frequency (the lowest frequency) is the same for both instruments (around $230\ \mathrm{Hz}$) so they are playing the same note, even though they will sound different.

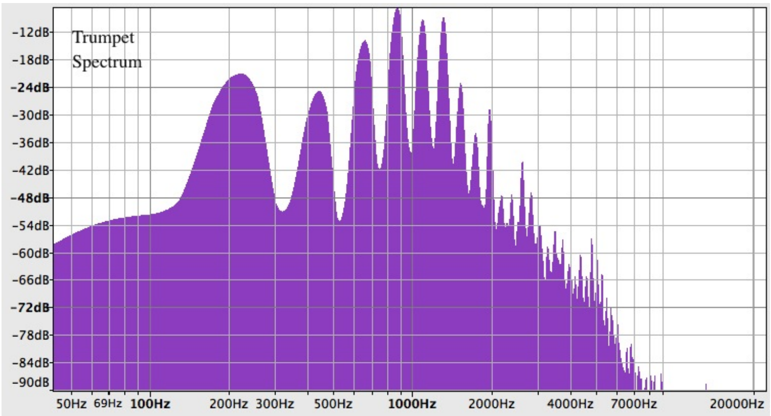


Figure 8.2.4.3

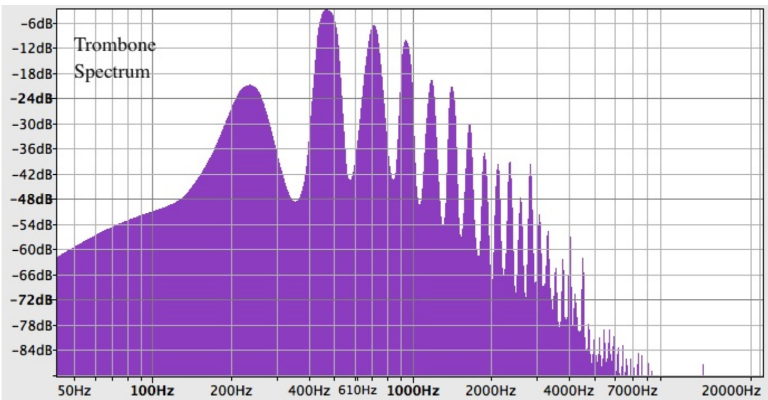


Figure 8.2.4.4

Most notes produced by musical instruments have higher frequencies that are multiples of the fundamental, or lowest frequency. When the higher frequencies are multiples of the fundamental they are called ***harmonics***. The more generic term for frequencies produced by an instrument or voice that are higher than the fundamental (whether multiples or not) is ***overtone***. We will discuss this in more detail in Chapter 11 on strings and stringed instruments.

Video/audio examples:

- Fourier spectrum of a clarinet and oboe and a trumpet.
- A 6 min. YouTube review of a few major points.
- Here is a Chrome Music lab demonstration of a few sound sources and their Spectrograms. A spectrogram shows the Fourier frequencies (plotted vertically) as they change over time (with time plotted on the horizontal axis). Audacity and some mobile apps (for example Spectral View) will show spectrograms.

Summary

The French mathematician Fourier discovered that you could build any repeating waveform by adding enough sine and/or cosine waves. This is called Fourier synthesis. It is also possible to analyze a repeating waveform to find out how much of each frequency is present. This is called Fourier analysis. For a repeating wave the extra frequencies present are all multiples of the fundamental and are called harmonic frequencies. The timbre of a musical instrument depends on the amount of each harmonic that is present.

Questions on Fourier:

1. What does the fundamental frequency of a musical instrument determine?
2. What is an oscilloscope? What does it do?
3. Explain Fourier synthesis.
4. Explain Fourier analysis.
5. What is a Fourier series?
6. In a nutshell, what is the difference between Fourier analysis and Fourier synthesis?
7. What are harmonics?
8. If the fundamental of a note is $220\ \mathrm{Hz}$, what is the frequency of the second harmonic? Third harmonic?
9. What is a normal mode?
10. Use a graphing calculator (or paste these into the online graphing calculator www.metacalculator.com/online/) to graph the following series. What shape do you get for each?
    a. $1*\sin(x)+(1/3)*\sin(3*x)+(1/5)*\sin(5*x)+(1/7)*\sin(7*x)+(1/9)*\sin(9*x)+(1/11)*\sin(11*x)$
    b. $1*\sin(x)-(1/9)*\sin(3*)+(1/25)*\sin(5*x)-(1/49)*\sin(7*x)+(1/81)*\sin(9*x)$
    c. $1*\sin(x)-(1/2)*\sin(2*x)+(1/3)*\sin(3*x)-(1/4)*\sin(4*x)+(1/5)*\sin(5*x)-(1/6)*\sin(6*x)+(1/7)*\sin(7*x)$
11. For each of the above series, add a few more terms following the pattern and re-plot the series. What can you say about adding more terms to the series?
12. When a trumpet and a clarinet play the same note, they are making a sound wave with the same fundamental frequency. Explain why, even though they are playing the same frequency the two instruments sound different. (Answer with more than just "their timbre is different".)
13. What is timbre and what causes it?
14. Why do different instruments or voices playing or singing the same note sound different?
15. Suppose you do a Fourier analysis of the sounds from a trumpet and a guitar, each playing the same note. What would be the same in the two analyses and what would be different?
16. Describe your experience with the program Audacity in the mini-lab. What did you do, what did you learn?
17. What is different about the Fourier spectrum of simple sound sources such as tuning forks and the spectrum of musical instruments?
18. What does a spectrogram show (this is one of the options in Audacity)?
19. In general terms, how do electronic synthesizers work?
20. How does an electric piano manage to sound like a piano?

8.3: Microphone Sound Analyzer

Using this software you can analyze a sound sample captured by your computer or mobile device microphone.

8.3.1: Fourier Analysis of Microphone Data

This is not a simulation but rather software for an alternative version of the Mini-lab on sound analysis using Audacity in the Video/audio examples of the previous section.

The software records the sound from the microphone of a computer or mobile device and displays its amplitude (bottom graph) and frequency (top graph) as calculated by a Fast Fourier Transform (FFT) algorithm. The recording length can be varied by changing the number of data points. However, the number of data points must be a power of two for the algorithm to work. The total range of frequencies is from $20\ \mathrm{Hz}$ to $20\ \mathrm{kHz}$. The minimum and maximum values displayed on the frequency graph can be adjusted to see a more narrow range of frequencies.

Note

The Microphone Sound Analyzer was developed using the Easy JavaScript Simulations (EjS) version 5.3 and is distributed as a ready-to-run html page and requires only a browser with JavaScript support. This model runs on all platforms, including mobile devices, that support the w3C Media Devices API.

https://developer.mozilla.org/en-US/docs/Web/API/MediaDevices

Simulation Questions:

Instructions: When the web page loads the software will ask for access to the device microphone. Click 'allow'. Make some sound (sing, talk, whistle) and click the play button to see the software at work. A second click on the play button will capture a segment of data.

1. Capture a sound sample while whistling a single note (or use a tuning fork if you have one). Describe the waveform of the amplitude (bottom graph) of this sound.
2. The bottom graph is an amplitude versus time (in milliseconds) graph. Click the 'measure' checkbox. Adjust the bars to measure the period of the wave (it is more accurate to measure the period of several waves and divide by the number of waves to get the period of one wave). What is this whistle's period in seconds?
3. Use the 'peaks' checkbox to find the main frequency of your whistle (upper graph). What is the frequency? (Note: You may notice two peaks in this part. Try taking a sound sample while blowing through your lips without making a whistle. Can you now explain what the lower peak is showing?)
4. Divide the period of the wave into one to get the frequency (change to seconds first). Does the frequency as calculated from the period match the frequency (in Hz) on the top graph?
5. Now capture a sound of your voice or a musical instrument playing the same tone as your whistle. What is different about the two graphs from those of a whistle or tuning fork? What is similar?
6. Use the peak checkbox and the period on the lower graph to compare the highest frequency shown in the top graph. How do they compare?
7. The lower frequency peaks are overtones. Are they harmonic? How do you know?
8. Now have a lab partner sing the same note (or use a different instrument). How does their waveform and frequency spectrum differ from yours?
9. Suppose a clarinet and a trumpet both play the same note (have the same fundamental frequency). What would be different and what would be the same for the set of graphs for each. Why is it that you can still tell them apart, even though they are playing the same note?
10. Write a brief definition of each of the following: Fourier Analysis, Fourier Synthesis, spectrogram, harmonics, overtones, timbre.

CHAPTER OVERVIEW

9: The Ear and Perception

According to what we have learned so far, the fundamental frequency of vibrations in the air is the biggest factor in determining the perceived pitch of a note and the energy per second per square meter (intensity) determines how loud the sound is. We also know that our perceptions do not perfectly match frequencies and intensities measured by laboratory instruments. In this chapter we look at how the ear turns vibrations into the perception of sound. Some of the exact details of this process are still not completely understood but the general picture of how we hear is fairly well established. Once again we will see that the human hearing mechanism gives us experiences that do not correspond exactly to laboratory measurements.

Key Terms:

Pinna, middle ear, inner ear, cochlea, timpanic membrane, ossicles, cochlea, semicircular canals, vestibular nerve, auditory nerve, inner hair cells, outer hair cells, place theory of hearing, timing theory of hearing, missing fundamental, auditory illusions, attack frequencies, critical bands, conductive hearing loss, otoacoustic emission, sensorineural hearing loss, presbycusis.

9.1: The Ear and Perception

We look at how the ear turns vibrations into the perception of sound. Some of the exact details of this process are still not completely understood but the general picture of how we hear is fairly well established. Once again we will see that the human hearing mechanism gives us experiences that do not correspond exactly to laboratory measurements.

9.1.1: Structure of the Ear

The image below shows the main parts of the ear. The outer ear or ***pinna*** captures sound and funnels it into the auditory canal where it reaches the ***timpanic membrane*** or eardrum. The eardrum changes the vibrations in the air into mechanical vibrations which travel through as small set of bones called ***ossicles*** (the ***malleus***, ***incus*** and ***stapes***) in the middle ear. These bones are connected to the ***cochlea*** which is in the inner ear. The cochlea changes the mechanical vibrations into nerve impulses that travel to the brain.

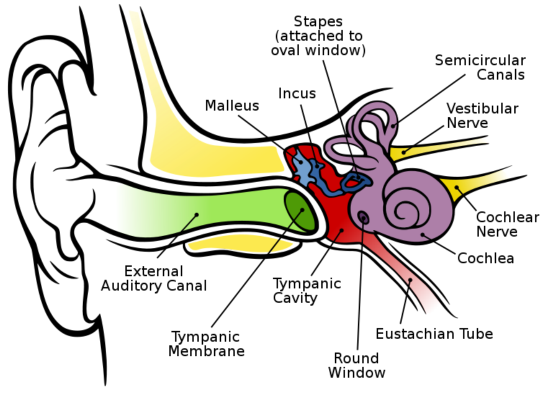


Figure 9.1.1.1

The outer ear or ***pinna***, the part of the ear you can see, funnels sound into the ***auditory canal*** which travels to the ***eardrum***. It is thought that the curvature of the outer ear allows us to judge the direction from which sound is coming. Experiments have been done with plastic models of a head with the ears attached backwards or with different shapes and microphones where the auditory canal is located. When recorded sounds from these microphones are played back to a real person using headphones, their sense of where sounds are coming from change drastically, indicating that the shape of the ear does have an effect on the perception of sound direction. A few more comments on how we perceive the direction to a sound source are found in Chapter 16: Acoustics.

When sound reaches the eardrum or ***tympanic membrane*** it starts to vibrate. A tiny bone called the ***mallus*** is attached to the back of the timpanic membrane and carries the vibrations to the ***incus*** and then to the ***stapes***. The stapes connects to the cochlea so that vibrations in the three bones cause fluid in the cochlea to move.

Air vibrates more easily than fluid in the cochlea so there is a problem getting energy to move from the air into the liquid inside the cochlea. The problem of getting vibrational energy to move from a medium to one with different transmission properties is called ***impedance mismatch*** and we will talk about this in more detail when we talk about musical instruments. The structure of the middle ear has two mechanisms to overcome this problem. The three bones (together called ***ossicles***) act like levers so that a small movement of the timpanic membrane becomes a larger movement of the stapes. The ratio is about 1 to 1.3, in other words, if the eardrum moves 1 mm the end of the stapes connected to the cochlea moves 1.3 mm. Recall from chapter one that pressure is proportional to area. The eardrum has 17 times the area of the base of the stapes. Because of the difference in size between the eardrum and the base of the stapes and the extra leverage in the ossicles the pressure at the cochlea is $17\times 1.3=22$ times the pressure on the eardrum. These two mechanisms overcomes the problem of impedance mismatch between the air at the eardrum and the fluid in the cochlea.

The three bones in the middle ear also have muscles attached to them. If a very loud sound comes into the ear the muscles tense up in an automatic reflex to dampen the vibrations of the three bones. This helps protect the inner ear from damage from sudden intense sounds.

The cochlea has two functions. The three ***semicircular canals*** are oriented at 90 degrees to each other and give us our sense of balance. The canals are filled with a fluid. When we move our head the inertia of the fluid causes it to remain stationary while the body moves. This relative motion causes small hairs at locations called ampulla inside the canals to be bent, sending signals to the brain via the ***vestibular nerve***. Two other locations, the utricle and the saccule have similar arrangements of small hairs that detect motion. There are three canals to account for rotational motion along three axes; left to right rotational head motion, up down head motion and tilting head motion to the left or right. The bending of the hair cells in the five ampulla due to motion of the head is shown in the picture on the right.

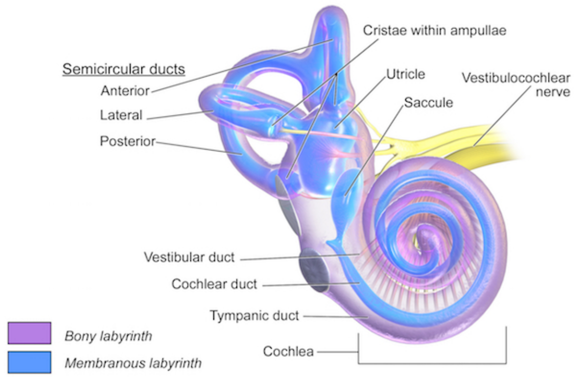


Figure 9.1.1.2

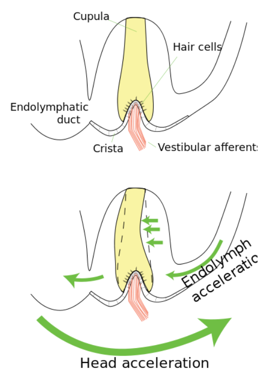


Figure 9.1.1.3

The second function of the cochlea is to convert mechanical vibrations of the fluid inside into nerve impulses that we perceive as sound. The cochlea is basically one long tube, folded over so that sounds go up one side (the scala visibuli) and back down the other (the scala tympani). Vibrations start at the oval window where the stapes is attached and travel through the tube to the apex, then go down the other side until they reach the round window. The tube is coiled and embedded in the bones of the head. The round window is a thin, flexible membrane that keeps pressure from getting too high inside the cochlea. The section between the two halves is called the ***basilar membrane***. The basilar membrane is thickest at the apex where the tube doubles back in the other direction and thinnest at the oval and round window end.

Below is a cross section of the basilar membrane which divides the two parts of the cochlea. If the right frequency is present this part of the membrane will undergo resonance and vibrate. When the membrane vibrates the part called the ***tectorial membrane*** tries to remain stationary because of inertia. The flapping together of the basilar membrane and the tectorial membrane bend the tiny hairs between them.

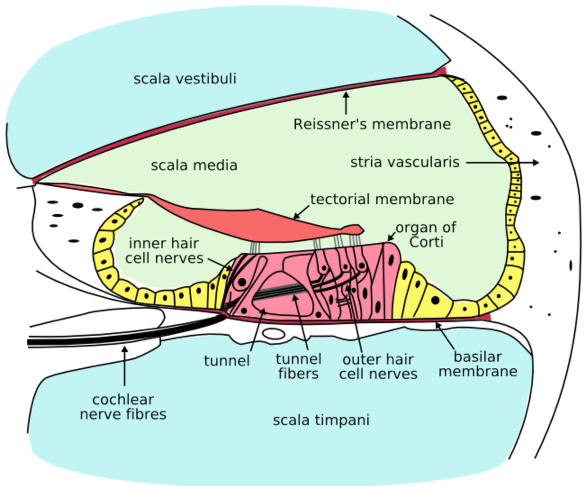


Figure $9.1.1.4$

There are about $16,000$ hair cells on the basilar membrane. Two sets of hairs are found at each location on the basilar membrane, both of which are connect to nerve cells. Bending of the hairs causes a chemical channel to open and the nerve sends a signal to the brain along the ***auditory nerve***. The ***inner hair cells*** are mainly responsible for hearing a particular frequency. The ***outer hair cells*** also bend and cause nerve impulses to be formed when the basilar membrane vibrates. These impulses, however, act to modify the signals being sent by the inner hair nerve cells. The result of this is a single signal being sent to the brain with information about the frequency being heard.

One interesting property of the outer hair cells is that they react to incoming sound by contracting. This contraction can actually produce sound that can be detected outside the ear. It is possible that the produced sound acts to reinforce and sharpen the response of the inner hair cells. The process of the ear producing sound is called ***otoacoustic emission*** and can be used to test for hearing loss. Sharp clicking sounds can be presented to the ear (including the ears of babies and children too young to respond to a hearing test) and if the ear is working correctly, there will be a sound response coming from back from the ear that can be detected with a small microphone near the ear.

Video/Audio examples:

- Overview YouTube video of how sound turns into nerve impulses.
- Simulated graphics of how sound is transmitted from the outer ear to the cochlea.
- Lectures from Interactive Biology YouTube series:
    1. How sound is transmitted from the outer ear to the cochlea (12 min.);
    2. How the cochlea works (9 min.);
    3. The organ of corti (8 min.);
    4. The function of the hair cells (5 min.).

9.1.2: The Place Theory of Hearing

The ear-brain system is a complex instrument. Currently there are two overlapping theories of how we hear; the place theory of hearing and the temporal theory of hearing. Neither of these concepts alone appears to be sufficient to explain the richness of auditory phenomena that we experience.

We know from the structure of the cochlea that different parts resonate at different frequencies; the end closest to the stapes resonates at high frequencies and the end furthest from the ossicles resonates at low frequencies. Nerves are connected to hairs located along the cochlea which are stimulated when vibrations are present. A logical conclusion is that each place in the cochlea corresponds to the perception of a given frequency. This is called the ***place theory of hearing***

How might the place theory work? Below is a screen shot from the driven springs simulation you experimented with in Chapter 4. There are five masses but only one (number two) has very much of an amplitude at a driving frequency of $1.5\ \mathrm{Hz}$. If we drive the masses at a different frequency one of the other masses will have a large amplitude but not mass number two. In other words mass number two has a resonance frequency at $1.5\ \mathrm{Hz}$. (Go back and experiment with the simulation if you don't remember what was going on in this simulation.).

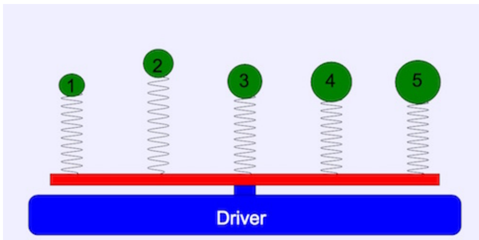


Figure 9.1.2.1

We can imagine this as a model of the cochlea, straightened out with the stapes end on the right. The less massive parts of the cochlea (to the left- opposite of the masses in the simulation) have high resonance frequencies and the part furthest from the ossicles (to the right in the picture) have lower frequency resonance. A given frequency presented to the cochlea only causes motion in one part of the cochlea. This part of the cochlea sends a nerve impulse to the brain and we perceive that frequency.

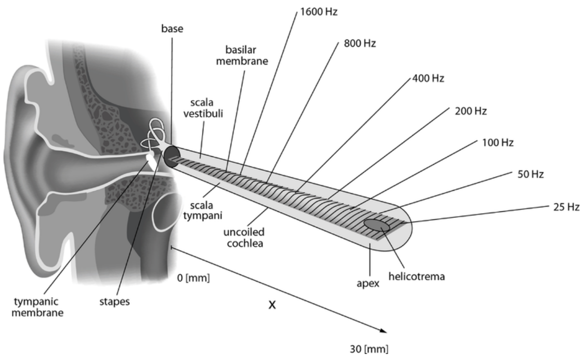


Figure 9.1.2.2

A problem with the place theory is that the resonance curves turn out to be very broad and they overlap, as shown in the graph below (compare with the resonance graph of amplitude versus frequency in Chapter 4). In other words the sections of the cochlea are low Q-factor resonators. This would seem to make it very difficult for the ear to pick out frequencies which are close together but we know that the just noticeable difference in frequency is about $1\ \mathrm{Hz}$ for frequencies lower than $1000\ \mathrm{Hz}$ for most people. If the place theory of hearing was correct we would expect that changing the frequency from, say $250\ \mathrm{Hz}$ to $240\ \mathrm{Hz}$ would shift the region of the basilar membrane that vibrates and trigger different nerve cells. But what actually happens is that the region that vibrates at $250\ \mathrm{Hz}$ overlaps with the region that vibrates at $240\ \mathrm{Hz}$ to such an extent that pretty much the same nerves are firing. The figure below shows four nearly overlapping resonance curves for four different frequencies.

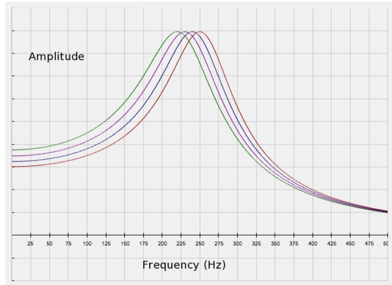


Figure 9.1.2.3

We also know from the uncertainty principle, discussed in Chapter 9, that a sharp resonance would mean less information about the duration of the sound. If sharp resonance peaks (high Q-factor for the cochlea) were the mechanism that enabled us to hear frequencies that are close together, we would not be able to hear sudden changes in frequency. But we are easily able to hear frequencies that change in less than a tenth of second so the resonance peaks cannot be very sharp which means the place theory cannot be the whole story of how we hear different frequencies.

A possible explanation to save the place theory might be that neighboring nerves are inhibited by the nerve firing at the center of the excited region. This is known to happen in touch and vision (for example our touch sensation is more localized than the actual number of skin cells that are activated) but has not been shown to happen in hearing.

9.1.3: The Temporal Theory of Hearing

A second theory of hearing is called the ***periodicity*** or ***temporal theory of hearing***. In this theory it is the timing of the firing of nerve impulses that carries information about the perceived pitch. A simple sine wave at 500 Hz repeats with a period of $1/500\text{ Hz} = 0.002\text{ s}$. The simplest form of the theory says that the vibration causes a nerve to fire every 0.002 s, sending a signal to the brain that is interpreted as a 500 Hz sound. Different sections of a complex sound waveform also repeat periodically. We would expect that hair cells might also fire with the same periodicity of sections of the waveform. For complex waveforms there might be more nerves firing for larger amplitude parts of the wave so that information about the wave shape is also transmitted to the brain.

One difficulty with this theory is that the nerves attached to the hair cells in the cochlea don't seem to fire as often as the theory would predict (and can't fire at a rate of $20,000\text{ Hz}$, at the high end of human hearing). For example instead of firing every 0.002 s for a 500 Hz signal the nerve might initially fire at a period of 0.002 s but then skip to 0.004 s intervals, 0.006 s intervals or some combination of these periods. This could still transmit the correct information, however. The brain could possibly interpret a sequence of 0.002 s, 0.004 s, 0.006 s firings etc. as multiples of 0.002 and conclude the real signal is 0.002 s. Another idea is that the nerves in the cochlea itself somehow filter and combine signals to send a message that is read by the brain as the signal for a 500 Hz signal. There are two locations of hair cells inside the cochlea. It is possible that impulses from the outer hair cells could affect signals from the inner hair cells in a feedback process to produce some combined signal that has information about frequency and waveform. Alternative schemes explaining how the temporal theory could still work with a nerve firing rate not equal to the period of the signal have been proposed but as yet there is not enough experimental evidence to settle the issue.

9.1.4: Hearing Loss

Hearing loss can occur at several points along the path between outer ear and nerve impulse sent to the brain. If the outer ear fills up with wax or is damaged, fewer vibrations will reach the ear drum. Very loud sounds may damage or even burst the eardrum. Although a burst eardrum usually grows back, it may grow back with scar tissue that changes its ability to vibrate properly. The inner surfaces of the middle ear can also be damaged by loud sounds and some diseases (for example inner ear infections). This may cause scars to form in the middle ear region that block or modify the transmission of vibrations along the bone passageway.

Another type of hearing loss involves problems with the three bones in the middle ear. If they become disconnected (which could happen because of a very loud sound) or if they get calcium deposits due to aging they may not effectively transmit vibrations to the cochlea. The above types of hearing loss are called ***conductive losses*** and they sometimes can be repaired by surgery.

Damage to the cochlea or the auditory nerve generally cannot be fixed by surgery. This type of hearing loss is called ***sensorineural loss***. If a very loud vibration causes the stapes to pierce the cochlea vibrations will no longer be efficiently transmitted to the hair nerve cells. Likewise a very large vibration inside the cochlea may break the hair cells so that they can no longer bend in response to vibrations of the basilar membrane. Aging also affects the hairs by making them gradually become stiffer. This means they do not react the same way to vibrations of the basilar membrane or may even break off.

Everyone gradually loses some of their hearing as they get older. Age related hearing loss is called ***presbycusis*** and is the most common form of hearing loss after damage done by loud sounds. Both exposure to loudness and aging generally affect higher frequency perception more than lower frequencies. The typical 25 year old male has lost about 15 dB for frequencies above $6,000\text{ Hz}$ and the 25 year old female has lost 5 dB in the same frequency range. Recall that doubling the loudness (two instruments instead of one) is a 3 dB increase so decreasing your hearing ability by 3 dB decreases the perceived intensity by about half. By the time you reach 55 years of age the average person has lost about 30 dB in loudness for frequencies above $4,000\text{ Hz}$ if you are male and 10 dB if female. On the other hand, frequencies under 1000 Hz are generally not affected much by aging.

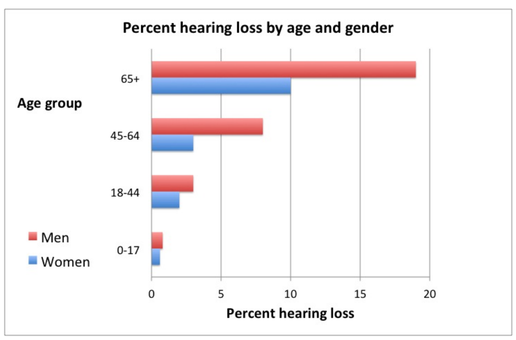


Figure $9.1.4.1$

Some types of hearing loss can be overcome by amplifying the sounds before they enter the middle ear. This is done by the use of battery operated hearing aids. Hearing aids generally only amplify sound and so may not be helpful if the loss of hearing is in a certain frequency range. For Sensorineural loss it may be possible to implant wires directly into the cochlea or the auditory nerve region and attaching an external microphone. This is called a cochlear implant.

Video/Audio Examples:

- Mayo Clinic on hearing loss hearing loss.
- YouTube discussion of tinnitus (permanent ringing or noise in the ear).
- An online hearing test from the University of New South Wales.
- Two tests of frequency range of hearing: one, two. Note: Your computer speakers won't be able to play the lower frequencies.

9.1.5: The Missing Fundamental

One important auditory phenomena that cannot easily be explained by either theory of hearing is the ***missing fundamental*** or ***virtual pitch***. Some sources refer to the effect as ***residue pitch***. As we saw in the chapter on Fourier analysis, musical instruments have overtones which are harmonic, meaning they are multiples of the fundamental frequency. When you hear a guitar play a note with a fundamental frequency of $100 \mathrm{~Hz}$ there are harmonic frequencies of $200 \mathrm{~Hz}, 300 \mathrm{~Hz}, 400 \mathrm{~Hz}$ etc. present in the sound. Your ear-brain system perceives this as a single pitch of $100 \mathrm{~Hz}$ instead of a collection of individual frequencies. An interesting auditory phenomena is that if you are listening to a series of frequencies of $200 \mathrm{~Hz}, 300 \mathrm{~Hz}, 400 \mathrm{~Hz}$ etc. your ear-brain system will ALSO perceive a single note of $100 \mathrm{~Hz}$ *even if the* $100 \mathrm{~Hz}$ *fundamental is missing*.

9.1.6: Missing Fundamental Simulation

The simulation is actually the same as the Fourier simulation in the previous chapter but you will do something different with it. Up to eight harmonics (five on mobile devices) of a sine wave are shown and the fundamental is set at $f_1 = 200\ \mathrm{Hz}$. The second harmonic is $f_2 = 400\ \mathrm{Hz}$, the third harmonic is $f_3 = 600\ \mathrm{Hz}$, etc.

Note

The Fourier Series and Sound JavaScript Model uses the HTML 5 Web Audio API. This API is still under development and may not be supported on all platforms. Press the Reset button to reinitialize the simulation if the sound does not play when the simulations is first loaded.

Simulation Questions:

1. With a fundamental frequency of $200\ \mathrm{Hz}$, push all of the sliders to the right. Listen to the sound (you may need to use external speakers or good headphones since the built-in speakers in your computer are usually not very good). Describe the sound, the timbre.
2. Now slide the first slider, A_1, to zero. You will notice a change in the timbre of the sound but does the perceived pitch (the note being played) change?
3. Try sliding both the A_1 and A_2 amplitudes to zero, leaving the other amplitudes at 1.00. The timbre will again be different but does the perceived pitch change?
4. Try leaving only the highest two frequencies at maximum amplitude and the others at zero. Has the perceived pitch changed?

The missing fundamental is what makes it possible to hear music over small speakers that cannot reproduce the full range of frequencies. Small speakers often do not produce the lower base note frequencies but you still hear them because the higher harmonics are present and your ear-brain system fills in the missing fundamental. As we will see, for many musical instruments, percussion instruments in particular, the missing fundamental is the note we hear when the instrument is playing.

Experiments done with pure tones through headphones show that in fact you only need two frequencies, for example $200\ \mathrm{Hz}$ and $300\ \mathrm{Hz}$, to perceive a missing fundamental of $100\ \mathrm{Hz}$ if the notes are in phase. Even stranger is the fact that you will perceive the missing fundamental when the individual harmonics are played to different ears. Somehow the brain combines signals from both ears to hear one note at a frequency that isn't present. Any theory of how the ear works must be able explain this and other curious auditory phenomena.

The idea that certain regions of the basilar membrane respond to certain frequencies as an explanation for sound perception seems to be basically correct. The place theory, however, cannot explain the missing fundamental phenomena. If a fundamental frequency is missing it cannot cause that region of the basilar membrane to vibrate yet we perceive the frequency as being present. Somehow the harmonics add up to give us that experience. In fact, we would expect that if the harmonics were exciting different regions we might perceive separate harmonic tones instead of one single note but we don't; nearly all listeners hear a single missing fundamental.

In support of the temporal theory, the missing fundamental shifts in a peculiar way if the harmonics shift. For example a missing fundamental of $200\ \mathrm{Hz}$ is heard when harmonics of $1800\ \mathrm{Hz}$, $2000\ \mathrm{Hz}$ and $2200\ \mathrm{Hz}$ are played. It would be tempting to say that the difference of $200\ \mathrm{Hz}$ between the harmonics causes the perception of a $200\ \mathrm{Hz}$ signal. But if these frequencies are shifted to $1860\ \mathrm{Hz}$, $2060\ \mathrm{Hz}$ and $2260\ \mathrm{Hz}$, also a difference of $200\ \mathrm{Hz}$, the ear-brain system perceives a missing fundamental of about $207\ \mathrm{Hz}$. The temporal theory can explain this because the peak to peak period of the second combination of frequencies is around $4.83\ \mathrm{ms}$ which would be the same as a pure tone of $207\ \mathrm{Hz}$. On the other hand, the temporal theory predicts a greater sensitivity to the phase of a wave than we experience (under most circumstances we are unaware of the phase of a given frequency). The place theory does not have this problem.

It is possible (and there is some evidence to support the idea) that both place and temporal mechanisms work but operate in different frequency regimes. The place theory seems to be the more likely mechanism for frequencies above $5000\ \mathrm{Hz}$ while the temporal theory seems to be a better explanation for frequencies under $5000\ \mathrm{Hz}$. Someone with perfect hearing can hear frequencies up to $20,000\ \mathrm{Hz}$ but our sense of pitch and ability to distinguish differences in frequency gets much weaker for frequencies above $5000\ \mathrm{Hz}$ as noted in Section 8.1.3 on just noticeable differences in frequency. These features of our ear-brain system fit with a dual system of frequency detection. Most likely the mechanisms for both theories combine somehow to give us our perception of musical sound.

9.2: Beats

The phenomena of beats occurs when two notes are close together in frequencies and we perceive one note which varies in loudness. A guitar string can be tuned by comparing a note with a known pitch and tuning the string until the beats disappear. What happens if the two frequencies get further and further apart?

9.2.1: Other Combination Tones

The ear-brain system can be easily fooled in many ways giving rise to what are known collectively as ***phantom tones***. The missing fundamental, described above, is one example of a phantom tone. As another example, under the right circumstances listeners will perceive a frequency which is the difference between two frequencies: $f = f_1 - f_2$, instead of the two individual frequencies. A similar effect can sometimes occur when the sum of two frequencies is heard instead of the two individual frequencies: $f = f_1 + f_2$. Since most musical instruments have harmonics and overtones, these combination tones (sometimes called Tartini tones) can have an effect on the timbre of a musical note, both for an individual instrument and for two or more instruments playing together.

The phenomenon of beats, discussed previously, occurs when two notes are close together in frequencies and we perceive one note which varies in loudness. A guitar string can be tuned by comparing a note with a known pitch and tuning the string until the beats disappear. What happens if the two frequencies get further and further apart? If the notes are a little further apart we still hear a single note but it sounds rough or wavering. Eventually we hear two separate notes instead of beats. The range of frequencies where the two frequencies are close enough to cause us to hear a single sound is called a ***critical band***. Notice this is not the same thing as just noticeable difference in frequency. We may still be able to tell one note from another if they played one after the other but when played at the same time our hearing mechanism combines them into a single note.

The following diagram shows the phenomena of critical bands. For a given frequency f_1 a second frequency is played at the same time but its frequency is changed to be first below and then above the frequency f_1.

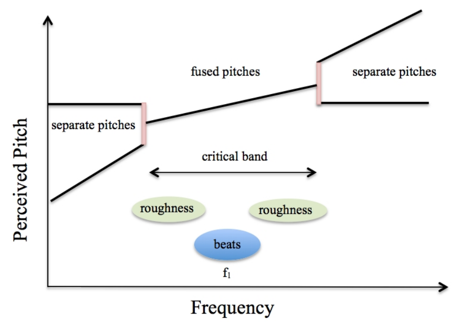


Figure 9.2.1.1

The critical band width is different at different chosen frequencies, f_1. It is about 100 Hz for frequencies below 1000 Hz and then steadily increases to about 3000 Hz at $10,000\text{ Hz}$. If you start with a frequency, f_1, of $10,000\text{ Hz}$ you will hear beats if a second frequency is introduced that is between $7,000\text{ Hz}$ and $13,000\text{ Hz}$. But if you start with a frequency, f_1, at 800 Hz you will only hear beats if the second frequency is between 700 Hz and 900 Hz.

The critical band also determines how well a tone is masked by noise (a random spread of frequencies). If there is one (or several) background frequencies (noise) but the frequency you are tying to hear is outside the critical band surrounding the noise then it is less difficult to hear the signal. If the signal frequency falls in the critical band, however, it is more difficult to perceive. Effectively this is because the ear mechanisms are already responding to that range of frequencies and the signal doesn't cause additional changes to the hearing system.

Dissonance is the musical term that is used to describe musical notes that have harmonics close enough to fall in the roughness regime but not close enough to form beats. When two notes are played on a piano an octave apart (the second note has a fundamental exactly twice the fundamental of the first or a 1 to 2 ratio) all of the harmonics line up and fall on top of each other and there is no dissonance. (For example a 200 Hz fundamental has harmonics at $400\text{ Hz}, 600\text{ Hz}, 800\text{ Hz}$, etc. and the note an octave above has frequencies at $400\text{ Hz}, 800\text{ Hz}$, etc.). However, when a perfect fifth is played (the fundamentals have a 2 to 3 ratio) the harmonics do not exactly line up. The match of harmonics is worse for harmonics of a minor third (ratio of 4 to 5) and progressively worse for a perfect fourth (3 to 4 ratio) and minor seventh (9 to 16 ratio). If the harmonics fall close enough together there may be beating or roughness which changes the timbre of the combination. This is the primary reason a minor seventh chord does not sound as pleasant to the ears as an octave chord or a perfect fifth; the harmonics produce more roughness in the minor seventh case. The effect is stronger when the harmonics of the instrument are stronger and weaker for instruments with fewer or weaker harmonics.

A more subtle from of beats, called waveform beats, can occur for a slight difference between two frequencies about an octave apart (one frequency is twice the other). So for example, a $201\ \mathrm{Hz}$ sound and a $400\ \mathrm{Hz}$ sound together will have a $2\ \mathrm{Hz}$ oscillation in volume. This is because the octave above $201\ \mathrm{Hz}$ would have a $402\ \mathrm{Hz}$ frequency which is $2\ \mathrm{Hz}$ from the $400\ \mathrm{Hz}$ signal. Interestingly, a $200\ \mathrm{Hz}$ plus a $401\ \mathrm{Hz}$ signal is heard with a beat frequency of $1\ \mathrm{Hz}$ because the octave above $200\ \mathrm{Hz}$ is $400\ \mathrm{Hz}$ which is only $1\ \mathrm{Hz}$ difference from $401\ \mathrm{Hz}$. The difference between the two cases has to do with the actual shape of the waveform and indicates that our ear-brain system does detect some information about the shape of the wave being heard, in addition to the frequencies (and is best explained by the place theory of hearing). This phenomena can have a slight effect on the timbre of a musical note.

9.2.2: Beats and Critical Bands Simulation

In this simulation you can hear and see two frequencies and their combination. You can view either the red signal or the green signal and/or their combination (in blue) using the check boxes at the top. The slider labeled Δ adds or subtracts an amount from f_A to modify f_B. The plot is what you would see on an oscilloscope connected to a microphone near the two sound sources with these frequencies. The delay slider sets the phase difference between the two signals so that you can see how they combine at different times which is equivalent to moving the microphone closer or further from the sources. The Scope Δt menu adjust the horizontal axis scale on the oscilloscope, allowing you to zoom in or out on the view. Time is measured in seconds.

Simulation Questions:

1. Listen to the combination $f_A = 300\text{ Hz}$ and $f_B = 305\text{ Hz}$ and describe what you hear.
2. The beat frequency is calculated by subtracting the two frequencies and taking the absolute value (to get a positive number). What is the beat frequency for the combination $f_A = 300\text{ Hz}$ and $f_B = 305\text{ Hz}$?
3. Change the Scope Δt measurement to zoom in and out on the signal. Why does the combination wave (blue) get larger and smaller?
4. Change f_B to be 302 Hz. Does the beat frequency get larger or smaller?
5. Now try a frequency of 310 Hz for f_B. What do you notice about the beat frequency?
6. It is also possible to hear beats between notes that are nearly an octave apart (double the frequency). This is called ***waveform beats***. First listen to a 201 Hz signal and a 200 Hz signal. Now listen carefully to a 201 Hz signal and a 400 Hz signal (this is a little hard to detect without earphones). Is it the same beat frequency?
7. In the previous example you should have heard a beat frequency of 1 Hz for the 201 Hz plus 200 Hz combination but a 2 Hz beat frequency for the 201 Hz signal added to the 400 Hz signal. This is because our ear/brain system hears the beating between $2 \times 201 = 402\text{ Hz}$ and 400 Hz. Now compare this to a 200 Hz signal added to a 401 Hz signal. What is the beat frequency in this case? In this case our ear/brain system compares $2 \times 200 = 400\text{ Hz}$ with 401 Hz and hears a beat frequency of 1 Hz.
8. Reset the simulation and very slowly adjust the Δ slider to hear different combinations. For small differences in frequency you will hear beats. For slightly larger differences you will hear roughness (as described in the text). As the frequencies get even further apart you will hear two notes (or a combination that does not sound rough). Adjust Δ until the perception of beats begins to sound rough. What is Δ at this point?
9. Now try to find the value of Δ where the perception of roughness starts to sound like separate notes or harmonious. What value ofΔ did you get for this transition?
10. If you repeat the previous two questions for a different starting frequency for f_A you may get a different critical band. Try this and describe your results.

9.2.3: Other Interesting Auditory Phenomena

Most musical instruments produce a combination of frequencies that change in amplitude as the note is played. These transitory pitches are called ***attack frequencies*** and they sometimes enable us to tell the difference between different instruments. When music is played backwards, particularly on stringed instruments it often sounds very strange because the attack frequencies occur in the wrong place. Our perception of timbre is determined not only by harmonics, intensity and duration as we have discussed previously, but also by the attack frequencies. Here are YouTubes of a piano played backwards; one, two.

The following is a list of other examples where our ears are fooled into perceiving something that isn't actually present. Definitions and examples of several of them can be found at Wikipedia or Richard Waren's site.

> **Note**
>
> Many of these effects can only be heard using high quality headphones.

- ***Shepard's Illusion***. The illusion of a continuously rising (or falling) scale of notes. YouTube example. The continuous version of Shepard's Illusion is called the ***Risset rhythm***.
- ***The Octave Illusion***. Two notes an octave apart are played alternating back and forth between each ear (eg. $400 \mathrm{~Hz}$ to left ear, $800 \mathrm{~Hz}$ to right ear, then switch). Most listeners perceive a the low tone in one ear and a high frequency tone in the other (instead of back and forth). Which ear depends on handedness (right handed people hear the high note in the right ear). Wikipedia (with sound sample).
- ***Deutsch's Scale Illusion***. If a two different patterns of notes which alternate up and down in pitch are played simultaneously to each ear the perception of a rising scale is heard.
- ***Glissando illusion***. An instrument or voice can be made to appear to change from one side of a stereo output to the other by simultaneously playing a sine wave that is shifting in pitch.
- ***Continuity of Tones***. If you interrupt a continuous rising tone with a brief burst of noise your perception will be that the tone did not stop during the break.
- ***McGurk effect***. Your vision affects what you think you hear. For example if you see a set of lips forming the sound "ga" while you are hearing the sound "ba" you will perceive the sound "da". There are many other examples where visual clues affect what you think you hear; YouTube example.
- ***Context effects***. The sound immediately preceding a sound may affect the how the second sound is perceived. For example "al" plus "da" sounds like "al ga" but "ar" plus "da" sounds like "ar da". Similar to this is prior information affects what you hear: YouTube example
- ***Chromatic Auditory Illusion***. Alternating pitches may be perceived as two scales going in opposite directions.
- ***Melody illusion***. Keeping the same notes but changing their octave can make a melody unrecognizable; YouTube example.

Video/Audio examples:

- Many auditory illusions are demonstrated in YouTube videos.
- The Auditory Neuroscience Website is a wonderful source of sound samples and videos of various auditory illusions, demonstrations of acoustic phenomena and indepth discussions of hearing loss.
- List of auditory examples demonstrated in class.
- A list of auditory CDs along with a description of each. You may be able to find others in a search on Amazon.
- Sonification is the process of making data audible in order to detect patterns in a large data set. Technical link: Sonificaiton Handbook.

9.2.4: Animal Hearing

Note

Much of the information about animal perception in this chapter and in Chapters 14 and 16 comes from the excellent book *Engineering Animals: How Life Works* by Mark Denny and Alan McFadzean.

Many animals, including dogs, can hear frequencies in the ultrasonic range (above $20,000\text{ Hz}$). Dog whistles used to train dogs have frequencies between about $23,000\text{ Hz}$ and $54,000\text{ Hz}$ so dogs (and many other animals) can hear them but humans cannot. As mentioned in Chapter 8, humans are about the best existing species in being able to hear frequencies in the range of 1000 Hz to 5000 Hz. In this frequency range, humans and elephants can hear softer sounds than dogs and cats, who can hear better than rats, who can hear better than horses and cows. Birds and fish generally hear a much smaller range of frequencies than mammals and also do not hear softer sounds as well. Some insects hear very specific bands of frequencies, for example crickets have a hearing range for signals generated by other crickets for communication purposes but they have a different, unconnected range of hearing to detect the echolocation signals coming from bats, which are a predator.

Animal	Hearing range in Hertz
Humans	$20-20,000$
Bats	$2000-110,000$
Elephant	$16-12,000$
Fur Seal	$800-50,000$
Beluga Whale	$1000-123,000$
Sea Lion	$450-50,000$
Harp Seal	$950-65,000$
Harbor Porpoise	$550-105,000$
Killer Whale	$800-13,500$
Bottlenose Dolphin	$90-105,000$
Porpoise	$75-150,000$
Dog	$67-45,000$
Cat	$45-64,000$
Rat	$200-76,000$
Opossum	$500-64,000$
Chicken	$125-2,000$
Parakeet	$200-8,500$
Horse	$55-33,500$

Table $9.2.4.1$

As a rule of thumb, small animals tend to make and hear higher frequencies and larger animals are more likely to make and hear lower frequencies, although there are many exceptions as can be seen in the chart. In general the shape of the outer ear is also related to which frequency range and the direction an animal is listening to. For example rabbits have tall ear lobes which makes them more sensitive to sounds coming from a horizontal direction because their predators are mostly terrestrial. Mice, on the other hand, have rounder ears which makes them more sensitive to sounds coming from above, since birds of prey are predators for them. Elephants have very broad, flat ears which are more receptive to lower frequencies with which they communicate. Owls and some other animals can change the relative orientation of their outer ears or will move their entire heads to achieve better accuracy in determining the direction from which sounds are coming. Many animals, such as deer, have muscles that can point the pinna in different directions, enabling better predator detection (humans have atrophied versions of these same muscles but they no longer produce much motion of the ear). Further discussion of animal hearing and its use for orientation and navigation will be postponed to Chapter 16: Acoustics.

Video/audio examples:

- Article by Peter L. Tyack in *Physics Today* about Human-generated sound and marine mammals.

Summary

The ear is composed of three main parts. The shape of the outer ear gives us information about the direction sound is coming from and funnels sound into the middle ear. The middle ear overcomes the impedance mismatch between air and liquid and transmits sound to the cochlea via the ossicles. The cochlea gives us our sense of balance and turns mechanical vibrations into nerve impulses. There are two theories of hearing, the place theory and the temporal theory. Neither can account for the richness of our perception of sound such as the missing fundamental, critical bands or other interesting aural illusions. Both theories may operate but in different frequency ranges. Age related hearing loss, presbycusis, generally lowers the loudness of high frequencies as we grow older. Loud sounds can also cause either conductive or sensorineural hearing loss.

Questions on Perception:

1. What does the fundamental frequency of an instrument determine?
2. Explain the difference between subjective measures of pitch, loudness and timbre as compared to the objective measurements of fundamental frequency, sound intensity and waveform.
3. Describe the structure of the outer ear, middle ear and inner ear. What is the function of each of these sections of the ear?
4. What is the function of the pinnae?
5. What do the ossicles in the middle ear do?
6. What does the cochlea do besides turn sound into nerve impulses?
7. What are the two types of hair cells (in the cochlea) and what does each do?
8. What is the basilar membrane and what does it do?
9. What is impedance mismatch? Why is this significant in the context of the ear?
10. How does the structure of the ear overcome the impedance mismatch between the air at the eardrum and the fluid in the cochlea?
11. Explain the place theory of hearing.
12. Explain the temporal theory of hearing.
13. Why do we have two theories of hearing?
14. What is virtual pitch (missing fundamental)?
15. An ear bud for a cell phone cannot produce a low frequency, long wavelength sound wave of $50\ \mathrm{Hz}$. Yet we perceive this frequency when we listen to music through an ear bud. How does this work?
16. There are two uses of the word ‘beats’ in sound. One has to do with the rhythm of the music. The other has to do with something that happens when two notes near the same frequency are played together. Explain the second use of the word ‘beats’.
17. Explain the concept of critical bands for two notes that are close together in frequency played at the same time.
18. What is dissonance? How would you demonstrate this if you had a way to make several different frequencies (like the critical band simulation)?
19. What are attack frequencies? What effect do they have on our perception of an instruments timbre?
20. What are the two main causes of hearing loss?
21. What are the two main types of hearing loss?
22. What is presbycusis, why does it happen and what are the results?
23. Which type of hearing loss can be corrected by surgery?
24. What can go wrong in the middle ear to cause hearing loss?
25. What can go wrong in the cochlea to cause hearing loss?
26. Pick two the following auditory illusions and explain them:
    a. Shepard’s illusion
    b. octave illusion

c. Deutsch's Scale Illusion
d. glissando illusion
e. McGurk effect
f. melody illusion.

CHAPTER OVERVIEW

10: Strings

In this chapter we start with resonance on a string and then look at resonances on two dimensional surfaces. A brief introduction of three dimensional resonance in an enclosed volume is given and will be expanded on in Chapter 16 on Acoustics. Once the basic behavior of waves on a string, surfaces and in an enclosed volume are examined we look at how these principles end up explaining how stringed instruments such as guitars and violins work.

Key Terms:

String resonance, nodes, anti-nodes, fundamental, harmonics, partials, overtones, conservation of energy (again!), nodal lines, holographic interferometry, Helmholtz resonance, f hole, bridge, wolf tone, sound post, bass bar, plectrum, tangent, piano sound board.

10.1: Driven String and Resonance

In this simulation a string is driven at one end by an oscillating driver. The result is that a wave will eventually form on the string. At certain frequencies the wave will become large and we refer to this resonance phenomena as a standing wave.

10.1.1: String Resonance

Recall that speed of a wave on a string is determined by the density and tension; $v = (T/\mu)^{1/2}$ where T is the tension in the string in Newtons and μ is the mass per length in kilograms per meter). We also know that once the speed is fixed, frequency (in Hertz) and wavelength (in meters) are inversely proportional; $v = f\lambda$. So the three parameters that determine the frequencies of a string are tension, density (mass per length) and length. The density of a string is determined by thickness and mass; a thick, heavy string is more dense so waves travel more slowly.

For standing waves on a string the ends are fixed and the string does not move. Places where the sting is not vibrating are called ***nodes***. This limits the wavelengths that are possible which in turn determines the frequencies since the speed is fixed and $v = f\lambda$. The lowest frequency is called the ***fundamental*** or first harmonic. For a string, higher frequencies are all multiples of the fundamental and are called ***harmonics*** or ***partials***. The more general term ***overtone*** is used to indicate frequencies greater than the fundamental which may or may not be harmonic. This can be a bit confusing because for strings there are only harmonics and the second harmonic is the first overtone, etc. The various harmonics (overtones) are also called the ***normal modes of vibration*** of the string.

What wavelengths will fit on a string of length L? There has to be a node on each end so it has to be the case that $L = n\lambda/2$ where n is a whole number. In other words, you can have half a wave on the string ($n = 1$), one wave ($n = 2$), one and a half waves ($n = 3$), etc. But you never have any other fraction of a wave because that would require not having a node at both ends.

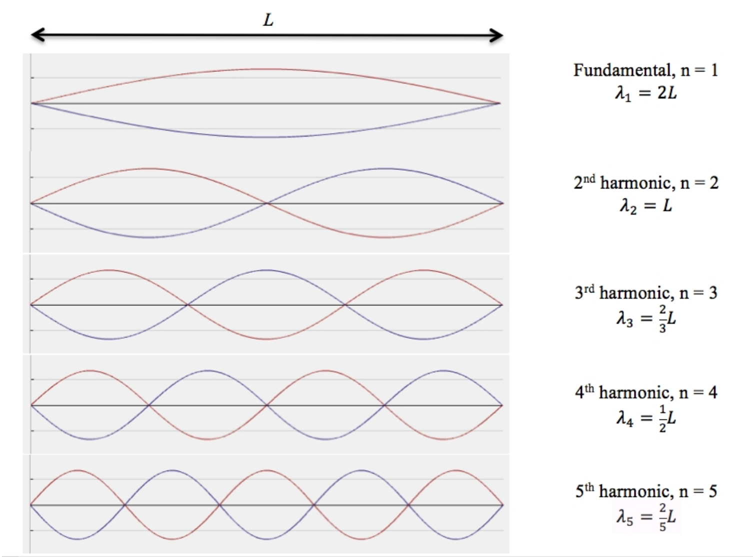


Figure 10.1.1.1

Notice that the higher harmonics have nodes in other locations besides just at the ends. In between two nodes is a region where the vibrations are a maximum. These are called ***anti-nodes***. The fundamental has one anti-node the 2nd harmonic has two anti-nodes.

Once the string density and tension are chosen the speed is fixed and the frequencies will depend on the wavelength as shown in the following table. Notice that the harmonics are all multiples of the fundamental.

Harmonic number	Wavelength	Frequency $f = v/\lambda$
$n = 1$	$\lambda_1 = 2L$	$f_1 = v/\lambda_1$
$n = 2$	$\lambda_2 = L = \lambda_1/2$	$f_2 = v/\lambda_2 = 2f_1$
$b = 3$	$\lambda_3 = \frac{2}{3}L = \lambda_1/3$	$f_3 = v/\lambda_3 = 3f_1$
$n = 4$	$\lambda_4 = \frac{1}{2}L = \lambda_1/4$	$f_4 = v/\lambda_4 = 4f_1$
$n = 5$	$\lambda_5 = \frac{2}{5}L = \lambda_1/5$	$f_5 = v/\lambda_5 = 5f_1$
n	$\lambda_n = \frac{2}{n}L = \lambda_1/n$	$f_n = v/\lambda_n = nf_1$

Table 10.1.1.1

In Chapter 4 we defined resonance to be the case when the amplitude of vibration got larger because the driving frequency matched the natural frequency. Here we see there are many natural frequencies. This means there are many resonance frequencies. For a string these are all harmonics (whole number multiples) of the fundamental. In a real string that is plucked or bowed, it is often the case that several of these resonance frequencies are present at the same time. This means that musical instruments usually produce not only a fundamental, but also harmonics that give the instrument its timbre.

10.1.2: Driven String Simulation

In this simulation a string is driven on one end by an oscillating driver. We can imagine the string as if it were made up of a row of identical oscillating masses, each undergoing simple harmonic motion (just like the row of masses in the transverse wave simulation). Because the string is driven from one end, each successive mass in the row will start later than the one before it (in other words each mass is slightly out of phase with its neighbor). The result is a wave will eventually form on the string.

As we know from the chapter on resonance, if a mass is driven at its natural frequency, its amplitude will increase to a maximum. So the masses that make up the wave on the string will have the largest amplitude only when driven at resonance. When resonance occurs the string has the largest amplitude standing wave on it.

There is one slight complication in this picture, however. There is more than one way for the phases between the masses to occur and still have the masses in resonance. These other resonances occur at multiples of the fundamental resonance frequency and are called harmonics.

Simulation Questions:

1. Start the simulation and wait a few seconds. This is the fundamental (resonance) frequency. What is the period of the wave? (Use the step button and the time to find the period.) What is the frequency of the wave? Is this the same number as the driving frequency?
2. The length of the string is $100\ \mathrm{cm}$. What is the wavelength of the fundamental frequency?
3. Reset the simulation and change the driving frequency to $0.3\ \mathrm{Hz}$. How does the amplitude compare to when the driving frequency is $0.4\ \mathrm{Hz}$?
4. Reset the simulation, set the driving frequency to $0.3\ \mathrm{Hz}$ and increase the driving amplitude to $0.09\ \mathrm{cm}$. Is the amplitude of the wave larger than when the driving frequency is $0.4\ \mathrm{Hz}$? Explain.
5. Reset the simulation and change the driving frequency to $0.6\ \mathrm{Hz}$. How does the amplitude compare to when the driving frequency is $0.4\ \mathrm{Hz}$?
6. Now try a driving frequency of $0.8\ \mathrm{Hz}$. How does the amplitude compare to when the driving frequency is $0.3\ \mathrm{Hz}$ or $0.6\ \mathrm{Hz}$?
7. This frequency, $0.8\ \mathrm{Hz}$, is the second harmonic; the masses are again in resonance but with different phases. What is the wavelength of the second harmonic?
8. Find the frequencies for the third and fourth harmonics. What are they?
9. What are the wavelengths for the third and fourth harmonics?

10.2: Plucked String

In this section a set of initial conditions for a vibrating string is shown. The first is the fundamental frequency of the string, the second is the second harmonic. The third and fourth initial conditions simulate plucking in the center and at a location one fourth of the way along the string.

10.2.1: Plucked String

Generally strings are either plucked or bowed. In both cases the string does not undergo the simple scenario of harmonics described above. Plucking a string at the center does not cause a nicely shaped sinusoidal wave; instead you start with a triangle shape on the string.

But we know from Fourier's work that any repeating shape can be formed from a series of sine waves. Plucking a string at the center emphasizes the fundamental but many other harmonics will be included. Plucking the string at a location one fourth of the way along the string makes the second harmonic a bit louder but other harmonics will still be present. The result of plucking (starting with a triangle shape) and plucking at different locations means the spectrum is not uniform; where you pluck the string determines which harmonics are emphasized.

In the plucked case the triangle shape immediately converts into a combination of sines and cosines, some of which die away quickly. If the string is bowed, however, the triangle wave is maintained since the bow continues to pull the string to one side at the point of contact. The triangle shaped wave travels to the bridge, reflects, and returns to the bow contact location. When the point of the triangle shape returns to the bow it causes the string to break loose from the bow. The wave continues and reflects off the fret end, returning to the bow again, now causing the string to stick to the bow. This slip-stick mechanism maintains a triangle shaped wave moving on the string, reflecting from each end. Once again, changing the location of the bow contact determines which harmonics are emphasized.

Video/audio examples:

- A YouTube of a standing waves on a driven string.
- Wikipedia on string resonance (lists many types of stringed instruments.
- A few stringed instruments, such as the Indian sitar have sympathetic strings. These are extra strings that are not normally played by the musician but vibrate due to resonance because they are tuned to the strings that are plucked or bowed.
- Slow motion YouTube of a bowed string.
- Slow motion YouTube of a plucked string.

10.2.2: Plucked String Simulation

In this simulation a set of initial conditions for a vibrating string can be chosen. The first is the fundamental frequency of the string, the second is the second harmonic. The third and fourth initial conditions simulate plucking in the center and at a location one fourth of the way along the string.

Simulation Questions:

1. Run the fundamental and second harmonic initial conditions. How are standing waves formed on a string?
2. How many sine wave Fourier components are present in the case of the fundamental?
3. Which Fourier components are present if the string vibrates in the second harmonic mode?
4. Now look at the case of the string plucked in the center. What is the initial shape? Based on your experience with the triangle wave forms in the Waveform simulation in Chapter 9, do you expect the string plucked at the center to sound different from the fundamental or second harmonic? Explain.
5. Now look at the case of the string plucked off-center. Which wave is this similar to in the Waveform simulation in Chapter 9?
6. Based on what you know from the Fourier simulation in Chapter 9, what would you expect the Fourier components to look like? (Hint; Do Fourier simulation question seven.)
7. Based on your experience with listening to the wave forms in the Waveform simulation in Chapter 9, do you expect the string plucked off-center to sound different from the fundamental or second harmonic? Will it sound different from the string plucked in the center? Explain.

10.3: Vibrating Plates Simulation

The Vibrating Plates simulation examines vibrational modes on a rectangular surface. The surface is fixed at the edges so the nodal lines occur in different places compared to a rectangle with free edges. The model assumes that the surface is very thin and very flexible; real surfaces which are stiffer will have slightly different nodal lines and anti-nodes.

10.3.1: Surface Resonances

If you fix a string at both ends in free space with no resonating body attached to it the string does not make very much sound. Connecting the string to a surface that vibrates allows the energy of the string to move to the surface and cause it to vibrate. A large surface can move much more air, resulting in a much louder sound.

You might think this violates conservation of energy; a vibrating string in air doesn't produce much sound but a string attached to a surface does. But conservation of energy is a fundamental law of physics that can't be broken. So what is happening? If you time how long a free string vibrates and compare it to how long it will vibrate if attached to a surface you find something interesting. When attached to a surface the string's vibrations die away much more rapidly. In other words, because the energy is being used to create lots of sound, it dissipates much faster. A free string can vibrate longer because it doesn't dissipate it's energy making sound.

Vibrating strings have resonances with different numbers of nodal points, places on the string which do not vibrate. These depend on the driving frequency. A similar thing happens with surfaces; there are resonance frequencies which result in places where not much vibration occurs. These locations are linked together in ***nodal lines*** which depend on the shape and thickness of the surface. One way to see these lines is to drive the surface with an oscillator and put powder or salt on the surface. The powder will not move from the nodal line but will be thrown off of the anti-nodal regions, as shown in the following example.

Video/audio examples:

- Chladni plate. Notice that as the frequency is increased different nodal lines occur, just like different nodes occurred on the string at different frequencies. If you listen carefully to this video you can hear that there is more sound when the plate reaches a resonance. This is because the amplitude is larger at the resonance frequency (as expected).
- Flat plates of various shapes called Bell plates, tuned to specific frequencies, have long been used as inexpensive substitutes for bells.

We know that stringed instruments have harmonic frequencies which are multiples of the fundamental. This is the case because the string is a fixed length; the longest wave that can exist on the string (with fixed ends) has a wavelength that is twice the length of the string. The next wavelength that can fit is the exact length of the string; the next wavelength that will fit is 1.5 times the length of the string and so on as we saw previously in the chapter on stringed instruments. Wavelengths in between these would not have a node at both ends and so can't exist on the string. Each of these different ways of vibrating is labeled by the number n; the fundamental is labeled $n=1$, the second harmonic is labeled $n=2$, etc. This number, called the ***mode number*** indicates how many anti-nodes are on the string.

Surfaces of various shapes (round, rectangular, square) however, are two dimensional and so will require two mode numbers, n and m, to label each mode of vibration. For a rectangular surface fixed at the edges we can label the two dimensions as x and y and there are sine wave shapes in the x-direction and in the y-direction with nodes at the edges, just like a string, as shown in the following simulation.

10.3.2: Vibrating Plate Simulation

This simulation allows you to examine vibrational modes on a rectangular surface. In this case, unlike the Chladni plate in the video, the surface is fixed at the edges so the nodal lines occur in different places compared to a rectangle with free edges. This simulation also models a surface that is very thin and very flexible (a membrane); real surfaces which are stiffer will have slightly different nodal lines and anti-nodes. The effect is very similar, however. Free edged surfaces, thin flexible surfaces and thick stiff surfaces all have nodal lines and anti-node areas.

The buttons at the top allow you to see the first few pure modes. The custom button allows you to look at higher pure modes and to add several modes together. To see a higher order pure mode in the simulation, click the reset button, click the custom button, choose n and m, enter (so the yellow disappears) and click 'add'. If you do *not* click the reset first button the selected mode gets added to the modes that are already present. The simulation can be viewed as a two dimensional surface or as a surface in three dimensions. In the 3D view you can grab the surface with the mouse and rotate it. The amplitudes of the vibrations are exaggerated compared to a real surface to make them more visible. Increasing Δt makes the simulation run faster but also less accurately, particularly for higher mode numbers.

Note

The relative mode amplitudes in the case of more than one mode are not properly normalized in this simulation.

Simulation Questions:

1. Describe the $n=1, \ m=1$ mode in the surface view and in the 3D view. If you sliced through the middle of the surface in the x-direction, which string mode would it look like? How about in the y-direction?
2. What is the frequency of the $n=1, \ m=1$ mode? What is the period?
3. Describe the $n=2, \ m=1$ mode in the surface and 3D view. (Don't forget you can rotate the view using the mouse to get a better angle.) Now if you slice along the x-direction in the middle of the surface which string mode does it look like?
4. For the $n=2$ and $m=1$ mode, how many anti-nodes are along the x-direction? How many anti-nodes are along the y-direction? Describe the location of the nodal line located.
5. What is the frequency of the $n=2, \ m=1$ mode? How does this compare to the fundamental frequency? Is this a harmonic? How do you know?
6. Now look at the $n=1$ and $m=2$ mode and describe the motion in the surface and 3D view. What is the frequency of this mode? How does this compare to the $n=2$ and $m=1$ mode frequency? This is an example of ***degeneracy***; two different modes end up with the same frequency because of the symmetrical nature of the system.
7. Look at the $n=2$ and $m=2$ mode. For a slice in the x-direction a quarter of the way up the y-axis, what string mode does this look like?
8. Where are the nodal lines for the $n=2, \ m=2$ mode?
9. What is the frequency of the $n=2, \ m=2$ mode? How does this compare to the $n=2$ and $m=1$ mode frequency? Is the $n=2, \ m=2$ mode a harmonic?
10. Now look at other pure modes by choosing the custom button, selecting n and m, 'enter' and 'add'. To change to a different pure mode, click reset before each 'add'. Try to find other cases of degeneracy. Which combinations of n and m are degenerate?
11. Find the number of anti-nodes in the x-direction and in the y-direction for the $n=3$ and $m=2$ pure mode (don't forget to click reset before adding the values of n and m).
12. Describe the nodal lines for the $n=3$ and $m=2$ pure mode.
13. If you do NOT click reset before adding a mode, the mode gets added to whatever modes are already present on the surface. Add the $n=1$ and $m=1$ mode to the $n=2$ and $m=2$ mode. Describe what you see in the surface view and the 3D view.
14. Add several higher modes (don't hit reset in between). What is the effect of having lots of different modes on the surface at the same time?

10.3.3: Other Surfaces

The following are a few pictures of nodal lines by Thomas Erndl from a web site with a discussion and collection of nodal surfaces for guitars and violins. Notice that they are similar to other cases of surfaces (in the simulation and videos above) but because of the shape of the guitar and irregular thickness of the surface, the nodal patterns are not exactly symmetric.

Figure 10.3.3.1

Figure 10.3.3.2

Figure 10.3.3.3

Figure 10.3.3.4

As was the case for a string, surfaces have many different resonance frequencies. For the string these other resonances are all multiples of the fundamental. However if you look carefully at the Chladni patterns of the plate and the guitar and violin you will notice that these resonance frequencies are not harmonics (multiples) of the fundamental. These higher frequencies are called ***overtones*** to distinguish them from harmonics. Strings and tubes (discussed in the next chapter) have overtones that are harmonic while surfaces typically do not. The rectangular plate simulation shows that a few of the overtones are multiples of the fundamental frequency but most are not. We also saw in the simulation that some modes are degenerate, meaning two different sets of mode numbers have the same frequency.

The Chladni plate resonance frequencies tend to be sharp or high Q-factor resonances, as are the circular and rectangular modes in the various simulations linked from this page. Because the bodies of most musical instruments do not have symmetric shapes and are not of uniform thickness (due to supporting structures inside), the resonances of musical stringed instruments are not sharp. As mentioned in Chapter 4 on resonances, the Q-factor is a measure of how wide the resonance curve is. In general, stringed instruments have surfaces with low Q-factor, in other words, broad resonance curves. The trade off is that, as previously mentioned, a low Q-factor also means more oscillations before being damped out. This means the sound will last longer before dissipating.

There is a second way to see the vibrational modes of a surface, called ***holographic interferometry***. In the chapter on wave behavior we found that if two waves arrive at a point and are out-of-phase they undergo destructive interference and cancel out. If they are in-phase they interfere constructively. Suppose we reflect laser light off of the vibrating surface of a guitar into a camera. Now shine a second, reference laser beam directly into the camera. If the reflected beam comes off a surface that has moved half a wavelength (at the instance the camera shutter opens) the two beams will be out of phase and not expose the film. If the surface has moved a whole wavelength (or two or three, etc.) the reflected beam will be in-phase with the reference beam and expose the film. The result is a contour map of the deflections of the surface from equilibrium.

In the following picture (from the BBC report In Pictures: Stringed Theory) the concentric lines show regions where the guitar surface is vibrating in and out. The center of each region of concentric circles is where the movement is a maximum. A dark center indicates movement that is out of phase (inward) from a center that is light colored (outward). It is also possible to make successive pictures into a video of the surface as it changes over time.

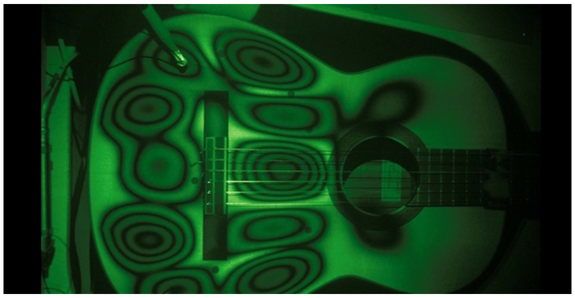


Figure $10.3.3.5$

Because different parts of the surface vibrate for each frequency range, the direction of sound emission is affected. For example, a violin in the frequency range $200\ \mathrm{Hz}$ to $500\ \mathrm{Hz}$ emits sound pretty much equally in all directions. But in the frequency range $550\ \mathrm{Hz}$ to $700\ \mathrm{Hz}$ more sound is emitted to the left and right of the performer than straight ahead or behind. In the 800 Hz range the sound is emitted forward and left and right but not behind. For frequencies between $1000\ \mathrm{Hz}$ and $1250\ \mathrm{Hz}$ more sound is emitted at an angle of about 70 degrees forward and to the right of the performer.

10.3.4: Volume or Helmholtz Resonance

Rooms, containers and many instruments have a confined volume of air of a certain size. If a small opening is made in a container or a window opened in a room the air can oscillate with a frequency that depends on the size of the container or room. The fundamental frequency produced is referred to as ***Helmholtz resonance***, as mentioned in Chapter 4. If you have ever blown across the top of a soft drink bottle to get a note you have excited a Helmholtz resonance. Another example is when you drive your car at just the right speed with one window partly opened and you hear a low frequency sound.

For a rectangular container the walls are parallel to each other and we can describe the higher frequency modes of the container as standing waves in three dimensions, much like the standing waves on a guitar string. As shown above, for a string the overtones are given by $f_n = v/2((n/L)^2)^{1/2}$ where L is the length of the string, v is the speed of a wave on the string and n is the mode number. Extending this to three dimensions, the equation for frequency modes inside a rectangular container with height H, length L and width W is $f_{n,l,m} = v/2((n/L)^2 + (l/H)^2 + (m/W)^2)^{1/2}$. Now there are three different mode numbers since sound can travel in three directions. You can explore these modes in this box modes simulation Applet by Paul Falstad (notice you can grab and rotate the box to see different modes from different angles).

Acoustic stringed instruments have a body that contains a volume of air and air holes in the body so that this air can resonate. The shapes are much more complicated than a rectangular box but there are still resonance modes. Along with the surface resonances these volume resonances modify the string harmonics to give an instrument its unique timbre.

10.3.5: Stringed Instruments

There are probably an infinite number of ways to connect a string to a resonating body to form a musical instrument and any number of body shapes. Here is Wikipedia's list of stringed instruments with links to information about each one. There are also several variations in how the string is excited and how the string vibration is coupled to the body of the instrument. It is also possible to turn the vibrations into an electronic signal for amplification, as we will see later.

Stringed instruments have strings with string harmonic resonances connected to a surface which has additional overtones, some of which are not harmonic. Most stringed instruments also have a hollow body with an opening (called the ***f hole*** or sound hole in a violin and some guitars) so there are air resonances associated with the body cavity; the Helmholtz resonances we saw in Chapter 4 and mentioned above. The shape of the hole developed over time by trial and error to provide the most sound as explained in this article.

It is this combination of string, surface and Helmholtz resonances that make each type of instrument, and in fact each instrument unique. A primary difference between expensive violins such as a Stradivarius build in the late 1600s, early 1700s (worth millions of dollars) and a cheaply made modern violin is that the resonance frequencies of the Stradivarius are more distinct, have a larger amplitude and are closer to the frequencies of the notes played by the string. These resonances make it easier to reach a given note and the note being played sounds louder. (It should be noted, however, that blind listening tests can rarely distinguish between a Stradivarius and a well made modern instrument selling for tens of thousands of dollars.)

For most stringed instruments the ***bridge*** transmits vibrations from the string to the body of the instrument (electrical pickups will be discussed later). Many guitar and violin bridges are carved with interesting shapes. This makes them more flexible so that they are closer to the resonance frequencies of the string and body. This also means, however, that they can act as filters since they do not transmit some frequencies as easily. Some bridges are slightly rounded so that the length of the string changes slightly as the string vibrates. This will also affect the frequencies emitted by the instrument. In the picture on the left below the bridge of a double bass is the light colored piece. On the right is the bridge of a modern guitar.

Figure 10.3.5.1

Figure 10.3.5.2

One problem that occurs for some stringed instruments, particularly the cello, is an unwanted resonance of the bridge, string, and body called a ***wolf tone***. The result is an unpleasant, wavering note. For certain notes, energy from the string causes a resonance in the bridge and body that feeds back into the string, making the string hard to control. Beginning performers often have more trouble controlling these unwanted resonances than experienced musicians. Various design changes have been tried to avoid this problem but it is very difficult to construct an instrument with resonances only where desired and no others.

When the string vibrations are up and down relative to the bridge, energy is transmitted to the body more efficiently so the sound dies away more quickly. If the string is plucked in a direction parallel to the body the sound is softer but lasts longer because the bridge is less efficient in transmitting the vibrations. The schematic below shows the effect of the direction in which the string is plucked or bowed (blue arrows) relative to the body surface of the instrument.

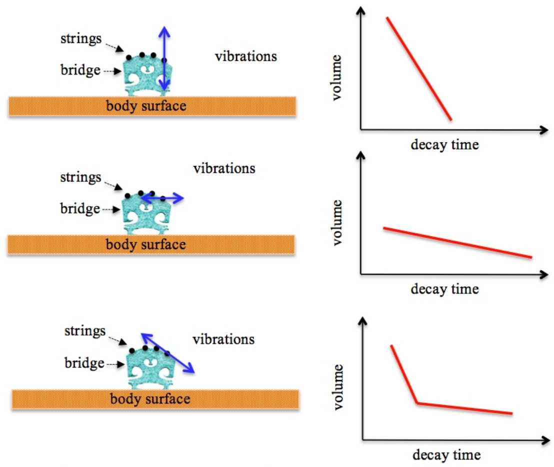


Figure 10.3.5.3

The top surface of most stringed instruments is smooth and flat but guitars, violins and other instruments have reinforcing bars on the inside, that strengthen the top and carry vibrations from the bridge area to the rest of the surface. For guitars this is called ***guitar bracing***, for violins the reinforcing bars are called ***bass bars***. Instrument makers generally try to construct these so that they do not cause unwanted resonances while reinforcing desired frequencies in the instrument but this is largely a trial and error process. The knowledge gained from this process has been handed down over many generations and there are at least half a dozen major bracing styles for guitars with many small variations of these. The following picture shows the bracing inside a guitar.

Figure 10.3.5.4

Many hollow bodied stringed instruments also have a ***sound post***. This is a small dowel inside the instrument wedged between the top to the bottom surfaces. The purpose is to conduct vibrations from the top surface to the bottom. Placement and size of the sound post is also determined largely by trial and error. Both sound bars and sound posts affect the resonance frequencies and overall sound of the instrument. Placing the sound post closer to the neck of a violin makes the tones louder; moving it further from the neck changes the tonal quality to be more rich by transmitting more overtone frequencies from the top to the bottom of the instrument. The sound post transmits lower frequencies more effectively with the result that high frequencies are emitted mostly by the top of a guitar or violin but low frequencies are emitted from both sides.

In the schematic below the cross section of a violin is shown. The bass bar runs the length of the violin and the sound post connects the top to the bottom surface. Notice that bass bar and sound post are positioned so that as the bridge rocks back and forth due to vibration of the strings the bridge will transmit energy to both the bass bar and the sound post. The asymmetric location of the sound post also means the top surface is more likely to vibrate with the left side out of phase with the right side (left side going up at the same time the right side is going down). This affects the Helmholtz resonances of the cavity inside.

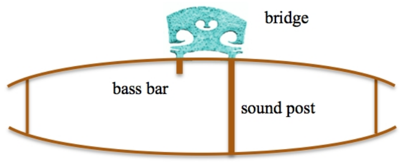


Figure 10.3.5.5

The Helmholtz resonances affect the volume of certain frequencies being produced. If the air is coming out of the f hole due to a Helmholtz resonance at the same time the instrument body top is vibrating downward (the two motions are 180 degrees out of phase), the exiting air will end up at a higher pressure and thus a higher volume. This effect is stronger for lower frequencies because the Helmholtz resonances tend to be lower frequency, around 270 Hz for a violin.

Another factor that affects the sound of a string instrument is the material of which it is constructed and the finish. More expensive guitars and violins are made from wood that is picked to be very dense and have very uniform and fine grain. The more uniform the wood grain the less likely there will be an unwanted resonance. It has been claimed that the famous Stradivarius violins have a special sound in part because the wood was treated with special chemicals. The process of how they were made was a secret that was lost hundreds of years ago and has not been replicated.

Strictly speaking keyboard instruments such as the piano, harpsichord and clavichord are classified as percussion instruments because there is a mechanism that strikes or plucks the string. In the clavichord, developed in the 15th century, a small metal piece called the ***tangent*** hits the string (sometimes two strings) when a lever is pressed by the performer. The harpsichord was developed in about a century later and uses a quill (originally from the feather of a bird) called the ***plectrum*** to pluck one or two strings when the key is pressed. A spring mechanism moves the plectrum out of the way while the mechanism returns to its starting point (see Wikipeda for details).

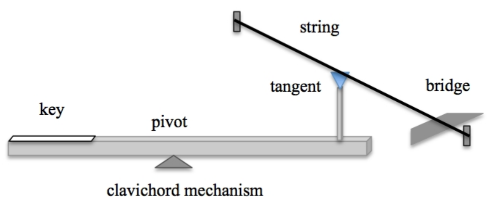


Figure 10.3.5.6

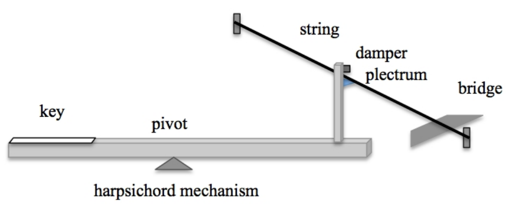


Figure 10.3.5.7

The piano was developed in the early 1700s This instrument overcame several limitations of the earlier harpsichords and clavichords. Because it was significantly louder it was originally called the 'piano forte' (forte is Latin for loud). Below is a very simplified diagram of the mechanism of the strike action of the piano (more details can be found in this animation of a piano mechanism). When the key is pressed a series of levers acts to throw the hammer upward so that it hits the string. The back check keeps the hammer from bouncing back and striking the string twice. As long as the key is held down another set of levers keeps the damper off the string.

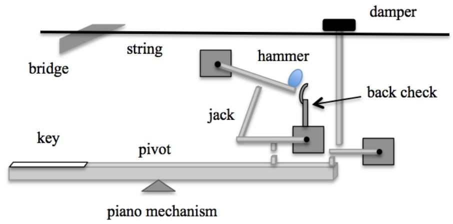


Figure 10.3.5.8

The clavichord, harpsichord and piano sound different in large part because the strings are set into vibration by different mechanisms. As we saw above, the initial shape of the string when it is plucked or struck determines which harmonics are more present. For the piano the harmonic that has an anti-node where the hammer strikes is dampened out because the hammer contact prevents the string from vibrating at that location, forming a temporary node. The piano also has a number of other innovations that give it a distinct sound. The mechanism of the piano allows the hammer to strike the strings at a different angle depending on how hard the key is pushed. The surface of the hammer is made of felt and different regions have slightly different stiffnesses of felt. So pushing a key harder not only causes the hammer to strike harder, it also causes a stiffer part of the hammer to strike the string, giving a louder sound and activating other harmonics.

Pianos also have multiple strings for some notes the piano plays. The treble pitches have three identical strings, the tenor range uses two strings, and the bass notes have only one string. There may be more than 230 strings used to produce the usual 88 notes on a piano. For treble notes all three strings are usually struck by the hammer but when the left foot petal is held down the hammer shifts so that only two strings are struck. In this case the third is activated by vibrations traveling from the other two through the frame that holds the strings. Since the strings can vibrate in different directions at the same time, transmitting energy at different rates and to different resonances in the piano body, the sound of a piano is much richer than instruments that sound only one string at a time. In the first picture below the strings are all moving up at the same time. In the second picture the third string is moving down while the other two are moving upwards. In the third picture the third string is moving in a direction perpendicular to the motion of the other two strings. Obviously there can be other combination of vibrations.

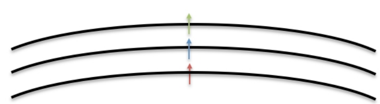


Figure 10.3.5.9

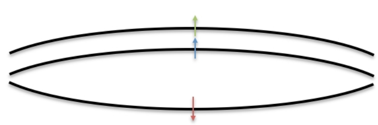


Figure 10.3.5.10

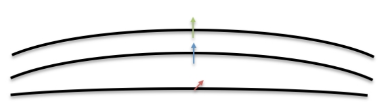


Figure 10.3.5.11

The piano also plays a larger range of notes than the harpsichord or clavichord. To archive the high notes the tension must be very high in the strings, as much as 1000 N for some strings. To withstand these forces the strings in a piano are attached to a cast iron frame. For strings under this much tension even slight displacements from equilibrium (when the hammer strikes the string) cause a significant change in the tension. This means that striking the note harder changes the pitch slightly because the string is stretched more. The effect is greater on higher harmonics because a higher harmonic has more bends along the length of the string, increasing the tension even more. The result is that the overtones are no longer exactly harmonic.

A second problem for a stringed instrument that plays very low notes is that for a given string density and tension, the strings have to be unreasonably long to play the lowest notes. In order to shorten the length of string needed for low notes the strings are made more dense by wrapping them either once or twice. Even with these adjustments a grand piano is quite large. Upright pianos put the strings into layers so that more strings can be packed into a smaller space. The wires shown below are double wrapped piano wires.

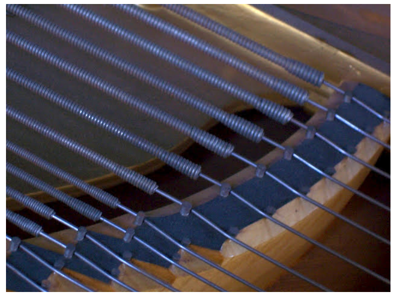


Figure 10.3.5.12

As with all stringed instruments the body of the piano acts as a resonator to amplify the sound of the vibrating strings (of course, at the cost of having the string vibrate for less time so that conservation of energy is obeyed). Pianos have a ***soundboard*** attached to the frame that holds the strings. Much like guitar bracing, the soundboard also has bracing that helps transmit the sound through the entire surface. The top of the piano serves two purposes. One is to reflect sound from the soundboard out towards the audience. The piano top also has resonant modes, just like those of a guitar or violin body that affect the frequencies being transmitted.

Video/audio examples:

- Information from Wikipedia and YouTube on a few well known stringed instruments:
  - The violin family.
  - The violin.
  - The guitar.
  - The harpsichord.
  - The clavichord.
  - The harp.
  - The following web sites have pictures, sound samples and information about modern musical instruments used in bands and orchestras: one, two, three.
  - Viola de gamba performance.
  - Clavichord performance.
  - Harpsichord performance.
  - Piano performance.
- How a guitar is made.
- Chladni patterns of a violin shape.
- Helmholtz resonance in a guitar body.
- Animation of violin vibration modes.
- Violin wolf tone.
- Project for class by Kelsey Krueger on the cello wolf tone.
- An NPR report which says that tests show even experienced musicians can't tell the difference between a Stradivarius and an ordinary violin.
- Animation of a piano mechanism.
- Simulation of a piano hammer hitting a string from Wolfram (you may need to download their plug in to play with this demonstration). Notice that the seventh harmonic is missing in the Fourier spectrum. This is because the hammer itself dampens this harmonic by touching the string where the anti-node would be.
- Frequencies of Piano notes.
- Formula for piano string frequencies.
- Some information and YouTubes about aeolian harps. Two examples: Ordinary harp in the wind and an aeolian harp in a fan. Why does this happen? Vortex shedding. More than you want to know. Notice that there is a natural frequency of the air oscillations going past the cylinder that depends on air speed and viscosity, and size of the object. If this vortex shedding frequency equals a natural frequency of the string, it will drive the string to vibrate (resonance). Here is another explanation of and demonstration of vortex shedding.

Summary

Strings have harmonic resonances at multiples of the fundamental frequency. The bridge of a stringed instrument transmits these harmonics to a surface that has its own resonance frequencies which are usually not harmonic. The surface can move more air than a string and so acts to amplify the sound of the string (at the expense of a shorter vibration time for the string). If the instrument has a hollow body the Helmholtz resonances of the cavity contribute to the frequencies that are present. The particular combination of string, surface and body cavity resonances give each individual instrument its timbre.

Questions on Stringed Instruments:

1. What are the three parameters that determine the fundamental frequency of a vibrating string?
2. What is the relationship between the speed of a wave on a string and the fundamental frequency?
3. What are harmonic frequencies?
4. If you have a fundamental frequency of 102 Hz, what are the next three harmonic frequencies above this?
5. What is the distinction between a node and an anti-node?
6. How many nodes and antinodes are there in the 3rd harmonic? The 6th?
7. Explain the difference between harmonics and overtones.
8. What is the relationship between the fundamental frequency of a vibrating string on a stringed instrument and pitch?
9. What effect does plucking a string at different locations have on the harmonics that are formed?
10. What is the difference in vibrational modes between a string that is plucked and one that is bowed?
11. Why is it that a vibrating string in air doesn’t produce as much sound as a vibrating string attached to a surface?
12. A string attached to a surface is louder than if it is not. How does this fit with conservation of energy?
13. Strings have harmonic overtones but surfaces typically do not. How does this effect the sound produced by a stringed instrument?
14. What is holographic interferometry and how does it show the vibrational modes of a surface?
15. Are Chladni plate resonance frequencies harmonic? Explain.
16. Explain Q-factor in terms of surface resonances for a stringed instrument.
17. Why do you want the body of a stringed instrument to have low Q resonance frequencies?
18. Why is uniformity in wood grain for a stringed instrument such as a guitar important with respect to resonance?
19. What does the bridge of a stringed instrument do?
20. What is the purpose of the sound post in a stringed instrument?
21. What is the purpose of bracing in a guitar?
22. What is the purpose of the bass bar in a violin?
23. What is the Wolf tone, and what causes it?
24. Which can play a note longer, a harpsichord or a piano and why?
25. Why is the piano louder than the harpsichord?
26. Using a reliable source, find out how much tension do the strings of a piano have. Using this, explain why pianos had to wait until the development of cast iron before the first piano was created.
27. What is a Helmholtz resonance? Explain what this has to do with acoustic string instruments.
28. Why do acoustic stringed instrument have hollow bodies with holes?
29. Find a reliable source and discuss the history of the f-hole.
30. Violins and guitars are both stringed instruments with hollow bodies and can play some of the same notes. Why is the timbre different between these two instruments?
31. On the Indian instrument, the sitar, why are there strings that are not plucked? Explain how this works.

CHAPTER OVERVIEW

11: Tubes

In this chapter we start with resonance in a tube. Once the basic behavior of pressure waves in a tube are explained we look at various instruments that are basically tubes, such as flutes, brass, woodwinds, and pipe organs.

Key Terms:

Displacement node versus pressure node, tube resonance, tube harmonics, impedance, reed, fipple, edge tone, embouchure, free reed aerophones.

11.1: Standing Waves in a Tube

This section shows a simulation that compares the fundamental, second, third and forth harmonics of standing waves on a string with standing waves in a tube. Notice that for a tube open on both ends the displacement nodes occur where the string has nodes and the displacement anti-nodes in the tube occur where the string has displacement nodes. The pressure nodes in a tube open at both ends occur in the same place as the string nodes.

11.1.1: Tube Resonance

In a tube based instrument waves are created at one end of the tube by something that vibrates and travel to the end of the tube and reflects as show in the animations at the bottom of this page. A standing wave is created by the waves traveling in each direction. Since the wave is traveling in air it will move at the speed of sound in air at the ambient temperature. It will also be true for waves in a tube that $v = f\lambda$ where f is in Hertz and λ is the wavelength in meters. As was the case for strings, the length of a tube determines the frequency of a standing wave in the tube. There are several complications, however, depending on if one or both ends are closed or open.

Strings always have displacement nodes at each end since the string is fixed there and cannot move. For a tube, however, if an end is open the air can move freely and so there is a ***displacement anti-node*** (the point of maximum air movement) at that end.

The pressure in the tube is different than the displacement, however. Since the air can move at the ends it does not build up any pressure so at these locations (the ends) there will be a ***pressure node*** where the pressure stays constant. At the center where the air cannot move the pressure has large oscillations and there is a pressure anti-node. The following two graphs show the displacement wave and the pressure wave of the fundamental for a tube open at both ends.

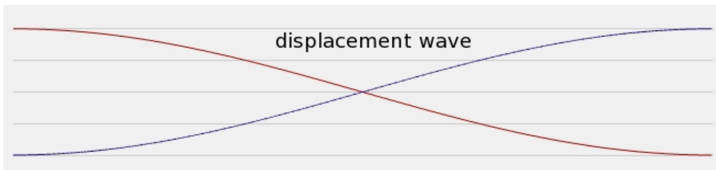


Figure 11.1.1.1

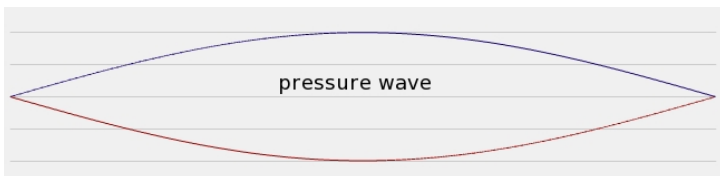


Figure 11.1.1.2

For a tube of length L to have a pressure node on each end it has to be the case that $L = n\lambda/2$ where n is a whole number. In other words, you can have half a wave in the tube ($n = 1$), one wave ($n = 2$), one and a half waves ($n = 3$), etc. But you never have any other fraction of a wave because that would require not having a node at both ends. The pressure in a tube *that is open on both ends* gives harmonics *exactly* the same as a string (look back to the previous chapter to verify this):

Frequencies for a tube open on both ends		
Harmonic number	**Wavelength**	**Frequency**
$n = 1$	$\lambda_1 = 2L$	$f_1 = v/\lambda_1$
$n = 2$	$\lambda_2 = L = \lambda_1/2$	$f_2 = v/\lambda_2 = 2f_1$
$n = 3$	$\lambda_3 = \frac{2}{3}L = \lambda_1/3$	$f_3 = v/\lambda_3 = 3f_1$
$n = 4$	$\lambda_4 = \frac{1}{2}L = \lambda_1/4$	$f_4 = v/\lambda_4 = 4f_1$
n	$\lambda_n = \frac{2}{n}L = \lambda_1/n$	$f_n = v/\lambda_n = nf_1$

Table 11.1.1.1

For many musical instruments that are made of a tube, the tube is closed at one end but opened at the other. If the end is closed the air cannot move so the pressure fluctuates and there is a pressure anti-node (see the simulation below). This same location is a displacement node because the air is not moving. The graphs below are of the displacement and the pressure waves for a tube closed on one end.

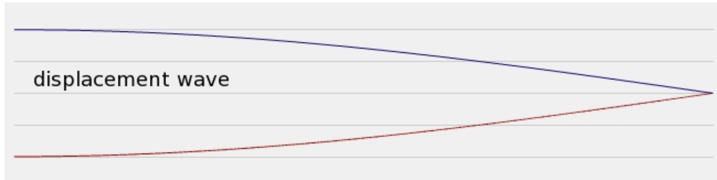


Figure 11.1.1.3

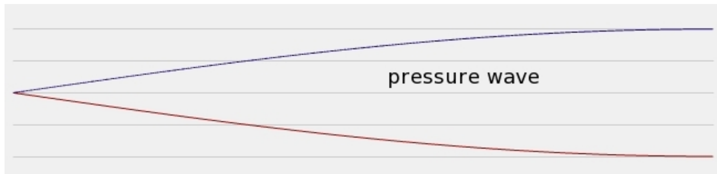


Figure 11.1.1.4

For a tube of length L to have a pressure node on one end but a pressure anti-node at the other the lowest possible wavelength is given by $L = n\lambda/4$ where n is a whole number. Notice this is half the lowest wavelength available to a pipe with two open ends. The requirement of having a pressure anti-node at the closed end means the even numbered frequencies will be missing from a tube closed at one end as shown in the following graphs of the first three harmonics available to a pipe closed on one end.

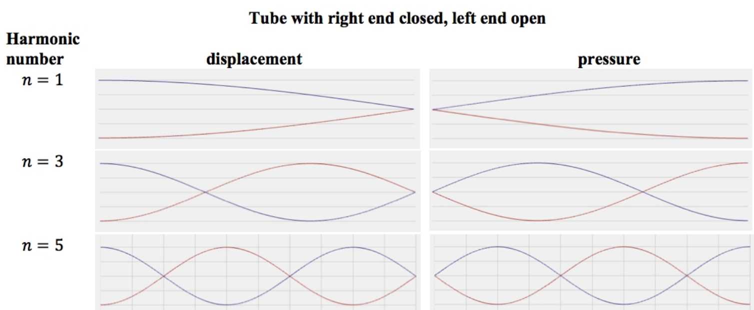


Figure 11.1.1.5

So for a tube open on *both* ends the available frequencies are the fundamental, $f_1, 2\times f_1, 3\times f_1,$ etc. But for a tube that is *closed* on one end only *odd* multiples of the fundamental f_1 are available: $3\times f_1, 5\times f_1,$ etc. The following table gives the frequencies available to a tube that is closed at one end.

Frequencies for a tube open on one end		
Harmonic number	**Wavelength**	**Frequency**
$n = 1$	$\lambda_1 = 4L$	$f_1 = v/\lambda_1$
$n = 2$	doesn't exist	missing
$n = 3$	$\lambda_3 = \frac{4}{3} = \lambda_1/3$	$f_3 = v/\lambda_3 = 3f_1$
$n = 4$	doesn't exist	missing
n (odd)	$\lambda_n = \frac{4}{n}L = \lambda_1/n$	$f_n = v/\lambda_n = nf_1$. n odd

Table 11.1.1.2

If you compare the above table with the one for a tube open on both ends you will notice two things. For a given length, L the tube with one end closed has a lower fundamental frequency (longer wavelength) and has only odd frequencies. This affects the available frequencies of the instrument and so affects the timbre. Most tube instruments are closed on one end because the musician playing the instrument has their mouth over one end. Pan flutes are interesting in that the bottom end is closed and the performer blows across the open end.

As mentioned earlier, the auditory canal of the ear acts as a tube with one end closed. In the chart for the harmonic $n = 1$ we have $L = n\lambda/4$ which determines the fundamental wavelength for the auditory canal. The fundamental frequency for this tube turns out to be around $3500\ \mathrm{Hz}$ and this is where human hearing is the most sensitive as we saw in the Sound Intensity Level chart in Chapter 8.

Unless the pipe is very narrow compared to its length, a slight correction to the above formulas needs to be made. We know a standing wave inside a tube is formed from waves being reflected from the ends (just like standing waves on a string). It turns out that for an open end, the wave doesn't reflect exactly at the end of the tube. In order to "feel" the pressure difference at the end and be reflected it goes a little past the end before being reflected. The effect is as if the pipe is just a little bit longer and this extra length depends on the radius of the tube. For a pipe of radius r, the extra amount turns out to be (after some sophisticated calculations) $0.61r$. The formulas for the tube open on one end can be corrected by replacing L in the table above with $L + 0.61r$. For a single open end the fundamental will be given by $\lambda_1 = 4(L + 0.61r)$. In the case of two open ends the extra length is added twice so the fundamental is $\lambda_1 = 2(L + 1.22r)$.

Unlike a stringed instrument where the speed of the wave on the string can be changed by changing the density and/or tension, the speed of sound in air at a given temperature is fixed. The fundamental frequency of a string can be changed by changing density, tension or length but the only way to change the fundamental frequency of a tube instrument is to change the length of the tube. There are several ways to do that, as we will see below.

Video/audio examples:

- Actually, the previous paragraph is not quite right; there is a way to change the frequency of a tube without changing its length. If the density of the gas inside changes, the speed of sound will change and this changes the frequencies in the tube. Tube instruments will be slightly out of tune on a cold day compared to a hot day due to the change in the density of air with temperature. The frequency will also change if the tube is filled with a gas other than air as shown in this YouTube of trombone and trumpet filled with different gasses.
- Animations of open ended and closed ended tubes.
- Visualization of sound in a tube using a Kundts Tube (filled with Styrofoam balls).
- Visualization of sound in a tube using a Rubens' Tube.

11.1.2: Standing Wave Simulation

The following simulation compares the fundamental, second, third and forth harmonics of standing waves on a string with standing waves in a tube. Notice that for a tube open on both ends the displacement nodes occur where the string has nodes and the displacement anti-nodes in the tube occur where the string has displacement nodes. The *pressure* nodes in a tube open at both ends occur in the same place as the string nodes.

For a tube closed at one end (as is the case for most musical instruments based on vibrating columns of air) the even harmonics cannot exist. A tube closed on one end has a different timbre because the even harmonics are missing.

Simulation Questions:

1. Play the standing wave simulation for the case of the fundamental. The length of the string is $3.14\ \mathrm{m}$. What is the wavelength of the fundamental?
2. Describe the fundamental of the tube simulation (bottom). Where do the dots (representing air molecules) move the most? Where do they form a displacement node? Assuming the tube is the same length as the string, what is the fundamental wavelength of the tube open at both ends?
3. Use the time in the simulations to find the period and calculate the frequency of the fundamental for both simulations. For a musical instrument this would be the frequency of the tone being sounded by the instrument when it plays its lowest note.
4. What is the wave speed of each of the component waves making up the fundamental (the speed determined by $v = f\lambda$)?
5. Now click the box for a tube closed at one end. What is the wavelength of the fundamental for a tube closed at one end? How is this different for the case of the tube open at both ends?
6. Reset the simulation and look at the second harmonic for the string and tube open at both ends. What is the wavelength and frequency of the second harmonic/first overtone for the string and tube open on both ends? What is the speed of the component waves?
7. Try the third and fourth harmonics for the string and tube open at both ends. What are the wavelengths and frequencies of these waves? What are the speeds of the component waves?
8. The formula for the wavelength as a function of the length of the string or open tube is given by $\lambda = 2L/n$ where n is a whole number and L is the length of the string. Verify this relationship with the numbers you got in the previous questions.
9. Now check the box for a pipe with closed end simulation and examine the harmonics. Describe the difference in the node and anti-node pattern. What are the wavelengths for these cases? The formula for the frequencies of a tube closed at one end are given by $\lambda = 4L/n$ where n is an odd whole number. Verify this relationship with the numbers you got in the previous questions.
10. Flutes are basically pipes with openings at both ends but clarinets, trumpets and trombones are basically tubes that are closed at one end. Why does this make a difference in the frequencies each instrument plays?
11. Pressure anti-nodes occur at places where the air is not moving (displacement nodes). What would be the effect of cutting a hole in the tube at the location of a pressure anti-node? Would the standing wavelength be affected? (This is the basis behind using finger holes in wind instruments to play different frequencies.)

11.2: Reflection of Waves at a Boundary

Waves on a string form standing waves because the wave reflects from each end of the string where there is a fixed node. How do standing waves form in a tube full of air? This section shows that waves reflect from the end of a tube and that this reflection can be of two types, depending on whether the boundary is 'soft' or 'rigid'.

11.2.1: Reflection of Waves

Waves on a string form standing waves because the wave reflects from each end of the string where there is a fixed node. How do standing waves form in a tube full of air?

11.2.2: Reflection Simulation

We already know that a wave which moves from one medium to another can change speed. In many cases this causes a bending of the wave called refraction. A second effect also occurs when a wave goes from one medium to another. There is usually a partial reflection of the wave, depending on how different the two mediums are. Reflection can be of two types, depending on whether the boundary is 'soft' or 'rigid'. If the wave is going from a more rigid medium into a softer medium the reflected wave will have the same phase as the incoming wave. Waves in a soft medium reflecting from a boundary with a stiff medium will change phase by 180 degrees. The simulation below shows this effect.

Simulation Questions:

1. Run the simulation to see how a Gaussian pulse reflects off the two different boundaries. How is a pulse reflected from a fixed boundary different from one reflected from a free boundary?
2. Now check the sine wave check box to see what happens when a sine wave hits the two types of boundaries. What is the end result in these cases? (Hint: Go back to chapter seven and the Adding Sinusoidal Waves simulation where you added two identical sine waves traveling in opposite directions.)
3. Although the reflecting sine waves in both cases interacts with the incoming wave to form a standing waves there is a slight difference between the two. Which case has a node at the boundary and which has an anti-node at the boundary?

11.3: Impedance

But why do sound waves reflect from the open end of a tube or when the size of tube changes abruptly? The resistance to the movement of a wave crossing a boundary from one medium into another is called impedance and occurs for waves on a string, sound waves in air and electronic signals in a circuit. When a wave tries to travel from a medium with one impedance to a region where the impedance is different, there will be a partial reflection.

11.3.1: Impedance

But why do sound waves reflect from the open end of a tube?

The resistance to the movement of a wave crossing a boundary from one medium into another is called ***impedance*** and occurs for waves on a string, sound waves in air and electronic signals in a circuit. When a wave tries to travel from a medium with one impedance to a region where the impedance is different, there will be a partial reflection. The reflection means that not all the energy of the wave is transmitted to the new medium.

There are several ways to try to avoid or minimize the effects of impedance. In general the idea is to ***match impedance*** between the two regions. For example, the speakers of a stereo system are chosen to match the impedance of the amplifier/tuner component. In electrical components impedance is measured in *ohms* (Ω) so if a speaker has an impedance of 10Ω an amplifier with 10Ω is chosen (or vice versa). If the impedance was not matched some of the energy from the amplifier would reflect back and not get to the speaker, making the system less efficient at turning electrical signals into sound. The three bones in your middle ear are responsible for matching impedance between the air outside your ear and the fluid in your cochlea as explained in this video on impedance matching in the ear.

Tube instruments have an impedance problem in that sound inside the tube will reflect off the ends, even an open end. Air inside the tube is confined by the walls of the tube but air outside is not. This difference in impedance causes some sound to reflect back into the tube at an open end and it is this reflected wave that sets up a standing wave inside the tube. Reflection is strongest if the tube diameter is less than a quarter of the wavelength of the sound wave in the tube. Open end reflection is weaker as the diameter gets bigger compared to the wavelength.

On the one hand this reflection is what causes the tube to have a fundamental frequency (resonance) but this also means less sound gets out of the instrument. Flutes are not corrected for impedance miss-match and are the softest woodwind instrument. Most other woodwinds are louder because they have bells that act as impedance matching devices between inside the instrument and outside. Brass instruments have even larger bells and so have better impedance matching and are even louder than woodwinds on average. As we will see, the bell of a wind instrument also affects the overtones present (and therefor the timbre) but the main function of a bell is to help sound waves exit the instrument by matching the impedance.

The mathematical symbol for impedance is Z, measured in ohms. In some references you will see the inverse of this number, the ***admittance*** which is defined as $Y = 1/Z$, measured in *siemens*. Here is a more complete definition of acoustical impedance. Here is a Java Applet that calculates impedance (reflected and transmitted sound energy) when sound is moving from one substance to another.

11.3.2: Impedance Simulation

In many cases a wave colliding with a boundary will partially reflect and be partially transmitted. The kind of wave reflected and the amount of energy transmitted depend on the properties of the material on either side of the boundary. This animation again simulates a string as a row of individual masses connected by invisible springs. In this case the mass of the string is different on the left as compared with the right. A similar reflection occurs when a sound wave confined to the inside of a tube encounters the open end of a tube where the air can move freely.

Simulation Questions:

1. Run the simulation and describe what happens when a pulse goes from a light string to a heavy string.
2. Click the 'Heavy to Light' button, run the simulation and describe what happens when a pulse goes from a heavy string to a light string.
3. In which case is the reflected pulse inverted? Based on what you learned about reflections from soft and hard boundaries in the previous simulation, explain this result.
4. In which case is the reflected pulse larger than the transmitted pulse?
5. In which case is the reflected pulse faster? Based on what you learned about how physical properties determine the speed of a wave, explain this result.

11.3.3: Woodwind Instruments

Probably the earliest instruments were simple flutes. The earliest known flute comes from Europe and is dated at about $35,000$ years old. Other woodwind instruments include the clarinet, oboe, bassoon and saxophone families of instruments.

In woodwind instruments (as in all instruments) there is something that vibrates to create sound and a resonating body which selects a fundamental frequency and overtones and amplifies them. From the above discussion we know that the tube itself is the resonating body. What creates the initial vibration? In the clarinet and saxophone family of instruments a single reed, clamped to a mouthpiece vibrates to create the initial sound. Most reeds, shown on the left, below, are made from a type of cane plant.

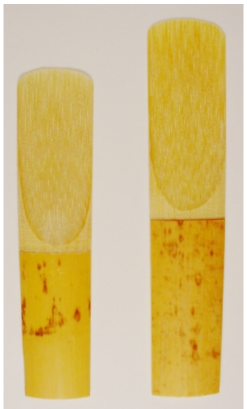


Figure $11.3.3.1$

Figure $11.3.3.2$

On the right above is a picture of a double reed. This is the vibrating device for oboes, bassoons and bagpipes. In most reed instruments the reed or double reed is clamped in the mouth so that the mouth can change the frequencies being produced. The adjustment of the performer's mouth to correctly produce the initial sound is called ***embouchure***. Even with this control, however, a reed mouthpiece without the instrument sounds basically like a duck call (and in fact many duck calls are based on a simple reed). Most reed instruments are blown closed, meaning that the air flowing past the reed tends to force it shut after which it springs open due to its own stiffness.

There does not appear to be anything that vibrates on a flute but we know vibrations are the source of all sound. So what vibrates? Most flutes have a sharp edged hole called a ***fipple*** that splits a stream of air. Initially the air stream goes into the flute cavity. As pressure builds up the stream will suddenly shift to flow to the outside of the flute. When the air pressure inside drops the stream re-directs back into the cavity. It is this back and forth motion of the air (which occurs very rapidly) that is the source of vibration for a flute. The sound produced (before resonance with the body of the flute) is called an ***edge tone***.

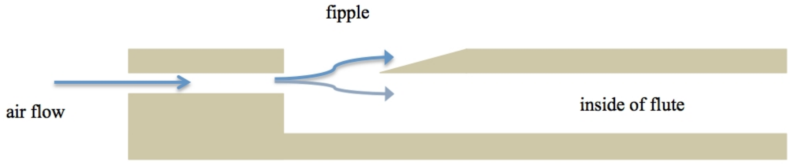


Figure $11.3.3.3$

Other kinds of flutes, including pan flutes and the modern orchestral flute, have a hole across which the musician blows. The edge of the hole acts the same way as the edge of a fipple with air first going into the flute, then building up pressure to make the flow go out. This rapid back and forth air flow causes the vibration that makes the initial sound. The picture on the right, below, shows an air stream being divided by a sharp edge. Notice the turbulent pattern generated as air first goes above and then below the edge.

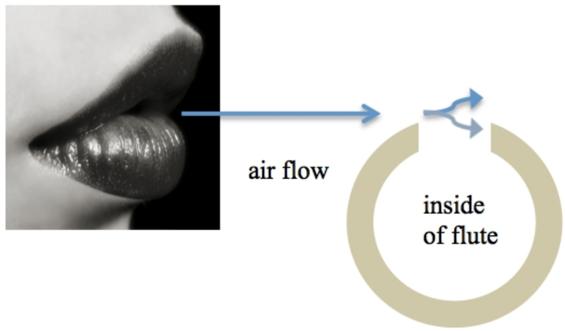


Figure 11.3.3.4

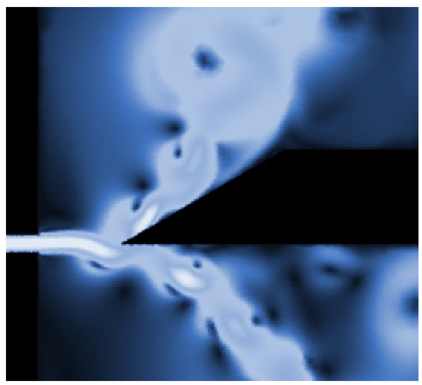


Figure 11.3.3.5

One way to change the fundamental frequency of a tubed instrument is to put holes in the tube which can be closed or opened with the fingers. As shown in the figure below, a hole allows a displacement anti-node of air to occur at the hole location, changing the resonant frequency. Flutes, clarinets and saxophones all use this method to change the fundamental frequency being played.

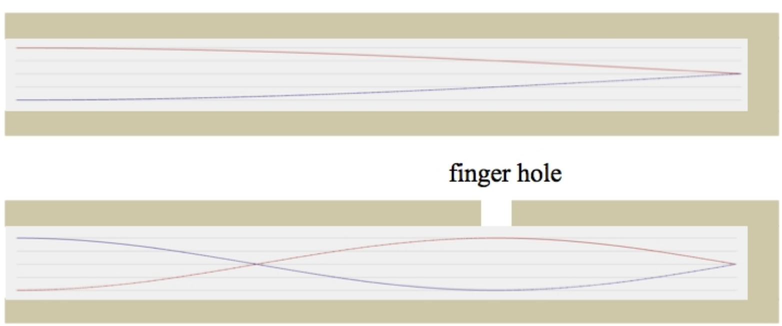


Figure 11.3.3.6

There are several problems with putting holes in a tube to get different resonant frequencies. The size of the hole affects the shape of the wave inside the tube which changes the overtones that are present. The result is small errors in pitch that vary with the range and volume of the notes being played. Early flutes and recorders from the Renaissance had holes of varying size to correct for these errors. Modern flutes have hole sizes that get slightly larger at the end furthest from the fipple which give more uniform changes in pitch. By using mechanical keys to cover the holes the hole size can be adjusted for sound quality rather than convenience of the performer. Covering and opening holes also changes the amount of impedance mismatch between inside and outside the instrument, so that different notes may sound slightly louder or softer, depending on the fingering.

A second problem for tube instruments with holes is that the human hand is of finite size. For a longer tube this becomes problematic; a normal hand cannot reach the hole locations necessary to make a specific note. Most modern instruments use levers called keys attached to pads on springs to cover holes that are further away than can be reached by a normal sized hand. A second reason to use keys is that more holes can be used. Human hands can only cover ten holes at one time because we only have ten fingers. To be capable of playing all 12 notes of an octave some keys are normally closed unless pressed to open them.

Flutes tend to have uniform diameter along their length but other wind instruments do not. The diagrams below show the difference in cross sectional shape between a flute, clarinet and saxophone. Bassoons are shaped more like a flute (the same cross section along the length) whereas oboes are more shaped like a saxophone with an increasing diameter and a bell on the end. Notice that the saxophone and oboe are flared along their entire length, the clarinet is flared only at the end and the flute has no flaring. As discussed above, the flaring changes the impedance mismatch between waves inside the instrument and sound transmitted away from the instrument. Notice that the lengths are different for these instruments as well. Not shown in the cross section diagram of the flute is the cork which provides an adjustable end to the left of the mouth (or embouchure) hole of the flue. The cork allows the resonances inside the flute to be tuned so that reflections from cork end reinforce sound generated at the embouchure hole.

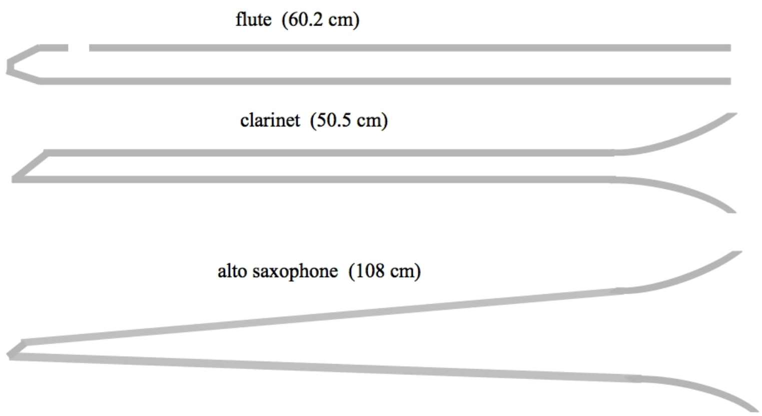


Figure $11.3.3.7$

What effect does the changing diameter have on the simple standing waves that we saw for a tube with constant diameter? The result, shown in the graphs below, is that the wave amplitudes are squeezed a little bit. Compare these results for the first few harmonics of a tube with constant diameter to those for the flute, above. We know from Fourier analysis that more overtones have to be present to get the more complicated wave shapes shown below. So a changing tube width will greatly affect the timbre of the instrument.

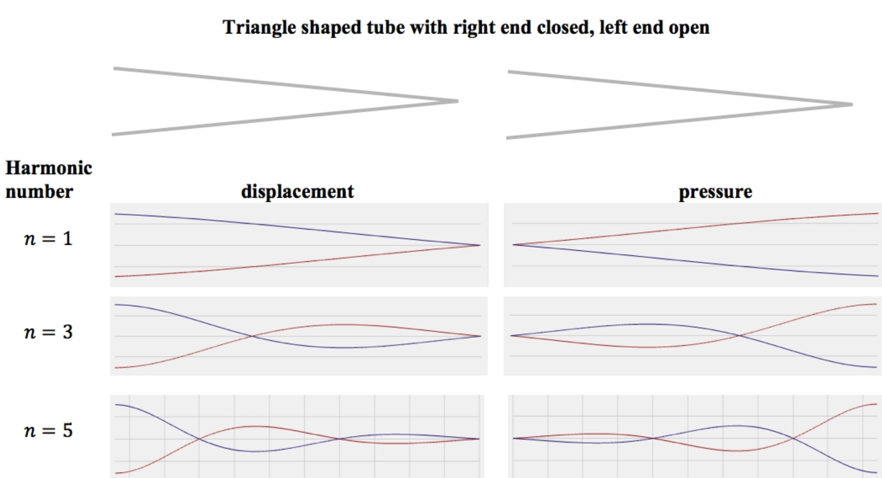


Figure $11.3.3.8$

Video/audio examples:

- Fipple turbulence video.
- Animation of a clarinet making sound.
- Bass flute performance.
- Tin whistle performance.
- Nose flute: history, modern.
- Wikipedia on flutes, clarinet, recorders, oboe, bassoon and saxophone family.

11.3.4: Brass Instruments

Brass instruments use the vibration of the player's lips with the proper embouchure to initiate the sound. As in the case of a reed instrument, buzzing the lips creates a broad spectrum of sound. Resonances in the instrument select one of these frequencies to be the fundamental of a note. Unlike most reed instruments in which the vibrating part is blown closed, for brass instruments the lips are blown open. Air rushing between the lips lowers the pressure and the lips close. Then pressure builds up and the lips open again. Blown open and blown closed vibrators both depend on the Bernoulli effect; moving air has lower pressure causing the reed or lip to close.

We know that tubes of different lengths have different resonant frequencies. Brass instruments change the length of the tube using one of two mechanisms. Slide trombones have a set of double tubes that slide over each other to make the length longer. Trumpets, french horns, tubas and other brass instruments use a set of valves to change the length of the tube. There are usually three and occasionally four valves on most brass instruments. The diagram below shows one of several ways to construct a valve that changes the effective tube length. Pushing the valve down causes the air path to detour through a section of pipe making the total path longer.

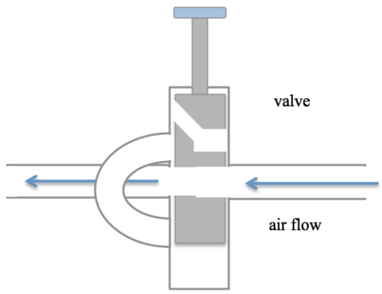


Figure 11.3.4.1

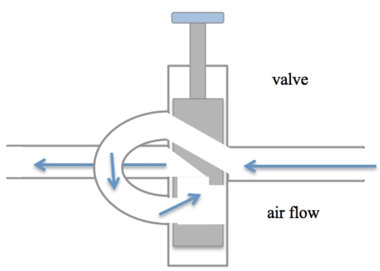


Figure 11.3.4.2

Brass instruments tend to have large bells and are louder than woodwinds on average because of better impedance matching as mentioned above. End effects due to the large bell also tends to shift the frequency spectra significantly from what would be expected from a simple tube. Although we expect a tube closed on one end to only have odd multiple overtones, the measured frequencies of a trumpet are harmonic. The table below compares the predicted frequencies of a simple tube of length 1.4 m (the approximate length of a trumpet) and the measured frequencies of a real trumpet. Notice that the predicted overtones are odd multiples of the fundamental but the real instrument has all harmonics.

Predicted	Measured
61.25	233.00
183.75	461.00
306.25	697.00
428.75	922.00
551.25	1158.00

Predicted	Measured
673.75	1388.00
796.25	1617.00
918.75	
1041.25	
1163.75	
1286.25	
1408.75	
1531.25	
1653.75	

Table 11.3.4.1

Why do the frequencies shift so much? For higher frequency overtones the standing wave extends further out into the bell region as shown in the diagram below. Since in effect higher harmonics 'see' a longer tube, the frequency of these overtones are shifted downward a little. The fundamental, because it 'sees' a shorter tube shifts up a little. The impedance mismatch for different overtones will also be different because of this effect. Higher frequencies escape more easily than lower frequencies because higher frequencies extend out further into the bell region of the instrument (as mentioned above, the amount of reflection at the end of a tube is affected by the diameter of the tube relative to the wavelength - so longer wavelengths reflect further out from the end of the bell). This means the overtones of a brass instrument are louder than in the case of a saxophone or clarinet. The effects of the bell on various overtones also gives brass instruments their unique timbre in addition to shifting the frequency spectrum.

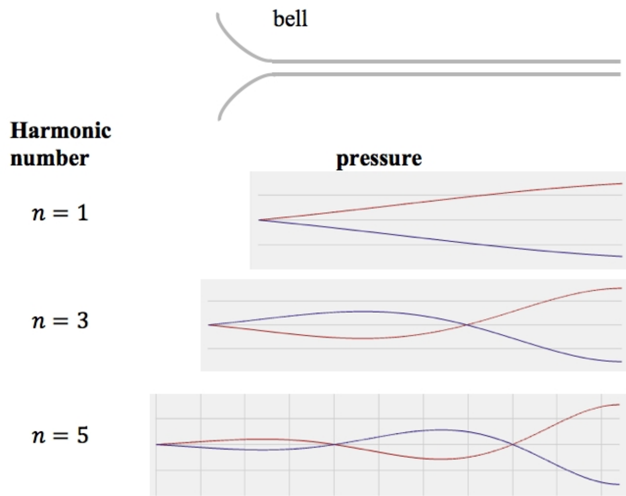


Figure 11.3.4.3

The mouthpiece also has an effect on the frequencies present in a brass instrument. The mouthpiece forms a small Helmholtz cavity with a volume resonance with a low Q-factor. As you may recall from Chapter Four on resonance, a low Q-factor means a very broad resonance. For modern instruments this resonances tends to weaken the lower harmonics (including the fundamental) so that the higher harmonics have a larger effect on the timbre.

As a final complication, brass instruments sometimes use mutes to change not only the volume but also the frequency spectrum radiated by the instrument. A mute does not seal up the bell end of a brass instrument but it does weakly produce extra frequencies corresponding to a tube that is closed on both ends. This changes the timbre of the instrument in addition to reducing the volume.

Video/audio examples:

- Slow motion of vibrating lips.
- How to buzz your lips.
- How valves work.
- How to make an instrument from a straw.
- Wikipedia on trombones, trumpets, french horns, tubas, and sousaphones.
- The following web sites have pictures, sound samples and information about modern musical instruments used in bands and orchestras: one, two, three, four.

11.3.5: Pipe Organs and Other Reed Instruments

Although technically a keyboard instrument, pipe organs are based on tubes. Instead of changing the length of a single pipe (with holes or other mechanisms) to get different frequencies there is a large collection of pipes of different lengths and a mechanism to blow air through them. The sound generating part may be either a fipple or a reed and some organs have different pipes with each type. The tubes may be made of wood or metal, may be open or closed on the end and have various shapes, as shown below. A variety of methods are used to tune an individual pipe by slightly changing its length or modifying the end of the pipe. As in the case of reed instruments different pipe shapes lead to different overtones that change the timbre of the note being played. Because a large number of pipes can be employed (the Wanamaker Pipe Organ has $28,604$ pipes), pipe organs can produce a wide variety of sounds. The longest pipes may be as long as 19.5 m which makes an 8 Hz pitch.

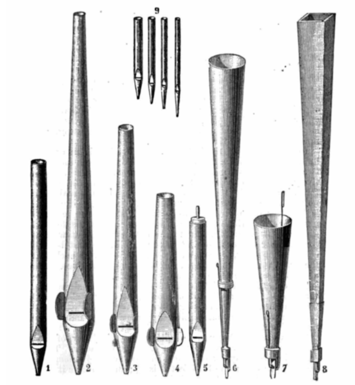


Figure $11.3.5.1$

A final class of instruments that use reeds but not tubes are called ***free reed aerophones***. These include harmonicas, accordions, bandoleóns and similar instruments. In these instruments a reed vibrates as air is forced over it, so they are blown open reeds, unlike clarinet or saxophone reeds. There is a reed with different tension to produce each note. The mouth and cupped hands form the resonating bodies for harmonicas. In accordions and similar instruments a set of bellows are used to force air over the reed and the air box forms the resonating cavity.

Video/audio examples:

- Wikipedia on free reed aerophones.
- A web site about harmonicas and other mouth blown, free reed instruments with sound samples.
- Performance of the Wanamaker Organ.
- Slow motion of harmonica reeds.
- One of many YouTube video of a didgeridoo. The didgeridoo is an instrument used by the aboriginal peoples of Australia and consists of a long tube of wood. The lips are buzzed to make a sound but the voice is also used. Because of the large diameter and length there are many interesting resonances that can be excited while playing the instrument.

Summary

As is the case for all instruments, tube instruments start with a vibrating sound source and use resonance to amplify the desired frequencies. For woodwind instruments a single or double reed vibrates. For flutes a fipple turns a smooth air flow into an oscillating flow which becomes the initial vibration. Woodwind instruments use holes to change the effective length of the tube to get different fundamental frequencies. Keys activated by springs and levers extend the reach of the human hand in opening and closing holes in the instrument. The cross sectional shape of the tube changes the overtone frequencies, giving the particular instrument its timbre. Brass instruments start with a pair of buzzing lips. The length of the tube is changed by either a sliding tube arrangement or a valve that shunts the air through tubes of different lengths. The impedance matching provided by the bell of a brass instruments enables them to produce more sound and also shift their overtone frequencies to give them their unique timbre.

Questions on Tubes:

1. How are the modes of vibration of a string similar to the modes of vibration (fundamental and overtones) of air in a tube and how are they different?

2. How are the modes of vibration of air in a tube that is closed at one end different than those of one open at both ends?
3. How is the length of a tube related to the fundamental frequency of vibration of air inside a tube open at both ends?
4. How is the length of a tube related to the fundamental frequency of vibration of air inside a tube open at only one end?
5. What is the difference between pressure nodes and displacement nodes in a tube instrument?
6. What are pressure anti-nodes and where are they located in a tube open on both ends for the fundamental and the first three harmonics?
7. How can the fundamental frequency of a tube instrument be changed?
8. Why does introducing finger holes in a flute affect the frequency played?
9. Holes on a tube instrument change the frequencies being played. List some problems with using holes for this purpose.
10. What is the general definition of impedance?
11. Why is some impedance necessary for tube instruments?
12. In a tube instrument, how are standing waves established in the tube?
13. What vibrates to produce the initial sound in each of the following instruments: Oboe, clarinet, flute, pipe organ, trumpet, trombone.
14. What is embouchure and why is it important?
15. Explain how a flute works without a reed to vibrate.
16. What is an edge tone?
17. Brass instruments are generally louder than woodwind instruments because they have bigger bells. Why does a bigger bell make the instrument sound louder (Hint: Think about impedance.)?
18. What other effect does a large bell have on a brass instrument besides making it sound louder?
19. Clarinets and trumpets are both tube instruments which can play the same note (same fundamental frequency) but they have very different timbre. What are some factors that cause the timbre to be different?
20. How do the following instruments change the pitch they are playing: Slide trombone, trumpet, flute, French horn, clarinet, saxophone, flute, pipe organ.
21. A didgeridoo is a long lube of wood, an instrument used by Aboriginals in Australia. How is the sound produce? Is there any impedance? Explain your answers.
22. Harmonicas, accordions, and bandoleóns all belong to what type of non-tube instrument? How do they work?

CHAPTER OVERVIEW

12: Percussion

Percussion instruments can be divided into two groups; those that make a distinct tone and those that do not. Pianos, xylophones, marimbas, steel drums and some tuned drums (for example tympani) are percussion instruments that have distinct tones. In other words, their overtones are harmonic. Cymbals, wood block, llaves and most drums, on the other hand do not produce a tone. This Wikipedia link has a long list of percussion instruments with links and explanations for each.

Key Terms:

Modes of vibration, mode numbers, nodal lines, strike tone, decay rate, anharmonic.

12.1: Percussion and Drumheads

This section allows you to see and manipulate the modes for a square drum head. You can change the modes using the sliders to change the mode numbers n and m. For a membrane there are nodal lines which do not vibrate similar to the nodes we saw on the string but now in two dimensions. You can rotate and enlarge the surface by dragging the mouse over the image. Just like the case for a vibrating string, more than one mode can be present on the two dimensional surface at the same time.

12.1.1: Vibrating Membranes

As we have seen string modes of vibration can be labeled by a single mode number, n, and the modes are harmonic. Each subsequent mode (or overtone) produces a frequency that is a multiple of the fundamental.

> **Note**
>
> There are slight variations to this rule when the string amplitude gets very large or if the string is very stiff but we won't worry about that here.

In the simulation of the square surface in Chapter 11 we saw that two mode numbers, n and m, are needed to specify a mode on a two dimensional surface. We also discovered that some of the modes were degenerate meaning two different combinations of n and m lead to the same frequency but in general the frequencies were not harmonic. This is generally true of most vibrating surfaces and membranes; they do not have harmonic overtones. As a result our ear-brain system does not detect a distinct pitch from most drums.

Video/audio examples:

- Pennsylvania State simulations of circular membranes.
- Wikipedia explanation of circular membranes (first admire the mathematics and then scroll to the bottom to see the simulations).
- YouTube of a membrane driven by a speaker.

12.1.2: Circular Drum Head Simulation

The following simulation allows you to see and manipulate the pure modes for a circular drum head (fixed around the edge). The simulation can be viewed as a two dimensional surface or as a surface in three dimensions. In the 3D view you can grab the surface with the mouse and rotate it. The amplitudes of the vibrations are exaggerated compared to a real surface to make them more visible. Increasing Δt makes the simulation run faster but also less accurately, particularly for higher mode numbers.

The shapes for a circular membrane are given by Bessel Functions multiplied by sine waves instead of sine waves as was the case in the rectangular surface (which is as much math as we need to know for now). For a rectangular surface the values of n and m start at one and one and are always positive. For a circular membrane n starts at one and must be positive but m can be zero and also negative.

> **Note**
>
> The relative mode amplitudes in the case of more than one mode are not properly normalized, mathematically, in this simulation.

Simulation Questions

1. Set $n=1$ and $m=0$, run the simulation and describe the motion in the surface view. Also describe the motion in the 3D view (if the surface is too large to fit in the window, use the mouse to shrink it to fit).
2. What is the frequency of the $n=1,\ m=0$ mode? What is the period?
3. Look at the $n=1$ and $m=1$ mode and describe the motion in the surface and 3D view. (Don't forget you can rotate the view using the mouse to get a better angle.) Describe the location of the nodal line.
4. For the $n=1$ and $m=1$ mode, what is the frequency? Is this a harmonic?
5. For the $n=2$ and $m=0$ mode, what is the frequency? Is this a harmonic?
6. Try several different pure modes. Can you find any harmonics of the lowest frequency mode?
7. Look at the pure modes for $n=1,2,3,4$ and $m=0$ (a pure mode has one of the two mode numbers equal to zero). What do the $m=0$ modes all have in common?
8. For any of the modes you have looked at, is there a cross section through the 3D view that would be a sine or cosine shape? The shape of the curve from the center out to the edge is given by Bessel Functions instead of sines or cosines.
9. Compare the $n=1,\ m=1$ mode with the $n=1,\ m=-1$ mode. What is the difference? Are these degenerate modes (the same frequency for two different choices of mode numbers)?
10. This simulation only shows pure modes. What would be the effect of having lots of different modes on the surface at the same time? (Go back to the rectangular plate simulation for a hint.)

12.1.3: Drums

The following diagrams are the first eight modes of a circular membrane, clamped along the outer edge (seen in the previous simulation). The mode numbers and frequencies are also given for each one. The figures are shown in the order of increasing frequencies rather than increasing mode numbers. In the figure the green area is moving towards you while the red area moves away; then these areas switch and go in the opposite direction.

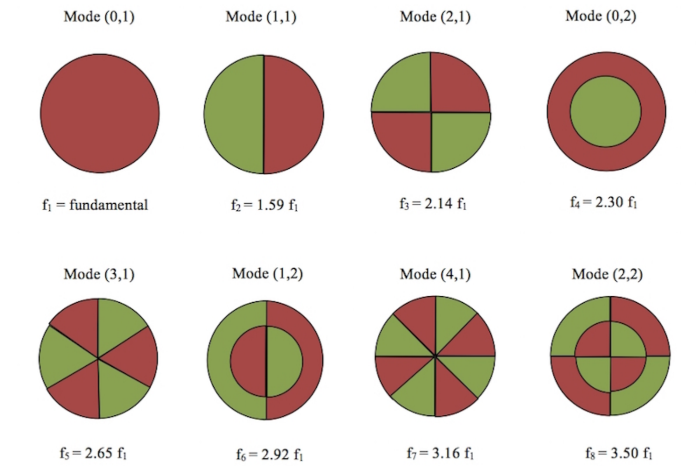


Figure 12.1.3.1

If you go back to the Chladni plate on YouTube and listen to the frequencies you will notice that, unlike strings or tubes, the resonances are not harmonics; they aren't multiples of the fundamental. Likewise in the square plate simulation, the circular membrane simulation and the above diagrams you will notice that the modes of vibration of a circular surface also are not multiples of the fundamental. The second frequency shown, f_2 (the first overtone) is 1.59 times the fundamental, f_1. This is an essential difference between instruments that make a perceptible pitch and those that do not. If you perceive a pitch from an instrument at least some of the overtones are harmonic (multiples of the fundamental) but this is not the case for many percussion instruments.

A second feature of percussion instruments is the short duration of the sound. As mentioned in Chapter 9 on Fourier analysis, the uncertainty principle comes into play for short pulses of sound. A short pulse requires many frequencies and few of these additional frequencies are multiples of a fundamental frequency as can be verified by a Fourier analysis of a cymbal or drum strike. Combinations of frequencies which are not multiples of some fundamental frequency usually sound like noise to our ears. Examples of sound sources that have lots of frequencies which are not harmonics include slaps, cracks, crashes, hand claps and thunder claps.

Thunder is formed when there is an electrical discharge from a thundercloud that suddenly heats air to nearly $30,000\ ^\circ\mathrm{C}$along the path of the discharge. The air can reach a pressure ten times that of normal atmospheric pressure in a few microseconds. The sudden expansion causes a shock wave to propagate away from the region of discharge. Due to irregularities in the path and heating of the air, a large range of frequencies centered around $100\ \mathrm{Hz}$ are produced. Variations in air density in the storm and local topographical features also affect the range of frequencies and what someone on the ground will hear. Some frequencies are below $20\ \mathrm{Hz}$ and cannot be heard by humans but can be felt. As a result we do not hear thunder as a discernible tone.

Just like in the case of a string, the location of the initial displacement of a membrane activates different modes. Below are some images from a simulation of a circular membrane. The modes which are activated are shown in the columns below the membrane; the first column is $n=0$ and m equal to one, two, three, etc.; the second column is $n=1$ and m equal to one, two, three, etc. In the first picture the mouse was used to "strike" the membrane in the center. Notice that mostly modes in the first column are activated (modes $(0,1)$, $(0,2)$, $(0,3)$etc. as you go from top to bottom) with a few in the third column (modes $(3,1)$, $(3,2)$,etc. This is the case of a drumstick with a relatively large end (a mallet) hitting the center (note that the displacement is broad). The second picture is the case where a drumstick with a smaller profile made an indention on the drum head at the center (the "poke" selection in the simulation). For this case the fundamental modes are almost exclusively activated (only the first column is lit up). In the third picture the drum head is struck off center. Now many higher modes are activated (as indicated by the colors in the squares at the bottom). This tells us that, even though we still may not hear a pitch (because the overtones are not harmonic) the drum will sound different because there are different overtones. This is very similar to the case of a string that is plucked or bowed; the initial shape determines which specific modes of vibration (overtones) are activated.

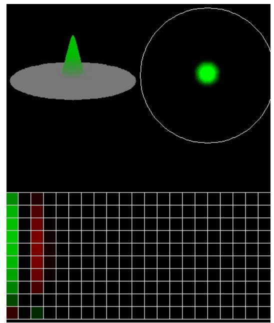


Figure 12.1.3.2

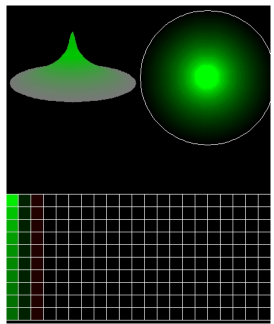


Figure 12.1.3.3

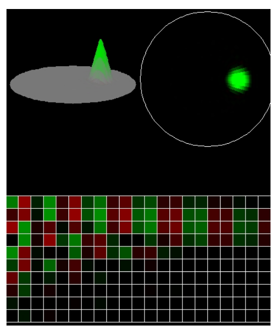


Figure 12.1.3.4

Most drum membranes are stretched over a hollow body, often cylindrical in shape. This configuration allows for two different types of resonances. The air cavity inside the drum will have a set of resonance frequencies determined by its shape and size. Like the air resonances in a guitar body or in a tube, this will emphasize some frequencies at the expense of others and this effect may even be strong enough to give the drum a pitch (see the next section). The body of the drum also has its own modes of vibration as shown in the following diagram of the cylindrical sides of a drum. Here, green is an area moving towards the center of the drum while the red area moves away from the center. This means, again, that some of the membrane's overtones will be reinforced by resonance while others are not, much like the body of a guitar or violin reinforce some (but not all) of the string harmonics.

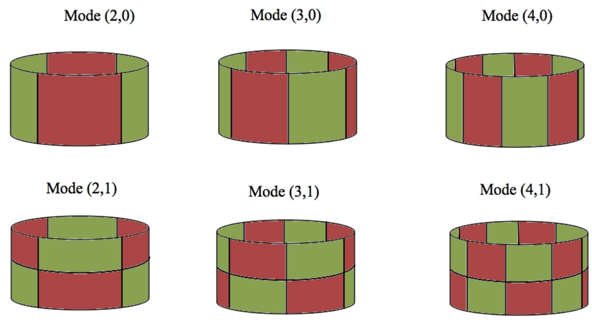


Figure 12.1.3.5

The body of the drum has one further function in that it controls how sound coming from the top surface of the membrane interacts with sound from the bottom of the surface. As shown in the following figure, with no body these sounds would tend to cancel due to destructive interference (as the surface moves it creates a pressure wave from the top that is exactly out of phase with the pressure wave from the bottom). If, instead, the sound from the underside has to travel through the drum body so that it comes out in phase (half a wavelength behind) then there is constructive interference. This will depend on the wavelength produced so a body of a given size will emphasize some frequency modes and tend to cancel others. This is why tall drums and drums with irregular shaped sides (like conga drums) sound different than shorter drums, even if they have the same drum head.

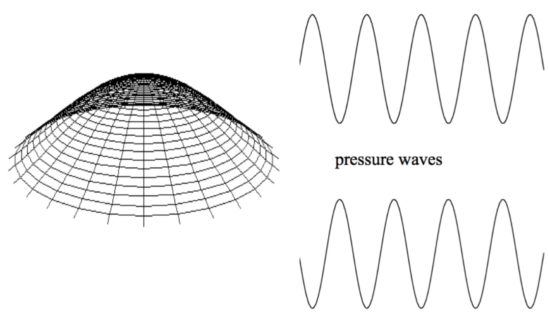


Figure 12.1.3.6

Here is a YouTube of a cymbal crash in slow motion. The modes are not the same as for a drum head because the edges of the cymbal are free instead of clamped and the cymbal is not flat. Because they are made of metal, which is stiffer than a drum head, the frequencies produced are higher. But the same sort of thing happens; the overtones are not harmonics so the cymbal does not have a fixed pitch.

Video/audio examples:

- Snare drum solo.
- Slow motion video of a snare drum being struck.

12.1.4: Harmonic Percussion Instruments

If we want a percussion instrument to have a perceptible pitch there are two choices. We can activate only one single mode so that there is just one frequency present or we can modify the instrument so that at least some of the overtones are harmonic. The tabla is an Indian drum where the head changes thickness from the outer edge inwards which can be done by rubbing material into the drum head. This shifts the overtones to be harmonic so that the tabla has a specific pitch, as can be heard in this tabla YouTube video.

The tympani or kettle drum used in symphony orchestras also has a distinctive pitch. These drums have a closed, bowl shaped body which may be hemispherical or parabolic in shape. The air trapped inside tends to lower the overtones because it cannot move freely. This plus the shape of the body ends up causing the drum to have harmonics at a fifth (ratio of three to two), a major seventh (a ratio of 15 to eight) and an octave (ratio of two to one) above the fundamental. These harmonics are enough to cause our ears to hear a pitch instead of noise.

Many percussion instruments are composed of a series of bars of different lengths that are struck. The xylophone, vibraphone, glockenspiel, triangle, and marimba are examples. As a first approximation of these instruments we consider the modes of vibration of a rod. In general the modes do not produce harmonics but if the main fundamental has a larger amplitude than the other modes it will be perceived as a pitch.

The following pictures are the first four modes of a free rod. The displacement is exaggerated; real xylophone bars, for example, are much too stiff to bend that much. The first mode has a node at the center so the left side will go down while the right side moves up and the middle stays fixed. The second mode has a two nodes at one quarter (0.25) and three quarters (0.75) of the way along the length. Mode three has nodes at about 0.16, $0.50$and 0.84 (meters if the rod is one meter long). You can see higher modes and the vibrational motion of these modes with the simulation of vibrations of simple systems from Wolfram (you may need to download their plug in to play with this demonstration). Click on the + sign under time in the simulation to view the controls for simulating the mode vibrations over time. Here is a simulation of vibrations of a bar.

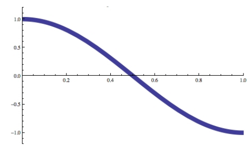


Figure $12.1.4.1$

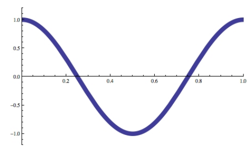


Figure $12.1.4.2$

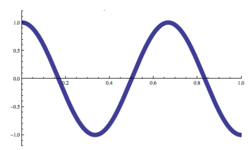


Figure $12.1.4.3$

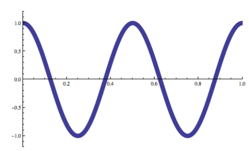


Figure $12.1.4.4$

The xylophone, vibraphone, glockenspiel and marimba have rectangular bars that have the similar modes as the ones shown above for a rod. The bars are of different lengths and are usually suspended at the nodes of the second vibrational mode. This helps damp out other modes so that each bar has a predominant frequency, usually the first overtone (second vibrational mode). The ***rates of***

decay of different modes are not the same for these instruments. This means that the initial tone may sound harsh because it has a large number of ***anharmonic*** overtones (not multiples of the fundamental) but the tone may improve in the next few milliseconds as some overtones fade away more quickly than the first overtone. Like strings and other instruments, striking the bar in the middle is more likely to excite the lower overtones. Marimbas and vibraphones also have resonating tubes of different lengths beneath each bar which acts as a resonator.

Marimbas originally were made with gourds as resonators but these have been replaced by tubes in modern instruments. Without these tubes the sound from the top of the bar would be exactly out of phase with sound from the bottom and tend to cancel since the top and bottom move in the same direction when struck (as was the case for a drum head without a body, mentioned above). The tubes create a resonance sound wave exactly out of phase with the sound from the bottom of the marimba bar, canceling it. This accentuates the fundamental frequency from the top of the bar and also produces harmonics for that frequency. This is part of the reason why marimbas and vibraphones are more melodic sounding but the xylophone, which does not have resonators, has a harsh sound. A final characteristic of the bars on a marimba that give it its distinctive sound is they are not precisely rectangular. The under surface of the bar is carved and tuned by ear to produce higher harmonics. The smaller mass of the bar where it is being struck also decreases the impedance mismatch between the bar and the mallet, allowing more energy to go into the bar from the mallet. The cross sectional shape of a marimba bar is shown below.

Figure 12.1.4.5

Video/audio examples:

- Tympani solo.
- Marimba solo.
- Xylophone solo.
- Glockenspiel solo.

The orchestral triangle is, in effect, a rod fixed at one end and bent into three sections. Here are the first four vibrational modes for a rod fixed at one end. Notice that the left end is fixed and stays does not move. The vibrations of these and higher modes can also be seen using the Wolfram simulation of vibrations of simple systems or with this simulation of a vibrating bar.

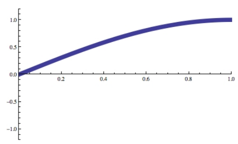


Figure 12.1.4.6

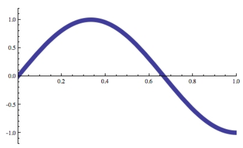


Figure 12.1.4.7

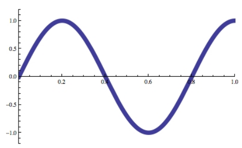


Figure 12.1.4.8

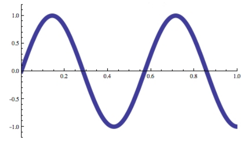


Figure 12.1.4.9

Orchestral chimes (sometimes called tubular bells) are tubes of different lengths that are struck at the top with a light, soft hammer. An interesting property of these instruments is that the fundamental frequency does not (and in fact none of the vibrational frequencies) correspond to the perceived pitch (called the ***strike tone***). The vibrational modes are similar to a free rod and it turns out that modes four, five and six are approximately in the ratio two to three to four (there are frequencies at $2f$, $2.9f$ and $3.9f$ of the perceived strike tone, with frequency f). This is an example of ***virtual pitch*** (or missing fundamental which was discussed in Chapter 10). Our ear/brain system interprets the harmonics as belonging to a fundamental frequency that is not actually present.

Bells have been around as musical instruments almost as long as flutes. The vibrational modes depend on the diameter of the bell, the thickness and the exact shape. These modes are tuned by trimming material from the inside of the bell so that there are harmonics of the fundamental frequency, giving the bell a distinct tone. Even so the modes of vibration are quite complicated as shown in these animations of bell modes. Many higher bell modes can be seen in these holograms of ancient Chinese bells and in this animation of ancient bell modes. In the hologram pictures the center of the circular regions are anti-nodes. Notice that some centers are dark at the same time other centers are light. This indicates that they are moving in opposite directions (are 180 degrees out of phase). Some of the bells pictured here were made as early as 1766 BCE. The different modes can be excited by striking the bell in different locations so that one bell can play more than one note.

Steel drums or steelpans are traditionally made by bending and stretching the top part of a large, 55 gallon oil drum to play different notes at different locations on the head. They became popular in the immigrant, former slave communities of the Caribbean in the 1930s. Each region of the surface of the drum is stretched until it has the harmonics of the desired note.

The gamelan is an orchestra of instruments from Indonesia that consists of a large number of mostly percussion instruments (occasionally stringed instruments and voice are included). Each player in the orchestra strikes one or more different instruments in turn to produce the melody. The tuning has five notes to a scale and so sounds odd to western ears (western music uses an eight tone scale).

Video/audio examples:

- Steel drum solo.
- Gamelan examples from YouTube: one, two, three.

12.1.5: Other Musical Instruments

There are many older instruments which only exist in museums today such as the shawm, viol, cornophone and crumhorn. Others, like the Hurdy Gurdy, have been around since the 11th century and still exist but are rarely heard. The following web sites describe various antique instruments and have sound samples for some of them: Renaissance Instruments, Guide to Medieval and Renaissance Instruments, Edinburgh University Collection of Musical Instruments, Zzounds renaissance instruments.

Musicians continue to make new instruments and modify older ones.

Video/audio examples:

- YouTube performance of a Hurdy Gurdy.
- YouTube of the Stroh violin.
- The Tromba-marina is an interesting instrument popular in the 18th century.
- YouTube of the wine glass harmonica.
- YouTube of several unusual instruments and a web page of a few more unusual instruments.
- Wikipedia on the Jew's harp.
- Chinese singing fountain bowl YouTube.
- Tibetan singing bowl YouTube. One with water.
- Here is a web site that discusses how to create and play a pitch-bending accordion.
- Aeolian percussion instruments at the Harmonic Fields.
- The Hang is a relatively new instrument similar to a steel drum but played by striking it with the hand. YouTube of a Hang being played.
- The Theremin is an interesting instrument that uses electrical capacitance to change an electronic circuit to make sound. The performer does not actually touch the instrument as seen in these YouTube videos: one, two.
- Here are a series of Ted talks about unusual instruments and performances.
- If you ever thought of making a unique instrument that never existed before you might want to take a look at the Experimental Instruments web site for ideas.

Summary

Percussion instruments tend to have overtones that are not harmonic (not multiples of the fundamental). If the anharmonicities are strong enough the sound is perceived as noise, not a pitch. Under certain circumstances the anharmonic overtones can be damped or tuned so that there are enough harmonics to cause the instrument to have a perceived pitch. This can happen with harmonics of a real fundamental frequency and also when the harmonics are multiples of a missing fundamental that isn't a vibrational mode of the instruments.

Questions on Percussion Instruments:

1. What are the two main categories of percussion instruments?
2. How are the vibrational modes of a drum head similar to those of a string and how are they different?
3. Why do you need two numbers to specify the vibrational mode of a drum head but only one number for a string?
4. Describe the first few modes of a square membrane, clamped at the edge.
5. Describe the first few modes of a circular membrane, clamped at the edge. How are they different from the square case?
6. Why does striking a circular drum at the center sound different from striking it off center?
7. Draw the $(5,0)$ and $(5,1)$ modes for the body of a cylindrical drum head.
8. What is the difference between instruments that make a perceptible pitch and those that do not?
9. Why don’t most drums have a specific pitch?
10. Drums are membranes stretched over a body. What would happen if there was no body to the drum (just a fixed circle holding the membrane)?
11. Why does a timpani drum have a perceived note but a snare drum does not?
12. How does the tension of the membrane change the vibrational modes of a drum head?
13. How do you change the pitch of a timpani drum?
14. Why does the tabla play a perceived note but a snare drum does not?
15. How does the thickness of a drum head modify its sound?
16. What is the purpose of the snare on a snare drum?

17. What are some differences between the vibrational modes of a string and of a rod that is free at both ends?
18. Draw the $5\mathrm{th}$ mode of a rod with one end fixed.
19. Which mode is emphasized by suspending a bar from the points one quarter and three quarters along its length? Explain.
20. What is different about a xylophone and a vibraphone that make them sound different?
21. What effect does different decay rates for various overtones of a percussion instrument have on the perceived pitch?
22. When playing a xylophone, why might it initially sound harsh before it smoothens out?
23. What is the purpose of resonance tubes below the bars of a vibraphone and marimba?
24. What is the purpose of trimming the bottom side of a marimba bar?
25. Why is the strike tone of a chime not the same as the perceived tone (Hint: Review the section in the perception chapter on virtual pitch.)?
26. What is a gamelan?
27. In class and in the book you saw some pictures of holograms of ancient Chinese bells. Explain what the holograms showed.
28. What is a Theremin and how does it work?
29. What is different about the vibrational modes of a cylindrical cymbal and a cylindrical drum head?
30. Define anharmonic.

CHAPTER OVERVIEW

13: Voice

As we know, musical instruments consist of a vibration which is amplified by resonance. The human singing voice is no different. The vocal chords are the vibrating part and the throat, mouth, nasal cavities and bronchial tubes constitute the resonance cavities that amplify these vibrations into sound. Because every person's combination of throat, mouth, nasal cavities and bronchial tubes is slightly different, we all sound slightly different. The fact that we can change the shape of some of these cavities at will enables us to produce a wide range of pitches, depending on the initial structure and training. Here is Wikipedia on the human voice.

Key Terms:

Vocal cords, vocal tract, Bernoulli's principle, resonance cavities, vocal formants, trachea, larynx, phonemes (plosive, fricative, vowels, diphthongs, consonants, semivowels, gliding consonants, liquids, nasal), throat singing, articulation (opera singers).

13.1: The Human Voice

The human singing voice is also a vibration which is amplified by resonance. The vocal chords are the vibrating part and the throat, mouth, nasal cavities and bronchial tubes constitute the resonance cavities that amplify these vibrations into sound. Because every person's combination of throat, mouth, nasal cavities and bronchial tubes is slightly different, we all sound slightly different.

13.1.1: The Vocal Tract

The main components of the vocal tract are sketched in the figure below. The ***nasal cavity*** (1) is actually a set of connecting chambers in the front part of the face, above the ***oral cavity*** (2) or mouth. The ***hard palate*** (3) separates the mouth from the nasal cavity. The ***soft palate*** (4) extends the hard palate and connects with the ***uvula*** (6), the part you see hanging down in your mouth when you look in the mirror. The teeth (5), lips (7) and various parts of the tongue (9, 11, 13, 15) are also shown. The ***pharynx*** (8) connects the mouth to the ***trachea*** (16) or wind pipe and the ***esophagus*** (not labeled). The ***epiglottis*** (10) closes the trachea when you swallow food so that the food goes down the esophagus and not into your lungs. The ***vocal cords*** (12) or folds are located at the ***glottis*** (14), the external part of which is called the ***larynx*** (17).

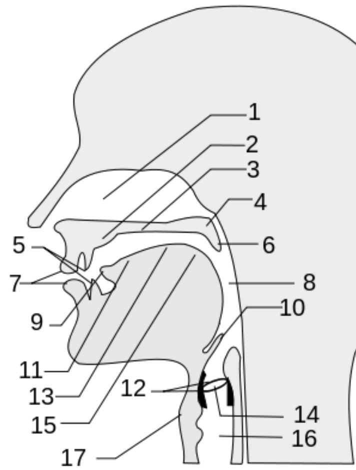


Figure 13.1.1.1

The ***vocal cords*** or vocal folds are actually two folds of skin that can be brought together to partially block the flow of air coming from the lungs. When they do this the Bernoulli effect causes them to snap shut until the air behind them builds up enough force to them open again. The picture on the left is a picture of the folds (looking down into the throat from the mouth) when the folds are relaxed and open. On the right is a diagram of the vibration process.

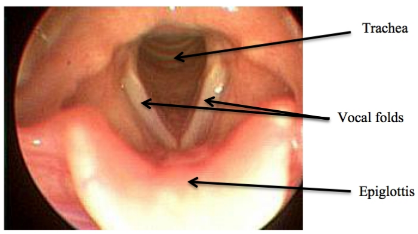


Figure 13.1.1.2

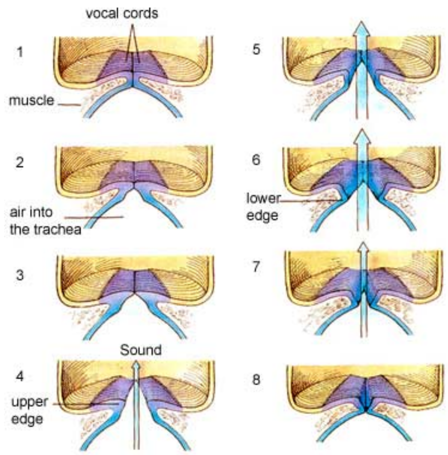


Figure 13.1.1.3

In the same way that a brass player's lips buzz, the vocal cords open and close to form a buzzing sound. By tightening the muscles around the flaps we can change the frequency of the buzz but only slightly. The following are videos of vibrating vocal cords in singers.

Video/audio examples:

- Video with explanation of vocal cords while singing.
- Web page with several videos of vocal cords while singing.
- YouTube of Mel Blanc's vocal cords.

13.1.2: Vocal Formants

As in the case of a brass instrument, the shape and size of the resonance cavities in the body select some of the buzzing frequencies but diminish other ranges of frequencies. The nasal cavity, the oral cavity, the larynx, the trachea and the lungs are all air cavities that can have an effect on the sound being produced. Lets look at a simple example; buzzing sound fed through three Helmholtz resonators.

The buzzing sound produced by the vocal folds has a wide range of frequencies between $60\ \mathrm{Hz}$ and about $7,000\ \mathrm{Hz}$ when overtones are included (technically it is called pink noise which will be discussed in Chapter 18). Below is a diagram of the frequencies (in hertz) of pink noise; all frequencies are present but the amplitude gets progressively smaller for higher and higher frequencies.

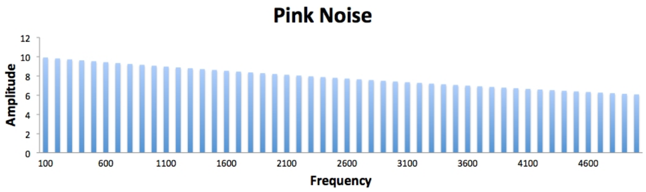
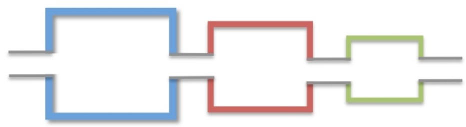


Figure 13.1.2.1

Suppose we have three Helmholtz resonance cavities, all linked together as shown in this figure:



Three Helmholtz cavities with different resonance frequencies.

Figure 13.1.2.2

These resonating cavities are of three different sizes so they will have three different resonance peaks (review Chapter 4 if you have forgotten what resonance is) as shown in the following figure.

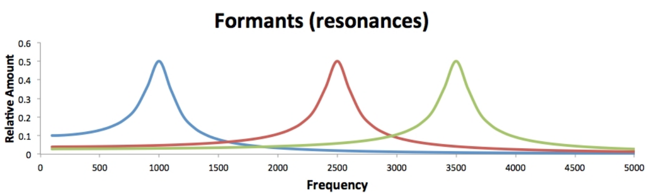


Figure 13.1.2.3

The result of feeding pink noise into the three different resonators is shown in the graph below (frequency in hertz). Notice that the available frequencies are now limited by an envelope as the result of the resonance cavities. This envelope imposed on the spectrum of available frequencies is called a ***formant***. Its particular shape is controlled by the shape of the resonance cavity.

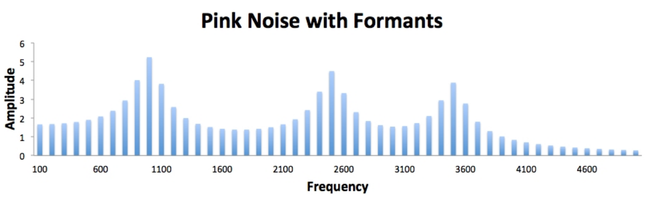


Figure 13.1.2.4

A simple model of the human vocal system is that of two resonance cavities, the mouth and the larynx. Both cavities change shape when you talk or sing but the mouth changes more. In the following figure the vocal tract is modeled for the sound 'ah' (written as /a/ in the International Phonetic Alphabet). If you think about what shape your vocal tract has when it makes this sound, the mouth is very open. Below the vocal tract figure is a two cavity model with the first cavity (the mouth) bigger than the second. A sketch of the formant is also shown on the right. Notice that lower frequencies are emphasized over higher frequencies.

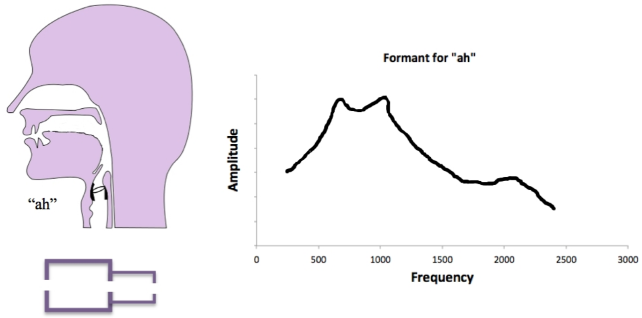


Figure 13.1.2.5

When you make the sound "ee" (written as /i/ in the International Phonetic Alphabet) the mouth closes up to make a small resonance cavity. The figure below shows the vocal tract, a two cavity model and the approximate formant. For this sound there is more sound at higher frequencies as seen by the three peaks on the right side of the graph.

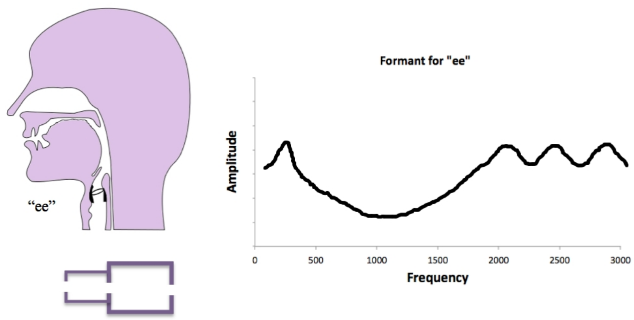


Figure 13.1.2.6

Here is a graph of formant vowel frequencies. The graph shows the first two formants of English vowel sounds. Notice that the /i/ sound ("ee" as in the word 'seat') for most speakers lies in the upper left of the graph so the first formant has low frequencies and the second formant has high frequencies. The /a/ ("ah" as in the word 'sought') for most speakers lies towards the lower right; high first formant frequencies and second formant frequencies lower than for /i/.

Because the mouth cavity and larynx are so flexible, a wide variety of formants can be made resulting in a wide variety of sounds. Try saying the words "heed", "hid", "head", "had", "hod", "hawed", "hood", "who'd". How does your mouth and larynx structure change for each sound?

Video/audio examples:

- Here is a video of a talking mouth, created by scientists at Kagawa University, Japan. You will notice something is missing in the sounds. This is because the robot only produces formants; it does not have a tongue or lips to produce the other phonemes needed for speech.
- Here is an interactive graph of formant vowel frequencies.
- The vowel formants, in Canadian English.
- Formants, explained with sample frequencies.
- Formants, explained with sample frequencies and MRI images.

13.1.3: Phonemes

The distinct sounds that make words spoken in a particular language are called ***phonemes***. Different languages often use distinct phonemes. We learn how to make these distinctive sounds very early in life by imitating the sounds around us. People who learn to speak a different language as adults often have an accent because they are using the phonemes they learned for their first language instead of the correct phonemes used by a native speaker. Phonemes in English can be grouped into five categories; plosives, fricatives, pure vowels, diphthongs and gliding consonants. A sixth category involving nasal sounds is sometimes included in this list.

As mentioned in the previous section, frequencies emphasized by the vocal formants play a role in each type of phoneme. In much the same way that attack frequencies for a musical instrument affect the identification of a musical sound by the ear-brain system, how the formant frequencies are initiated and terminated affect the way they are heard. For example ***plosive*** consonants are formed when the air flow through the vocal system stops completely. Examples in English would include the letters 'p', 't', and 'b'. The sounds 'g' as in 'gun' and 'k' as in 'kick' are also included as plosives. Changing the way the air stops or starts can affect the transient frequencies at the beginning and end and this changes the perceived sound. For example the sounds of the letters 'p' and 'b' have basically the same formants but different initial plosives.

Fricatives (sometimes also called sibilants) are another variation in sound in addition to the formant that has to do with the way it is produced. Examples are 'f', 'v', 'th' (as in 'thin'), 'th' (as in 'these'), 's', 'z', 'sh', and 'zh' (as in 'measure'). These sounds include frequencies that are caused by air turbulence in the mouth, in addition to the frequencies created by the vocal cords.

Most ***vowels*** are phonemes created by the action of formants on the vocal cord frequencies. For some vowels or vowel combination and for some consonants the sound changes while being made. In the case of ***diphthongs***, two sounds are put together with a glide connecting them. Examples in English are found in the words 'eye', 'hay', 'boy', 'low', and 'cow'. A similar sort of thing happens for the ***gliding consonants*** which can further be broken down into ***semivowels*** such as 'w' and 'y' and ***liquids*** such as 'l' and 'r'. In all of these cases the formants change during the creation of the sound.

A final category of phonemes are the ***nasal*** sounds. For these phonemes the nasal cavities are used to change the vocal formants to produce sounds such as 'm', 'n' and 'ng' (as in English).

The phrase "SAY BITE AGAIN" has several phonemes. "S" is fricative "AY" is a diphthong, "B" is a plosive"I" is vowel, "T" is a plosive, "AG" is a plosive, "AY" is a diphthong and "N" is nasal. Here is a sound file of "SAY BITE AGAIN" spoken normally.

The following diagram show a picture of the recorded sound wave and corresponding spectrogram for the phrase "SAY BITE AGAIN" spoken normally.

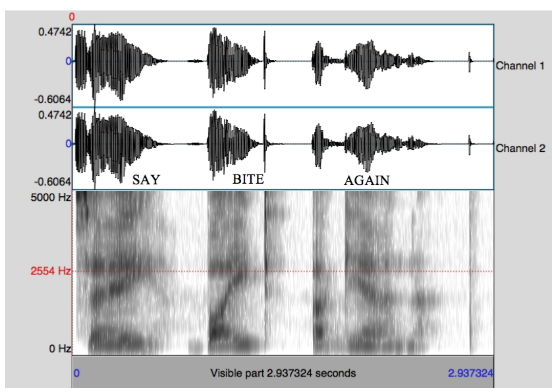


Figure 13.1.3.1

Here is a sound file of "SAY BITE AGAIN" spoken so that the phonemes are separated in time.

The following diagram show a picture of the recorded sound wave and corresponding spectrogram for the phrase "SAY BITE AGAIN" with each phoneme spoken separately ("S", "AY", "B","I","T", "AG", "AI", "N").

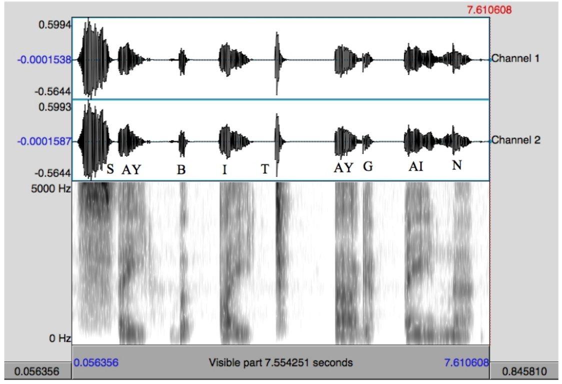


Figure 13.1.3.2

Video/audio examples:

- An x-ray video of various phonemes being pronounced.

13.1.4: Singing

The nasal, trachea and lung cavities also add resonances that change the timbre of a person's voice. Some people are gifted with internal air cavities that lead to formants which emphasize frequencies that we find pleasant to listen to; for example a fundamental and several overtones which are exact harmonics. If they are further gifted with (or can train) an ear-brain system that can distinguish frequencies well, they can become a good singer. Training of the voice can improve the range and ability of a singer. Opera sopranos, for example, undergo a long period of training to be able to sing higher notes and to sing loud enough to be heard over the orchestra.

Being heard over the sound of an orchestra seems like an impossible task but opera singers are helped by the fact that the formants involved emphasize many frequencies between 2000 Hz and 3000 Hz. Trained singers create this formant by lowering the larynx and expanding the throat region above the larynx. Orchestral instruments are louder but in a lower frequency range. That plus the fact that humans hear better in the range of frequencies that the voice uses makes it possible for an opera singer to sound as loud as the orchestra without using a microphone (although many performances today do use artificial amplification).

Although the formants used by an opera singer makes the sound more easily heard, they also have an effect on the way certain phonemes sound. The result is that in opera singing, particularly in the higher register of singing, it is harder to distinguish the words being sung. Here are examples of an opera soprano singing scales using a given phoneme. Notice that in the higher pitches it becomes difficult to distinguish different phonemes.

Because there are two or three predominant vocal formants, some singers can emphasize a note and a higher partial at the same time so that they can be heard as distinct tones. This is often called overtone or throat singing. Singers in Mongolia and some other regions use this technique as demonstrated in this video of Mongolian throat singing. About 20 seconds into the video you will hear a whistling sound singing a melody along with a lower drone sound at the same time. Both sounds are coming from the singer's vocal system. This is something that can be learned by almost anyone, as demonstrated by the dozens of YouTube videos on "Tuva Throat Singing".

We know that the speed of sound changes if the density of the medium in which it is traveling changes. We also know that since the speed of sound is related to wavelength and frequency by $v = \lambda f$, changing the speed of a wave will change its frequency (the wavelength remains fixed). You may have heard the voice of someone change when they inhale helium. This occurs because the speed of sound changes and therefore so does the frequency. The effect is as if the formants have shifted upward. Here is a video of a voice with helium and with sulfur hexafluoride. Sound in sulfur hexafluoride is slower so the frequencies of the formant will be lower.

Note

Do *NOT* try this at home. Helium and sulfur hexafluoride do not have oxygen in them and it is possible to suffocate. In particular sulfur hexafluoride is heavier than air and is difficult to get out of the lungs.

Video/audio examples:

- An MRI video of a singing soprano, and the scientific paper that describes why they did it.
- A young girl throat singing "Amazing Grace" (you may have to listen twice to hear the melody).
- An very good explanation of polyphonic singing.
- Recordings and description of Katadjak', a form of Inuit throat singing and voice competition.
- A little science behind "beatboxing". An example without the help of a microphone, Bobby McFerrin.
- What happens to a musical instrument filled with helium or sulfur hexafluoride?

13.1.5: Animal Sounds

Many animals use sounds either to attract mates or warn away competitors. A few animals use sounds to navigate (discussed in Chapter 16: Acoustics). In general, smaller animals make higher frequencies and larger animals make lower frequencies but there are plenty of exceptions. Elephants and whales use infrasound (below $20\ \mathrm{Hz}$) to communicate over long distances but both mammals can also make higher frequency sounds.

Most mammals have vocal cords much like humans. Some marine mammals such as dolphins have phonic lips in their nasal cavities instead of vocal cords in their throats. These phonic lips act the same way vocal cords do, vibrating when the dolphin exhales and are controlled by muscles in the nasal passage. Whales do not have vocal cords but can pass air between two large internal cavities in their bodies and the passage way vibrates, allowing the whale to make sounds while submerged without exhaling.

Most insects and some crustaceans make sounds by rubbing body parts together. This form of making sound is called stridulation. Insects also use resonating body parts such as wings and hollow body cavities to amplify the sound. Cicadas have an air filled abdominal cavity ribbed with cartilage and muscles and it is the contraction of these muscles which makes the sound we hear. Cicadas are much louder than other insects which generate sound by stridulation because of their unique method of making sound. A few insects such as crickets use leaf structures or tunnels in the ground to amplify their calls. The treehopper, an insect, communicates to others of its species by sending vibrations through the branches of the plant it is on.

Fish generally make sounds by contracting the muscles around their swim bladder if they have one. This typically produces a broad spectrum of low frequency sounds (under $1000\ \mathrm{Hz}$) and is called drumming. At least one type of herring emits gas through its anus to produce sounds.

Most birds have two sets of syrinx, one on each of their two trachea. These small areas of the trachea can vibrate and because there are usually two regions, some birds can make two sounds at the same time. This is part of the reason birdsong can be very complex. This arrangement also allows birds to produce short sounds with durations of $1/200\mathrm{th}$ of a second, 10 times faster than humans. This is probably why birds also have absolute pitch (their ear/brain system hears exact frequencies) whereas humans have relative pitch. Some birds and some frogs have resonating sacs that amplify the sounds being produced by vocal cords or syrinx.

Video/audio examples:

- Snapping shrimp, and more snapping shrimp.
- Ant communication.
- Recordings of soundscapes (natural ambient sound- there is more there than you think!).

Summary

Like all instruments, the human voice relies on resonance to emphasize a particular fundamental frequency and overtones. For talking, humans use the mouth and larynx to form two resonance cavities which produce vocal formants that shape the buzzing of the vocal folds into distinct sounds. Other phonemes require attack and ending frequencies produced by starting or stopping the air flow with the lips or tongue. Singers train their vocal folds, larynx and mouth to have very clear resonances. They are helped by being gifted with nasal and lung cavities that accentuate pleasant overtones.

Questions on Voice:

1. Describe the vocal cords and explain how they work.
2. Explain how the Bernoulli effect relates to the movement of your vocal chords.
3. How are the muscles of the vocal chords similar to a brass player’s lips buzzing?
4. What range of frequencies (in Hz) do your vocal cords make?
5. List the main parts of the vocal tract and their function.
6. What actually happens when you swallow and something “goes down the wrong pipe?” Do we actually have two different “pipes” in our throats?
7. Musical instruments all have a vibrating part that acts as a source of sound and resonators that enhance certain frequencies. For the human voice, what acts as the vibrating sound source and what acts as resonators?
8. What other parts of the body function as resonating cavities for the human voice?
9. What is pink noise?

10. What are vocal formants?
11. Explain, in terms of formants, what is occurring when a Mongolian throat singer sings.
12. What happens to the human voice if you inhale helium? Why?
13. What happens to the human voice if you inhale sulfur hexafluoride? Why?
14. Why is inhaling sulfur hexafluoride more dangerous than inhaling Helium?
15. Would helium and/or sulfur hexafluoride change the sound of a stringed instrument? What about a tube based instrument? Explain.
16. The distinct sounds that make up the words spoken in a particular language are called _____________.
17. What is a diphthong?
18. Name the six categories of phonemes in the English language and how they are formed.
19. Describe the difference in what we do with our mouths when we say a long o, an ah, and an ee sound.
20. Without a tongue or lips, are we able to create all phonemes needed for speech? Explain.
21. What are plosives and fricatives? Give three examples of each.
22. How are the phonemes used by opera singers different from ordinary speech?
23. Why are the phonemes used by opera singers different from ordinary speech?
24. What factors allow an opera singer to be heard over an orchestra, even without electronic amplification?
25. How does impedance matching apply to operatic singing?

CHAPTER OVERVIEW

14: Musical Scales

Unlike the human voice (which can make any pitch in a large range of pitches) for many instruments once the decision about how to construct the instrument is made, there are a limited number of pitches it can produce. For example, once the holes in a flute are drilled the flute can only make a certain set of notes (although for some instruments the musician can often 'bend' a note by changing the position of the lips and or fingers). This set of notes is called a scale and we would like to have a scale (or set of notes) where we get the greatest number of combinations that sound good together. We also would like to standardize the scale in such a way that if we build other instruments, two instruments playing together can play the same pitch. This turns out to be more difficult than it would seem. The choice of scale is arbitrary; we can choose any combination of notes that we like and in fact some cultures have chosen scales very different from the ones used in western music. However there are advantages and disadvantages for any of these various choices, as explained in this chapter.

Key Terms:

Scale (Pythagorean, Ptolemaic, Equal-tempered), mode (or key), septatonic, pentatonic, the problems of changing key and changing octave, the problem of fixed tuning for pianos and tube based instruments, tone, semitone, cents, temperaments, Railsback curve.

14.1: Musical Scales

The notes on musical instruments are organized into scales and we would like to have a scale where we get the greatest number of combinations that sound good together. We also would like to standardize the scale in such a way that if we build other instruments, two instruments playing together can play the same pitch. This turns out to be more difficult than it would seem. The choice of scale is arbitrary and different cultures have chosen differently.

14.1.1: The Pythagorean Scale

The ancient Greeks, who had only simple stringed instruments and flutes, noticed two things about pitches produced by a vibrating string. They noticed that a string of half the length of another but with the same tension and thickness sounded similar. For example if the original string played a frequency of 880 Hz a similar string of twice the length would play a note of 440 Hz, an ***octave*** lower. They couldn't measure these frequencies but they could hear that there was a pleasant relationship between the two pitches. The same thing happens by holding the string down in the center; each half will sound a note and octave higher than the full length. The ancient Greeks also noticed that holding a string down at $2/3$ of its length would produce two notes (by plucking each side) that sounded pleasant together. We call the interval between the note played by the $1/3$ length and the note played by $1/2$ the length of the same string a ***perfect fifth*** (if you sing the children's song 'Baa Baa Black Sheep' the first Baa and Black are a fifth apart). The ratio of frequencies is $3:2$. Two other notes that sound good together are the notes produced by the long part of the $2/3$ of the string and the note formed from holding the string down at its center. The interval between these two notes is called a ***perfect fourth*** and the frequency ratio between them is $4:3$ (the first two notes of 'Hark the Herald Angels Sing').

As we have learned, the frequency of the fundamental is the main factor that determines the pitch of the note we hear but not the only one. In the following you will notice that the frequencies are often not exactly harmonics nor whole numbers. In most cases this does not matter because our hearing is not accurate enough to detect the difference of a few Hz (remember that the JND in frequency is about 1 Hz for tones below 1000 Hz). The modern, equal temperament scale divides an octave into twelve equal steps (called ***semitones***). Each semitone is divided into 100 *cents*. On the equal-tempered scale (see below) this is about 0.3 Hz for the notes in the octave starting at middle C (261.63 Hz). Because the frequency spacing for different octaves is not the same the frequency change per cent will be different in each octave. Trained musicians can hear a difference in frequencies of five cents (1.5 Hz) under test conditions. Normal music notes may be off from their intended frequency by as much as 20 cents (about 6 Hz) but we don't notice unless the sounds are isolated or there is beating.

Is there a reason that notes with a ratio of $3:2$ (a perfect fifth) and other ratios sound good together? Maybe. In the case of fifths the harmonics overlap (a note whose fundamental is 300 Hz has harmonics at 600 Hz and 1200 Hz which overlap with the harmonics of a fundamental at 200 Hz which are $400\text{ Hz},\ 600\text{ Hz},\ 800\text{ Hz},\ 1000\text{ Hz}$and 1200 Hz). In other words, there is less dissonance (roughness produced by overtones which are close to but not exactly the same frequency- see Chapter 10). It is not clear why humans tend to find notes with overlapping harmonics pleasing but this seems to be true in most cultures (although the notes and scales used may be different). The western world inherited Greek preferences for scales but many non-western cultures include other preferences which sound odd to the western ear. An appreciation for these sounds can be learned, however.

The realization that the ratios $3:2$ and $2:1$ (octaves) sound good together led the Greek philosopher and mathematician Pythagoras to come up with what is now known as the ***Pythagorean scale***. To construct this scale we start with a note or frequency. If we double it we have the same note an octave higher. If we multiply or divide by $3/2$ we have notes in between. In the example below we start with the note D at 147 Hz and apply the rule (the frequencies are rounded off to whole numbers in this example). The first five notes generated by this procedure are called the ***pentatonic scale*** which has been used for two thousand years. If we generate two more notes we have a ***septatonic scale*** or seven notes between the two notes an octave apart. The procedure can be continued to find other notes in this octave.

Frequency	Note
147 Hz	D
$147\times 2=294\text{ Hz}$	D an octave higher
$147\times 3/2=220\text{ Hz}$	A
$294\times 2/3=196\text{ Hz}$	G
$220\times 3/2=330\text{ Hz}$	E
$196\times 2/3\times 2=261\text{ Hz}$	C

Table 14.1.1.1

It should be noted that Pythagoras only codified this scale mathematically; musicians were already using it because it sounded good. It seemed to be interesting and important to Pythagoras and other early mathematicians that what sounded good to the ear could be explained as simple mathematical ratios between the lengths of the string on a harp.

A particular choice of starting note and system of generating a note scale is called a ***mode*** (note that this is **not** the same as modes of vibrations in a tube or on a string or membrane). If we start with the note F and multiply/divide by $3/2$ and 2 each time we generate a scale called the ***Lydian mode***. If we start with the note B and apply the rule the seven note scale is the ***Locrian mode***.

A later mathematician and astronomer, Ptolemy, added notes to the Pythagorean scale by including ratios of $5:4$. Using the same procedure as above (multiplying or dividing by $5/4$ and 2) will also produce a seven note scale but with slightly different frequencies. So for example $261.63\text{ Hz} \times 5/4 = 327.04\text{ Hz}$ which would be the note E but it doesn't quite have the same frequency as the E in the Pythagorean scale (331.12 Hz). However most people cannot hear this small of a frequency difference unless a direct comparison is being made with another tone (in which case you can hear beats) so a song played or sung with this set of frequencies sounds about the same as the Pythagorean scale. For the Pythagorean scale, it doesn't matter which note you start on, each scale sounds similar because the spacing is similar because the ratio is always $3:2$. This turns out not to be true for the scale created by Ptolemy because some ratios are $4:5$ and others are $3:2$. So shifting keys or modes (starting at a different note) does make a difference in the Ptolemaic system but not in the Pythagorean.

Musical Ratios Used by Various Groups			
Use or Name	**Musical Interval**	**Frequency Ratio**	**Size in Cents (Pythagorean Scale)**
	Unison	1 : 1	0
Pythagorean & Renaissance	Octave	2 : 1	1200
Pythagorean & Renaissance	Fifth	3 : 2	701.95
Pythagorean & Renaissance	Fourth	4 : 3	498.05
Renaissance	Major Third	5 : 4	386.31
Renaissance	Minor Sixth	8 : 5	813.69
	Minor Third	6 : 5	315.64
	Major Sixth	5 : 3	884.36
Indian and Modern	(no name)	7 : 4	968.83
Indian and Modern	(no name)	11 : 8	551.32

Table $14.1.1.2$

In the 14th century music began to get more complicated. Instead of everyone singing the same note in a choir, harmony became popular. Prior to this time most music in Europe was church music and was governed by strict rules (they were supposed to always use the Pythagorean scale). As folk music became more popular some of these rules were relaxed. Rounds (like Row, Row Row Your Boat), where each singer sings the same song but starts a bar or so later, also became popular. In order for this to sound good most of the notes of the song should harmonize with each other (so that when they overlap there is harmony; in other words many of the harmonics overlap). What performers discovered was that the notes in the Ptolemy scale allowed more harmonizing because there is more of a variety of notes. Eventually this scale replaced the more rigid Pythagorean scale. We now call the Ptolemaic scale the ***Just scale***.

14.1.2: Equal Temperament

Seems simple enough; start with a note at a given frequency, apply the multiplication rules to get the other notes in the scale. So multiplying by 2 gives a note an octave higher, multiplying by $3/2$ a note in between, and so forth. What could go wrong?

We already know the rules for the Pythagorean scale do not give exactly the same frequencies as the rules for the Ptolemaic (or Just) Scale. Let's stick with Pythagoras and decide we want to fill in the notes between C (at the modern day frequency of $261.63\ \mathrm{Hz}$) and C an octave above ($261.63\ \mathrm{Hz} \times 2 = 523.15\ \mathrm{Hz}$). We start with $261.63\ \mathrm{Hz}$ and multiply by $3/2$ to get $392.44\ \mathrm{Hz}$ which is G. We take G ($392.44\ \mathrm{Hz}$) and multiply by $3/2$ again to get $588.67\ \mathrm{Hz}$ which is an octave above so we divide by 2 to get $294.33\ \mathrm{Hz}$ which is D. We can get A at $441.49\ \mathrm{Hz}$ by multiplying $294.33\ \mathrm{Hz}$ by $3/2$. Multiplying by $3/2$ and dividing by 2 gives $331.12\ \mathrm{Hz}$ which is E. Multiplying by $3/2$ gives $496.69\ \mathrm{Hz}$ which is B. Repeating this procedure (multiplying by $3/2$ and dividing by 2 if it is outside the octave) two more times gives $\mathrm{F}^{\#}$ (F sharp) and $\mathrm{C}^{\#}$ (C sharp). So far so good.

Now we run into a problem. Because if we use this $\mathrm{C}^{\#}$ at $279.39\ \mathrm{Hz}$ and multiply by $3/2$ we get $419.08\ \mathrm{Hz}$ *which is not a note on the scale!* In fact it is so close to A^{b} (A flat) at $412.42\ \mathrm{Hz}$ that we probably would mistake it for A^{b}. Likewise if we do the procedure again we get $314.31\ \mathrm{Hz}$ which is very close to E^{b} at $310.08\ \mathrm{Hz}$ but is not on this scale. In other words, the Pythagorean system of generating notes finds an infinite set of notes between C at $261.63\ \mathrm{Hz}$ and the C an octave above. The Pythagorean solution was to stop after five notes (a pentatonic scale) and not use any more. But the idea of using the simple mathematical rule of $2/3$ for generating more notes where all the notes harmonize doesn't work. The dilemma is that ratios of $3:2$ sound good to the ear but using a ratio of $3:2$ does not result in a closed, limited number of notes.

As music became more complicated, musicians started to want to switch keys during the piece of music. When a musician changes key they are essentially starting with a different note (say B at $496.67\ \mathrm{Hz}$ in the Pythagorean scale) and using an octave based on that new starting point. But starting on a different note and applying the rule to generate the scale results in new notes which are not close to previous notes (it almost works in the Pythagorean scale but not in the Just temperament). If you have a musical instrument, say a piano, set up so that the notes correspond to a Just scale based on F (the key of F) for example, changing to a key of B isn't possible unless you add extra strings to the piano because there are new frequencies in the key of B. The new notes also do not harmonize with the old notes. To maintain the Just scale requires a different instrument, one tuned differently for each key you want to play in (or many extra strings).

This problem gets worse when you try to generate notes an octave above or below the initial octave using the rules; the higher octave doesn't harmonize with the lower octave if they are generated using the rules. Notes an octave above should be twice the frequency of the note in the lower octave but using the formula to generate the note doesn't give a frequency twice that of the lower octave. For singing and even violins, which can make any desired frequency (within a given range), this isn't a problem because the performer can just shift the note a little so that it sounds right. But for the piano, organ, flute and stringed instruments with frets this is a big problem because the notes are determined by the construction of the instrument and cannot be changed without re-tuning it. So it appears there is a choice between making instruments so they harmonize well in one octave and scale but not another; or have different instruments for each octave and each scale; or re-tune the instrument every time you want to change scales or octaves.

A scale system that tries to correct for the problem of being able to play in any key is called a ***temperament***. The chosen starting note of a scale is called the ***tonic*** and a musical ***fifth*** consists of the tonic and the fifth note of the scale.

Historically several scale adjustments generically known as ***well-tempered scales*** were proposed and eventually the ***even-tempered scale*** won out (see Well versus Even Temperament). In this scale the frequencies between octaves are chosen to be equally spaced in which case the ratio between notes is 1.059463 (the notes can be generated by multiplying or dividing by either 2 and/or 1.059463). This divides an octave into twelve notes. Another way to write this mathematically is $f = f_o \times (2^{1/2})^n$ where f_o is generally chosen to be $440\ \mathrm{Hz}$ in modern times. $2^{1/2} = 1.059463094\ldots$and each value of n generates a new note.

Johann Sebastian Bach was the musician who managed to get the world to pay attention to alternative scales by writing music (this was in the 1700's) specifically to sound good for instruments tuned a certain way. Eventually most western cultures adopted the even-tempered scale (some other cultures did not which is why, for example, music from India often sounds very different than western music). Because the ratios between frequencies are not exactly $3:2$ or $5:4$ you can often hear chords where there is beating and dissonance. The chords also do not sound as pure as Pythagorean tuning because they are not exactly harmonic however the greater flexibility to change keys and to (almost) harmonize over several octaves makes equal temperament a useful compromise. Below is a diagram of the frequencies of various scale systems (temperaments).

Frequencies of Various Scale Systems									
Pythagorean		Just Major		Mean-tone		Werckmeister		Equal-tempered	
Note	Hz	*Note*	Hz	*Note*	Hz	*Note*	Hz	*Note*	Hz
C	523.25	C	523.25	C	523.25	C	523.25	C	523.25
B	496.67	B	490.55	B	489.03	B	491.67	B	493.88
B^b	465.12	B^b	470.93	B^b	468.02	$B^b/A^\#$	465.12	$B^b/A^\#$	466.16
A	441.49	A	436.05	A	437.41	A	437.05	A	440.00
A^b	413.42	A^b	418.60	A^b	418.60	$A^b/G^\#$	413.42	$A^b/G^\#$	415.30
G	392.44	G	392.44	G	391.21	G	391.16	G	392.00
$F^\#$	372.52	$F^\#$	367.92	$F^\#$	365.62	$G^b/F^\#$	367.51	$G^b/F^\#$	369.99
F	348.83	F	348.83	F	349.92	F	348.83	F	349.23
E	331.11	E	326.03	E	327.03	E	327.76	E	329.62
E^b	310.08	E^b	313.96	E^b	312.98	$E^b/D^\#$	310.08	$E^b/D^\#$	311.13
D	294.33	D	294.33	D	292.50	D	292.37	D	293.66
$C^\#$	279.39	$C^\#$	272.54	$C^\#$	273.37	$D^b/C^\#$	275.62	$D^b/C^\#$	277.18
C	261.63	C	261.63	C	261.63	C	261.63	C	261.63

Table $14.1.2.1$: Frequencies of various scales, based on $C_4 = 261.626$ Hz.

Modern musical scales in western culture are different in one other way from older classical music. At the time of Bach the scales were based on the note A being about $415\ \mathrm{Hz}$. In Handel's time the frequency of A was $422.5\ \mathrm{Hz}$ and today it is $440\ \mathrm{Hz}$. This gradual shift upward was made possible by stronger materials for pianos and guitars which could withstand the greater tension in tighter strings.

As mentioned previously, the constraints of making a piano with a wide range of notes but is not too large requires using strings that are dense and under a great deal of tension. This makes them slightly nonlinear, meaning that the overtones are not exactly harmonic. If the piano was tuned in a precise mathematical way (using any of the scale systems) these an-harmonic overtones would not sound well when two notes an octave apart are played. What is generally done is to tune the piano by ear so that it sounds correct, starting with low notes and working up to higher pitches. The result of this process is that, for a modern piano, low notes are slightly flat (lower) while high notes are slightly sharp (higher). The figure below (called a ***Railsback curve***) shows the difference from perfect tuning of a modern piano.

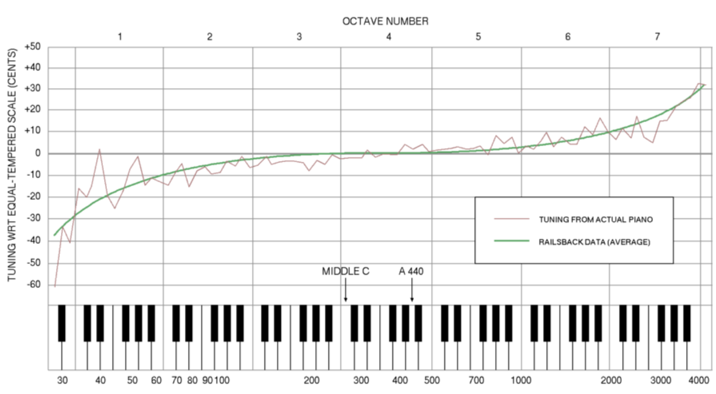


Figure $14.1.2.1$

14.1.3: Temperament Simulation

The simulation shows an octave of notes on a keyboard starting with C which has a modern day frequency of $261.63 \mathrm{~Hz}$. You can click on each key to hear the frequency corresponding to note shown on the key. The buttons across the top allow you to select the tunings of four different temperaments; choosing a different temperament will change the frequency of the notes between C and C an octave higher to match the chosen temperament. The notes sound mechanical because they are pure sine wave with no overtones and so have no timbre. The scale button plays the notes on the scale and the $5 \mathrm{th}$ button plays a perfect fifth; C and G for this scale. Clicking on the overtone checkbox will add the second harmonic (twice the fundamental) at half the volume of the fundamental and the third harmonic with a volume one third the fundamental to any note being played. These overtones are louder than would be normal for a stringed instrument in order to better demonstrate beating and dissonance in the different chord combinations (Go back to chapter 10 to review beats and dissonance if you have forgotten). Three chords used in popular music can also be played. The C Major chord plays the notes C, E and G; the D major chord plays the notes D, $\mathrm{F}^{\#}$ (F sharp) and A; the G major chord plays the notes G, B, and D.

Note

Press Reset if sound does not play when the simulation first loads.

Simulation Quesions:

1. Play a few notes by clicking on the keyboard and then change temperaments and try the same notes. Can you hear the difference?
2. Because the Just Noticeable Different (JND) for most people is a few hertz, you probably were not able to hear the difference between the same note in two different temperaments if they are played separately. Push the scale button and listen to the scale in different temperaments. Can you detect the difference for a string of notes?
3. Again, since the difference between the same note in different temperaments is less than three Hertz you probably can't tell from individual notes played one after the other which temperament you are playing. Now try the fifth in each of the temperaments. What do you notice? Which temperaments have the least beating for the fifth?
4. You can hear the difference in temperaments by listening to a $5 \mathrm{th}$ or chords of three notes. Try each of the three chords in different temperaments. You may hear some beating for some temperaments that you don't hear in others. Describe what you hear for each temperament.
5. Now click the overtone check box and play a few individual notes. What is different?
6. Only the first two overtones are added and they are much louder than would occur in a real instrument so the timbre still isn't pleasant like a real instrument would be. Now play the fifth and the chords in different temperaments with overtones. Which temperaments have less beating and dissonance?
7. You may have noticed less beating and dissonance for the Pythagorean and Just temperaments with overtones. Recall that strings with length ratios of $3/2$ and $4/3$ are thought to sound good together because the frequencies in these two scales have exact ratios. Harmonics of these frequencies also have the same ratio so they also sound good together; the harmonics of one note are also harmonics of other notes. This will not be the case for an equal temperament scale because the 12 notes of the scale are equally spaced in the octave which results in the $5 \mathrm{th}$ note not being an exact $3/2$ ratio of the tonic (it is in fact a ratio of 1.4953 between C and G for the equal temperament instead of $3/2=1.5$ as is the case for the Pythagorean temperament). How does the choice of temperament change how musical scales sound?

Video/audio examples:

- Web site on how to read music.
- Brief explanation of music theory from a musicians' frame of reference.
- Wikipedia on musical tuning systems (towards the bottom) has samples as well.
- The choice of mode is actually a bit more complicated than simply choosing a starting note. See The Physics of Musical Scales by Durfee and Colton.
- A more in-depth discussion of Harmonic Series as it relates to musical scales.
- Wikipedia diagram of harmonic series intervals.
- A more detailed explanation of musical intervals and scales.
- The circle of fifths is a way to see how different musical keys are related. Here is a circle of fifths simulation from Wolfram (you may need to download their plug in to play with this demonstration).

- Pierre Lewis gives a detailed explanation of temperaments.
- Dallin Durfee's Temperament Studio , a Java applet that plays scales and chords in different temperaments.
- Pierre Lewis's Java Tuner, an applet that plays scales in different temperaments. If security blocks the Java you can still download it and use it locally (scroll to the bottom of the page).
- Wolfram's demonstration of musical scales and temperaments (you may need to download their plug in to play with this demonstration).
- Music by Bill Sethares using alternate tunings (click on mp3s for listening).
- Article in *American Scientist* by Cook and Hayashi about The Psychoacoustics of Harmony Perception.
- Human voices are not restricted to sounding a specific note like most musical instruments; the voice is very flexible in the choice of tones it can make. The so called barbershop seventh is an example. The notes in a typical chord have the ratios $4:5:6:7$ which are notes in the Just scale. These frequencies also cause a missing fundamental frequency so that four singers will create five perceived notes. The chord makes use of a tuning scale that cannot be reproduced on a piano tuned to an equal tempered scale. Here is an example from The Music Man (skip to 40 seconds to hear the first chord).

Summary

Musicians even before recorded history have made music and musical instruments that sounded good to them. The oldest flute found is about $35,000$ years old and is made of the bone of a giant vulture. Ancient Greek mathematicians realized that the notes on a string that sounded pleasant together have specific length ratios. For a long time after that mathematicians thought they could codify the scales used by musicians into set ratios of frequencies. With the advent of modern music which uses more than one octave, much more harmony and instruments that have fixed scales that are not easily changed, compromises in the 'perfect' ratios had to be made. The slight differences between Pythagorean and equal tempered scales can be heard by a trained ear but are not enough to change the overall feel of a musical piece.

Questions on Scales:

1. What is the Pythagorean scale and how was it developed?
2. What is an octave?
3. If you hold a guitar string down in the middle each side of the string will be a ________ lower (or half) than the original sound of the full length.
4. In what way is the Pythagorean scale limited? What problems arise from its limitation?
5. Define a perfect fifth and a perfect fourth.
6. What is the ratio of a string's length that produces a perfect 5th? A perfect 4th?
7. Why do notes with a ratio of $2/3$ (a perfect fifth) and other ratios sound good together?
8. Explain the contribution Ptolemy made to the Pythagorean scale.
9. Explain how the Pythagorean formula eventually calculates notes which are not on the desired scale.
10. Why is it that most people cannot hear the small frequency differences between notes on the Polemic scale and the Pythagorean scale (Hint: Review the chapter on perception.)?
11. What is a mode?
12. A particular choice of starting note and system of generating a note scale is called a __________?
13. What is a temperament?
14. What is a semitone?
15. How many semitones are in an octave? How many cents are in a semitone?
16. Why did the question of music scales become more complicated beginning with the 14th century?
17. How are modern musical scales in Western culture different from older classical music?
18. Why and when was the equal tempered scale developed? What problems did it solve?
19. Who popularized the equal tempered scale?
20. On the equal temperament scale _______ cents is about 0.3 Hz.
21. What is the current scale that most modern composers use? What is the compromise of using this scale?
22. When tuning a piano, how are the higher notes tuned as compared to the lower notes?
23. What is a Railsback curve?
24. Why are the overtones in a piano not exactly harmonic? What is the result of this mismatch?
25. Barbershop quartets, generally, have a very rich sound and use a specific type of scale. What scale do they use? Why is it easier for the quartet to use this scale compared to a piano and what happens when their frequencies combine?

26. The note A gradually changed in frequency over time from $415\ \mathrm{Hz}$ to $440\ \mathrm{Hz}$. What made this possible?

CHAPTER OVERVIEW

15: Acoustics

The study of what happens to sound in an enclosed space or as the result of interactions with large objects such as buildings is called ***acoustics***. Humans have been trying to improve the acoustics of auditoriums and other public spaces since the time of the ancient Romans. Marcus Vitruvius Pollio worked as an architect and engineer in the first century BCE and is credited with installing resonating cavities called Echea in Roman amphitheaters. These were made of brass or clay and placed under the seats of the auditorium to enhance the voices of actors on stage. None of these devices still exist and it is not clear if they actually worked or not.

The rules presented previously (Chapter 7) of how waves interact with objects, such as reflection, refraction, path difference, diffraction and interference, govern how sound will behave inside rooms and auditoriums as well as outdoor concert pavilions. Although much is known about the topic, it is still difficult to know exactly what a given band or orchestra will sound like in a newly designed auditorium.

Key Terms:

Acoustics, reverberation, comb effect, echo, direct sound, reverberation time, absorptivity, standing waves, dead spot, feedback, binaural hearing, reflection (specular and diffuse), exponential decay, attenuation, infrasound, matching room acoustics to type of music.

15.1: Acoustics

The study of what happens to sound in an enclosed space or as the result of interactions with large objects such as buildings is called acoustics. Humans have been trying to improve the acoustics of auditoriums and other public spaces since the time of the ancient Romans. Reflection, refraction, path difference, diffraction and interference will govern how sound behaves inside rooms, auditoriums and concert pavilions.

15.1.1: 16.1.1-Acoustic Qualities

There are several general properties that are desirable in an auditorium or concert space. We would like every listener to have exactly the same sound experience. Any noise external to the room should be excluded or minimized. The sounds coming from the performers should be heard clearly and distinctly. If there are any echoes they should be below some threshold relative to the total sound so as not to interfere with the sound being listened to and should not change the sensation of where the music is coming from. The performers should not hear distracting echoes. One of the largest factors impacting how the shape, size and construction of an auditorium affect the sound of a musical performance is reverberation.

15.1.2: Reverberation

Reverberation is the combined effects of multiple echoes in a room or concert hall as heard by a listener. Nearly all indoor spaces have reverberation but it is often missing at outdoor venues. If the walls of the room are hard there will be repeated echoes which may begin to overlap and cause the musical sound to be mushy or muddled. If the walls absorb too much of the sound the room will sound dead. We rely on reverberation to give us a sense of how large the room is; large concert halls have a different reverberation time than a small room.

Video/audio examples:

- Here is a web site with sounds of reverberation in an empty room versus one with carpet.
- Before and after example of unwanted reverberation.

In most of the simulations, animations and drawings of waves in this book (and others) it is assumed that the amplitude (and hence the loudness) of a sound wave doesn't change much as it travels. If the wave front is from a distant source it will be approximately flat (a so called plane wave) and the amplitude does not change as the wave propagates. However for a local source the wave front spreads out in a spherical shape and the further away you are from the sound source the softer it sounds. In other words, the amplitude of a small sound source obeys an inverse square law as it spreads out into open space (Chapter 8). This means the sound intensity decreases as $1/r^2$ where r is the distance from the source. As a result of the inverse square law the ***direct sound*** arriving at the listener will be slightly less loud than the sound at the source and all reflected sound will be even softer. Doubling the distance from the sound will cause the amplitude to be $1/2^2 = 1/4$ as intense. Likewise tripling the distance means the intensity will be $1/9^{\text{th}}$ as much.

In the room without carpet in the demo (above) a sudden sound (a clap) was made in one location. It takes sound, traveling at $345\ \text{m/s}$ a few milliseconds to reach the receiver (a video recorder in this case). This is the ***direct sound*** and is slightly less loud than the original sound at the source. Split seconds later echoes start to arrive at the receiver. Each echo arrives later and with lower volume than the previous one because it travels a bit further. For example the first echo might come off the ceiling, the next from the nearest wall, the next from the wall behind the receiver, etc. A graph is shown below of the initial sound (blue at t_1) followed by the direct sound (green at t_2) and then the echoes (red, $t_{reflected}$ and afterward). Only the first few echos are shown.

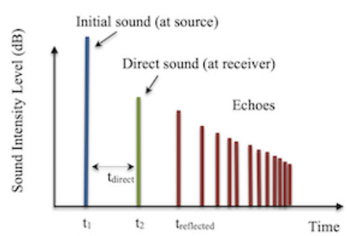


Figure 15.1.2.1

Suppose a continuous sound is made at the front of an auditorium for some length of time instead of a clap like in the demonstration video. The first graph below shows the sound at the source, starting at time t_1, ending at time t_f and lasting a length of time, T. A listener will experience a slight delay until the sound reaches them (t_{direct} in the second graph) and hear the sound as starting at t_2. A split second after the listener hears the sound directly from the source they will begin to hear echoes off walls, floor and ceiling, causing the sound to be slightly louder (the stair steps in SIL starting at $t_{reflected}$). The listener does not hear the sound start to decay until a time t_{direct} after the sound is turned off (sound is still traveling to the listener for a brief time after the source is turned off). After a length of time T the listener hears the direct sound stop (at time t_{decay}) and only the echoes remain and they start to decay. The time it takes the sound to drop in volume by 60 dB from when it starts to decay is called the ***reverberation time*** (T_r in the graph).

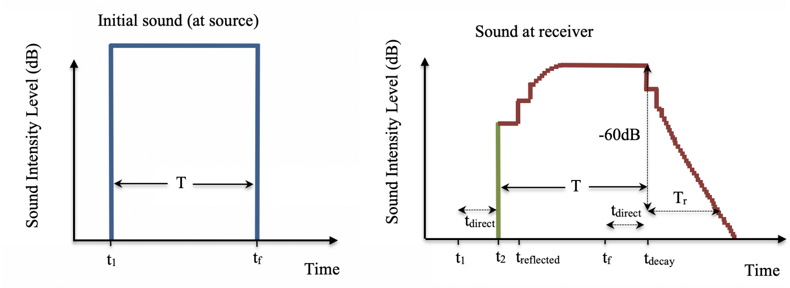


Figure 15.1.2.2

The decay of sound in the graph is fairly linear; it pretty much drops off in a straight line (the SIL scale is logarithmic so an exponential decay in intensity appears as a straight line in an intensity level graph). Real decay curves can have this shape or may be curved or may be irregular shaped. In part this depends on the shape of the room and the materials present which we investigate next.

An approximate equation (due to Wallace Sabine) for calculating reverberation time in seconds is $T_r = (0.16\ \mathrm{s/m})V/S_e$ where T_r is the reverberation time, V is the volume of the room in m^3 and S_e is the effective absorption area in m^2. The effective absorption is $S_e = a_1 S_1 + a_2 S_2 + a_3 S_3 + \ldots$ where S_1 is the surface area that has absorptivity a_1 etc. If the ***absorptivity*** or absorption coefficient is zero the surface is a perfect reflector of sound. Absorptivity is measured in units of *sabin*. A perfect absorber would have an absorptivity of 1.0 S. Below is a table of the absorptivity of various surfaces, in Sabins. Notice that even painting a concrete wall has a significant effect on the absorptivity. This is because unpainted concrete and brick are porous and as a result can absorb more sound.

Typical Acoustic Absorptivity of Different Materials

	Frequency (Hz					
Surface	125	250	500	1000	2000	4000
Acoustic tile, ridged mount	0.2	0.4	0.7	0.8	0.6	0.4
Acoustic tile, suspended	0.5	0.7	0.6	0.7	0.7	0.5
Drywall, gypsum, 1/2 inch on studs	0.3	0.1	0.05	0.04	0.07	0.1
Plywood, 1/4 inch, on studs	0.6	0.3	0.1	0.1	0.1	0.1
Concrete block, unpainted	0.4	0.4	0.3	0.3	0.4	0.3
Concrete block, painted	0.1	0.05	0.06	0.07	0.1	0.1
Concrete, poured	0.01	0.01	0.02	0.02	0.02	0.03
Brick	0.03	0.03	0.03	0.04	0.05	0.07
Vinyl on concrete	0.02	0.03	0.03	0.03	0.03	0.02
Heavy carpet on concrete	0.02	0.06	0.15	0.4	0.6	0.6

Typical Acoustic Absorptivity of Different Materials						
	Frequency (Hz					
Padded carpet	0.1	0.3	0.4	0.5	0.6	0.7
Window glass	0.3	0.2	0.2	0.1	0.07	0.04
Drapes, medium	0.07	0.3	0.5	0.7	0.7	0.6
Upholstered seats, unoccupied	0.2	0.4	0.6	0.7	0.6	0.6
Upholstered seats, occupied	0.4	0.6	0.8	0.9	0.9	0.9
Wood or metal seats, unoccupied	0.02	0.03	0.03	0.06	0.06	0.05
Wood or metal seats, occupied	0.4	0.4	0.7	0.7	0.8	0.7

Table $15.1.2.1$: Actual values will vary depending on the exact makeup of the material. Sources: D. E. Hall, *Musical Acoustics*, 2002

As we saw in Chapter 7, a path difference can cause destructive or constructive interference due to a phase difference. If there is strong reflection it can be the case that reflected sound and sound coming directly from the stage may reach a listener out of phase and thus cancel destructively for a particular wavelength as shown in the diagram below.

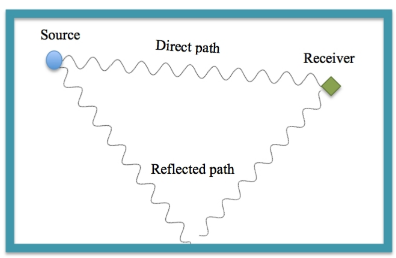


Figure $15.1.2.3$

The effect depends on the amount of delay and can result in peaks and dips in the frequency spectrum at the listener location. This is sometimes called the comb effect and can be produced artificially by adding a recording to itself but with a slight delay. This can also lead to locations in a concert hall called ***dead spots***, where the range of frequencies in the original sound source is not heard correctly. As a result of interference between the direct sound and reflected sound, different kinds of music sound better with different reverberation times. Medieval music was often performed in cathedrals with very long reverberation times, sometimes as long as 13 seconds. The composers and musicians of the time gradually modified the music to fit the venue. This type of music is usually very simple and uses lots of harmonic notes so that the echoes would harmonize with and reinforce the notes being sung. Pipe organ music generally sounds better with more reverberation for the same reason. In general, symphony music sounds best with less reverberation than pipe organs but with more reverberation than chamber music. Symphony halls typically have reverberation times of two to five seconds; 2.5 seconds is considered optimal. Speech usually sounds best with very little reverberation; too many echoes makes the speaker difficult to understand. Opera houses and speech auditoriums are designed to have reverberation times of 0.8 to 1.5 seconds. A typical living room in a house has a reverberation time of 0.4 seconds.

Reverberation can be artificially added to recorded and live music using electronic equipment (sometimes referred to as comb filters). This may be done, for example to a voice to give it a more full bodied sound or to make an outdoor concert sound like it is indoors. Reverberation may also be added to give a recording a sound as if the listener was in a larger auditorium when in fact they are listening from headphones. The sound tracts of most modern recorded music, movies and TV programs are heavily processed to overcome deficiencies in the sound quality due to the acoustics of the recording location.

Video/audio examples:

- A short YouTube about reverberation.
- A short YouTube about Balloon Pop with and without reverberation.

15.1.3: Other Acoustical Effects

As we saw in Chapter 11 on stringed instruments, the equation for the Helmholtz frequency modes inside a rectangular container with height H, length L and width W is $f_{n,l,m} = v/2((n/L)^2 + (l/H)^2 + (m/W)^2)^{1/2}$. Here v is the speed of sound in air and n, l, and m are mode numbers. For a given set of mode numbers there will be standing waves in the container. You can hear standing waves when you sing in a rectangular shower stall. Stand in the shower stall and, starting with a low tone, gradually increase the pitch. You will find certain notes that sound louder; these are the standing waves for that enclosure. In a rectangular auditorium the possibility of standing waves results in two problems. For a piece of music certain notes will be enhanced while others will not. You can explore these modes in this box modes simulation Applet (notice you can grab and rotate the box to see different modes from different angles).

A second problem when there are standing waves is the occurrence of nodes and anti-nodes. You can often detect these nodes by moving your head slightly to the left or right in a room where there are parallel walls and some ambient noise; you will notice the sound changes a little bit as you move your head. A listener seated at a node hears less of that frequency while a listener at a anti-node hears more. You may have noticed that auditoriums are not perfectly square but have sloping floors and ceilings and walls at an angle to the stage as seen in the picture below of Ríos Reyna Hall in Caracas, Venezuela. This is in part to eliminate or reduce standing waves. However the effects of a non-rectangular room on reflected sound can be complicated and unpredictable. As a result, acoustical design is still somewhat of an art.

Figure 15.1.3.1

Newly constructed concert halls may be built with floating floors, mounted on springs and dampeners to isolate the space from external vibrations. The walls and ceilings may also be isolated from the surroundings using acoustically isolating connectors to attach them to external walls and ceilings. This 'box within a box' construction greatly reduces interference from external noise.

If the building is already in place it may be possible to improve the sound and remove at least some of the standing waves by using absorbing barriers of various sizes and shapes. Recall from Chapter 7 that, although sound obeys the law of reflection, if the surface has irregularities of the same size as the wavelength the reflection will be diffuse instead of specular. Obviously objects like chandeliers, seating, and lights will also interact with sound wavelengths of the same size causing diffuse reflection. Diffuse reflection tends to help prevent standing waves and dead spots.

Although it is not known if the Echea of Marcus Vitruvius Pollio actually reinforced the sound of the actor's voices onstage, a similar type of resonance cavity is used today to *reduce* unwanted sound from fans in air ducts. If a resonance cavity is attached to an air duct and has a resonance frequency just below the fan frequency the cavity will produce a sound wave exactly out of phase of the sound from the fan. This tends to cancel the unwanted noise from the fan.

As a final aid to getting sound from the stage of a concert hall to the listeners, a orchestral or band shell is often used. As we saw in Chapter 7, waves emitted from the focal point of a concave mirror are reflected outward in parallel beams. An orchestra shell is half a concave parabolic mirror. The orchestra or band performs at or near the focal point of the curved surface and the surface projects the sounds out in even parallel paths to the audience. These shells can be used indoors or, as in the picture below, outside.

Figure 15.1.3.2

15.1.4: Sound Reinforcement

As if the problem of sound reflection wasn't complicated enough, most auditoriums today use amplified sound in the form of microphones and speakers. If presented with two different sources (for example a singer and the amplified singing from a speaker) our ear-brain hearing system may be confused about where the sound is coming from. If the sounds arrive more than 30 milliseconds apart (the case when the speaker and singer are more than 10 m apart) we will hear an echo. If the speakers are closer to the listener than the original source the sound will appear to come from the speaker, not the singer because our ear-brain system identifies whichever signal getting to the ear first as the source. This is called the ***precedence effect*** and can be manipulated to fool the listener into thinking the sound comes from the singer and not the speaker. This will work, even if the direct signal is much weaker than the sound from the speakers. To arrange this, a very slight delay is often introduced in the amplification system so that the direct sound arrives just ahead of the amplified sound of the speakers. If there are speakers at different locations in the auditorium this requires careful arranging of speakers and the amount of delay as shown in the following diagram.

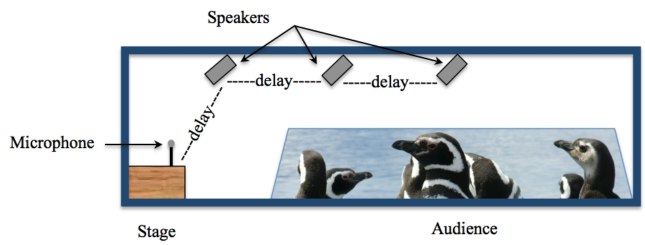


Figure 15.1.4.1

A common problem with amplified sound is ***feedback***. Feedback occurs when sound from the speaker is fed back into the microphone (usually accidentally). The amplifier system tries to amplify the sound again (and again and again). Once the amplifier reaches its maximum capacity to amplify, the sound is distorted and an unpleasant, high pitched squeal results (although a few musicians have actually used feedback as part of their performance).

Video/audio examples:

- Wikipedia on acoustic feedback.
- Jimmy Hendrix was a rock star in the 1960s who helped make the use of feedback in music popular.
- Two discussions of acoustics: serious, for the cinema.

A sound sample of feedback:

15.1.5: Spatial Perception

Our perception of distance and direction also strongly affects our experience in a concert hall. We determine the distance to a sound source in part by how loud it is. Part of our clues about distance also come from reverberation. If there is little or no reflected sound it is difficult to determine how far way a sound source is.

In addition to ***reverberation*** our ear-brain hearing system uses four other methods to determine the direction and distance to a sound source. The distance between your ears is about 15 cm, causing as much as a 0.4 milliseconddelay for sounds coming from one side. If a sound arrives at the left ear a little before the right ear we perceive the sound as coming from the left. This ***time delay*** gives us clues about left and right but does not help us distinguish front from back or up from down. It works best for short wavelengths of sound that have to travel around our head to reach the other ear.

Our ear/brain system is also capable of detecting ***phase difference*** in some special cases. Recall from Chapter 3 that the phase of a wave tells us when it starts relative to some arbitrary starting time. If the phase of the peaks in the sound wave arriving at the left ear is slightly ahead of the phase of the peaks arriving at the right ear our ear-brain system interprets the sound as coming from the left. Notice this is different from a time lag between ears; phase differences can be detected even for a continuous tone that does not start or stop. This ability occurs most prominently for sounds below 1500 Hz. The picture below shows a difference of phase between a wave going to the right ear and a wave going to the left ear.

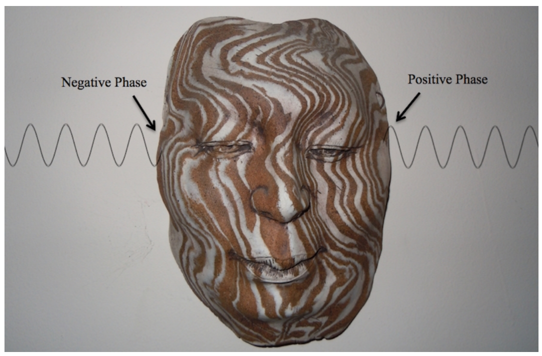


Figure $15.1.5.1$

For higher frequencies we may be able to detect slight ***intensity differences*** of sounds coming from different directions. Frequencies between 2 kHz and 3 kHz have wavelengths that are larger than your head so they diffract (see Chapter 7) around your head, making the time delay negligible. But the intensity difference still provides a clue about direction. If the left ear perceives a louder sound we experience the sound as coming from the left.

The above three methods of determining direction to a sound source work well for left-right determination and require two ears. This makes them ***binaural*** methods for determining direction. A forth method of determining direction relies on the ***shape of the outer ear***. For frequencies above 5 kHz the shape of the outer ear reflects sounds coming from different directions with slightly different time delays. Slight movements of the head change the amount and time delay of these frequencies relative to other frequencies which gives us a sense of whether the sound is coming from in front or behind us.

For humans our left-right determination of sound direction is stronger than our ability to determine whether sound is coming from above or below. This has the practical result that sound from speakers in a concert hall which are too far apart in the horizontal direction do not sound right; we are not easily fooled into thinking the sound is coming from the center of the stage where the performers are. Because our up-down discrimination is not so refined we are less bothered by speakers hanging overhead. For the most part we tend to perceive the sound as coming to us from the front even when the speakers are above us.

15.1.6: Animal Acoustics

Note

Much of the information about animal perception in this chapter and in Chapters 10 and 14 comes from the excellent book *Engineering Animals: How Life Works* by Mark Denny and Alan McFadzean.

Animals use sounds to communicate to others in their species, to navigate and to find prey. As mentioned previously, small animals tend to make and use high frequency sound to communicate whereas larger animals tend to use lower frequencies but there are exceptions. The size of an animal also affects which of the methods in the previous section it might use to detect the source of a sound. Animals with small heads tend to use phase difference to locate a sound source because the timing difference is too small if your ears are closer together.

Under normal circumstances low frequencies travel larger distances with less ***attenuation*** or loss. Elephants use this fact to communicate over distances of a few kilometers using ***infrasound*** (frequencies below 20 Hz). Woodpeckers use hollow trees to create lower sounds than they sing in order to attract mates and establish territory over longer distances. There is some evidence that pigeons can use the infrasound acoustic signatures of particular land features for long distance migration. Because higher frequencies attenuate sooner than low frequencies, particularly in forests, the timbre of a birdsong will change depending on the distance from the source to the listener. Some birds appear to be able to make use of this fact to locate mates and competitors in places where visibility is limited.

Water carries low frequencies much better than air and whales and dolphins use low frequency sounds to communicate. In the ocean the speed of sound is affected by temperature, salinity and pressure. Temperature decreases with depth but pressure increases with the result that there is a layer of the ocean at a depth of about 0.7 km where the speed of sound is the slowest. At this particular depth, called the Sound Fixing and Ranging channel (SOFAR), the speed of sound is about 1480 m/s (compared to as much as 1540 m/s at other temperatures and pressures). The layer acts as a wave guide; sound that tries to leave the layer refracts back into the layer because of total internal reflection, just like light trying to exit a fiber optic cable at a small angle. Because sound is trapped in this layer it can travel long distances with very little attenuation. The navy has investigated the SOFAR channel as a potential way to detect sounds of enemy submarines at large distances. Some whales appear to be able to communicate over hundreds of kilometers by emitting low frequency sounds in the SOFAR layer.

Most birds do not emit sound signals to navigate or find prey but instead use intensity, phase difference or timing differences of ambient sounds as described in the previous section. The ears of owls are located asymmetrically at different heights on the head. Owls can also change the orientation of the feathers around the ears to get slightly different phase information from a sound source. Based on laboratory experiments we know some owls use timing to establish the azimuthal (up and down) angle to a prey but they use phase shifts to determine the horizontal location of a sound source. This probably gives them something like a two dimensional picture of what is around them using sound instead of light.

A few birds navigate by ***echolocation***, emitting signals and listening for their return. Oilbirds and swiftlets can navigate by timing how long a signal takes to return, the least sophisticated form of echolocation. This allows them to find their nests in dark caves where they nest and avoid collision with other birds but they do not use echolocation to find prey.

Bats and toothed whales are best at echolocation. Higher frequencies have shorter wavelengths which increase accuracy so most echolocating animals use high frequency chirps or clicks to echolocate. Because sound attenuation in water is less than in air, whales can echolocate over much larger distances than bats. Bats use sound to communicate with others of their own species over distances of 50 m to 100 m but must be within about 5 m to use echolocation to hunt prey.

About 800 species of bats echolocate and some have abilities beyond what humans can do with radar or other electronic forms of echolocation. Bats often have unusual face and nose shapes in order to funnel the emitted sound into a narrow beam which helps avoid spurious echoes from background sources. They also have sophisticated brain circuitry which can measure the Doppler shift (Chapter 7) to not only locate their prey but also determine how fast and in which direction the prey is traveling. There is some evidence that some bats can even use the very small Doppler shift from the flapping of the wings of a moth to identify which kind of moth is present. Things are not all bad for the moths, however. Tiger moths can detect the signals emitted by its bat predator and take evasive action and they can even emit a 'jamming' signal to confuse the bat.

Whales and dolphins emit higher frequency clicks, chirps and other sounds which they use for echolocation in addition to lower frequency sounds for communication. It has been shown that dolphins can emit clicks of around 50 micro seconds in duration which gives them up to a $1 \mathrm{~cm}$ resolution at a distance of $100 \mathrm{~m}$. Many of these marine mammals have structures in their foreheads which focus the sounds into a beam pointing forward. When dolphins are closer to their prey they change the frequency of the chirps which gives them a $0.5 \mathrm{~mm}$ resolution at a distance of $1 \mathrm{~m}$. Because some of these sounds penetrate into the fish they are hunting it is very likely dolphins can identify the type of fish present by forming basically a three dimensional picture of the fish, including its skeleton and inner organs using only sound.

One might wonder how it is that echolocators don't deafen their delicate ears with the sounds they emit. We know sound intensity dies off as an inverse square law (Chapter 8) so by the time a signal returns after reflecting off a prey it must be very weak, in fact as much as 10^{20} times weaker. The bat or whale wants to maximize the output signal but have ears as sensitive as possible for the return signal. How is this possible? Most echolocators have mechanisms that mechanically disconnect their hearing apparatus when they are emitting signals.

15.1.7: Other Acoustical Applications

The science of acoustics has been and is being applied to a large number of different problems. The following is a list of interesting applications of acoustic techniques.

Video/audio examples:

- A YouTube example where they manipulate your hearing to mimic the effect of spacial changes in the location of the source (requires headphones).
- Sound tourism; a list and samples of interesting sounds in the world by Trevor Cox.
- Interview with a blind person who uses sound to navigate.
- Putting a fire out with sound.
- Mechanical sound cancellation using acoustic metamaterial (Ghaffarivardavagh, Nikolajczyk, Anderson, and Zhang).
- In the phenomena of thermoacoustics sound is can be turned into heat or heat into sound. Using this conversion it is possible to make a thermoacoustic heat engine that acts as a heat pump or refrigerator.
- Levitation of small objects with standing sound waves. Another YouTube. Here is an article in Science about uses of sonic levitation.
- Moving Helmholtz resonators.
- Making standing wave patterns in liquids with sound. A second example with a cornstarch solution. A third example, more controlled.
- Explanation of earthquake booms, Seneca guns and other sounds.
- Discussion in *Physics Today*, vol. 71, issue 8, 2018, of infrasound as a tool for the remote sensing of thunderstorms, volcanoes, nuclear bombs and more.
- Turning scientific data into sound: Samples from a group in Germany; Wikipedia article on Sonification; Sonification of Tohoku earthquake, Japan; Sonification of particles from the sun hitting satellites; Sonification of Voyager data.
- Singing sand dunes. Here is a YouTube of booming sand.
- Here is a list of software that allows you to modify sound files to create your own acoustical effects (I have only used Audacity; I cannot vouch for the others):
    - Audacity.
    - Gold Wave.
    - Adobe Audition.
    - Praat.
    - Max and several other sound products.
    - Melodyne.
    - Fleximusic.
- Various resources on musical acoustics from the University of New South Wales.
- Dan Russell's page of simulations and animations on acoustics.

Summary

Auditorium design generally includes gently curving surfaces, diffuse reflection and controlled absorptivity of sound. Large flat surfaces with strong, specular reflection are avoided. Still, auditorium design is only partially scientific; a great deal is left to trial and error. Our perception is sensitive to reverberation, echoes, timing, intensity, phase and other clues about the direction and distance to a sound source. These effects have to be taken into account in order for a performance to sound natural, especially when electronic amplification is used.

Questions on Acoustics:

1. Define acoustics.
2. What are some of the acoustic qualities to be considered when planning a music or concert hall?
3. Who was Marcus Vitruvius Pollio and what did he do?
4. What is the principle behind Echeas?
5. Explain reverberation and how it is measured.
6. Why would you not want to get rid of reverberation entirely?
7. What does it mean if the absorptivity of absorption coefficient is zero?

8. Absorption is measured in what units?
9. If the distance to a sound source is doubled, that happens to the sound intensity?
10. What is the Comb effect and why is it useful?
11. Explain feedback (what is it, when does it occur, why does it occur).
12. What is a way to minimize echoes in a concert hall?
13. What is one of the largest factors that impact the sound of a musical performance?
14. Why do auditoriums have walls and floors that are slanted and angled?
15. In addition to reverberation our ear-brain hearing system uses four other methods to determine the direction to a sound source. Write a brief description of each.
16. What is the precedence effect?
17. How does our perception of the direction of sound from a vertical angle differ from a horizontal angle?
18. How are we fooled at a concert to think the sound we hear is coming from the singer and not from the speakers on each side?
19. A person is singing on stage, but their voice is amplified by speakers that are closer to the audience. How can you make it so the audience perceives the sound originating from the stage and not the speaker?
20. Why do speakers overhead not cause us to think the musicians are on the ceiling?
21. Is it possible to design a perfect concert hall using only scientific principles? Explain your answer.
22. Why are multiple speakers aligned in a plane going away from the stage each given a slight delay?
23. Name a famous rocker that used feedback from his guitar as part of his musical performance.
24. What two problems do standing waves pose for sound in a room?
25. How can dead spots be reduced in a room?

CHAPTER OVERVIEW

16: Electricity and Magnetism

This chapter is a very brief overview of some key concepts in electricity and magnetism. Most of our modern technology is based on these concepts. In the next chapter we will use these concepts to explain how microphones, speakers and digital recording work.

Key Terms:

Ohm's law, charge, coulombs, current, amperes, electrical potential, volts, resistance, ohms, magnetic field, compass, tesla, gauss, electromagnet, force on an electric charge in a magnetic field, electric motor, Faraday's law, electric generator.

16.1: EandM- Ohm's Law

The electrical resistance of an electrical conductor is a measure of the difficulty to pass an electric current through that conductor. It is measured in Ohms and the relation between resistance (R), current (I) and electrical potential (V) is Ohm's law: $V = IR$. Ohm's law says that a larger voltage makes more current flow if resistance is fixed. Or if resistance is lower at the same voltage, more current will flow.

16.1.1: Ohm's Law

Atoms are made of protons, neutrons and electrons. The protons and neutrons (each of which is 1800 times heavier than the electron) are found at the center of the atom in the nucleus. The electrons surround the nucleus and occupy most of the space in the atom. ***Charge*** is measured in ***coulombs***; a proton has a positive charge of $+1.6 \times 10^{-19}$ C while the electron has a negative charge of -1.6×10^{-19} C and the neutron has no charge. The nuclear force holds the protons and neutrons together; the electrical force between the protons and electrons keeps the electrons from wandering off. The electrons associated with metal atoms such as iron, aluminum or copper can move freely from one atom to another in a solid made of these atoms. A flow of electrons is called an electrical ***current*** which is measured in *amperes*. An ampere is a flow of one coulomb of charge per second.

What is required for a current to flow? We already know there must be a metal conductor so that there are free electrons. There also must be energy to make them start moving. Recall earlier in the semester we said that if you raise a mass up to a certain height it would store gravitational potential energy, measured in joules. In electricity, instead of measuring potential energy we measure the potential energy *per charge* which is called *voltage*. A volt is a joule of potential energy per each coulomb of charge. When electrons flow through an electrical component such as a light bulb, toaster or radio they give up some of their energy but the same number of electrons per second flow out as are flowing in (charge is conserved; we don't lose any of it). Although the same current (electrons per second) flows out of the component as flows in, the total current in the circuit is controlled by the ***resistance*** of the circuit. Resistance is measured in *ohms*. The relation between resistance (R), current (I) and electrical potential (V) is ***Ohm's law***: $V = IR$. Ohm's law says that a larger voltage makes more current flow if resistance is fixed. Or if resistance is lower at the same voltage, more current will flow.

A good analogy for Ohm's law is water flowing in a system of pipes. Current is like the water, voltage is like the pressure and resistance is analogous to the size and number of pipes. There is pressure in the pipes (potential energy in the battery or power supply) even if the water isn't flowing right at the moment. Flipping a switch in an electrical circuit is equivalent to turning a valve on in a water flow system. If the pipes are bigger or there are more of them (less resistance) there is more flow (more current) at the same pressure (same voltage).

Video/audio examples:

- Ohm's law simulation (turn in answers on a separate sheet of paper): Electrical circuits. Requires JAVA.

> **Note**
>
> There is a simplified HTML5 version of this simulation also available from PhET.

1. Download or run the simulation from the web page. Drag a battery, light bulb, switch and several wires from the panel onto the screen. Connect them so that you get current to flow and the light bulb to light. Draw a sketch of your circuit. Hint: The different items connect together automatically; you can separate them by using a control click on the junction.
2. Click the box for adding a non-contact ammeter. Drag it over different parts of the circuit. What is the current in the circuit? Is it the same everywhere? If the current isn't used up in the circuit, what is?
3. Click on the box to add a voltmeter. You will notice there are two leads to the voltmeter; this is because it measures the potential energy *difference* between two locations. What is the voltage difference across the battery? What about the voltage across the light bulb? What about the voltage from one end of a wire to the other end?
4. Separate one of the junctions and add a resistor. What happens to the current compared to the circuit with no resistor? What happens to your voltage readings?
5. Replace the battery with the AC source. What is the difference between a battery (which provides DC or direct current) and the AC (alternating current) source?
6. Add a current chart to the circuit. Move the probe over various parts of the circuit. Describe or sketch what the current is doing, according to the current chart.

- The basics of Ohm's law.
- It's not the voltage that kills you .... or is it?

Questions on Ohm's Law ($V = IR$):

1. Suppose you have a wire in a circuit that has 10 A flowing in it and it branches into two other wires. If there is 7 A in one of the wires, how much must be flowing in the other?

2. Which is more like an electrical circuit, the cooling system of your car or the plumbing in your house? Explain.
3. Why is it incorrect to say that voltage flows around a circuit? (Hint: start with definitions of current and voltage.)
4. A person standing on an insulated stool touches a charged insulated conductor. What happens?
5. Birds sit on high-tension wires and do not get electrocuted, even when the wire is bare, yet a squirrel which steps from a bare wire to a pole or to another wire dies instantly. Why?
6. What happens to the brightness of a light bulb if more current flows through it?
7. What is the difference between current (in Amperes) and voltage (in Volts)?
8. If you want make a brighter light bulb, do you want to increase the resistance or decrease the resistance of the filament? (Hint: The brightness increases if more current flows.)
9. What is the difference between AC and DC?
10. What do batteries supply to an electric circuit?
11. When a battery dies, is it out of electrons or out of energy? Explain.
12. If the same amount of current (electrons per second) flows into a light bulb flow out of it, what is being 'used up' in the circuit?
13. If the same current flows into your house as flows out of it, why do you have to pay for electricity?
14. Is it current or voltage that cause electric shock?
15. Wet feet reduce the resistance between you and the ground. From this fact explain why the same 120 V outlet is much more dangerous if you have wet feet than if you have dry feet.
16. Why do warnings on power relay stations say 'Warning, High Voltage' when it is the current that is dangerous?

16.2: EandM- Currents and Magnetic Fields

In this section we study the magnetic field of either a permanent magnet or the field produced by a flow of current in a coil. Field is measured in Gauss. The compass, magnet and coil are all draggable. The earth's magnetic field can also be demonstrated.

16.2.1: Currents Cause Magnetic Fields

Out of the some 117 elements known in the world, three are special. Cobalt, nickel and iron have a special arrangement of electrons that causes them to have a magnetic field. All permanent magnets are composed of one or a combination of these metals. Magnetic fields are measured in either *tesla*, $\mathrm{T}=\mathrm{Vs}/\mathrm{m}^2$, or *gauss*, where $1\ \mathrm{G}=10^{-4}\ \mathrm{T}$.

The magnetic field of individual cobalt, nickel or iron will tend to line up with each other over limited regions called domains. As a result of this alignment the magnetic field due to all the atoms in a particular domain points in one direction. If the domains line up with each other in a piece of metal you have a magnet; if the domains do not line up the magnetic field cancels and the piece of metal is not a magnet, although the individual atoms still cause each domain to have a magnetic field. One way to destroy the magnetic field of a magnet is to drop it or heat it so that the domains reorient in such a way that the total field is zero.

Technology would be very limited if the only way to get a magnetic field was from one of the three special metals. Fortunately we can arrange a current flow of electrons in a special way so that we can mimic the magnetic field found naturally in these three elements. All currents create magnetic fields but when the arrangement of the current is specifically for the purpose of making a magnetic field we call it an ***electromagnet***. Electromagnets can create exactly the same magnetic fields found in permanent magnets but have the added advantage that they can be turned off or reversed by turning off or reversing the current flow through them.

16.2.2: Magnetic Field Simulation

In this simulation you examine the magnetic field of either a permanent magnet or the field produced by a flow of current in a coil. Field is measured in gauss. The compass, magnet and coil are all draggable. The earth's magnetic field can also be demonstrated.

Simulation Questions:

1. First leave the magnet stationary and move the compass around by dragging it at its center. Which end of the compass points towards the north pole of the magnet? What happens at the south pole of the magnet?
2. Describe the magnetic field around the magnet. What happens to the field strength as you get further away? What happens near the middle of the magnet?
3. Reset the simulation. Now leave the compass stationary and move the magnet around. Describe the behavior of the compass needle. Is your description of the magnetic field different from the case of moving the compass? Explain.
4. Click the 'Earth' radio button. What do you notice about the magnetic field of the earth? Is the geographic north different from the magnetic north? Explain.
5. Reset the simulation and click the 'coil' radio button. Leave the coil stationary and move the compass around. How does the magnetic field of the coil compare to the magnetic field of the bar?
6. Now try different currents in the coil using the slider. What happens when the current is zero? Explain what happens when the current is negative (which means it flows in the opposite direction).

16.3: EandM- Electric and Magnetic Forces

In this section we study electric and magnetic fields with different orientations to see their effects on neutral, positive and negative charges. For the electric field case the particles have zero initial velocity. In second case with a magnetic field in the x-direction the initial velocity is zero but there is a check-box so that you can give the particles an initial velocity in the $+x$ direction. In the third case the magnetic field is rotated so that it points into the screen.

16.3.1: Currents in a Magnetic Field MAY Experience a Force

Positive charges (protons) repel each other as do negative charges (electrons). Unlike charges (electrons and protons) attract. Most material contains equal amounts of positive and negative charge because of the strong attraction between protons and electrons. However, you can demonstrate this attractive electric force by rubbing various things like plastic and fur together. For example, the clothes in your dryer will stick together when they come out because in the process of tumbling, some electrons get rubbed off onto one piece of clothing from another. The piece of clothing that loses electrons has too few electrons and is positive. The article of clothing that picked up the electrons is now negative. So the two pieces stick together. These are *electrical forces* and have **nothing** to do with magnetism.

We also know that magnetic north poles attract magnetic south poles but repel other magnetic north poles. These are *magnetic forces* and have **nothing** to do with electrical forces. So what happens to an *electrical* charge when it encounters a *magnetic* field? Stationary electric charges in a magnetic field do not move; there is neither attraction nor repulsion. Charges initially moving parallel to a magnetic field do not change their motion and they also feel no force due to the magnetic field. Only if the charge is moving and tries to cut across a magnetic field does it feel a force as demonstrated in the simulation below.

An aurora (for example the aurora borealis (norhtern polar lights) or the aurora australis (southern polar lights)) is an interesting example of the force on moving charges. Positive charged particles come streaming in from outer space but may interact with the earth's magnetic field if they are moving in the right direction. At the equator they cut across the magnetic field of the earth and so are deflected into orbiting bands of particles called the Van Allen radiation belts. At the poles the charges are moving parallel to the field and so are not deflected. When they hit the upper atmosphere they give off energy in the form of light as they collide with air molecules and slow down. It is this light that forms the auroras borealis and australis.

The most important application of the force on a charge crossing a magnetic field is the electric motor. In an electric motor, current in a wire flows through a magnetic field. The magnetic force on the charges moving in the wire causes a force on the wire. By carefully arranging the wires and fields we can get the force to make a current carrying coil of wire turn in a circle. This is the basis behind an electric motor, as shown in the diagram below.

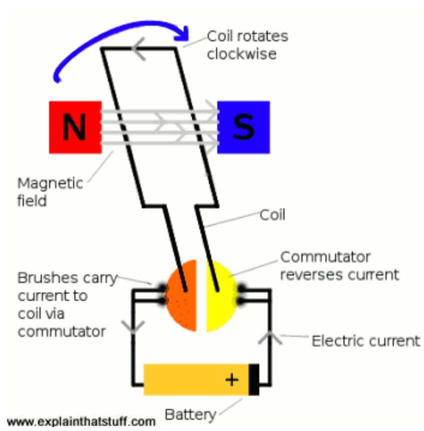


Figure 16.3.1.1

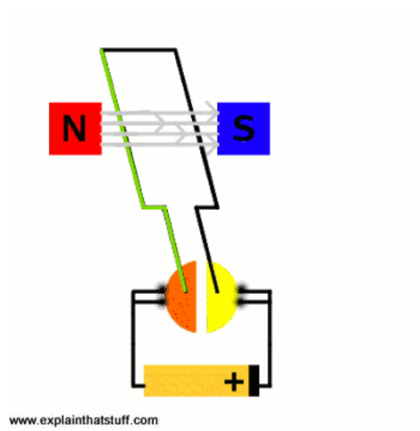


Figure $16.3.1.2$

In the picture on the left, above, current flows from the battery at the bottom up the right side of the coil (black wire; real motors have hundreds of turns but the drawing only shows one for clarity). As the current passes the south end of the magnet (blue) it feels a downward force. The current continues across the top of the coil and comes downward past the north end of the magnet (red) where it feels an upward force. The current then returns to the battery. The upward force on the left half of the coil and the downward force on the right half of the coil cause the coil to rotate 180 degrees. What keeps it going on around? Once the left side has flipped around to the right (and the right side to the left) notice that the current will still be going up on the right and down on the left. The commutator has reversed the current flow. So the coil continues to rotate in the same direction (blue arrow). See if you can follow the sequence in the animation on the right.

16.3.2: Electric and Magnetic Forces Simulation

In this simulation you can select an electric field and two different orientations of a magnetic field to see their effects on neutral, positive and negative charges. For the electric field case the particles have zero initial velocity. In second case with a magnetic field in the x-direction the initial velocity is zero but there is a check-box so that you can give the particles an initial velocity in the $+x$ direction. In the third case the magnetic field is rotated so that it points into the screen (which is now the x direction) and the particles have an initial velocity in the $+z$ direction. For this case a black arrow shows the direction of the force on the particle. Light grey arrows show the velocity of the particle.

Simulation Questions:

1. With the electric field case selected, run the simulation for each of the three particles (neutral, positive and negative) and describe what happens (reset the simulation between each choice). Do the charges have a constant velocity or are they accelerating? Hint: The length of the gray arrow is proportional to the speed of the particle.
2. Reset the simulation and choose the second case with a magnetic field in the x-direction and the particles with a zero initial velocity. What happens to the particles? Can the magnetic field cause them to accelerate if they are initially stationary?
3. For the magnetic field in the x-direction, use the check-box to give the particle an initial x-velocity. Try the other charges, resetting the simulation each time. Describe the motion of the particles. Do they accelerate in this case? (Again; the length of the gray arrow is proportional to the speed.)
4. Reset the simulation and choose the magnetic field pointing into the screen (the third case at the top). Each of the particles start with the same initial x-velocity, indicated by the grey arrow. Describe what happens in each case.

Questions on Magnetic Fields:

1. Electric fields come from charges. Where do magnetic fields come from (your answer should include more than just saying “magnets”)?
2. In what sense can we say that the ultimate source of all magnetic fields (even permanent magnets) is moving charge?
3. What is an electromagnet?
4. Suppose you float a magnet in a bowl by attaching it to a piece of Styrofoam. Will it drift towards the north side of the bowl due to the attraction of the north pole of the magnet? Why or why not?
5. Opposite magnetic poles attract each other. So why does the north end of a compass point north?
6. If a compass needle could point in any direction (north, south, up, down, etc.) which way would it point if it were located at the earth’s geographical north pole?
7. Does a compass still point north if it is in the southern hemisphere, or does it reverse?
8. Do electrical charges always feel a force due to a magnetic field? Explain.
9. If a charged particle moves in a straight line through a region of space at constant speed, can you say that the magnetic field in that region is zero? Explain.
10. What is the origin of the Aurora Borealis (the Northern Lights)? Why are the not usually seen near the equator?
11. Residents of Alaska get hit by more cosmic rays (charged particles from space) than residents of Panama. Why is that?
12. Explain how an electric motor works.

16.4: EandM- Faraday's Law

If a changing magnetic field is present near a wire that is part of a circuit it will cause current to flow in the circuit. This is known as Faraday's law and is the basis for a lot of modern technology. Electric generators, metal detectors, the read head on a computer hard drive, credit card readers, cassette tape readers. We will see several applications for sound reproduction.

16.4.1: A CHANGING Magnetic Field Can Cause Current to Flow

If a *changing* magnetic field is present near a wire that is part of a circuit it will cause current to flow in the circuit. This is known as ***Faraday's law*** and is the basis for a lot of modern technology. Electric generators, traffic detectors embedded in the road, metal detectors, the read head on a computer hard drive, credit card readers, cassette tape readers, and transformers (both the ones on the utility pole outside your house and the little boxes that plug into the wall to run electronic gear) all use Faraday's law to operate. We will see several applications for sound reproduction in the next chapter.

One important application of Faraday's law is an electrical transformer. A transformer consists of a piece of iron with two separate coils wrapped around it. One coil is called the primary, the other coil is called the secondary. If there is an alternating current (AC) in the primary there will also be a changing magnetic field. This changing field will induce a current in the secondary due to Faraday's law. Although the wires of the primary are not physically connected to the wires of the secondary, a current flows in the secondary if an alternating current flows in the primary.

Why have a transformer? Why not just connect the secondary to the primary directly? The voltage and current can be adjusted between the primary and secondary by changing the number of loops in each. If the primary has 100 loops and the secondary 10 loops the voltage in the secondary will be $1/10$ of the voltage in the primary but the current in the secondary will be 10 times that of the primary. The voltage ratio between primary and secondary is proportional to the ratio in the number of turns of wire in each. This makes it easy to step up voltage (while stepping down current) or step down voltage (while stepping up current).

Why change voltage with a transformer? Electricity is delivered to your neighborhood at high voltage low current. Low current means less loss to resistance between the power plant and your neighborhood. Another transformer near your house steps the voltage down and the current up so that more of current is delivered to your home. You want more current but less voltage so that more energy is available for your electrical appliances.

16.4.2: Faraday's Law Simulation

In this simulation you can experiment with Faraday's law. One important application is also shown; a moving coil microphone. Magnetic flux is defined to be the amount of magnetic field passing through a given area and has units of webers where $1\ \mathrm{Wb} = 1\ \mathrm{Tm}^2$. The technical definition of Faraday's law says that a changing magnetic flux (ϕ) causes an electromotive force (***emf*** in volts). If there is a circuit, that emf will cause current to flow, just like a battery. A second law, ***Lenz's law***, says that the current flow will be in a direction opposite to the change that induced the current flow. The current direction is shown two different ways; one is by the arrow in the meter which is attached to the coil. The other is an arrow which appears below the coil when the current flows.

Simulation Questions:

1. Run the simulation and slowly drag the magnet back and forth through the coil. What happens to the emf (shown in the graph and in the box below) when the north pole of the magnet moves into the coil? Which way does the needle on the meter point?
2. What happens to the emf when the north end of the magnet moves away (to the right) from the coil? Which way does the needle point in this case?
3. What happens to the emf if the magnet is not moving?
4. When the magnet is close to the coil but not moving, the emf is zero. What about the flux (ϕ)? This is the essence of Faraday's law, the flux has to change for an emf to occur.
5. Choosing the reverse button allows you to drag the coil instead of the magnet. Run the simulation after clicking the reverse button. Does it make a difference in the graphs if you move the magnet or the coil? Explain.
6. Now choose the coil button and run the simulation. As we saw earlier in this chapter, a coil with current flowing in it will have a magnetic field. Is there a difference in the emf graph if you drag a current-carrying coil through the stationary coil instead of a magnet? Explain.
7. Now choose the microphone button and run the simulation. This is a simulation of a moving coil microphone. In this type of microphone a small coil is attached to a flexible diaphragm. When sound waves hit the diaphragm it causes the coil to vibrate near a stationary magnet. The changing flux in the coil causes a current to flow due to Faraday's law. It is this current that is amplified and sent to loudspeakers or recorded on tape or digital recording (see the next chapter). You can change the sound wave frequency. How does this affect the frequency of the emf shown in the graph?

Summary

Ohm's law says voltage acts as a kind of potential energy that will cause charge to flow if there is a path (a circuit). The number of charges in a circuit remain fixed but they carry electrical energy which is turned into other useful forms of energy (light, heat, sound, etc.) by components in the circuit (light bulb, toaster, stereo). Currents cause magnetic fields. If a current finds itself in a magnetic field caused by some other source (magnet or other current) it will feel a force unless it moves parallel to the magnetic field. This is the basis of an electric motor. Faraday's law says a *changing* magnetic field through an area (or equivalently, a changing area with constant field) will cause a voltage. This is the mechanism behind electric generators, credit cards, metal detectors, computer hard drives, etc.

Questions on Faraday's Law:

1. Why is a generator coil harder to turn when it is generating electricity than when it does not?
2. A magnet falling through a narrow copper tube will slow down, even though copper is not magnetic (your instructor may have demonstrated this in class). Explain why this happens.
3. Does your car burn more gas when you run the head lights than if the lights are off?
4. When you swipe a credit card the reader gets information from a strip on the back of the card. Explain how that works.
5. Most traffic lights are connected by a small computer chip to a wire embedded in the road which detects the presence of a car. How does this work?
6. The metal detectors at airport security can detect non-magnetic metals such as aluminum. Explain how that works.
7. Information is contained on your computer hard drive as a series of small magnetic fields (hard drives have iron particles embedded in them so that different regions can be magnetized). The read head consists of a small coil of wire that is located very close to the disk and can be moved around to reach different parts of the disk. Explain how the read head detects the information. Would this work if the disk were not spinning?
8. Explain how a generator works.

9. What is the difference between an electric motor and an electric generator?
10. Why don't transformers work with direct current (DC)?
11. Why is power transmitted at high voltages (and low current) over long distances?
12. What is Faraday's law?
13. A transformer with 10 turns in the primary and 100 turns in the secondary will convert an AC voltage of 5 V to 50 V. Explain why this doesn't contradict conservation of energy.
14. A pickup for an electric guitar consists of a small metal coil of wire wrapped around a magnet. Due to Faraday’s law current is induced in the coil if the magnetic field near the pickup changes. Will this type of pickup work with nylon or other, non-metal strings? Explain.
15. Suppose a metal string is vibrating at 100 Hz in front of the pickup described in the previous question. What frequency will the induced current in the coil have as a result of Faraday’s law?

CHAPTER OVERVIEW

17: Electronics

This chapter explains several electronic devices used in sound recording and reproduction. Concepts needed to explain the function of these devices were introduced in the previous chapter.

There is no way to record, transmit and replay sounds perfectly so that they sound exactly as they were heard originally. All recording processes fail to capture the full range of audio frequencies present when the original sound was made. Microphones have limited ranges of sensitivity as do speakers and amplifying systems. These problems are referred to generally as ***distortion*** and are discussed below. The problems of sound reproduction are further complicated by the fact that all the devices involved in recording, transmitting and reproducing sound add unwanted frequencies called ***noise***. Pure noise can generally be defined as a sound sample which contains some of all frequencies. The amount of each range of frequencies present determines the ***color*** of the noise. Here are Wikipedia pages on different colors of noise: white noise, pink noise, Brownian noise, grey noise.

Key Terms

Noise, speakers (magnetic, electrostatic, tweeter, bass), microphones (dynamic; 3 types: electrostatic, piezoelectric, carbon), vinyl, magnetic tape, analog, analog to digital, digital to analog, binary number system, sample rate, bit rate, digital recording (divots, CD, DVD, Blu-ray), audio compression, lossless, lossy, MPEG3, AM, FM, digital, distortion (amplitude, harmonic, frequency, phase), MIDI.

17.1: Electronics

It is not possible to record, transmit and replay sounds perfectly so that they sound exactly as they were heard originally. This chapter explains several electronic devices used in sound recording and reproduction using concepts that were introduced in previous chapters.

17.1.1: Microphones

The earliest sound recordings were made without the use of electronics at all. A large cone channeled sound to a diaphragm attached to a needle so that the needle would vibrate at the same rate as the sound vibrations entering the cone. To make a recording the needle was placed on a rotating cylinder coated in hot wax or plastic which was slowly moved past the needle as it vibrated. The vibrations of the needle were recorded as a single continuous fluctuating groove in the wax. The wax was hot initially so that it was soft enough to record the vibrations of the needle. Because the process was entirely mechanical it was not very efficient; singers had to practically shout into the cone to produce enough vibrations to form a recording.

Modern recording of music starts by turning sound into electrical impulses. This is an ***analog process***; the electrical current varies in strength and frequency in proportion to the loudness and frequency of the sound wave. In other words, when the sound varies more rapidly (high frequency) so does the electrical signal and when the sound is louder, the electrical signal has a higher amplitude in proportion to the sound. The conversion from sound to electrical signals is accomplished by a microphone and there are six major types based on three different physical processes.

The ***dynamic microphone*** is based on Faraday's law which we learned about in Chapter 17. The first figure below on the left is the ***moving-coil dynamic microphone***. A flexible diaphragm is attached to a coil so that vibrations in the diaphragm cause the coil to move. If the coil is close to a permanent magnet, movement of the coil in the magnetic field will cause a current to flow in the coil due to Faraday's law. Sound waves cause the diaphragm to vibrate so the sound vibrations are converted into a fluctuating current flow. In theory the same arrangement can be either a microphone (vibrations input, oscillating current output) or a speaker (fluctuating current input, vibrations output) as we will see.

The moving-coil microphone became popular in the late 1930s and is designed to pick up sound from one direction. This and the fact that it is a bit more durable than some of the other microphones makes it a favorite choice for live performances where there may be a lot of other stray sound present on stage. Because it can withstand loud sound without much distortion it is popular with blues and hip-hop singers.

The second arrangement is the ***magnetic microphone***. In this microphone a movable cone is attached to a small magnet or piece of iron which can move back and forth between the poles of a permanent magnet. A coil is wrapped around the permanent magnet. Sound causes the cone to vibrate which moves the small magnet. The moving magnet changes the magnetic field in the permanent magnet and therefore in the coil, creating a current that oscillates in proportion to the sound oscillation.

Faraday's law is also the physical principle behind the ***dynamic ribbon microphone*** shown in the third picture below. A thin piece of metallic foil vibrates when sound hits it. The principle is the same as the moving-coil microphone, the vibrating metal sheet is moving in the magnetic field of a permanent magnet and so a current that matches the variations in the vibrating ribbon is created. In the first two types the leads (not shown) are attached to the coil.

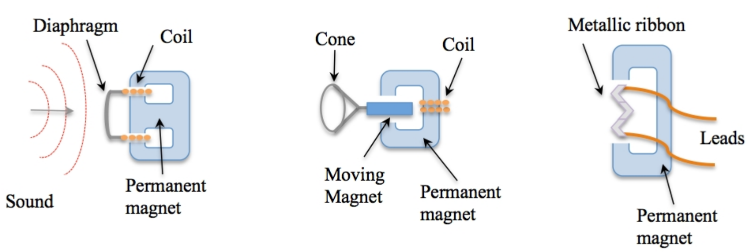


Figure 17.1.1.1

The ribbon microphone was invented in the early 1920s. The range of frequencies that it responds to is limited; it does not capture high frequencies well. However, it has a frequency range well suited for the human voice and was popular in studio recordings of the 1920s. Some microphones (such as the carbon microphone, below) produce crackling sounds (noise) when trying to record higher frequencies. The ribbon microphone produces a smooth response which attenuates at higher frequency so less noise is present. The ribbon microphone also exhibits a proximity effect, boosting lower frequencies a bit more as the source moves closer to the microphone, making the voice sound more intimate. The proximity effect made the crooner styles of singers like Nat King Cole, Bing Crosby, and Rosemary Clooney popular.

One problem of the above three types of microphones is that the phase of the electrical signal lags the phase of the original vibrations by a quarter of a wavelength. A close examination of Faraday's law shows that the voltage in the coil is a maximum when the coil is moving the fastest which occurs as the diaphragm passes its equilibrium point. When the diaphragm is pushed in to its furthest point it stops moving and the voltage produced by Faraday's law drops to zero. Suppose we have a sound wave from a tuning fork which we know is pretty close to a pure sine wave. An oscilloscope would show a sine wave like the red curve in the figure below. But if we could plot the voltage across the microphone we would see the blue curve which is ninety degrees ($\pi/2 \text{ radians}$) out of phase with the sound.

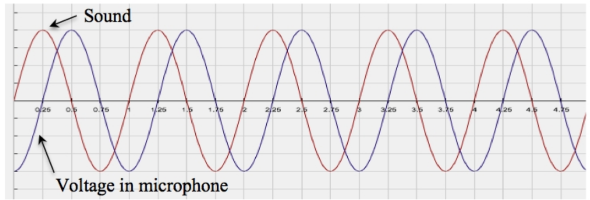


Figure 17.1.1.2

We know from Chapter 16 that one of the four ways the ear-brain system determines direction is the relative phase of the signal between one ear and the other. If the speaker or recording system does not correct for this quarter phase error in the microphone the reproduced sound may not sound right to our ears. A further complication is that other types of microphones (described below) do not have this quarter phase difference. So a recording session that uses multiple types of microphones may result in recorded sound that has a mixture of phase errors.

Three other techniques for making a microphone are described next. These microphones produce voltage changes based on physical mechanisms other than Faraday's law.

Physically squeezing certain types of crystals will cause a voltage to form across the crystal that is proportional to how much force is applied. A crystal that behaves this way is called a ***piezoelectric crystal*** and is the basis for a type of microphone called a ***piezoelectric microphone***. Because a crystal of a given size has a natural frequency of vibration that is very constant, piezoelectric crystals are also used in watches to keep time. The oscillations are driven by a circuit that is tuned to a certain frequency by the crystal oscillations acting as a voltage source for timing purposes. The first figure on the left is a schematic of a piezoelectric microphone.

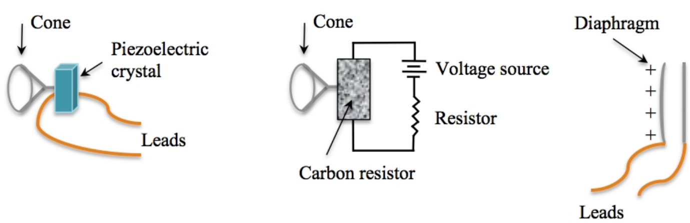


Figure 17.1.1.3

Another type of microphone, called a ***carbon microphone***, is based on the change in resistance of a container of carbon granules (a carbon resistor). As the microphone cone vibrates it squeezes the container so that the carbon granules are closer together, causing a change in the resistance of the container. Ohm's law, from the previous chapter, tells us that the current will be different for a circuit with a fixed voltage if the resistance changes. Fluctuations of the cone cause fluctuations in resistance, turning sound into a fluctuating current flow through the second resistor. The second diagram above is for a carbon microphone.

The carbon microphone was the first type of microphone used in recording music electronically and was invented in the early 1900s. It has a very narrow frequency response and some noise which gives it a distinctive sound, familiar from movies depicting old radio broadcasts or analog telephones. With the advent of the carbon microphone singers no longer had to shout as they did for mechanical recordings directly on wax and a style of music called whispering became popular.

The third figure above shows the ***electrostatic or condenser microphone***. In this microphone a flexible diagram is charged and placed close to a fixed plate with the opposite charge. The arrangement of two oppositely charged surfaces brought very close together but not touching is called a ***capacitor***. The voltage between the two surfaces depends on the amount of charge on one side and distance between them. As the diaphragm vibrates the voltage between the two surfaces oscillates. This oscillating voltage is the input signal to an amplifying circuit.

The condenser microphone was invented in 1917 but was not popularly used for recording until better versions were created in the late 1940s. The condenser microphone is better at higher frequencies than other microphones and has a slightly better response for the range of human vocalization than ribbon or carbon microphones. Frank Sinatra was one vocalist who helped popularize the sound of the condenser microphone.

Video/audio examples:

- A technical YouTube on how to check the phase of a set of microphones.

17.1.2: Recording

Vinyl

As mentioned above, earliest sound recordings were made without the use of electronics at all; grooves made by a vibrating needle were made in a cylinder of hot wax and these recorded the oscillations of sound being funneled through a large cone. Once sound was recorded on the cylinder it was cooled to solidify the wax or plastic. To play the recorded sound, the needle was returned to the beginning of the now cool cylinder and the movement of the cylinder was duplicated. Now, however, the fluctuating groove would cause the needle to vibrate which would cause the diaphragm to vibrate, recreating the sound.

Cylinder recordings were popular from the 1880s until the 1920s when they were gradually replaced by disk shaped ***vinyl records***. Using basically the same idea as the cylinder recording, a rotating disk made of vinyl recorded the fluctuating groove. Electronic amplification of the needle vibrations eventually replaced the mechanical diaphragm and cone system to reproduce the sound. Motion of the needle is detected by a coil and magnet system, again using Faraday's law. A metal copy of the original disk is made and used as a mold to make multiple copies of the same record. An electron microscope picture of the grooves in a vinyl record is shown below. Vinyl records are analog recordings; the grooves in the vinyl vary the same way the original sound did. Faster vibrations of sound produce grooves with more oscillations and louder sounds produce bigger oscillations.

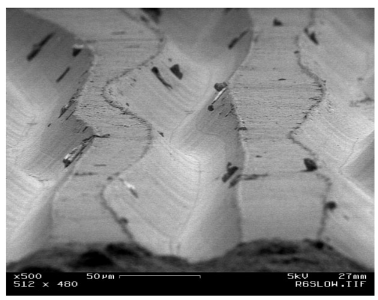


Figure 17.1.2.1

For ***stereo recordings***, each side of the groove records the fluctuations coming from a different microphone. It is also possible to record four different fluctuations from four microphones in a single groove, a process known as quadraphonic sound recording.

Vinyl records have several disadvantages as a sound recording medium. They are somewhat fragile in that they can easily be scratched, broken or melted. Leaving a vinyl record in your car on a sunny day (or even in a sunny spot in front of a window) generally means you will not be able to play it again due to warping of vinyl in the heat. Stacking many records on top of each other or under books will also warp vinyl records. Scratches on the vinyl will be translated into sounds as hiss and pop which interfere with the recorded music. Records must be kept dust free to avoid having the needle skip over dirt in the grooves. Over time the needle will wear away the vinyl, reducing the accuracy of the recording. The needle also deteriorates over time; even the best diamond needles eventually wear out and have to be replaced.

Video/audio examples:

- Wikipedia on vinyl records.
- Web page with more electron microscope pictures of grooves in vinyl.
- Somewhat long YouTube on how to fix a scratch on a vinyl record.

Sound sample of a scratch vinyl record:

Tape

Magnetic tape recording was developed in Germany before the Second World War but was not available commercially until around 1950. A magnetic tape is a long thin piece of plastic, embedded with iron compounds (ferric oxide; Fe_2O_3) in powdered form. Recordings are made on the tape by passing it over a magnetic ***write (or recording) head*** that is receiving fluctuating electrical signals from a microphone. The write head is basically an electromagnet that magnetizes the iron compound on the tape in a pattern identical to fluctuating current in the head. The recorded signal is analog; the magnetic field of the iron in the tape varies with amplitude and frequency just like the sound did. In the diagram below a schematic of a write head is shown with a top view of the tape. The tape is moving from right to left, unwinding from a reel on the right (not shown) and rewinding on a reel on the left (not shown). The left and right stereo channels are recorded side by side.

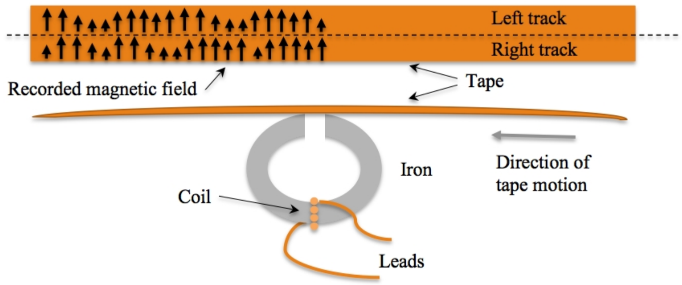


Figure 17.1.2.2

The ***read head*** of a tape player does the reverse of the write head. As the tape with a magnetic signal passes over an iron loop it induces an oscillating magnetic field in the iron. The changing magnetic field in the iron causes current to flow in a coil wrapped around the iron (Faraday's law again). The schematic for this process would look identical to the diagram above (and in fact some tape players have a single read/write head that performs both functions).

Magnetic tape became a recording industry standard because the sound from many microphones could be recorded simultaneously on a wide piece of tape as separate tracks -as many as 16 tracks could be recorded simultaneously. The same technique made it possible to record video on one track and audio on another. Video recorders began to be available in the 1960s. Being able to move the tape at different speeds is also an advantage. Using tape moving past the write head at a faster speed allows higher frequencies to be recorded more accurately. The trade-off is that more tape has to be used for the same amount of recording.

There are several drawbacks with using magnetic tapes as a recording medium. The plastic can stretch, break or melt. The size of the compound metallic grains in the tape means that very rapid fluctuations in magnetic field cannot be recorded. As a result, magnetic tape does not record high frequencies very well except at very high tape speeds which introduce other problems. If the tape is exposed to a magnetic field, the information is changed or lost. Gradually the magnetized iron may lose its field as the magnetic fields of neighboring regions of tape interact with each other. Most magnetic tape is wound onto a reel so that one layer can affect the layers above and below, creating "ghost sounds". As the tape moves past the read head it will pick up randomly oriented magnetic fields of the iron compounds, even on regions of the tape where no sound is recorded. This is heard as ***tape hiss*** and is especially noticeable in quiet sections of music.

Several clever ways to suppress noise on magnetic tapes have been developed. The most common method is to amplify softer parts of the music when they are recorded and then reduce their volume as they are played back. As shown in the diagram below, the loud parts are not amplified when recorded and are played back at normal volume but sounds with lower amplitude are first amplified before recording.

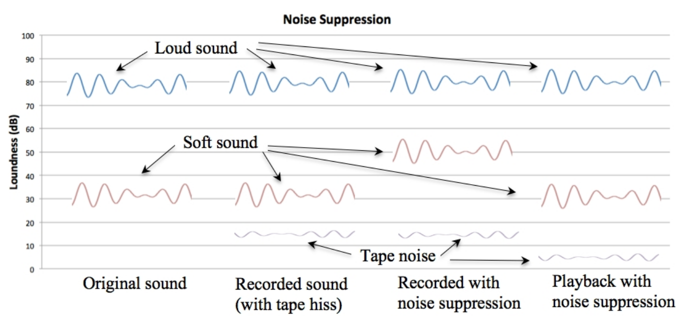


Figure 17.1.2.3

Video/audio examples:

- Wikipedia on reel-to-reel tape.
- Wikipedia on cassette tape.
- Wikipedia on noise reduction.

Sound Sample of tape hiss:

Sound sample of tape distortion:

Sound sample of tape distortion:

Digital

Because vinyl records and most magnetic tapes are used to capture the actual amplitudes and vibrational frequencies of the sounds they are recording they are known as ***analog recordings***. The grooves in the record or the magnetic fields of the iron compounds on the tape have variations proportional in size and frequency to the music they have recorded. An entirely different way to record sound called ***digital recording*** was developed beginning in the late 1950s.

Let's look at a sine wave voltage (the blue curve in the figure below). This could represent the signal coming from a microphone which is picking up the sound of a tuning fork. The signal varies from 1000 millivolts (mV) or 1 volt to $-1000\ \mathrm{mV}$ $(-1$ volt). Instead of recording the actual shape of the curve, suppose we sample the amplitude (in millivolts) of the curve at many different times. So for example, we could record the voltage every 0.1 millisecond (ms). This would give us the red, stair shaped curve in the figure below. For the first $0.1\ \mathrm{ms}$ the recorded voltage is zero, at $0.1\ \mathrm{ms}$ the voltage is $250\ \mathrm{mV}$, at $0.2\ \mathrm{ms}$ the voltage is $500\ \mathrm{mV}$, at $0.3\ \mathrm{ms}$ the voltage is $750\ \mathrm{ms}$, and so on. This list of numbers $(0,\ 250,\ 500,\ 750$ etc.) with the times they were taken ($0.1\ \mathrm{ms},\ 0.2\ \mathrm{ms},\ 0.3\ \mathrm{ms}$, etc.) would be a rough representation of the original curve in numerical form.

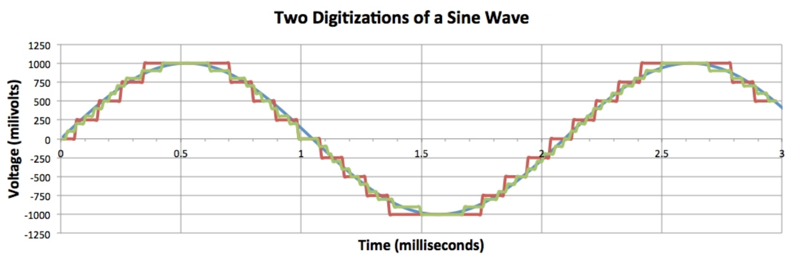


Figure 17.1.2.4

How can we get a set of numbers that is closer to the original sine wave? Suppose instead of sampling every $0.1\ \mathrm{ms}$ we sample twice as often or every $0.05\ \mathrm{ms}$? This is the green curve in the figure above. So at $0\ \mathrm{ms}$ we still have $0\ \mathrm{mV}$ but at $0.05\ \mathrm{ms}$ we get $100\ \mathrm{mV}$, at $0.1\ \mathrm{ms}$ we record $200\ \mathrm{mV}$, at $0.15\ \mathrm{ms}$ we record $300\ \mathrm{mV}$, etc. Now we have more numbers and the jumps are smaller ($100\ \mathrm{mV}$ increases instead of $250\ \mathrm{mV}$ increases). What if we want to get even closer to the original curve? In fact we can make as accurate a representation as we want just by taking more points at a shorter ***sample rate***. This is the first step in the process called ***analog to digital conversion***; we convert an analog signal to a series of numbers.

There are a couple of other details to the process of recording in a digital format. Computers can only work with ***binary numbers***; in other words, numbers that are either one or zero. This is because the electronic states inside a computer chip are either on or off. This isn't really a problem because there is a binary number for every ordinary number. Below is a table of binary numbers from one to 15.

Number	Binary Equivalent
0	000000
1	000001
2	000010
3	000011
4	000100
5	000101
6	000110
7	000111

Number	Binary Equivalent
8	001000
9	001001
10	001010
11	001011
12	001100
13	001101
14	001110
15	001111

Table 17.1.2.1

Another limitation is the number voltage steps available for dividing the amplitude of the signal. Sampling more often doesn't do any good if the voltage steps cannot be made small enough. In early analog to digital converters the voltage step size was limited by the number of ones and zeros (bits) that could fit into a memory slot and this was called the bit depth. A larger number of bits (larger bit depth) means you can divide the voltage of any given sample into smaller steps and thus have a more accurate picture of the sound wave. Most voltages are now divided into the number of steps represented by the largest number that can be stored using 16 bits (in other words the largest binary number with 16 digits which turns out to be the number 65535). The ***bit rate***, usually measured in kbps (thousand bits per second) is the number of bits per sample (the bit depth) times the sample rate. For sound sampled at 44.1 kHz with a 16 bit A to D converter produces a bit rate of $44.1 \text{ kHz} \times 16 \text{ bits} = 705.6 \text{ kbps}$ (for two channel stereo the bit rate would be twice this).

Once we have a string of binary numbers recorded, how do we get the sound back? To play back the digital recorded wave we feed the list of numbers to a device that produces a voltage equal to the number it reads. The changing voltages are amplified and fed to a speaker to reproduce the sound. This is called ***digital to analog conversion***. Notice that this means the reproduced sine wave will not be exactly the same as the original. Instead it will now be one of the stair step waves shown above. However, if the reproduced wave is close enough to the original our ear-brain system is fooled.

Computer disks, both the outdated floppy disk and current hard drive technology record information using the same method as magnetic tape. A plastic medium is embedded with iron compounds which can me magnetized as they pass underneath a coil. The information is read (Faraday's law again) by a coil held very near the surface of the disk (a computer crash originally meant literally that either the read head or the write head hit the surface of the disk). Instead of recording analog information (variations that are proportional to the sound variations), the data is stored as either on (a magnetic field) or off (a reversed magnetic field). In other words, the data is stored as binary information.

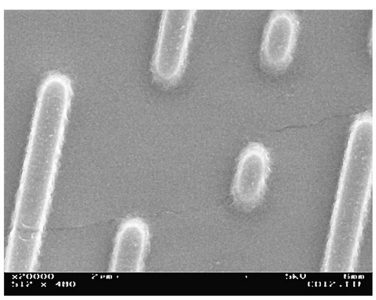


Figure 17.1.2.5

A ***Compact Disk*** or CD records the digital information as a series of divots called pits (shown above in an electron microscope picture) that are burned into the surface of a plastic disk. The flat regions between pits are called lands. A short pit might represent the binary number zero and a longer pit the number one. Three (or more) laser beams, slightly offset from each other are used to record and read the data. In the reading stage the center beam reflects off the disk into a photo detector as the disk turns below it. The reflection is detected as a beam that is alternately broken for a short period of time (a short pit) or a slightly longer period of time (a long pit). Two beams to either side of the read beam keep the center beam aligned on the row of divots as shown in the diagram below.

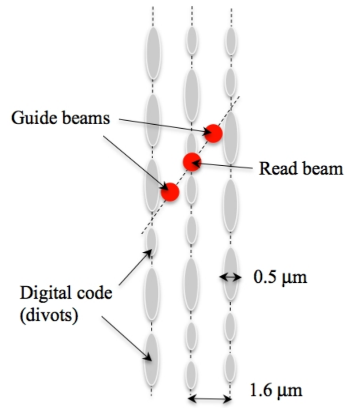


Figure 17.1.2.6

CD technology needed the development of the laser in order to work. The first solid state lasers were infrared, followed by red lasers. Lasers in other colors took longer to develop because of technical difficulties. Blu-ray disk technology, which uses a blue laser, wasn't available until the early 2000s after the development of blue lasers. The reason these disks hold more information is the pits are smaller and closer together. Recall from Chapter 7 that waves interact with objects close to the size of their wavelength; laser light doesn't diffract through a doorway because the opening is much larger than the wavelength. The wavelength of red light is too long to read the tiny pits in a Blu-ray disk but the pits can be read by the shorter wavelengths of blue laser light. Regular CDs use light with a wavelength of $780\ \mathrm{nm}$, DVDs use wavelengths of $650\ \mathrm{nm}$ and Blu-ray uses $405\ \mathrm{nm}$ light.

One obvious problem with digital recording is the trade-off between ***sample rate*** and ***bit rate***. Suppose the sine wave in our example above is oscillating at $60\ \mathrm{Hz}$ (60 oscillations per second). If the sample rate is $60\ \mathrm{Hz}$ (60 samples per second) each sample will catch the same point on the sine wave so the list of numbers will be constant and the signal is not recorded. In general you have to sample a sine wave at least twice a cycle in order to record the variation (in the sine wave above this would be every $0.5\ \mathrm{ms}$ which would record a peak followed by a trough followed by a peak, etc.). And even then the playback voltages would constitute a triangle wave of the same frequency as the original sine wave rather than a sine curve. The minimum sample rate needed to record a given frequency is called the ***Nyquist rate***. The highest frequency that can be recorded is one half the Nyquist rate and is called the ***Nyquist frequency***.

Humans with perfect hearing can hear up to $20,000\ \mathrm{Hz}$ so a sample rate of $40,000\ \mathrm{Hz}$ should be sufficient for most recorded music. The recording industry settled on a sample rate $44.1\ \mathrm{kHz}$ ($44,1000\ \mathrm{Hz}$) as the industry standard for CD recordings. However the recording rate used in music studios is usually $48\ \mathrm{kHz}$ or higher. Higher sample rates are also used for non-audio signals, for example, DVD and Blu-ray audio sample rates are sometimes $96\ \mathrm{kHz}$ or $192\ \mathrm{kHz}$.

Since most people cannot hear frequencies above $15,000\ \mathrm{Hz}$ very well, audio sampled at lower sampling rates often does not sound very different. Likewise dividing the sample into 65535 voltage steps isn't always necessary to capture changes in the signal, especially if the signal does not change quickly. So either the sample rate or the bit depth can be lowered without degrading the quality of sound enough to notice for most people. Most software for recording (ripping) a CD to put music onto a MP3 player (for example iTunes) lets the user choose the bit rate so that the size of the files can be adjusted to allow more music to be put onto the storage device. In most cases the sample rate stays fixed but the number of bits used to determine the voltage step size is modified. iTunes, for example, allows the user to select bit rates of $320\ \mathrm{kbps}$ down to $64\ \mathrm{kbps}$. A reduction from $320\ \mathrm{kbps}$ to $64\ \mathrm{kbps}$ will reduce the file size of a typical song recording to one third its initial size since not as many voltage steps are being recorded per sample. For a lot of music the lower sound quality of a lower bit rate is not noticeable.

A second way to record digital music using less computer memory or CD space is by using ***compression*** software. Although some of software details used commercially are not made public, the general idea behind MP3 (MPEG-3) for audio, JPEG for pictures and movies as well as other compression algorithms is fairly simple. Suppose you digitize an analog signal into a stream of (binary) numbers. As you look at the stream you notice there just happens to be a sequence of ten number 2s in a row. You could simply record the ten numbers onto the CD or computer drive and be done. Or you could record 10×2 to indicate a repeat of the number 2, ten times. This latter way takes up less space because you only have to record two numbers instead of 10. When the recording is decoded the software produces the stream of 10 number 2s when it reads the code 10×2 so that the correct voltage is played in the speaker. Other strategies for compression include eliminating sounds that are not likely to be audible to a human ear and using pattern recognition to predict the frequencies that will occur rather than accurately recording all of the patterns in the sound sample. Compression is ***lossless*** if all the original data is recorded. In a ***lossy*** compression some data that is assumed not to affect sound quality is discarded.

There is one special type of digital signal that is used internally in electronic instruments such as keyboards, drum machines, music sequencers and computers connected to these devices. MIDI stands for Musical Instrument Digital Interface and is an industry standard for communicating between electronic music devices. A MIDI recording is a bit like recording the sheet music to a piece of music instead of the actual sound. When a key is pressed on an electronic keyboard, information is collected about how long the key is pressed, how hard and possibly other physical information about the movement of the key. This information is digital in form (binary) and can be recorded by a computer, manipulated by a computer program, or sent to an output device that converts the digital signal into an analog signal that can be amplified and sent to a speaker or headphone. Because the output is computer controlled, the key sequence can be used to control any sound, for example flute sounds or trumpet sounds, etc. MIDI files are usually much smaller than audio files which can be an advantage. A disadvantage is the full range of analog musical frequencies cannot be recorded this way.

Video/audio examples:

- A comparison of CD, DVD, HD DVD and Blu-ray.
- Sound samples of a song from a CD, recorded at different sample rates. Original file was AIFF format, 52.5 MB. All recordings were made at fixed speed. The first number is the bit rate in kilobites per second, the second number is the sample rate, the third number is the file size.
- The Wikipedia history of the choice of 44.1 kHz.
- A more detailed explanation of MIDI.
- A MIDI file. You can download this, open it with Audacity and compare the file with a regular mp3. It will also open with GarageBand and other software.

17.1.3: Amplification

In the diagram on the left, below, we can imagine a small flow of water (labeled Signal) controlling a larger flow of water. This is the general idea of ***electronic amplification***; a large current flow is controlled by a small current or voltage signal. Originally this was done using a ***vacuum tube***. As shown in the second figure below, the tube has a positive end (***anode***) and a negative end (***cathode***) and all air has been removed. A large voltage is applied from anode to cathode so that electrons will stream from the cathode to the anode (just to be clear, recall that conventional current, I in Amperes that we talked about in the last chapter is labeled to flow in the direction opposite to the electron flow). A wire grill called the ***grid*** is placed in between the anode and cathode. If the grid is neutral, the electron flow from cathode to anode passes through the grid and a steady, direct current is produced. If a small signal is applied to the grid it affects the much larger current flow between anode and cathode. So for example, a low voltage sine wave applied to the grid becomes a large current sine wave flowing through the tube. The ratio of the output signal to the input signal is called the ***gain***. This can be measured as voltage gain (ratio of voltage out to voltage in) or current gain (ratio of current out to current in).

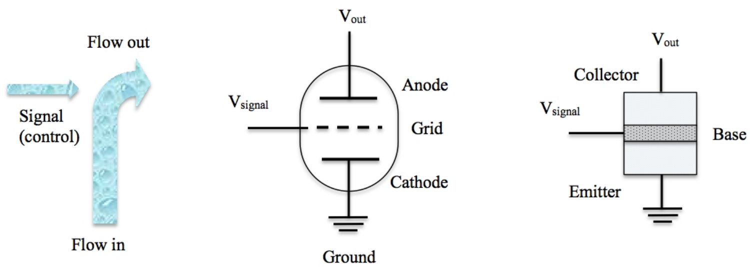


Figure 17.1.3.1

Although they take much more energy to run than modern solid state amplifying circuits, tube amplifiers are still used by the military for radar and high power radio transmission. Some audiophiles claim that tube amplifiers produce a better sound than solid state amplifiers. A disadvantage of tube amplifiers is the tubes take up a lot of space, are heavy, use much more energy than transistors and will stop working if the vacuum seal breaks.

Transistors were developed in the 1940s and gradually replaced most vacuum tubes in electronic applications. Transistors work basically the same way as a vacuum tube. Two crystals (made of silicon or germanium) with slightly different electrical properties are put together separated by a third crystal as shown in the figure on the right, above. The ***emitter*** end of the transistor is made so that it has electrons ready to flow and the ***collector*** end will accept these electrons. When a voltage is applied across the transistor a current flows through the transistor and electrons flow from emitter to collector. The function of the grid in a tube is performed by a crystal called the ***base*** which has slightly different electrical properties from the emitter or collector. When a small signal is applied across the base, the larger current flowing from emitter to collector is controlled to have the same variations as the signal. The terms emitter, collector and base are replaced by source, drain and gate for some other types of transistors using different crystals with different electrical properties.

Transistor amplifiers take much less energy to run than tube amplifiers and generally can take much more physical abuse (shock, temperature changes, etc.) and still work. In theory a transistor should last forever because it is a solid crystal although some transistors can be destroyed by electrostatic sparks. They can also be made incredibly small (today's computer chips contain billions (10^9) of transistors), making cell phones, mp3 players and other small electronic devices possible.

In addition to being used as amplifiers, transistor (and tubes) can also be used as ***electronic switches***. In this application, the signal to the base either allows current to flow from emitter to collector or turns the current completely off. Transistors that function this way can be used to represent ones (all the way on) or zeros (all the way off) in a digital computer.

All amplification suffers from ***distortion***. A perfect recording and playback process would exactly reproduce the original sound but this is never possible; there are always some unwanted modifications to the sound. Microphones, recording devices, speakers and amplification systems all produce some distortion of the signal. The difference between noise and distortion is that noise is extra, unwanted signals whereas distortion is the unwanted modification of the signal.

One form of distortion is when the voltage output is already at the maximum possible for the circuit and the signal amplitude increases. The result is the peaks of the output signal are cut off as shown in the diagram below. The original signal is the blue curve but the amplifier has a maximum output of $0.6 \mathrm{~V}$ so it clips off any part of the signal above $0.6 \mathrm{~V}$ (the red curve). The amplified signal will not sound right because the volume doesn't increase the way it was suppose to in the original recording. This is called ***amplitude distortion***.

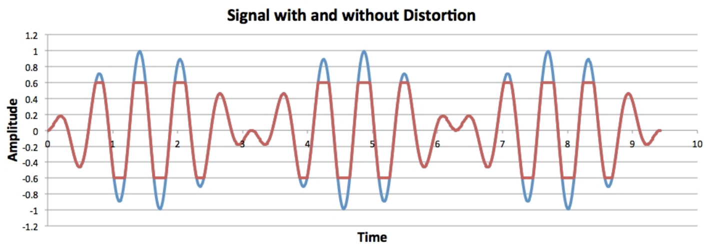


Figure $17.1.3.2$

When there is amplitude distortion the effect of flattening the top of the output wave when it reaches the speaker is to introduce unwanted overtones. We know from Fourier analysis (Chapter 9) that the sudden change from a smooth curve to a flat top requires many extra harmonics to describe the new wave. These new harmonics change the sound produced by the amplifier/speaker system. This type of distortion is called ***harmonic distortion***. Harmonic distortion also occurs if different frequency ranges are amplified by different amounts (in any part of the recording to playback sequence), even if the amplitude is not flattened.

Most amplifiers (and microphones and speakers!) do not treat all frequencies the same. For example, the amplifier may be able to amplify high frequencies better than low frequencies. Or it may not be able to react fast enough to amplify the high frequencies accurately. This is called ***frequency distortion***.

We have already mentioned that microphones do not produce a current with the same phase of the original sound. Amplifiers also can have this problem depending on how they are constructed. If the phase of the output wave is not the same as the input source there is ***phase distortion*** in the output.

Sometimes electronics and software are used to intentionally modify a signal to make it sound different, as mentioned in Chapter 16. One example is adding in reverberation to make the recording sound more natural, as if it were recorded in a concert hall instead of a studio. The following two videos explain in detail one way reverb can be added to a recorded signal to make it sound like it was recorded somewhere other than a studio: Altiverb web site and The 3-D Audio and Applied Acoustics (3D3A) Laboratory at Princeton University web site. Commercial software is also available, for example Room EQ Wizard. Similar techniques can be used to improve the sound of an auditorium if electronic amplification is being used; signals from the microphones can be delayed or modified to provide a better interaction with the room acoustics.

17.1.4: Speakers

A loudspeaker takes an electronic signal and turns it back into sound. How does a speaker reproduce sound from an electrical signal? We know from the previous chapter that the current flow in a coil of wire will produce a magnetic field. This electromagnet will attract or repel other magnets as if it were an ordinary magnet. One difference, however, is that we can easily reverse the polarity of the electromagnet by changing the direction of current flow. An oscillating current in the coil will result in an oscillating force if the coil is near a permanent magnet. This is the basis of a ***dynamic speaker***. A coil surrounded by a permanent magnet is attached to a cone made of cardboard, plastic or other flexible material. When an oscillating electrical signal flows through the coil there is an oscillating force on the coil and cone. This makes the cone vibrate back and forth, producing sound waves in the air next to the cone.

Dynamic speakers have two basic designs depending on the range of sound frequencies being reproduced. ***Bass*** speakers or woofers have large, flat cones attached to the coil and are designed to produce low frequencies. A sketch and cutaway picture are shown below. The voice coil is attached to the cone or diaphragm. Electrical signals enter from the leads and go through thin wires which are glued to the spider, a lightweight flexible support system for the coil. Oscillating current in the voice coil causes an alternating magnetic force between the coil and the permanent magnet. This alternating force on the coil is transmitted to the cone which causes air to vibrate, creating sound.

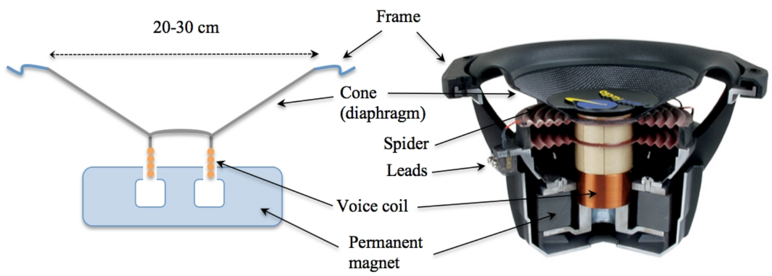


Figure 17.1.4.1

Tweeters are designed similarly to bass speakers but have a small (2 cm to 5 cm), dome shaped vibrating part in place of the cone. Because they are smaller and lighter they are better able to produce higher frequencies. In both types of speakers the permanent magnet does not move. This is because magnets are heavy; if the permanent magnet was attached to the cone instead of the coil, the extra inertia would keep the cone from responding to fast variations in the signal.

It is also possible to make an ***electrostatic speaker*** that uses electrical forces, rather than magnetic forces. For this type of speaker, a thin sheet of flexible plastic (a diaphragm) with a metallic coating is suspended between two thin metal screens or grids. There is a small air gap between the grids and the plastic sheet. The sheet is charged by an external source and the oscillating signal is sent to the grids. The oscillating voltage between the grids causes the diaphragm to vibrate, producing sound. This is exactly the opposite process that occurs in an electrostatic microphone, described previously. An advantage of the electrostatic speaker is that it is very light and so responds very well to high frequencies. Another advantage is that, unlike a magnetic speaker where the sound comes from a small area, the area of the diaphragm can be quite large. The large area constitutes a less localized sound source making the sound seem more natural.

The idea of a speaker system is to try to trick the ear-brain system into perceiving the sound as if it were the original performance. Because we have an ear-brain system that incorporates sound from two ears, most sound recording and reproduction is in ***stereo*** which means two separate sound tracks are used. Sound is recorded from two or more microphones and mixed into two tracks which have slightly different sound. When the final recording is played back these tracks are fed to two separate speakers with the listener sitting somewhere between them. One potential problem with a two speaker system is phase. If the speakers are wired incorrectly, the left speaker can be out of phase with the right speaker causing destructive interference between the two sound sources. Some stereo systems have jacks for the speaker cables so that it isn't possible to connect them incorrectly. But if you are using free wires to connect your speakers you should connect them one way, listen to the sound, then reverse the leads to one of the speaker cabinets and listen again. The correctly wired case will sound louder, particularly for lower notes.

It is difficult to make a single speaker that can reproduce all frequencies equally well. To overcome this limitation bass and tweeters are often combined. The most common design is to have a pair of speaker boxes with one bass and one tweeter in each box. This requires a ***crossover*** circuit that separates the incoming signal into high frequencies (which go to the tweeter) and low frequencies (which go to the bass). Other designs have a third speaker to handle mid-range frequencies and a three-way crossover. Because low frequency sound has very large wavelengths it is difficult for the ear-brain system to accurately identify where the source is located. Modern speaker arrangements exploit this fact by having a single bass speaker (called a woofer or subwoofer) but a set of two mid-range tweeters to maintain the stereo effect. Modern speaker systems also often add two more speakers that are placed behind the listener to provide artificial reverberation. As we discussed in Chapter 17, our experience of sound direction depends to some degree on reverberation so that adding speakers behind the listener with a slight delay adds depth to the perception of sound from a stereo system.

Free-standing speakers do not sound very good, particularly for lower frequencies. When the speaker cone moves forward to produce a compression, the back side of the cone produces a rarefaction (a low pressure region). As the speaker vibrates compressions and rarefactions on the front side are exactly out of phase with the rarefactions and compressions coming from the back of the speaker. When sound from the back combines with sound from the front, the waves are out of phase and tend to cancel due to destructive interference as seen in the first figure on the left, below. This is very similar to the problem of the drum body that we saw in the previous chapter. The effect is more noticeable for lower frequencies which have longer wavelengths. This is especially a problem for the electrostatic speaker which typically does not have an enclosure and so does not produce low frequencies well.

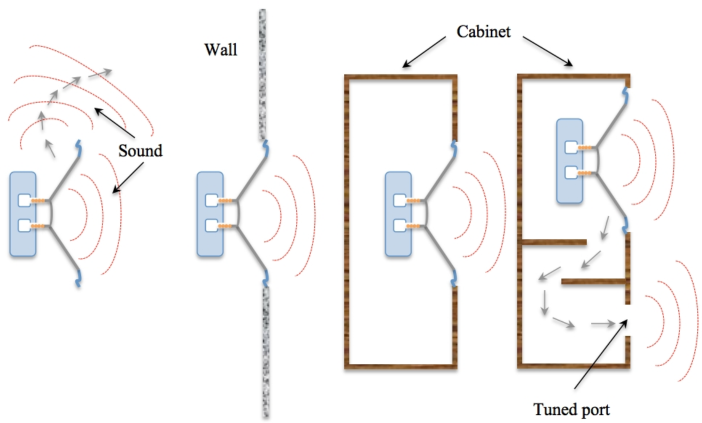


Figure 17.1.4.2

There are several potential solutions to the problem of destructive interference between the back and front of the speaker. If the speaker is mounted in a wall or ceiling the sound from the back is trapped and cannot cancel sound from the front. This arrangement is called an ***infinite baffle*** and is shown in the second figure above. A second solution is to put the speaker in a box or ***speaker cabinet*** which has sound absorbing material inside so that the wave from the back is again trapped. This is shown in the third figure above. A third solution is to cause the sound from the back of the speaker to travel an extra half wavelength before it recombines with waves from the front. This is done by making a hole in the front of the speaker cabinet (sometimes called a ***bass reflex port*** or bass port or tuning port) and putting barriers inside the cabinet so that the sound has to travel an extra distance in order to exit as shown in the figure on the right, above. Obviously this extra distance can't be exactly half a wavelength for all the different wavelengths being produced but it can be arranged to emphasize a frequency range, typically lower notes where the destructive interference of a free standing speaker is more noticeable.

Another way of explaining how a bass port works is to think of the bass port as an opening to an enclosed volume of air. As we saw in Chapters 4 and 16, the opening will make a Helmholtz resonance frequency possible which depends on the size of the speaker cabinet. Choosing the right sized cabinet will reinforce certain low frequency notes in the music, giving a better bass response.

Video/audio examples:

- Here is a simulation of a vibrating speaker from Wolfram (you may need to download their plug in to play with this demonstration).
- A YouTube explanation of signal clipping (amplitude distortion) and its effect on a speaker.
- Here is a YouTube video showing how a loudspeaker is made.
- Here is a YouTube of how a electrostatic speaker is made.

17.1.5: Transmission

Electronic signals may be transmitted as current flows in wires, pulses of light in fiber optic cables and by electromagnetic waves. There are three basic formats in which to send the information; amplitude modulation (AM), frequency modulation (FM), and phase modulation (digital).

As we saw earlier, a radio wave is emitted from an antenna when charge in the antenna is up and down as shown in the simulation of radio waves in Chapter 5. We can change the frequency of the emitted wave by changing the rate at which we move the charge in the antenna back and forth. The amplitude of the wave can be changed how far the charge moves up and down. Suppose you want to send a sine wave signal to the receiving antenna. The easiest way would be to simply move the charge in the antenna with the same frequency and amplitude as the sine wave you want to send. So far so good, the receiving antenna and amplification system can be tuned to that frequency. The problem comes when you want to send a different frequency. Then the receiving system has to tune to a new frequency. We know from Fourier analysis that sound (and therefor the electrical signal carrying it) can be thought of as a combination of many sine waves, each with a different amplitude and frequency. This causes a difficulty for the receiving antenna system because it can only be tuned to one frequency at a time.

In ***amplitude modulation*** (AM) a carrier wave with a constant frequency has its amplitude changed to encode the signal. The changing envelope of the carrier wave represents the signal being sent. The receiving antenna is tuned to the carrier frequency but the electronics on the receiving end decodes the amplitude as the signal and ignores the carrier frequency.

In ***frequency modulation*** (FM) the carrier's frequency is adjusted slightly up or down to indicate the signal being sent. The electronics on the receiving end interpret the slight variation in frequency as changes in the signal being transmitted. A simulation of both is shown below.

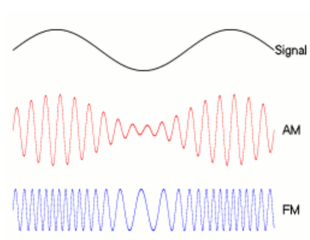


Figure 17.1.5.1

There are advantages and disadvantages to both AM and FM. The AM carrier signal uses longer wavelengths and these waves will reflect off a charged region in the upper atmosphere called the ionosphere. FM, which has shorter carrier wavelengths, does not. As a result, an FM station has to be in direct line of sight in order for the waves to be received. Cell phone and TV signals are also shorter wavelengths than AM and have the same problem; you have to be within sight of a cell tower or a TV tower to receive a signal. AM signals, on the other hand, because of their longer wavelengths (lower frequencies), will reflect off the ionosphere and can be received at long distances from their source, as seen in the diagram below. Shortwave radio and ham (amateur) radio use wavelengths only slightly shorter (wavelengths between AM and FM) than broadcast AM and can be transmitted around the world. Because the ionosphere is higher at night due to temperature changes and the effect of the sun on the layer, the range of AM, ham and shortwave reception improves at night.

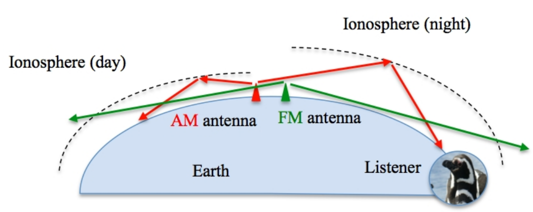


Figure 17.1.5.2

An advantage of FM over AM is that FM can transmit a higher quality signal. If a signal varies too rapidly the signal amplitude might vary faster than the carrier frequency for AM. This does not happen with FM because the carrier wave has a higher frequency. The fact that the amplitude remains constant for FM also means the strength of the signal does not vary whereas a weak AM source might be lost at points where the signal amplitude is low.

A final difference in the quality of AM versus FM signals has to do with the effect of static from other sources such as lightning, car spark plugs and other random noise. Random electromagnetic noise will change the amplitude of both AM and FM carrier signals. But because the information resides in changes to the frequency of the carrier for FM, the quality of the signal is not as affected by changes in amplitude due to static noise.

Digital signals such as WiFi and Bluetooth involve sending a signal where the phase of the carrier wave is shifted relative to a reference wave. A code or key is set up in advance so that a particular shift in phase indicates a predetermined binary number. The phase shift is measured using a device called a lock-in amplifier that compares the arriving signal to a local reference frequency. Some digital signals involve sending several phase shifted frequencies at the same time at different carrier frequencies and there are several different ways to set up the key that matches the phase to a binary number. Digital signals typically use carrier waves of much higher frequencies than FM or AM so that information can be transmitted at a higher rate. As a result of the higher frequencies, however they suffer some of the same limitations of FM; the antennas must be in direct line of sight in order to communicate. Digital WiFi signals are also sent at much lower power than radio transmission and so are of short range, typically $30\ \mathrm{m}$ indoors and up to $100\ \mathrm{m}$ outdoors.

Summary

Recording generally starts with sound waves that are analog in nature. A microphone converts the vibrations in the air into an analog electrical signal; the amplitude and frequency of the electrical signal matches the sound amplitude and frequency. Vinyl and tape are analog recordings of the sound; the amplitude and frequency of the groove vibrations or the amplitude and frequency of the magnetic fields in the tape vary the same way as those of the sound source. For digital recordings the analog signal from the microphone is converted to a binary digital signal which is recorded as ones and zeros on a CD, DVD or disk drive. Electronic instruments create digital signals directly (MIDI) which can be recorded or manipulated with a computer. Digital signals are converted into analog signals, amplified and then sent to a speaker to convert them from electrical analog signals to sound vibrations. Here is a schematic of electronic recording and transmission.

Index

D

Index

D

Glossary

Sample Word 1 | Sample Definition 1

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Book: Sound - An Interactive eBook (Forinash and Christian)

Webpages: 187

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 3.0: 96.3% (180 pages)
- Undeclared: 3.7% (7 pages)

By Page

- Book: Sound - An Interactive eBook (Forinash and Christian) - *CC BY-NC-SA 3.0*
  - Front Matter - *CC BY-NC-SA 3.0*
    - TitlePage - *Undeclared*
    - InfoPage - *CC BY-NC-SA 3.0*
    - Introduction to Sound - *CC BY-NC-SA 3.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Basics - *CC BY-NC-SA 3.0*
    - 1.1: Basics- Work and Energy - *CC BY-NC-SA 3.0*
      - 1.1.1: Motion and Force - *CC BY-NC-SA 3.0*
      - 1.1.2: Newton's Three Laws - *CC BY-NC-SA 3.0*
      - 1.1.3: Density and Pressure - *CC BY-NC-SA 3.0*
      - 1.1.4: Energy and Power - *CC BY-NC-SA 3.0*
      - 1.1.5: Two Energy Laws - *CC BY-NC-SA 3.0*
      - 1.1.6: Conservation of Energy Simulation - *CC BY-NC-SA 3.0*
    - 1.2: Basics- The Molecular Basis of Matter - *CC BY-NC-SA 3.0*
      - 1.2.1: Molecular Dynamics Simulation - *CC BY-NC-SA 3.0*
  - 2: Vibrations - *CC BY-NC-SA 3.0*
    - 2.1: Vibrations - *CC BY-NC-SA 3.0*
      - 2.1.1: Simple Harmonic Motion - *CC BY-NC-SA 3.0*
      - 2.1.2: Period, Frequency, Amplitude, Restoring Force, Phase - *CC BY-NC-SA 3.0*
      - 2.1.3: Simple Harmonic Motion Simulation - *CC BY-NC-SA 3.0*
  - 3: Resonance - *CC BY-NC-SA 3.0*
    - 3.1: Resonance - *CC BY-NC-SA 3.0*
      - 3.1.1: Resonance Examples - *CC BY-NC-SA 3.0*
      - 3.1.2: A Few Other Examples of Resonance - *CC BY-NC-SA 3.0*
      - 3.1.3: Harmonic Motion and Resonance Simulation - *CC BY-NC-SA 3.0*
    - 3.2: Resonance Springs - *CC BY-NC-SA 3.0*
      - 3.2.1: Driven Springs Simulation - *CC BY-NC-SA 3.0*
    - 3.3: Quality Factor - *CC BY-NC-SA 3.0*
      - 3.3.1: Quality - *CC BY-NC-SA 3.0*
      - 3.3.2: Q-Factor Simulation - *CC BY-NC-SA 3.0*
  - 4: Wave Types - *CC BY-NC-SA 3.0*
    - 4.1: Transverse Waves - *CC BY-NC-SA 3.0*
      - 4.1.1: Transverse Waves - *CC BY-NC-SA 3.0*
      - 4.1.2: Transverse Wave Simulation - *CC BY-NC-SA 3.0*
    - 4.2: Longitudinal Waves - *CC BY-NC-SA 3.0*
      - 4.2.1: Longitudinal Waves - *CC BY-NC-SA 3.0*
      - 4.2.2: Longitudinal Wave Simulation - *CC BY-NC-SA 3.0*
    - 4.3: Other Waves - *CC BY-NC-SA 3.0*
      - 4.3.1: Torsional Waves - *CC BY-NC-SA 3.0*
      - 4.3.2: Examples of Waves - *CC BY-NC-SA 3.0*
      - 4.3.3: Water Wave Simulation - *CC BY-NC-SA 3.0*
    - 4.4: Electromagnetic Waves - *CC BY-NC-SA 3.0*
      - 4.4.1: Electromagnetic Waves - *CC BY-NC-SA 3.0*
      - 4.4.2: Antenna Simulation - *CC BY-NC-SA 3.0*
  - 5: Wave Speed - *CC BY-NC-SA 3.0*
    - 5.1: Wave Speed - *CC BY-NC-SA 3.0*
      - 5.1.1: Speeds of Different Types of Waves - *CC BY-NC-SA 3.0*
      - 5.1.2: Speed of a Wave Simulation - *CC BY-NC-SA 3.0*
  - 6: Wave Behavior - *CC BY-NC-SA 3.0*
    - 6.1: Doppler Shift - *CC BY-NC-SA 3.0*
      - 6.1.1: Doppler Shift - *CC BY-NC-SA 3.0*
      - 6.1.2: The Doppler Effect Simulation - *CC BY-NC-SA 3.0*
    - 6.2: Mirrors and Reflection - *CC BY-NC-SA 3.0*
      - 6.2.1: Reflection - *CC BY-NC-SA 3.0*