PHYSICAL ACOUSTICS

OF

MUSIC PERCEPTION

R.B. Moore

McGill University

CONTENTS

CHAPTER 1
INTRODUCTION

These notes are intended to help music students understand physical acoustics. The material assembled here was done so at the request of the Faculty of Music of McGill University for students intending to enter a graduate program in music recording. While this means that the subject matter is approached from a definite point of view, it is hoped that it does not mean with a limited perspective. An attempt is made to emphasize the broad fundamentals of physical acoustics, particularly in their importance to understanding the acoustic environment of a human listener.

Why is an understanding of the fundamentals of physical acoustics so important to music recording? A simple analysis would indicate that music recording is straight-forward. The perception of live sound involves a source, a medium to propagate the sound from the source to the listener, and the receiver (the human ear-brain system).

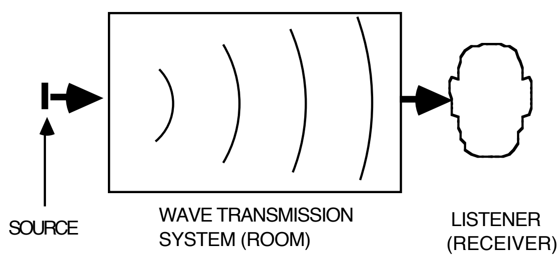


Fig. 1.1 System analysis of a live sound experience.

Recording the sound for future playback could be done by replacing the human listener with a dummy head of the same physical consistency as a real head and in which the human ears are replaced by sensitive microphones, the electric output of which are led to an electronic recorder.

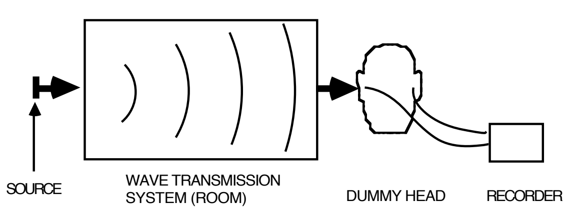


Fig. 1.2 System analysis of sound recording.

The original acoustic experience can then be duplicated by a playback system which recreates the original sound pressure patterns using headphones attached to a real persons head.

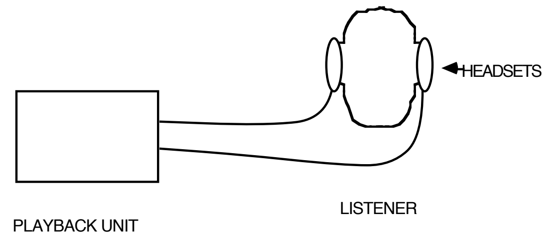


Fig. 1.3 System analysis of listening to recorded music.

By careful engineering of the dummy head, the microphones, the recorder and the playback system it would seen that we should be able to exactly recreate the sound pressure sequences in the original music, at least to within the human ability to detect any difference.

From the systems point of view the process can be represented by a simple diagram.

Fig. 1.4 "Black Box" representation of live musical experience.

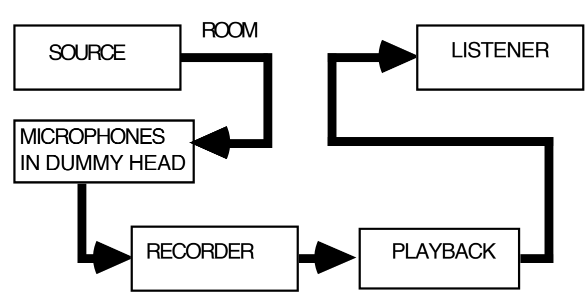


Fig. 1.4 "Black Box" representation of live musical experience.

A lot of engineering effort has indeed been directed at constructing the necessary devices and very faithful reproduction of sound signals is now possible. Therefore it would seem that there is no need for the typical recording engineer activity of placing a myriad of microphones and a tangle of wires in a symphony hall and of tampering with the various recorded signals with complex electronics until a satisfactory result is obtained.

In my opinion, there are similarities between music recording and the recording of a visual experience, such as by movie or video cameras. We accept very easily that the human eye cannot be substituted by a television camera. There is much more to a live visual experience than there is to looking at a television replay or even a good cinematic reproduction. This is not to say that television or movies are bad; they are just not the real thing. For certain purposes, such as close up viewing and the removal of extraneous and distracting visual material, the reproductions may be even better than the "real thing" and certainly good television and movie producers use this fact to great advantage. (The movie can be "larger than life".) To try to exactly duplicate the original visual experience would be a misdirection of effort.

However, it is not important in recording engineering to rate the live visual experience against the reproduced one. Indeed sometimes the reproductions become more "real" to people that a real-life experience. I once tried to explain the stroboscopic effect by pointing out that the apparent backwards rotation of wagon wheels often seen in movie western scenes was a stroboscopic effect due to the intermittent nature of the image on the screen. It surprized me to find that most of the audience were convinced that this was not just a movie illusion but that real wagon wheels in bright sunlight would appear to do the same thing! Some were convinced that indeed they had seen this effect themselves. Obviously they had never seen a real wagon wheel on a moving cart in bright sunlight, or if they did, had not seen it often enough to notice what it was really like. For many of us, the recorded and reproduced image is more "real" than the live visual experience. At the cost of modern symphony tickets, the reproduced experience will be all that many people will experience.

What is important in recording engineering is that both live music and reproduced recorded music each can be a pleasurable experience on its own merits. Because of the difficulties which should soon become apparent, it is as forlorn a hope to exactly duplicate a live musical experience as it is to duplicate a live visual experience. What remains then, is to try to create via the recording medium, an experience which is in itself delightful, intriguing and stimulating, as any music should be.

Of course, this will involve subjective feelings about music and, as with any subjective feelings, there is a wide variety of tastes. While a considerable number of people might agree that a particular music record is delightful, intriguing and stimulating, there would likely be many more who do not consider it so. For example a good recording of Gregorian chant is not likely to appeal to a "rock" music fan. However, to a large extent the subjective feelings are related to the type of music rather than the quality of the recording. Given a particular type of music there will be a great deal of agreement among those who favour that type of music as to whether the record is a good one or not.

Of course a good record requires a good original performance; a bad performance can never yield a good record. However a good performance, as many performers know, does not always mean a good record. A lot of skill is required of the recording engineer in order to make a good record of a typical live concert. Just as in any good movie the camera is often used to give an enhanced view of something of interest, in a good music recording the balance of sounds and the "presence" of the instruments can often be better than in a typical seat in the recording hall. To make a good recording we must therefore know something about creating a good musical experience.

A musical experience is a multifaceted thing. At a fundamental level, it is a pattern recognition problem involving decoding the changing pitches, timbres and volumes of the sounds produced by musical sources. However, in identifying the music with the sources, there is another important aspect of a musical experience that is often overlooked; that of the role of the room in which the music is heard. Musical instruments when played in an anechoic chamber (a room designed to have no sound reflections from its walls) produce very boring sounds, even when played by very accomplished musicians. In fact the musicians themselves find it difficult to play the instruments properly in such a room. A great deal of the quality of a musical experience therefore comes from the sound reflected from the walls of the room in which the music is performed.

The science of psychology has not firmly established why this should be but for purposes of visualizing its importance we can imagine that auditory clues from reflected sounds have some very primitive functions in survival and in the sense of well being that comes from being surrounded by protective walls. For example it has recently been established that the preferred musical halls of the world have walls which are closer to the listeners than the ceiling. Apparently this has to do with the listener preferring the first reflected sounds to come horizontally as from a source such as another person at the listeners level, rather than from above such as would come from a clap of thunder or from someone or something higher than you. Why this preference occurs is not clear but what is clear is that the ear-brain system has the capability to exercise this preference. This means that it is capable of extremely fine discrimination of time of arrival and direction of arrival of complicated sound patterns.

The recording engineer thus has a problem; how to get some of the interest and quality of the reflected sounds onto the record without destroying the balance of pitches, timbres and volumes of the direct sounds from the instruments.

Again one is tempted to turn to the dummy head solution. Certainly, when played on a good stereo headset, this would give the same audio input as experienced by a listener seated at the dummy head location. However there are several serious drawbacks to such an approach. The most obvious is that many listeners do not like to wear headsets. Another that is fairly obvious is that listening to music in such a way is very solitary. This may be preferred by some people, particularly those that have had a great deal of recent stress from dealing with other people, but it is not the most generally preferred way to listen to music.

There is another reason, however, which is more fundamental and not so obvious. One of the ways a person perceives the direction of a source of sound is by an almost unconscious movement of the head. With such a movement, the relative phase of the sounds that arrive at the two ears changes and the brain seems to be capable of perceiving this phase change. With headsets, movement of the head gives no such effect. In essence, when you move your head, the whole room of sound seems to turn with you.

Thus, for a variety of reasons many people do not like to listen to music over headsets. There will always be a large number of listeners who will want to play the music over loudspeakers in a typical room of a home. For listening to a record playback in a room, care must be taken with the recording of reflected sounds in the auditorium since in listening to a record in a room, the reflected sounds within the living room itself add to the overall sound. Many good records and even many good stereo systems have their performance severely degraded by living room acoustics.

To understand the problems is not necessarily to have a solution. There is no handy set of prescriptions for how to get a record that will satisfy the greatest number of listeners. However, the more you know of the basic principles of acoustics, the better your chances of achieving the best solution.

Also, as a general interest, knowing how sound propagates, reflects and absorbs is basic to understanding the phenomenon of music itself and should increase your appreciation of music as well as help you as a performer.

Exercises and Discussion Points

1. What are some of the inherent differences between listening to "live" music compared to listening to recorded music? Discuss this from the point of view of what can be perceived in the live performance and what can be perceived in the recording. In what aspects is the live performance superior and in what aspects can the recorded music playback be superior? Use the analogy of a good movie compared to experiencing the "real thing".

2. Why is the production of a good music recording as much a matter of art as it is a matter of recording engineering?

3. What are possible reasons why the "dummy head" recording technique is not the universal solution to sound recording problems?

CHAPTER 2
THE PERCEPTION OF SOUND SOURCES IN A ROOM

2.1 The Distinguishing of Direct from Reflected Sound

One of the amazing subtleties of the live sound experience is the perception of sound coming directly from a source as distinct from sound that arrives at the ears by reflection from nearby surfaces. This can be demonstrated by setting up two speakers, one to an extreme left and the other to an extreme right, and putting separate sharp pulses of equal power into each speaker (see fig. 2.1). When the pulses are fed to the speakers in synchronism, observers to the right of the center line between the speakers will perceive a sharp click coming from the right speaker with the other speaker providing a sort of "stereo" effect telling the observer that there is sound also arriving from the direction that speaker. Observers to the left of this center line will experience the role of the speakers as reversed.

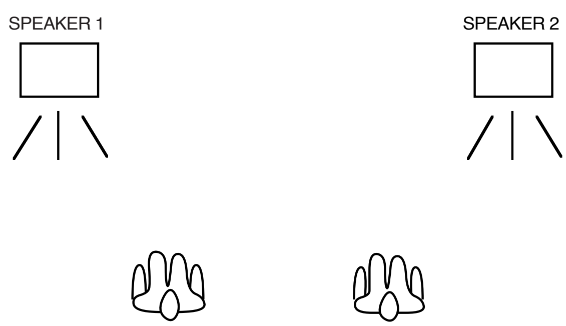


Fig. 2.1 Two observers listening to two speakers both emitting sharp sound clicks at exactly the same time. The observer on the left perceives the click as coming from speaker 1 and the observer on the right observes the click as coming from speaker 2.

This is not surprising. The speaker closer to the observer is the one which delivers the most sound and should therefore appear to be the source. However, when the power pulse to one of the speakers is delayed relative to the other with no change whatever in the power pulse levels, a surprizing effect occurs; the observer closer to the delayed speaker will observe the far speaker to be the apparent source of the sound!

This phenomenon is illustrated in fig. 2.2 where the pulse going to the speaker on the right is delayed relative to the pulse to the speaker on the left. Observers on the right of the room will discern the source of sound to be the speaker on the left. The more the pulse is delayed, the farther from the center line will be the observers who discern this apparent switch in sound source.

The same phenomenon would occur in reverse if the pulse to the left speaker is delayed relative to the pulse to the right speaker (fig. 2.3). Clearly the aural perception system is picking out the first sound to arrive at the ears and is using that to determine the direction of the source. A distant source with a weak sound at the observer will be perceived even in the presence of a stronger nearby source if the sound from the nearby source arrives later than that from the distant source. Furthermore, the perception system can discern time differences in the millisecond range. On an ordinary human scale this is an incredibly short time; a blink of an eye is about 100 ms. What is happening on such a short time scale?

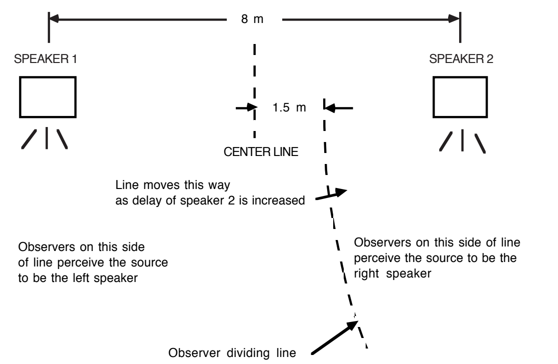


Fig. 2.2 The separation of observers into those perceiving different speakers to be the source of a click sound when one speaker is delayed relative to the other. In the case shown, speaker 2 is delayed by about 10 millisecond (ms) relative to speaker 1.

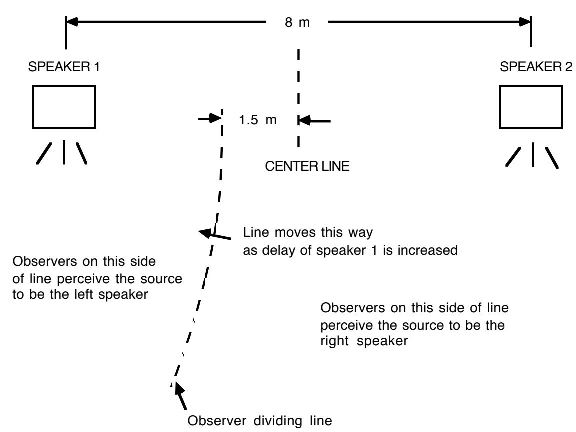


Fig. 2.3 The separation of observers into those perceiving different speakers to be the source of a click sound when the left speaker is delayed relative to the right. In the case shown, speaker 1 is delayed by about 10 ms relative to speaker 2.

The physics of sound propagation gives a clue as to what is happening. The velocity of sound in normal room air is about 340 meters per second corresponding to about 3 milliseconds for travelling 1 meter. When one source is delayed by 10 ms relative to the other, you would have to be 3.4 m closer to the delayed source than the to the other for the delayed source to be the apparent source. In the demonstration cited above, an observer who is 2 meters closer to one speaker than the other will therefore observe the closer speaker to be the source of the sound until that speaker is delayed by more than 6 ms relative to the farther speaker.

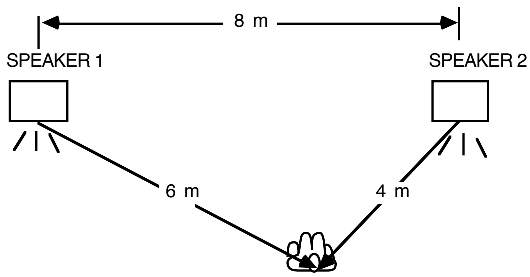


Fig. 2.4 The observer will discern speaker 2 as the source of a click sound when two identical pulses are fed into the speakers, until the delay of speaker 2 relative to speaker 1 is greater than about 6 ms.

In the branch of psychology called psychoacoustics, this perception phenomenon is called the "precedence phenomenon". Just how powerful it is can be demonstrated by having an observer about 5 times closer to one speaker than the other (see figure 2.5).

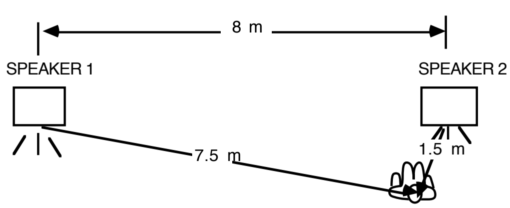


Fig. 2.5 The observer will discern speaker 2 as the source of a click sound when two identical pulses are fed into the speakers, until the delay of speaker 2 relative to speaker 1 is greater than about 18 ms. With delays of speaker 2 of more than 18 ms, speaker 1 will be discerned to be the source of the sound even though the sound intensity from speaker 1 is about 25 times weaker at the observer than the sound from speaker 2.

To show that timing effects are more important than loudness, one speaker can be put at about 25 times lower intensity than the other. Observers near the center line between the two speakers when the speakers are in synchronism will sense the sound as coming from the louder speaker. The intensity information is then being used to determine the direction of the sound source. However, as soon as the louder speaker is delayed relative to the weaker one, the sound is perceived as coming from the weaker speaker showing that the brain regards this information as being more important in determining the direction of a sound source than is the loudness information.

This does not mean that the loudness information is ignored. As in all perception activity, the brain appears to integrate all the information it gets. In an activity so important to survival as determining the direction of a sudden source of sound, the brain could not afford to do otherwise. Therefore even such clues as changes in visual patterns and general foreknowledge of the nature of the surroundings will be used in estimating the direction of the source. However, timing information about the first sound to arrive at the ears is regarded by the brain as perhaps the most important information of all. That this should be so is easily understood from simple physics. Curved reflecting surfaces can concentrate and focus sound power from a single source so that a reflected wave can deliver more power that the direct wave. What is always true however, is that the reflected wave arrives later than the direct wave; it will always have farther to travel because a straight line is the shortest distance between two points.

It is therefore very important for the brain to divide the perceived sound into two parts; direct sound coming from the source and indirect sound which arrives later and which is therefore deduced as coming from a reflecting surface. That is what the brain is doing in perceiving the click sounds from the two speakers; the sound from the speaker which is delayed is perceived as a reflected sound wave from a nearby surface. This can be seen by turning off the delayed speaker and noting the distinct change in the stereo image of the perceived sound. This comes about because the brain uses the direction of the delayed sound to perceive an imaginary wall which is delivering this delayed sound by reflection. That this "reflected" sound is different from an echo effect can be shown by increasing the delay of the second speaker to about 40 milliseconds whereupon there would clearly be an echo effect.

A very important point for recording engineers now arises. The typical environment in which recorded music is played through speakers does not have the reflecting surfaces to produce the "liveliness" caused by the reflecting surfaces that will be in the typical live performance hall. A satisfactory stereo image can therefore not usually be produced by simply putting different sources of sound into each speaker such as, say a voice in one and an accompanying piano in the other. Rather it involves putting both sound sources into each speaker but with different relative timings. Thus, for example, if the singer is to be perceived as being on the right and the piano on the left than the right speaker should receive the singers voice first and the left speaker should receive the piano sound first. The relative timing and amplitude of the two sources in each speaker can be manipulated to give a pleasing "stereo" effect.

2.2 The Perception of Sound Direction

The "precedence phenomenon" show that the brain picks off the first millisecond of sound from a new source to determine the direction of that source. But now consider what it does with this short segment of sound. What information can the brain have that will tell it the direction from which that sound is coming?

Here, just as it is very important for stereo vision that we have two eyes, it is very important for stereo aural images that we have two ears. It is the differences in the sound perceived by the two ears which seems to give the most important information for creating a stereo image in the brain.

Again there are two important pieces of information; the relative loudness and the relative timing in the two ears. The importance of the relative loudness information is obvious; sound which appears louder in the left ear will be coming from the left and sound which appears louder in the right ear will be coming from the right. Furthermore, the shape of the outer ear itself seems to be such that relative perceived loudness in the two ears and changes in perceived loudness due to head movements is can be used to perceive the direction of a sound.

However, recent experiments in psychoacoustics have shown that the brain also uses the relative time of arrival of a sound at the two ears as an important clue to the direction of that sound.

To understand this phenomenon, first consider a simple case of a sound pulse arriving at a listener from the right (fig. 2.6). The only timing information presented to the brain by the direct sound source itself is the time difference of the sound arriving at the two ears. Given the average distance between the two ears of a human being to be about 15 cm, the time interval between the arrival of the direct sound at the right ear and it arrival at the left ear is about 0.45 milliseconds (450 µs). If the sound source was directly to the left of the observer, the sound would of course arrive at the left ear 450 µs before it arrived at the right ear. The possible relative timings of the sound in the two ears therefore range over 900 µs or about 1 ms.

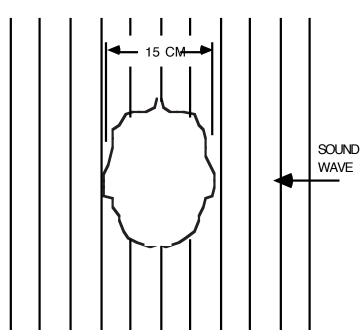


Fig. 2.6 Sound wavefronts falling on a listeners head from directly to the right.

This is a very short time range for the very simplest of tasks; that of determining whether the sound comes directly from the left or directly from the right. For the more demanding job of locating a direction somewhere in between even shorter time intervals are involved. For example, a sound at 45° to the right of straight ahead (fig. 2.7) would travel about an extra 10 cm to get to the left ear, corresponding to a time delay of only 0.3 milliseconds for that ear.

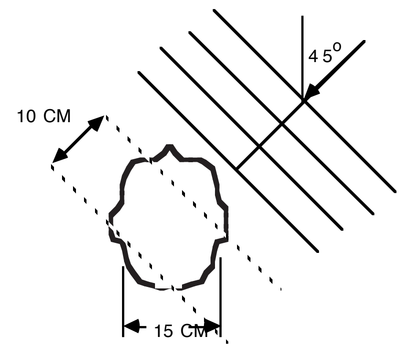


Fig. 2.7 Sound wavefronts falling on a listeners head from 45° to the right of directly in front.

The accuracy of determination of the direction of a sound source by timing has been rather thoroughly tested by varying the timing of sharp pulses applied separately to stereo headsets. It appears to be about ±30° for sound sources directly facing the observer. This means that the accuracy of the mechanism used by the brain for this timing is about ±200 µs.

Again, these timing intervals are incredibly short by normal human standards. To get an idea of the scale, a professional baseball player can direct a ball to a selected part of the outfield by timing his swing relative to the pitch; an incredible feat but one that only involves timing accuracies of about ±10 ms. In 200 µs, the tip of the bat of a major league hitter will move less than a centimeter.

The mechanism that achieves this timing is not very well known. About all that has been firmly established is that the aural nerves from the two ears are connected together where they meet and that the neural pulses generated by the "hair cells" of the cochlea and set along the aural nerves are timed to the arrival of a particular phase of a sound wave at the ears (see Roederer). The brain therefore has the basic tools it needs to make a comparison of the time of arrival of the sound at the two ears. It perhaps does this by having a set a "timing" neurons in which the neurons for testing a particular time interval have the appropriate synaptic connections for testing this time interval.

As a very simplified view of how this could be carried out, imagine that there was a neuron with an excitory synaptic connection from the right ear placed at the end of a dendrite of the axon (fig. 2.8). Suppose that as well it had an inhibitory synaptic connection from the left ear on a dendrite that was shorter. A pulse that arrived at this connection later than that at the excitory connection would still result in killing the action of the excitory synapse. In fact the pulse must arrive later to carry out this function.

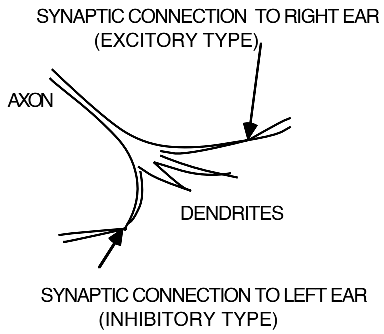


Fig. 2.8 Sound wavefronts falling on a listeners head from 45° the the right of directly in front.

The typical velocity of neural pulses is about 100 meters per second. For a delay of 200 μs the length of neural material involved would therefore be about 2 cm. It is easy to visualize that such a network of neural connections of various lengths could have evolved.

The actual mechanism that is used by the brain to discern sound direction by timing is not of importance for this course. What is important is that this mechanism exists and is an important mechanism in perceiving an acoustic environment. This means that it must be taken into account in any recording of the sound for future playback. Simply placing a microphone on the left side of the room to pick up the sound sources on the left side and another microphone on the right side to pick up the sound sources on the right side will not be enough to give a good "stereo" quality to the sound when it is replayed in a typical living room. One may have to pay more careful attention to the balance of sound from both sides falling on each of the microphones in order to get the speakers to give a pleasing sense of sound reflections within the living room listening area.

2.3 The Perception of Reverberant Sound

2.3.1 The Importance of Reverberant Sound

The brain perceives two important types of sound in a room; the direct sound which it uses to determine the direction of a sound source, and reflected sound which it perceives as coming later and giving clues as to the size and geometry of the room. However, there is one more important perceived type of sound in a room; reverberant sound. The importance of this type of sound in recording is easily shown by making two simple monaural recordings of a person speaking in a room; one with a microphone close to a speakers lips and the other with the microphone as far away from the speaker as possible. Playing back these recordings, except for perhaps a slightly different balance in the high to low frequency levels, the sound from the first recording is not very different in quality from the original sound from the person who is speaking. However the second recording produces a sound as if you were listening to the person from inside a barrel.

This is perhaps very familiar. Anyone who has tried to record the sounds of a party by placing a microphone in the middle of a room filled with people will have noticed the hollow sound of the recording; a sound which is distinctly different from the sound a person would hear if the persons ears were exactly at the position of the microphone used for the recording.

What is causing this effect? The microphone must be receiving the same sound as a human ear at the same location. Why is the microphone apparently hearing the sound differently than a human being? What is the cause of the resonant background sounds which make it difficult to hear individual voices on the recording?

This phenomenon occurs because a monaural recording cannot retain any directional information. The recording is nothing more than a record of the sound levels that fell on the microphone as the person is talking. When this recording is played back, all the sounds that fell on the microphone are played back through the loudspeaker. Directional information in the original sound is now completely distorted or even missing; all of the sound is coming from the loudspeaker. In the original sound there was direct sound from the speaker's mouth and reflected sound coming from the walls of the room. The brain could separate these sounds by their directional features. In the sound played from the record, this directional information is lost and the two sounds are muddled together.

There is also another compounding effect. The sound from the speaker also produces reflections from the walls of the room. The brain now perceives these sounds as the genuine room reflections, further adding to the perception that all the sounds heard as coming from the loudspeaker are in fact direct sounds. Thus, in the playback, the room reflections onto the microphone are perceived as direct sounds. It is these reflections that give the booming hollow sound. When the microphone is held close to the mouth, the direct sound on the microphone is much more powerful than the reverberant sound. The booming, hollow sound due to room reflections become imperceptible.

While the power of the reverberant sound is not so obvious in "live" listening, it is obvious from the recording of sound with a distant microphone that there is in fact quite a lot of power in this sound. How powerful is this typical reverberant sound compared to the direct sound from a source?

Again, this can be shown in a simple demonstration using an electronic noise generator, an audio amplifier with a single speaker and a sound level meter (fig. 2.6). The speaker should be placed somewhere in a room not too close to any of the walls.

The sound level produced with the noise generator on should be turned up so that the level at a point in the room as far from the speaker as possible without being very close to a wall, is about 60 dB. (This would be appropriate for a room in which the normal background sound level is 45 dB or less.) The point of this is that everywhere in the room the noise from the speaker should swamp any other background noise in the room.

The sound level is then measured at various points throughout the room. In a typical lecture room designed for 50 students, it will be found that the sound intensity at 3 meters or more from the speaker will be uniformly about 60 dB.

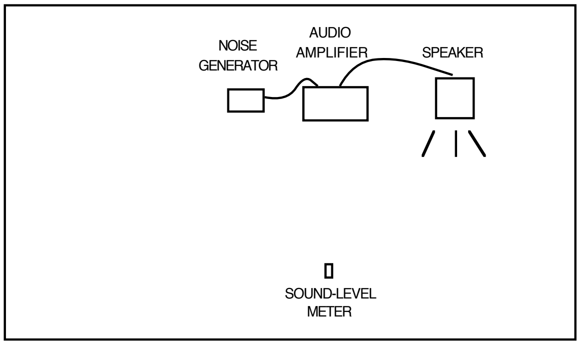


Fig. 2.6 Demonstration set-up for observing the relative importance of reverberant sound and direct sound.

This might already seem a little strange since one would expect that the sound level should continue to drop the farther you are from the speaker, even if you are more than 3 meters from the speaker. In the typical room being considered, the sound level at a half a meter from the speaker will be about 75 dB. From elementary physics, the sound intensity from a single radiating source will fall off by the inverse square law, meaning that a doubling of the distance from a source will cause the intensity to fall to one quarter. On the decibel scale a fall to one half is very near to a fall of 3 dB. A fall to one quarter is therefore a fall of 6 dB. (One quarter is one-half times one-half and decibels being logarithms add for multiplications.) The noise level at 1 meter from the speaker should therefore be about 69 dB which will be the case in the room will are considering.

Continuing away from the speaker, the noise level at 2 meters should be about 63 dB, at 4 meters should be about 57 dB and at 8 meters (which is as far away from the speaker as it would be possible to get in the type of room we are considering) the noise level should be about 50 dB.

What is observed however, is that the noise at the farthest distance from the speaker is actually 60 dB. This means that the observed noise intensity is 10 times greater than what would be expected from a simple inverse square fall off of the sound from the speaker!

The direct sound from the speaker will indeed fall off with the inverse square law and therefore have a level of about 50 dB at the farthest distances from the speaker. The noise which is amounting to 60 dB must therefore be sound which comes from reflections from the room walls. This strong sound from room reflections is called "reverberant sound". That this sound is ten times as strong as the direct sound is the reason for the booming "resonance" in the simple monaural recording made at the distant point.

Again, a human ear at the distant point also experiences the same relative strength of direct sound to reverberant sound. However, normally it does not seem that the reverberant sound is that much stronger than the direct sound, particularly for people with two good ears. The ability of the brain to use directional information to pick out the relatively feeble direct sound from the preponderance of reverberant sound is truly amazing.

2.3.2 Room Radius

The relative strength of the reverberant sound in a room to the direct sound is often expressed in terms of the "room radius". This is defined to be the distance from a source at which the direct sound and the reverberant sound are equal. For the case of the room that has been under consideration here, the reverberant sound level is 60 dB while the direct sound level is 63 dB at 1 meter and 57 dB at 2 meters. The direct sound level should be about 60 dB at a distance which is the square root of 2 or about 1.4 meters. In other words, you have to be as close as 1.4 meters to the speaker for the direct sound to be as strong as the reverberant sound.

Thus to enhance the direct sound over the reverberant sound in a recording, the microphone must be placed inside the "room radius" of the speaker. By placing the microphone at half the room radius, the direct sound will be 6 dB stronger than the reverberant sound and by placing it at 1/4 the room radius (35 cm) it will be 12 dB stronger.

This will be easy if there is only one source of sound to be recorded, such as a persons voice. However, in music recording there are often many performers of equal importance. To get a proper sound recording from each performer requires, in principle, a microphone well inside the room radius of each performer and a separate recording made of the sound picked up. For large orchestras, this is clearly not feasible and so compromises have to be made. The placing of the microphones and the use of the recording channels available in recording a full symphony orchestra concert is clearly an art which requires a high degree of experience as well as knowledge of the music and the instruments with which one is dealing.

At first it might appear that reverberant sound is a nuisance. After all, the information you want to hear is all in the direct sound. The problem is that the direct sound from a persons mouth, unless that person is an opera singer or a hog-caller, is insufficient for easy discernment beyond about 2 meters from the person. To follow what a person is saying in a room even as small as a typical living room, you need the extra sound power coming from the reverberant sound. This sound power, while lacking the directional clues as to the location of the speaker provided by the sharp attack components of sound, still contains important information as to the vowel sounds that are uttered and to some other components of human speech such as the hisses and burring sounds that often modify these vowel sounds. Similarly, the reverberant sound provides information about the tonality and harmony in musical sounds. The reverberant sound is therefore important.

However, there should not be too much reverberant sound. If there is, then it will smother the direct sound entirely and it will be very difficult to follow speech or quick music passages. (You really will start to get the booming resonant quality of the tape recording in the live perception of the sound.)

Achieving the correct balance of reverberant sound to direct sound in a room is a matter of acoustic engineering. This correct balance will depend on the type of acoustic use intended for the room. If there is to be fast speech or complicated music then the reverberant sound should be reduced. If on the other hand there are to be performances of slow choir music such as Gregorian chant, then the reverberant sound should be increased. In fact, it is largely the balance of the reverberant sound to direct sound that determines the desirability of the acoustics of a music performance hall.

Exercises and Discussion Topics

1. Discuss the importance of sharp transients in a sound for the brain to be able to locate the direction of the source of that sound. What physical information in the sound wave is apparently being used by the brain as the overriding factor in determining its source? Why do the transients (such as the "fricatives" in speech) provide this information more than does the usually following sustained sound (such as the vowels) which contain much more sound power than do the transients?
2. Why is the perception of a changing direction of a sound source best when the face is pointing toward the source?
3. What is the "precedence phenomenon"? What conceivable mechanism can be in the ear-brain system which can achieve this perception effect?
4. Discuss the problem of recording sounds in a "live" room such as a living room compared to a "dead" recording studio. What are the problems of recording in a "dead" studio?
5. What are the three types of sound in a room relating to the perception of a sound source and the reflecting surfaces in a room? Explain the perceptual role of these three types of sounds with some explanation as to how these separate perceptions are important to the human organism.
6. A "pink" noise generator is used to drive a loudspeaker near the center of a room. ("Pink" noise is noise which contains equal noise power within all octave intervals. It is the type of noise which appears capable of masking all frequencies of sound with equal capability and is used for purposes such as masking the telephone conversations of other people on other lines that sometimes sneak through by electronic "cross-talk" onto your line. The following measurements were taken with a sound level meter at various distances in a particular direction from the speaker.

Distance (meters)	Sound Level (dB)
0.5	102
1.0	96
2.0	92
3.0	91
4.0	90
5.0	90
6.0	89

By drawing graphs of this data, estimate from the graphs;

a) the sound level of the reverberant sound in the room (Watts per square meter and decibels)

b) the "room radius" for pink noise in this particular direction from the speaker

c) the ratio of direct sound to reverberant sound intensity (db and actual fraction) at the back of the room

d) If the loudspeaker radiates uniformly in all directions estimate the total sound power it radiates.

To deal with this problem, you need to know the connections between sound level in dB and actual intensities I in Watt/m^2;

$$dB = 10 \, \mathrm{Log}_{10}\left(\frac{I}{10^{-12}}\right)$$

$$I = 10^{\left(\frac{dB}{10}\right)} \times 10^{-12}$$

(See notes for Physics 224 or any good elementary physics textbook.)

CHAPTER 3

AN INTRODUCTION TO CONCERT HALL ACOUSTICS

Ever since people began congregating in enclosures it must have been realized that the enclosure itself considerably changes the sound from a source. As music developed, it was also realized that music in a room sounded very different from the same music in open air and that different rooms made the music sound differently. Rooms for listening to music evolved much the same as musical instruments; the features of rooms that had a good effect on the sound were copied and those of rooms that did not were not copied. Music, musical instruments and the rooms in which these instruments were played to make the music, evolved together.

It is not surprizing then, that the terms used to describe the musical effects of a room are similar to those used to describe music and musical instruments. As in describing any art, the words are often meant to evoke an emotional response in the reader similar to the emotional response felt by the writer. Schroederer gives a list of 56 words commonly used in Germany to describe the musical properties of various concert halls. Rather unfairly to North American readers, he does not give any English translation of these words, merely mentioning that they are as meaningless in a rational sense in German as they are in their English translations. To show the type of words they are, they are given here with as close a translation into English as possible.

auf dringlich	intrusive as in people or perfumes
auf richtig	sincere, upright as a person
ausgewogen	well-balanced as in a persons opinions
begeisternd	inspiring, thrilling as in a speech or artistic performance
betäubend	deadening as in a drug
bezaubernd	charming, enchanting as in a woman or story
brillant	brilliant as in human intelligence
deutlich	clear, distinct as in a view
empfindlich	sensitive, touchy (handle with care) as in a person
erhebend	elevating, uplifting
erheiternd	amusing as in anecdote
erschreckend	frightening, startling
erstaunlich	amazing
glasclar	clear as glass
glorios	glorious
hallig	like a large hall
hart	hard
heikel	difficult, delicate as of a subject or person
herrlich	marvelous, splendid
hinreissend	enrapturing, thrilling
intim	intimate
jämmerlich	miserable, wretched
kalt	cold
kranfhaft	morbid
lebendig	lively, vivacious
lieblich	charming, sweet as in a maiden
prächtig	magnificent
reich	rich
ruinös	ruinous as in business
schmal	narrow, small as in person
schillernd	iridescent
schön	beautiful
schrecklich	dreadful
schrill	shrill
temperamentvoll	vivacious
trocken	dry
überwältigend	overwhelming
unbarmherzig	pitiless, unmerciful
undeutlich	inarticulate
undurchsichtig	opaque
unerbitlich	inexorable, merciless
unheimlich	terrifying
verschmiert	smeared
verschmitzt	sly as in grin
verschmolzen	blended, merged
verworren	confused, muddled
vollkommen	perfect
volltönig	full-sounding
vorzüglich	outstanding
wahrhaftig	sincere
warm	warm
widerhallend	echoing
wohltönend	melodious
wohltuend	pleasant, producing a sense of well-being
wunderbar	miraculous

What these words clearly indicate is that people can get very emotional about room acoustics. People who have heard a favorite piece of music in a hall which to them gave a very favorable impression can get very upset when they hear the same music in not so favorable circumstances.

There is, of course, a great deal of subjectivity in such descriptions and, just as in art, there will be a wide spectrum of likes and dislikes. However, for a given type of music in a given hall there will be a fair degree of agreement among people who like that music, as to whether or not the hall is a good place for the music. Since these people will be the ones that will be expected to pay for the tickets, it is important that they be satisfied with the hall acoustics.

Just as people over the years have established the art of making good musical instruments, it would be expected that people would have established the art of building good music halls. However, the problems are not at all similar. Musical instruments can be made by copying proven designs and techniques. A modern violin does not differ significantly from those made three hundred years ago. On the other hand, no one would seriously consider building a modern music hall by copying one from three hundred years ago. Building codes, safety considerations, availability of materials, costs and required seating capacity have all changed drastically in three hundred years. Modern halls have to be built to achieve satisfactory acoustics using materials and building designs that have not had the centuries of development that have been applied to musical instruments.

One approach to the problem of using new materials and techniques is the empirical; build something according to some general ideas and then modify it on the basis of the actual performance achieved. This is how many new musical instruments originate. However, the cost and time involved would seem to prohibit such an approach in building a new concert hall. Nonetheless, it would seem that that is the approach actually taken in the building of many modern concert halls. Perhaps the most infamous case is the Avery Fischer Hall in New York, but, just to show that even the Germans can make disastrous design errors, there is also the case of the Rheingold Hall in Mainz. Both halls needed extensive restructuring after their official openings before the acoustics were judged as acceptable. What is clearly needed is an engineering approach based on scientific principles that would allow the prediction of the acoustic properties of a concert hall while it is still just a design on the drawing board.

The science of acoustics is well established and so is the engineering of buildings. It is therefore possible to design and build a concert hall with specified acoustics. Why then are mistakes made? Certainly a large part of the fault is not enough careful engineering or the acoustic engineers not having the final say in the interior design. However, there is also a large component of not knowing exactly what the desired acoustics of a hall are in scientific and hence engineering terms. Words meant to convey emotions felt about a hall are practically useless when trying to design better acoustics. The problem then becomes; what are the desired acoustic properties of a concert hall? To what physical properties of a hall are people responding when they use words such as "well-balanced", "charming" or "intimate"?

3.1 Desired Properties of Concert Halls

3.1.1 Reverberation Times

Sabine, at the start of the 20th century, was the first to establish a connection between a physical property of a hall and its acoustic impression. He proved that the acoustics of the auditorium in the newly opened Fogg Art Museum in Boston, which would probably described as inarticulate, or even dreadful, and which made speech practically unintelligible, was due to an excessively long exponential decay time constant for the sound in the room. To bring such an abstract concept to a more concrete, understandable level, he translated this decay time into a "Reverberation Time" which is 13.86 times the decay time and which turns out to be the time for the sound in the room to fall by 60 dB when a source is turned off. 60 dB is about the short-term range of hearing level for the average person and so the Reverberation Time is about the time it would take for a human being to perceive the sound to disappear. In a commonly used very rough test, the ringing sound following a sharp hand-clap will appear to last for about the reverberation time. By measuring halls with good acoustics, Sabine established that good auditoria for speech had reverberation times of one second or less whereas Fogg Auditorium had a reverberation time that was much longer.

From the physics of sound wave propagation, Sabine could relate the reverberation time of the room to the simple ratio of the volume of the room and the effective sound absorbing area of all the surfaces in the room. He showed that the relationship could be expressed by the simple formula (see Chap. 4);

$$\boxed{T = 0.165 \times \frac{V}{A_{eff}}}$$

where T is the reverberation time in seconds, V is the volume of the room in cubic meters and A_{eff} is the effective absorbing area of all the surfaces of the room in square meters. By measuring the sound absorbing properties of various building materials, he could predict their contribution to the absorbing area of the room and hence their effect on the reverberation time. He then recommended how Fogg Auditorium could be refurnished to get the desired reverberation time.

Sabine's success on Fogg Auditorium was impressive and when Boston planned a new symphony hall he was hired as its acoustic engineer. Boston Symphony Hall, from its very opening, has been regarded as one of the outstanding concert halls of the world and so concert hall acoustics began to be regarded as a science.

Following Sabine's work, there has been a great deal of data gathered on the reverberation times of various halls. The results indicate a wide spread of desired reverberation times, depending on the size of the hall and the type of music being played. The general consensus is shown in Fig. 3.1

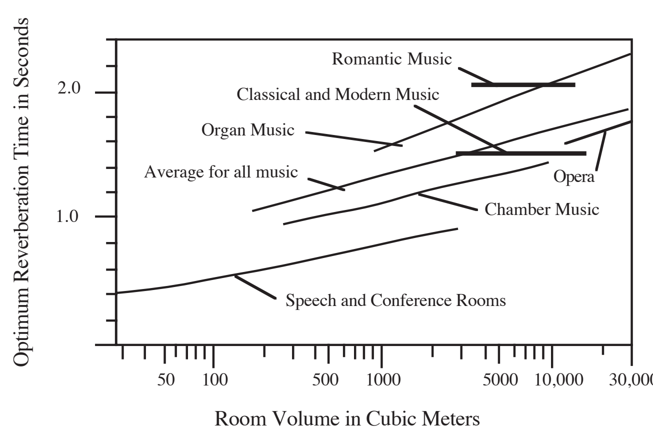


Figure 3.1 Desired Reverberation times for various uses of halls of different sizes.

Fortunately, there is a relationship between the size of a hall and the type of music that will normally be played in it. Chamber music is meant to be played in smaller halls than symphonic works. However, there are some modest sized halls of around 3000 m^3, (roughly 22 × 16 × 8.5 m) meant for about 1000 people, where all types of music and theater are likely to be performed. Such auditoria will then be most acceptable on the average when they have reverberation times which are a compromise of the best for the various uses (perhaps about 1.3 seconds). By increasing the reverberation time to 1.5 seconds, such a hall could become an excellent hall for classical music and by increasing it to 2.1 seconds it could be excellent for symphonic music. However the hall would then be practically useless for theater or the music of smaller groups.

In modern society, a hall seating only 1000 people could not support a symphonic orchestra and so small halls will generally be used for purposes requiring a shorter reverberation time. Small halls will therefore be judged harshly if they have a long reverberation time. However, such halls can usually have their acoustics easily modified if they are judged unsatisfactory.

Symphonic works on the other hand, are usually performed in larger halls built specifically for such music. Such halls are expensive undertakings, generally meant to be showpieces of major cities. Having the acoustics of such halls judged as excellent is of great consequence and there are rancorous debates between music critics, orchestra leaders and building architects when the acoustics are judged unsatisfactory. One othe most cited examples is that of Avery Fischer Hall in New York, opened with much fan-fare as The New York Philharmonic Hall in 1962 and immediately panned by the music critics. After many modifications to as late as 1975, the hall was judged as unsatisfactory and a total reconstruction of the interior was undertaken. Finally it appears that the acoustics are judged as acceptable.

The New York Philharmonic Hall is cited in practically every modern text on concert hall acoustics and severely damaged the reputation of the acoustic consultants involved in the original design. This is somewhat unfair since these acoustic consultants were not, it seems, responsible for the final decisions on the architectural features important to the acoustics of the hall. This is unfortunately a common situation in large buildings where the architectural features necessary for good acoustics add greatly to the already seemingly excessive cost. However, more important in the long run has been the loss of public esteem for the science of acoustics itself. There were many snickers when the recent ill-fated proposal for a Montreal Concert Hall at Berri-DeMontigny Metro station was promised to be "An Acoustic Gem".

Again, this is unfair to the science of acoustics. Modern acoustics is capable of predicting quite accurately how sounds will propagate throughout a concert hall. How then do such disasters as the New York Philharmonic Hall come about? How did Sabine succeed with very crude measuring tools and a slide-rule when later workers with the benefit of much more data, much more precise tools such as oscilloscopes and sound level meters and powerful computers, fail?

It was realized very early that Sabine's formula was not a complete prescription for the acoustics of a concert hall. In fact it gave wrong results if the hall was very large or if the absorbers were not uniformly scattered throughout the hall. More accurate formulae were developed but did not seem to provide, in themselves, a solution to good concert hall acoustics. The solution to good concert hall acoustics was not as simple as having a formula for the right average reverberation time.

3.1.2 Variation of Reverberation Time with Frequency

One of the factors realized very early was that good concert halls should have different reverberation times for different frequencies of sound; the preferred halls having longer reverberation times for low frequencies. As an example, the variation of reverberation time with frequency for the Musikvereinssaal is shown in Fig. 3.2.

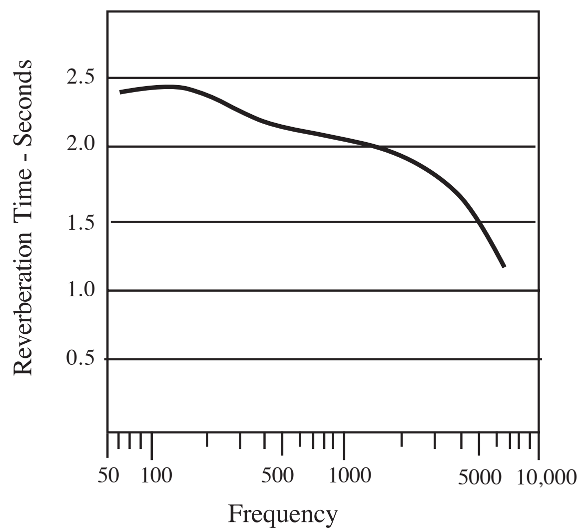


Figure 3.2 Variation of Reverberation times with frequency for the Musikvereinssaal (occupied).

The reverberation time at 100 Hz is 1.2 times that at 1000 Hz and twice that at 6000 Hz. The need for longer reverberation times at the low frequencies is explained by the human ear having a much smaller loudness range for low frequencies around 100 Hz than for frequencies around 1000 Hz. Whereas a 1000 Hz note will seem to disappear when it has dropped by 60 dB, a 100 Hz note of 100 dB will seem to disappear when it has dropped by only 40 dB. For a balanced timbre in the reverberant sound (i.e for the tone not to get harsher or brighter as it fades, the 100 Hz note should <u>appear</u> to last as long as the 1000 Hz note. For this to be, it must <u>actually</u> last longer! It was found that one of the greatest problems with the New York

Philharmonic Hall was the absence of bass reverberation due to absorption in the spaces between the ceiling reflectors.

However, while the correct reverberation time for all frequencies appears to be a necessary condition for excellent concert hall acoustics, meeting just that condition still does not seem to guarantee excellent acoustics. Human beings must be responding to something other than just the reverberation times of the sounds in the room.

One of the aspects of Sabine's work that is often overlooked is that he analyzed such renowned halls of the time as Carnegie Hall in New York, the Academy of Music in Philadelphia and the Gewandhaus in Liepzig, since destroyed but still regarded by many as the all-time greatest hall for classical music acoustics. He also toured with the Boston Symphony to judge the concert halls in which it played. It was this experience, probably much more than his simple formula, that led to his success with the Boston Symphony Hall. He seems to have copied features from the halls which had good acoustics. It is also perhaps significant that he seems to have had control over the interior design. The hall may be acoustically excellent but offends the eye of many modern viewers and would probably not have been allowed if an interior decorator or architect had veto power over the design.

What did Sabine achieve in his design, other than the correct reverberation times for the hall?

3.1.3 The Importance of Early Sounds

The question returns; what are the physical properties of the sound in a concert hall that make good acoustics for music performances?

Winckel, a renowned German professor of the physics of music undertook an inquiry in 1950 to 1955 with internationally renowned orchestra leaders to identify and study the best regarded concert halls of the world. The results for the top five at the time are listed below, in order or merit.

Concert Halls	Date of Construction	Volume m^3	Number of Seats	Mean Reverberation Time
1. GMV	1870	14,600	1680	2.05
2. TCBA	1908	20,550	2487	1.8
3. CA	1887	18,700	2206	2.0
4. SHB	1900	18,700	2631	1.8
5. KG	1935	11,900	1371	1.7

GMV - Grösser Musikvereinssaal Vienna
TCBA - Theatro Colon Buenos Aires
CA - Concertgebouw Amsterdam
SHB - Symphony Hall Boston
KG - Konzerthus Gothenburg

These figures show that there can be a large variation in reverberation times for excellent acoustics and that this variation is not correlated with room volume as indicated by the simple graph in Fig. 3.1. The Musikvereinshaal Of Vienna, one of the smaller halls has the longest reverberation time and is rated number one while Theatro Colon of Buenos Aires, the largest, is rated number two yet has a reverberation time 1/4 of a second shorter. This is even more puzzling when it is realized that tests have shown that a good music critic can detect the effects of a change of 1/20th of a second in the reverberation time of a hall. Furthermore, the Musik-vereinshaal, which is especially favored for romantic music (probably because of its longer reverberation time) is also claimed to be excellent for classical music which is normally regarded as being better in a hall with a shorter reverberation time. Also puzzling is that the Concertgebouw in Amsterdam is not regarded as being as resonant as the Theatro Colon even though it has 1/5th of a second longer reverberation time. Finally, the no longer existing Old Gewandhaus hall in Leipzig with a volume of 2100 m^3, has been calculated to have had a reverberation time of only 1.2 seconds in the occupied condition, yet is regarded as the best hall that ever existed for classical music. Reverberation times are therefore not the only factors in determining the quality of concert hall acoustics.

One fact not explained by the reverberation times of halls is that the most satisfactory acoustics for a rectangular hall seem to occur when the hall is about 3/4 as wide as it is deep and a little more than half as high as it is wide (ratios of length to width to height of 4:3:1.6). Also, this rule itself breaks down for halls larger than about 15,000 m^3 when they should be narrower than 3/4 of the depth. Winckel also thought it was noteworthy that the preferred halls of the world were mostly rectangular but decorated in the neo-classic style with columns, sculpture, coffered ceilings etc., copies of which would not be allowed to be built by any modern taxpaying public. It appears that the detailed geometry of a hall is as important in determining acoustic quality as are the overall reverberation times.

From the physics of sound propagation, it is easily seen that the detailed geometry of a room affects mostly the patterns of the first reflections that arrive at the listeners. Multiple reflections over many surfaces average out these detailed effects. Even in large halls, sound arriving after 1/2 a second, will have undergone 5 to 10 reflections. Modern concert hall acoustics has therefore concentrated on the early sound, often defined as the sound which arrives in the first 50 ms or 1/20th of a second which, in a typical concert hall, contains only the direct sound and one or two reflections.

What seems to be important is that there be a proper balance in sound arriving in the first 50 ms compared to what arrives later. This, of course, is closely related to the ratio of direct sound to reverberant sound, a concept which is in turn closely related to the reverberation time of a hall. However, the demands on the sound arriving in the first 50 ms are not just that it have a certain intensity compared to what follows. The details of how this first 50 ms of sound arrives at the listener seem to be very important.

The first sound to arrive is, of course, the direct sound. The importance of this sound has been discussed in Chapter 2. How the direct sound arrives has been discussed in Chapter 3. It is roughly independent of the geometry of the hall, providing there are no obstructions or very near surfaces; the sound radiates straight to the listener with an intensity which falls off by 6 dB for each doubling of the distance of the listener from the source. There is some enhancement of the direct sound by diffraction around the seats and peoples' heads in the foreground of the listener; an effect that was very important in ancient Greek ampitheaters. (There was only direct sound and no reverberant sound in these open air theaters.) The floors of the seating areas in halls should take this into account. Fortunately, the requirement of good line of sight of the performers generally means that this acoustic requirement is also met.

However, the sound that arrives at the listener in the next 50 ms following the direct sound seems to be also very important. First of all, some reflected sound should arrive in this first 50 ms or there will be a pronounced echo. It is also important that none of the subsequent reflections stand out from the others so as to give an echo effect.

It is thus clear why smaller halls, or halls with lots of columns, sculptures and other clutter, would be preferred. In a cluttered hall, there are many surfaces nearby that can give an early reflection to anyone in the hall and this early reflection will be followed by many small reflections from more distant objects. The direct sound is followed by an early reflected sound indicating intimate surroundings followed in turn by a multitude of small reflections building up to a reverberation which allows easy observation of the tonality and timbre of the music.

Why then are rectangular halls preferred with relatively high ceilings?

Firstly, it is not true that perfectly rectangular halls are preferred when the clutter is removed. It appears rather that the side walls should diverge a little. For example, the Theatro Colon in Buenos Aires is not rectangular and not decorated in the neo-classic style. This indicates that the side wall reflections must be of importance. The importance of the side wall reflections can also be related to the requirement of a hall height to width ratio of 1.6:3. When this condition is met, the ceiling is always father away from a listener than a wall. This means that the first reflection will then always come from one of the walls.

The desired acoustics of a concert hall from the point of view of a listener in the audience appears then to be that the direct sound should come cleanly through the hall, followed by the first reflected sound which comes from some vertical surface such as a side wall, followed by a host of smaller reflections which build up into a reverberant "bath" of sound coming from all directions in the room and appearing to last for an appropriate time without a significant change in its timbre as it decays away.

3.1.4 Stage Geometry Considerations

It has recently been established that the acoustic properties of the hall as observed by the performers is also of great importance in the average quality of the music that will be played in the hall and that these acoustic properties are somewhat different than those demanded by the listener. Generally, music performers need the immediate reflections from nearby walls to sense the presence and the music of the other performers and to keep the music together. However, they do not need as long a reverberation time as the audience. Jordan has shown that in preferred halls, performers hear a higher fraction of the sound in the first 50 ms then does the audience.

This means that not all of the sound of the orchestra should be radiated out to the audience but that some of it should be reflected immediately back to the performers themselves by appropriately placed reflecting surfaces or walls. However, after one or two of these reflections, the sound should have radiated out into the rest of the hall.

The detailed placement of these reflecting surfaces will depend on the actual seating arrangement of the performers and their styles of play as well as the tastes of the orchestra leader. For large halls and symphonic orchestras, the reflecting surfaces will generally be permanent fixtures or walls in the stage area.

3.2 Achieving Good Hall Acoustics

Achieving good concert hall acoustics in modest sized halls with adequate ceiling height is generally not a problem. However, the economics of staging music performances require large halls for symphonic orchestras. Furthermore, the ceiling of these halls cannot generally by high enough to allow the first reflection to be from a side wall for all members of the audience. The problem then becomes, how to deal with the sound reflections from the ceiling?

One solution would be to make the ceiling of sound absorbing material. This is the approach taken in typical modest sized lecture halls. However, the effect of covering the ceiling with sound absorbing material which completely eliminates ceiling reflections is to shorten the reverberation time. The larger the hall, the shorter will be the reverberation time for a completely sound absorbing ceiling. To obtain a sufficient reverberation time, large halls need the ceiling reflections.

3.2.1 Diffuse Reflector Ceilings

A new development in concert hall acoustics is the use of computers to find complex ceiling designs that break up the sound reflections over a broad frequency range and scatter the components in all directions rather than in the one direction that comes from sharp mirror image. The optical analogy would be that the ceiling looks like a white sheet of paper rather than a mirror. A sheet of good white paper and a mirror will reflect about the same fraction of light, but the mirror will produce an image source from its reflections (specular reflection) while the sheet of paper will not (difuse reflection). The sheet of paper is much easier to look at under a bright light than a looking glass mirror in which one sees the mirror image of the light. Similarly, a diffuse reflector for sound would produce a "softer" reflection than a broad flat surface giving specular reflection.

The general principle of a diffuse reflector is that the reflecting surface must have randomly spaced depressions of random depth, the spacings of the depressions and their depths covering the wavelengths of the waves to be dispersed. For light, this means only about 1 micron or a thousandth of a millimeter. Such a surface would appear flat unless looked at under a powerful microscope. For sound however, the spacing have to cover a range up to 3 or 4 meters to disperse all the important frequency components of a symphony orchestra.

A typical array of staggered blocks that achieves specular reflection of sound from a ceiling is shown in Fig. 3.3 (see Schroederer).

Figure 3.3 A Typical diffuse reflector ceiling, cross-sectional view. The pattern can have a repetition interval as shown.

With such a ceiling, there will be no discernible single reflection from the ceiling. Rather, the reflected sound will be dispersed in all directions with a great deal of it thrown to the side walls of the hall. The first discernible reflection will then be that which has had a single reflection from the side walls, the sound which scattered off the ceiling coming later from all directions and enveloping the listener.

Usually a computer is used to find an array of such depressions in a ceiling that will be adequately random. Such an approach was taken in the design of the Town Hall in Wellington, New Zealand which opened in 1983. Unlike many other new halls, this one was met with immediate critical acclaim.

The problem still remains though for very large halls; how to get a horizontal reflection within the first 30 ms so as to prevent an echo effect. An extreme example would be a performance of Aida in the Montreal Olympic Stadium. One could take the purist approach and say that such halls should never be used for serious music but people will pay for such extravaganzas and it is up to the acoustic engineering profession to provide the service of designing the best acoustics for the given situation. The solution to a problem of this scale will no doubt involve electroacoustical apparatus.

3.2.2 Electroacoustical Apparatus

One example of the use of electroacoustical apparatus in a large concert hall is the Royal Festival Hall in London which opened in 1951 with a volume of 27,000 m^3 and a seating capacity of 3000. This hall had a severe problem in that it had a short average reverberation time for such a large hall (1.45 sec). Also, it had no enhancement of the reverberation time for the bass. An electroacoustical system was installed which created artificial reverberation through speakers scattered throughout the ceiling. To create the impression of reverberant sounds, the speakers were fed by direct sound picked up from the stage and electronically processed to get the correct delays. The before and after reverberation curves for the hall are shown in Fig. 3.4.

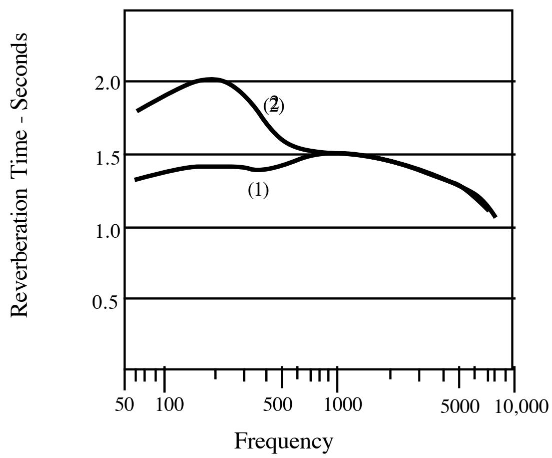


Figure 3.4 The reverberation times for different frequencies in the Royal Festival Hall of London. (1) before electroacoustical assistance (2) after electro-acoustical assistance.

This installation was not announced (the speakers were hidden in the ceiling) until audiences and critics noticed a substantial improvement in the hall acoustics.

Unfortunately a similar installation in a very large hall (Centenary Hall in Hochst, opened in 1963; volume 74,700 m^3, reverberation time without electroacoustics 1.2 sec, with electroacoustics 1.9 sec) did not yield satisfactory acoustics for symphonic performances and there is a tendency for critics, conductors and serious performers to deny the possibilities of electronic enhancement of the acoustics of concert halls.

Perhaps the reason for the failure on Century Hall is that the hall is too big to give satisfactory results by just modifying the reverberation times. What would be required would be to create some artificial sounds simulating horizontal reflections from side walls but arriving early enough not to give the echo effect. How this could be done without cluttering up the hall with suspended speakers is not clear but certainly the electronics technology now exists to create such effects. This is because of the tremendous impact of the so-called "digital revolution" on music and the availability of cheap powerful computers that can create a multitude of acoustical effects in a hall.

3.3 The Use of Computers in Concert Hall Acoustics

3.3.1 Diagnosing Concert Hall Acoustic Problems

One of the first uses of a computer to make a serious impact on the science of acoustics was by Schroederer in 1963 in analyzing the deficiencies of the New York Philharmonic Hall. He and a team from Bell Laboratories used a computer to generate precisely tailored tone bursts and to analyze the sound picked up in various seats in the hall from these tone bursts, included the seat which was regarded by the ushers (all students of the Juilliard School of Music in New York) as the best seat in the hall. By running comparison checks on the sound signals received at all these locations and using time and directional analysis of the differences, the major cause of the deficiencies was identified as the cloud reflector in the ceiling which absorbed far too much of the bass frequencies.

3.3.2 Determining Listener Preferences in Concert Hall Acoustics

The work on the New York Philharmonic Hall was carried out before computers became as cheap and as powerful as they are today. In the early 1970's, Schroederer (then a professor in Germany) again used the power of electronics and computers to do the first scientific comparison testing of audience preferences in concert halls on the level used by advertising agencies to determine preferences in consumer products. The actual techniques used in this work are of interest because they involve not only the computer but also technology of relevance to music recording in general.

The essence of scientific comparison testing of consumer preferences is to bring it down to a simple comparison of two items at a time, side by side and keeping score of the results. By using a mathematical technique called "multidimensional scaling" it is possible to find correlations of consumer preference with physical factors such as color, sweetness or shape of bottles.

In making such a comparison for concert halls this is a major problem; the concerts heard are often months apart, and the pieces played are not necessarily even similar. What is needed is the sort of pair comparison used in audio stores for the selection of loudspeakers, where the same music can be played with a very short interval on competing sets of equipment.

The difficulties of making such comparisons of concert halls are obvious with the distances and time requirementsof moving between different halls. It would be desirable to "bring the hall to the listener".

Schroederer realized that with modern technology there is a way that this can be done with reasonable success. The steps he used were:

1) Record a symphonic piece being played by musicians in an anechoic chamber.
2) Play this recording using large powerful loudspeakers on the stage of the hall to be studied. This effectively produces a "standard" symphony performance.
3) Record the sound produced in the hall using microphones in the ears of a dummy head.
4) Play the dummy head recording to a listener, also in an anechoic chamber, in such a way that the signal which went into an ear of the dummy goes into the corresponding ear of the listener.

The anechoic recording and playback ensure that only reflections in the concert hall are evaluated by the listener. It might be assumed that stereo headphones would be a simpler solution than the procedure in step 4 but headphones in general do not give a sense of realism: the sound seems to come from within the head of the listener because the stereo image is 'locked' to the listener's head, rotating with it as the head turns. To get a realistic stereo image without headphones, a technique called "holophonic sound projection" was used. The essential features of this technique are shown in Fig. 3.5.

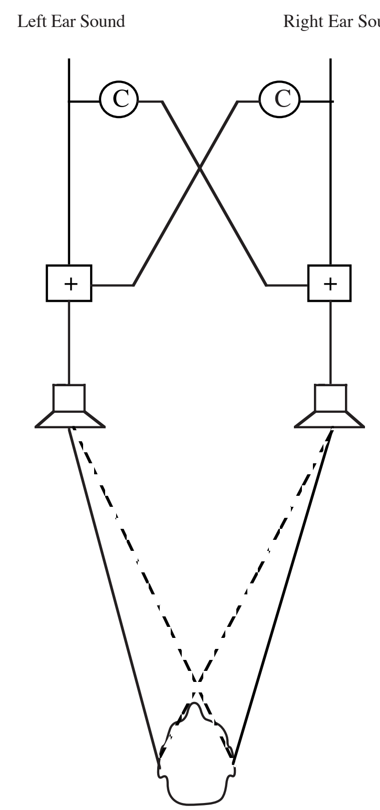


Figure 4.5 "Holophonic" stereo sound projection system. A portion of the signal that is sent to the left speaker ear and meant for the left ear is sent through an electronic circuit b(compensation filter C)) and added to the sound for the right speaker with an adder circuit (+) in such a way that it cancels the sound of the left speaker that arrives at the right ear. A similar circuit is used for the sound reaching the left ear from the right speaker.

In this arrangement, the playback signal which 'leaks' to the wrong ear by diffraction around the head is cancelled by a deliberately added antiphase signal in that channel. This requires tailoring both the amplitude and phase of the correcting signal. The result is said to be startlingly realistic. (Although such a system is now commercially marketed, it seems unlikely that it will become popular in home use; the controls are complicated, and the desired effect extends over only a very small region in a typical livingroom.)

With this technique concert halls can be compared in simple pair comparisons. The music recorded in one hall and the music recorded in another hall were both made available to an observer by a simple switch. By operating the switch the observer could judge the two samples against each other. Since it was exactly the same music in both cases, any differences in scoring was due to preference of the hall acoustics.

In Schroederer's study, 22 European halls were compared. The data were analyzed using multidimensional scaling. Correlation tests are carried out to see what acoustic parameters seem to be important in the consensus preference. The results showed a high correlation of preference with reverberation time. However, another more surprising factor appeared; a quantity called 'interaural coherence' is anticorrelated with consensus preference. Interaural coherence refers to the similarity of the sound at the two ears. In simple terms we prefer the actual waveforms of the sound arriving our two ears to be different.

The main factor which provides waveform dissimilarity at the two ears is the difference in arrival time (and hence phase) of the direct sound and the earlyreflections. Since wall and not ceiling reflections will be mainly responsible for providing binaural differences, this is another affirmation of the already mentioned preference for halls with ceilings more than half as high as the hall width. The difference now is that there is a scientific study confirming this fact and which also gives information about how much interaural difference is preferred.

3.3.3 Digital Processing of sound

Perhaps the greatest impact on the science and engineering of hall acoustics in the near future will come through the so-called "digital revolution" in music. Certainly it has had an impact on almost every other aspect of music, particularly music recording.

The basic principle of "digital music" is that the waveforms of the pressure oscillations making up sounds are measured and stored as numbers. More precisely, the voltage waveforms out of the microphones and their amplifiers are measured and stored. The music can then be restored later by using these numbers to recreate the original voltage waveforms as inputs to a stereo amplifier system and its loudspeakers.

The process of measuring the original waveform is called Analog to Digital conversion (AD), a standard technique in electrical engineering. It is usually done by a process of successive approximation. As an example illustrating the principle, suppose a voltage of 7.248 volts is to measured in a system where the voltage can be as high as 10 volts (See Fig. 3.6)

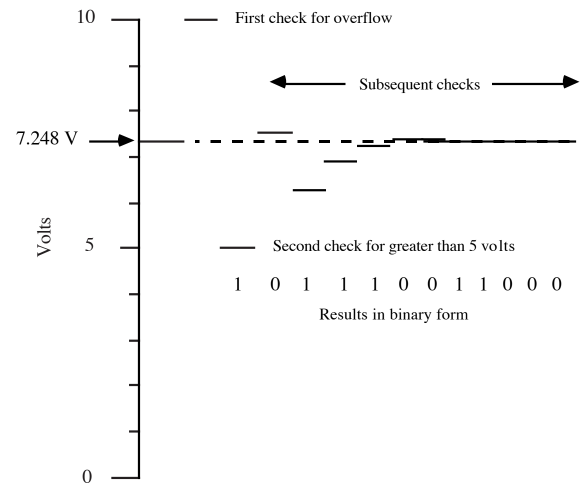


Fig. 3.6 Successive approximation measurement of a voltage. In the example shown, the voltage to be measured by the "digital" circuit is taken to be 7.248 Volts. The results to 12 bit accuracy is shown.

The first question asked of the electronic logic circuit is whether the voltage is greater than 10 volts. This can be done by a simple comparitor circuit with a yes-no answer, creating a 1 for a yes and a 0 for a no. In this case, if the answer is 1 than the circuit indicates an "overflow" at the input.

If the answer to the first question is a no, then the circuit next ask the same question again. The answer will now be no (0) and the circuit halves he difference, now downward, to 6.25 volts and again does a comparison. This sequence is repeated until the desired accuracy is obtained. In the example of 7.248 volts, the result would be 101110011000 (to an accuracy of 0.002 volts).

Thus twelve "bits" of information represent the measurement of the voltage. This is, of course, the measurement in "binary" arithmetic or arithmetic in base 2. To convert the number to the more familiar digital arithmetic (in base 10) the values of the bits are added as their values in base 10 (2048, 1024, 512, 256, 128, 64, 32, 16, 8, 4, 2 and 1 for the first twelve bits). The result is 2968 meaning that the voltage is 2968/4096 times 10 or 7.246 volts. (Remember that the accuracy of the measurement was only to 0.002 Volts).

Because it only involves a simple on-off storage, binary arithmetic is the form used in computers. The so-called "digital revolution" is therefore more of a "binary" revolution.

The operations involved in obtaining the binary measurement of a particular voltage in a waveform is therefore a very simple repetitive process, the sort of process for which computers are ideal. Modern logic circuits can do tens of million such operations a second and so a measurement to 12 bit accuracy of the type shown above can be carried out in a millionth of a second.

The pressures in a sound involve oscillations that can be heard by human beings at up to 20,000 per second. This sound is adequately measured for human hearing when it is sampled about 50,000 times per second. The accuracy required to cover the full dynamic range of sound that can be heard is about 1 part in 100,000. This requires about 16 bits in binary arithmetic. This has become the standard level of accuracy for the "digitization" of the waveform of audible sounds.

Ten bits in binary arithmetic is roughly a factor of 1000 (actually 1028). Twenty bits therefore represent about a million (actually 1056784). 16 bits actually represent a factor of 65536. This represents a dynamic range of 96 dB (20 times Log_{10} 65536 - remember it is pressure, not intensity that is being measured). Digital music is often quoted therefore as having a 96 dB dynamic range.

The operations involved in storing music waveforms as digital information is easy for modern electronics. However, the volume of information generated is enormous. At 16 bits 50,000 times per second, a million bits are generated in 1.25 seconds. A modern microcomputer holds about 1 million bytes, a byte being 8 bits, and so could only store about 10 seconds of music. The modern laser disk, which encodes the bits of the digital measurements of the sound as tiny specks under a transparent plastic covering, must contain almost 3,000,000,000 such specks for each hour of music.

The advantages of digital storage of music are obvious. Once the waveform has been saved as a set of numbers, there will be no deterioration of the stored waveform as long as the numbers are kept intact. Modern systems of storing numbers twice and checking numbers for errors due to loss of bits in the storage medium, allow virtually perfect storage of numbers. Laser disks store the music twice on different regions of the disk and the playback system continually compares the information with error checks so that even very dirty disks can give virtually error free return of the numbers encoded.

To recreate the waveform of a sound that has been digitally encoded, a device called an DA converter is used. This device simply adds voltages according to the binary information and sends the result to an output circuit as a voltage.

There is some controversy at present as to whether digital music is as good as many claims that are made for it. Many people seem to think that it is brighter or even more grating on the senses than a good analog recording while the engineers point out that the digital process does not add anything to (or take anything away from) the music that can be detected by human beings.

Perhaps the source of this conflict is the way the technique of digital music has been used. The tremendous range of power in digital music (up to 96 dB) compared to analog music (to about 65 dB) is quite often deliberated demonstrated by having a loud sound come out suddenly from a very quite background. Unfortunately, people do not seem to like such sudden intensities in the sounds that they hear. In real life such a sudden change in sound level is usually a signal for disaster. Maybe what is required in digital music is to make sure that there is some "floor level" of background sound, such as audience noise in the recording hall, that will substitute for the inherent background hiss of about 40 dB in analog media that are being used to play 100 dB music. Perhaps to resolve this conflict, an accurate pair comparison set of experiments should be carried out of the sort used by Schroederer to compare concert halls.

The power of the digital system is not just in the permanent, safe storage of sounds. The numbers played back do not have to be the same numbers as were originally stored. The information originally stored can be "processed" by a computer. Thus frequencies can be enhanced or removed. Computer methods exist for looking for patterns in numbers and removing or changing those patterns. For example, the characteristic pattern of the sound of a scratch on an old gramophone recording can be detected and removed. Furthermore, artificial interaural differences can be created so that a "stereo" record can be made form an old monaural recording. Such techniques are already highly developed for pictures such as those sent to earth from distant spacecraft where the incoming picture information, sent as digital information to overcome errors in the long-distance transmission of very weak signals, is processed by a computer. On a more mundane level, old movies are being restored to even better than original quality and sometimes even artificially colored. It is possible that soon there may be simple hardware and computer programs for personal computers at home to fix up the sound on old records and resave it in digital form.

This ability to modify the sounds by operating on the stored numbers leads to the possibilities of electronic enhancement of concert hall acoustics. Already there are systems for digitizing the analog sound from a home stereo system and processing it to reproduce sounds which have the characteristics of reverberant sound (multiple repetitions of the original sound with constantly changing frequency characteristics) and playing these sounds through smaller speakers scattered throughout the room. The results can quite significantly enhance the acoustics of an average living room. Systems for large concert halls would necessarily have to be much more complex, involving no doubt computer operation and control. However, the cost of such systems could easily be justified if the results were significant. As more and more is learned about how human beings respond to music in a concert hall, the engineering of such systems becomes more feasible. Schroederer himself, somewhat wistfully it seems, claims that someday there might even be intelligible public address systems.

3.3.4 General Uses of Computers in Music

There are, of course, many other areas in which computers are having an impact on music. To give a perspective of this rapidly changing field, some of the present uses are listed here.

3.3.4a Computer Generation of Music

The analog signals (actual voltages which vary in a pattern with time) that produce sounds do not have to use information that comes from the digitization of original sound; the number cans be easily generated by a computer itself. Thus waveforms that are impossible to create with an acoustic musical instrument, or even difficult to create directly with an electronic circuit can be created by having a computer generate the numbers corresponding to a desired waveform. This technique is used a great deal in research in the hearing processes of humans and animals and was the basis of the technique used by Schroederer to diagnose the problems in the New York Philharmonic Hall. One interesting application of this technique is that by the McGill Recording Studio which has produced a packaged set of digitizations of the sounds of standard musical instruments as starting points for people wishing to use a computer in any way to simulate sounds. As with any new musical instrument, how this technique will be used in the future depends a great deal on the experience, imagination and creativity of the artists who pick it up. There is no doubt, however, that a whole new set of possibilities have been added to the creation of music, research into hearing and even the analysis of the behavior of musical instruments and music halls.

3.3.4b Computer Operation of Music Systems

One of the possible application of computers in music is the operation of classical musical instruments and music hall acoustic adjustments for special effects. An example is the instrumentation of the pipe organ in the Sydney concert hall to record the actual operation of the keys in a live performance and to then exactly duplicate this operation on demand. While the use of such a system to give an organ "recital" seems to be the stuff of horror movie films, the possibilities for analyzing and teaching of organ playing technique are intriguing.

Another use of computers is in adjusting the moveable panels which are becoming more and more a feature of modern multipurpose music halls. Such a system has been installed in the new Roy Thompson Hall in Toronto. The advantages of computer control is that, once a pattern for the panels has been developed for a particular type of performance, this pattern can be recorded and easily repeated at a later time or even altered by a simple computer program to produce effects that might lead to even better acoustics.

3.3.4c Analysis of Sound

Once a sound has been stored in digital form, it is possible to use a variety of mathematical techniques to analysis the sound for particular characteristics. (Again, this was the basis of the technique used by Schroederer in analyzing the acoustics of the New York Philharmonic Hall). Two of the techniques of particular importance are "Fourier Analysis" and "Transfer Function Analysis". Both are highly mathematical techniques that depend critically on the availability of the information in digital form and high-speed computing.

Transfer Function Analysis is the comparison of two inputs for amplitude and phase relationships (how one of the inputs could be mathematically transformed so as be the exact duplicate of the other). This technique is of great importance in analyzing the possibilities of the stereo image that would be created by the two sounds meant for the two ears. It was the technique used by Schroederer in determining that it was interaural coherence (or lack of significant difference in the sound of the two inputs) that led to audience lack of preference of the sound of certain halls.

The technique of Fourier Analysis, which is mathematically related to Transfer Function Analysis but a little simpler, presents a detailed spectrum of the frequencies in a monaural sound. This spectrum is related to the timbre of a musical note and is the pattern that is most easily related to the quality of a sound that is heard. When a picture of the Fourier Analysis of a sound is presented, it is very easy to relate the picture to the sound being heard; harsh, high pitched sounds fill the upper part of a Fourier spectrum, while mellow, low pitched sounds fill the lower part;

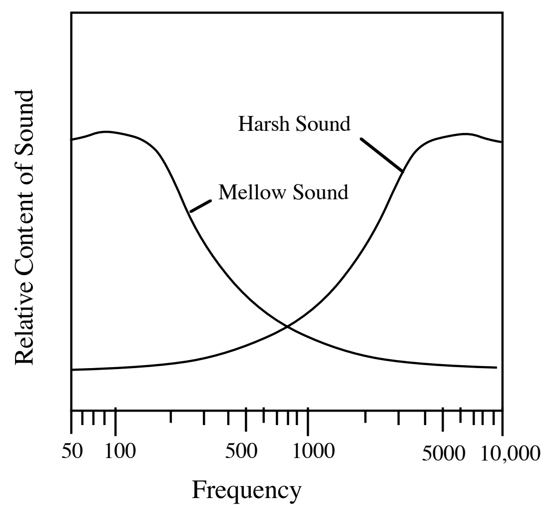


Figure 3.7 A typical result of the Fourier Analysis of sounds. The sound with the spectrum peaked in the low frequency region will be more mellow than the sound with the spectrum peaked in the high frequency region.

With modern computing techniques it is possible to digitize the output of a microphone and to do a Fourier Analysis (sometimes called a Fourier Transform) for 500 or more frequencies in less than 1/20th of a second. Furthermore, the electronics to do this can be easily mounted at modest cost on an expansion board for a personal computer and the personal computer used to display the results. This allows a continuous display of the frequency spectrum of the sounds as they are being heard by a listener and rapid learning of the connection between the patterns appearing on the screen and the sounds being heard.

This system has great potential in such areas as the training of deaf children how to speak. By trying to mimic the patterns on the screen for sounds correctly produced by a teacher or even those of a recording, the child can learn the muscle controls needed of the tongue, throat and face to produce real speech. It also has great potential in the teaching of music performance. The frequencies produced in singing, for example, can be displayed as they are being produced. Furthermore, the actual pitch of the note produced could be displayed, making it considerably easier for the student to learn to sing "in tune" as well as to produce the right timbre of note by noting the presence (or absence) of significant frequency components. As an example, the "operatic format", a cluster of frequencies around 4000 Hz that are sounded when a singer adopts an operatic style, can be clearly seen in a Fourier analysis of the sound. Developing the constriction in the vocal tract that produces this formant is a very difficult art that could be considerably aided by evidence on a screen that one has indeed come close to producing it.

3.3.4d Analyzing Human Auditory Systems

How the human auditory system decodes the frequency pattern of sounds falling on the ears into a recognizable pattern and responds to this pattern in about a tenth of a second, is one of the great mysteries of psychophysics. The question is not only of great importance in the theory of music but also in such practical questions as the nature and cause of hearing deficiencies associated with neural damage in children. Research in this area has begun to use computers programmed to behave like certain models of what the central nervous system might be doing to the nerve pulses from the cochlea. If the computer makes the same errors as those of a child with a certain hearing defect, then one has some indication that the computer is processing the incoming data in a way similar to that of the nervous system.

The work in this field is extremely complex, involving knowledge of biology, psychology, linguistics, mathematics, artificial intelligence and the physics of sound. The research is beset with many problems associated with the mysteries of how the brain itself is physically constructed. However, some overall principles seem to be emerging. These generally follow principles discovered many years ago in electrical engineering. There seem to be differential amplifier set-ups wherever possible; two eyes, two ears, two vestibular organs etc. It has been known in electrical engineering for a long time that such a system based on the comparison of two similar input devices is inherently more capable of picking out the desired signal from noise on the inputs. In the case of the ears, the determination of the direction of a sound source seems to be based on a highly developed mechanism for comparing the inputs from the two ears.

Another remarkable feature of the human central nervous system is one that has only recently become generally accepted in modern computer networks (but again which has been known as a good general principle for some time by electrical engineers); that of so-called "distributed computing". In such a scheme, the most effective use of computing power is achieved by distributing the load appropriately among the various elements in the system. Thus small jobs are best done by small local computers (often disguised as "smart-terminals") while the big powerful number crunchers are only called up by these local devices when the memory capacity, larger programs and sheer brute force of the larger computers are necessary. The central nervous system of humans seems to be designed along this same principle. Perhaps the best established example is color encoding by the eyes. There appears to be a set of microencoders just behind the retina to preprocess the color information generated by the cones in the fovea before sending this information to the brain for more sophisticated pattern recognition. Another well established example is the tight neural connection between the eyes and the vestibular organs. Most of the time the eyes and the vestibular organs work together, with no bothering of the brain, to control the position of the eye during head movements so that we do not have blurred vision whenever we move our heads. It is only when there is some confusion between the two such as can be produced by blindfolding a person when on a revolving chair, that the brain will be called in to clear up the giddiness and general confusion that has be created. (For more information on this matter, one of the world centers of authority in this field is the McGill School in Aviation and Space Medicine.)

The systems used for aural decoding are not so well understood. All that can be said with some certainly is that they must be based on some incredible local timing comparison microencoders. What they are, or even where they might be in the central nervous system (how far up the ladder to the central cognitive brain structure) is not even known. There is certainly a great deal of room for research here and computers are being used more and more as tools in this research. With the general invasion of computers into our modern society, it is perhaps not surprising that they invaded modern musical acoustics as well.

References

Technical Papers

Schroederer, M.R.; Progress in Architectural Acoustics and Artificial Reverberation; Concert Hall Acoustics and Number Theory"; J. Audio Eng. Soc. **32** (1984) pp 144-203

Schroederer, M.R.; "Towards Better Acoustics in Concert Halls"; Physics Today, October 1980 pp 24-30

D'Antonio, P. and Konnert, J.H.; "The reflection Phase Grating Diffuser: Design Theory and Application"; J. Audio Eng. Soc. **32** (1984) pp 228-238

Jordan, V.L.; "Acoustical Criteria and Acoustical Qualities of Concert Hall" ; "MUSIC ROOM AND ACOUSTICS", A collection of papers given at a seminar organized by the Royal Institute of Technology in Stockholm by the

Royal Swedish Academy of Music, Center for Human technology and the Center for Speech Communication Research and Musical Acoustics, April 1975.

Knudsen, V.O.; "Achitectural Acoustics"; Scientific American, November, 1978

Textbooks

Meyer, J., "Physics and the Performance of Music"; Verlag Das Musikinstrument Frankfurt/Main 1978 pp 103-135

Hall, E.; "Musical Acoustics", Wadworth Publishing Co. 1980 (Chapter 15)

Backus, J., "The Acoustical Foundations of Music"; 2nd Edition, W.W. Norton and Co. 1977 (Chapter 9)

Benade, A.H.; "Fundamentals of Musical Acoustics"; Oxford University Press, 1976 (Chapter 12)

Winckel, F., "Music Sound and Sensation", Dover Publications, 1967, (Chapter IV)

Olsen, H.F., "Music Physics and Engineering", 2nd edition, 1967, pp 266-324

Exercises and Discussion Topics

1. What are the desirable features of concert hall acoustics and how are they related to the geometrical properties of the hall? How do these features vary with type of music or performance. (Discus the relationship for a least three distinct types of music or performances)
2. Discuss the acoustic "tools" an architect has available to adjust and fine tune the acoustics of a hall. Explain in as short a statement as possible what these tools are and what they can be expected to accomplish.
3. Outline the steps used by Schroederer for a scientific evaluation of the preference in concert hall acoustics. Explain the reason for the various other steps taken by Schroederer.
4. Discuss in very general terms the acoustic properties of the various types of acoustic environments in which various types of performed music may be heard; in particular concert halls, opera halls, theaters, chamber music halls, churches and open air environments.
5. Show that the binary number in the text of this chapter actually represents the voltage value given.
6. Discuss the roles of computers in modern concert hall acoustics.
7. Discuss the possible roles of computers in music and acoustic research in general.

CHAPTER 4

THE SOUND OF A POINT SOURCE IN A ROOM

Just as sports seem to be the playful use of mental and physical skills that have developed to help us survive as a species, listening to music seems to be the playful use of the aural nervous system that has also been extremely important in primitive survival. Because of its importance to survival of the species, this sensory system has evolved so as to get the maximum possible information from sound at the ears. The physics of how sound propagates from a single source to the ears can help in understanding what this system does and what it is capable of doing.

One of the most important aspects of a sound is that it involves a pressure oscillation in a medium. The medium can be a gas such as air, a liquid such as that in the cochlea of the ear, or a solid such as steel or the bone matter in the head or, to take an extreme case, nuclear matter of the sort that exists in neutron stars. All that is required for sound propagation is that the medium have its mass somewhat evenly distributed and that it be elastic (i.e able to bounce back from a temporary distortion). These conditions exist to a high degree in the air in a room.

The pressure oscillations involved in sound are overpressures and underpressures relative to the normal (quiet) atmospheric pressure. The overpressures and the underpressures in sound on the average cancel to zero, leaving the average pressure to be the normal atmospheric pressure of quiet air.

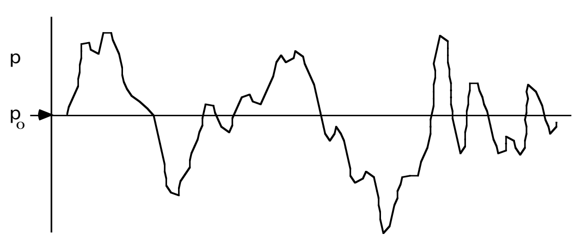


Figure 4.1 Example of the type of pressure oscillations that occur in sound.

Listening to sound is responding to these pressure oscillations. Hearing different sounds in different positions in a room means that the pattern of pressure oscillations is different at these different positions. In particular, a large part of the pleasure of listening to music seems to be in responding to the difference in the patterns of pressure oscillations at the positions of the two ears.

For normal room sounds, the overpressures and underpressures are very small compared to the normal atmospheric pressure of about 100kPa. A sound of 94 dB, corresponding to about the loudest sound you would want to hear from an orchestra, involves a pressure oscillation of amplitude about 1 Pa, i.e. between 100,001 Pa and 99,999 Pa. Pressure oscillations at the threshold of hearing are miniscule; about 20 millionths of a Pascal. A sensory system that can respond to such pressure changes, and at the same time look for differences at the two ears, has to be very highly developed indeed.

The aural sensory system, in responding to the pattern of pressure oscillations at the two ears, uses the information in these patterns to discern a source for the sounds. In also uses information in these patterns to discern things about the geometry of the room. This means that the brain must be capable of discerning the differences of the sound at the ears due to different ways the sound can be propagated from the source to the ears. To understand what the brain is doing, it is therefore important to understand how sound pressure oscillations propagate from a source to various points in a room.

A thorough coverage of the physics of sound propagation would include the mathematical description of wave motion in general and the derivation of the wave equations in a medium from the physical properties of that medium. Most good books on acoustics include this. However, many do not include a good introduction to the subject in words and pictures that help in understanding what the equations are describing. As an introduction, what will be presented here will be the words and pictures without the mathematics; the "how" of sound propagation will be presented with the "why" only being answered to the extent that is possible in words and pictures. For the benefit of those who may be interested in the basic equations of wave motion, they are summarized in the appendix to this chapter.

To understand anything new it is best to start with the simplest possible example. The simplest possible example of sound propagation is that from a flat wall which suddenly starts moving against the air in front of it. What will propagate from such a wall is a pulse of sound in the form of a pressure "wall" or "plane wave", the "plane" being referred to being the plane of the edge of the pressure zone and which is parallel to the wall but at some distance from it. This plane will be moving directly away from the wall in a direction perpendicular to it and the sound pulse associated with it will be the "direct" sound from the wall. What follows is a description of this direct sound and how it propagates.

4.1 The Direct Sound Wave

4.1.1 Plane Sound Waves

Suppose that a wall, which was at rest, suddenly starts moving at a uniform speed forward. Immediately, the air in front of the wall resists this motion and whatever is pushing the wall will have to apply a force to sustain the motion. In other words, the air will exert an overpressure on the wall, resisting the wall's motion.

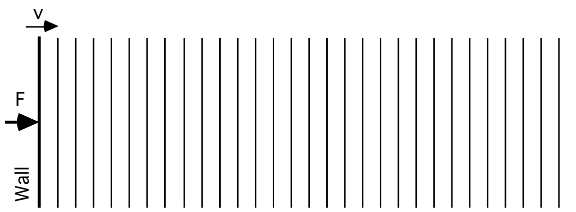


Figure 4.2 Pictorial representation of the air in front of a moving wall. The vertical lines are used to indicate position of the air at the instant of the picture. They might be thought of as very thin sheets hanging vertically and moving with the air as it moves. At the instant shown, the wall has just started to move at velocity *v*. At the start of the wall motion the air immediately resists the motion, requiring a force *F* to sustain it.

The overpressure of the air immediately in front of the wall will also push forward the air further from the wall and compress it as well. If the motion of the wall persists, there will soon be a region of compressed air extending in front of the wall;

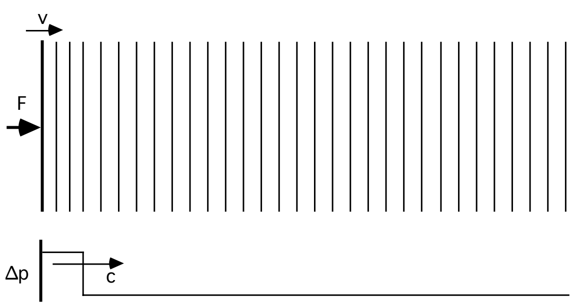


Figure 4.3 At the instant shown, the wall has moved forward and compressed the three region of air in front of it designated by the three vertical lines. This overpressure is shown on the pressure graph beneath the pictorial representation of the air. The air ahead of this region is still unaffected.

Here is where a very important phenomenon in sound propagation arises. For reasons related to the physics of how air moves under compression, the edge of the region of compression of the air will propagate forward at a very definite speed, independent of the speed of the wall motion. This speed is indicated as *c* in the diagram.

At a time later corresponding to twice the time for the diagram of Fig. 4.3, the region of compression will extend twice as far;

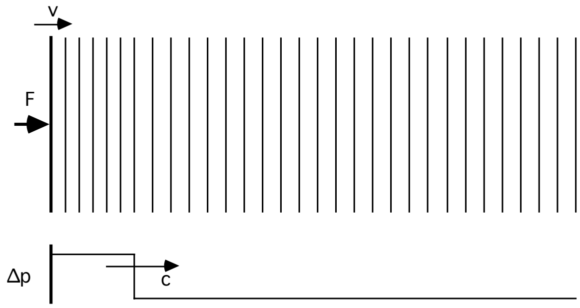


Figure 4.4 At the instant shown, the wall has been moving for twice as long as in fig. 4.3 and has compressed a region of air in front of it which is twice as thick as in fig, 4.3. This overpressure is shown on the pressure graph beneath the pictorial representation of the air. The regions ahead of the region of overpressure is still unaffected.

Now suppose the motion of the wall suddenly stops at the instant shown in fig. 4.4. Here arises yet another important phenomenon related to the physics of how air moves under pressure. It turns out that the momentum of the moving air just in front of the moving wall is just sufficient to cause it to move forward away from the wall the right amount to exactly relieve the overpressure. In other words, as soon as the wall stops, the air stops pushing against the wall and a region of normal pressure develops. This is shown in fig. 4.5

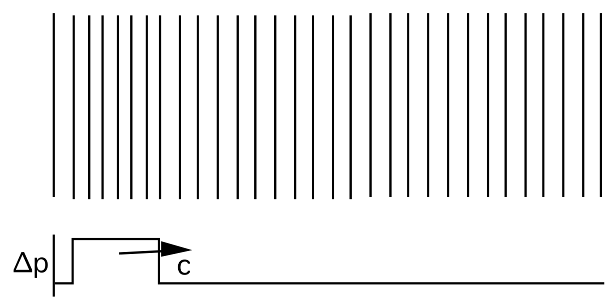


Figure 4.5 The air distribution and its overpressure a short interval after the wall has stopped moving. The overpressure is shown on the pressure graph beneath the pictorial representation of the air. The region of overpressure has the same extent as in fig. 4.4 but has moved away from the wall. The region between the overpressure region and the wall has returned to normal pressure. The region ahead of the overpressure region is still unaffected.

As time progresses, the region of overpressure will propagate forward at the speed of sound c;

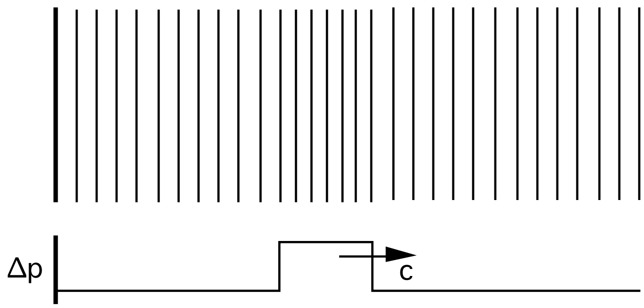


Figure 4.6 At the instant shown, the overpressure region has moved further to the right. It will continue to propagate in this fashion at a very definite speed; the speed of sound.

Now imagine a person to the extreme right in fig. 4.6. Such a person will experience an overpressure as the overpressure region passes by. The time it takes for the overpressure region to pass by is exactly the time it took to create the overpressure region in the first place, that is the time for which the wall was moving. In other words, the sound due to this overpressure will be perceived as lasting exactly as long as did the wall movement. Thus the person hears the movement of the wall but delayed by the time it takes for the leading edge of the overpressure region to reach the ear. For a distance from the wall of, say, 5 m at c = 343m/s this will be 5/343 = 14.6 ms.

If the wall moved originally with twice the speed, taking half the time to complete its motion, the force required for the wall movement would be twice as great, the region of compression would be half as wide and have twice the overpressure but the region would still propagate forward at the same speed (see fig. 4.7)

The same sort of thing happens when the wall moves backward, away from the air in front of it (fig. 4.8). Suppose, from its original position, the wall started to suddenly move backwards away from the air at a velocity v. This would immediately create a vacuum in front of the wall and the air in front of the wall would start to move to fill this vacuum. The region of underpressure would extend forward an amount that depended on how long the wall had been moving backward. When the wall stopped moving, the momentum of the thinned out air (now moving backward) would cause it to pile up in front of the now stationary wall. Again from the physics of the motion of air under pressure, in this case actually a vacuum, the air will move just the right amount to bring itself to a stationary state at normal pressure.

Now there will be a vacuum pulse moving away from the wall and an observer to the right will hear a negative sound pulse as coming from the wall. Even though the motion that caused this pulse was to the left, and all air motion involved in the sound pulse is to the left, the sound pulse actually travels to the right.

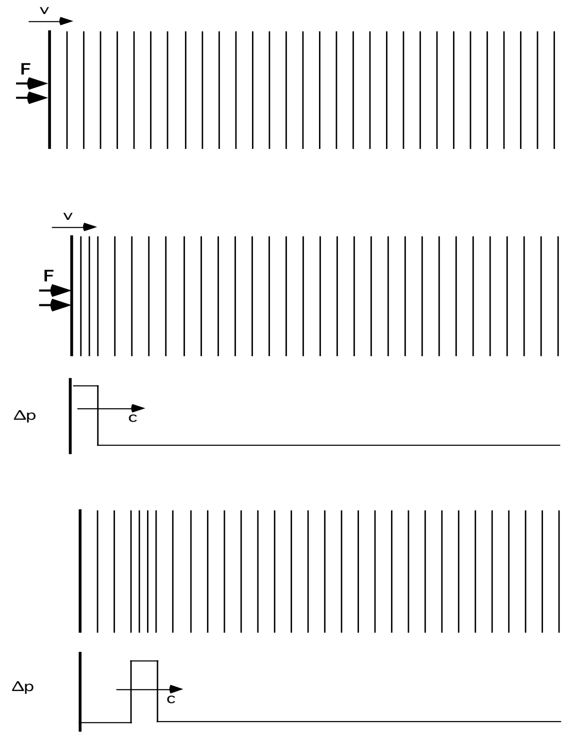


Figure 4.7 Diagrams of air motion with a wall moving at twice the speed for half the time of Figures 4.2 to 4.6

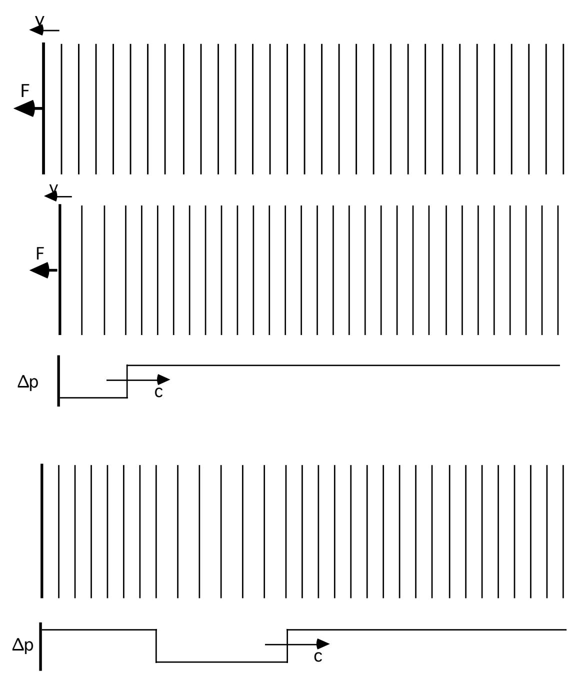


Figure 4.8 Diagrams of air motion with a wall moving backwards, away from the air. Note that the vacuum region still propagates forward away from the wall.

This is the remarkable thing about sound and is what makes it such a valuable aid in determining the direction and distance of any disturbance of the air by such things as a breaking twig or a falling stone. All disturbances that will cause a local overpressure or underpressure will be propagated directly away from the source at the same definite speed *c*.

In passing, it is perhaps interesting to note that there is an obvious limit to the propagation of a disturbance in this fashion. If the wall were to move forward faster than the speed of propagation of the disturbance, then the region of compression would never be able to get away from the wall. What this means is that all the air just masses against the wall, soon forming an insurmountable barrier to the wall motion. This phenomena does occur when objects move through air faster than the speed of sound and was first experienced by dive-bomber pilots in the second word war when, in a steep dive, their planes sometimes reached the speed of sound. The pilots described their experience as being similar to running into a brick wall, hence the origin of the expression "sound barrier" to such motion.

A related phenomenon occurs with motion away from the air. Such motion can occur in high power ultrasonics such as used in ultrasonic cleaning. Here a surface is made to move away from the medium (usually a liquid) at a speed greater than that of the propagation of the sound wave. What this means is that the medium in front of the surface cannot move fast enough to fill in the void created by the surface movement and a real cavity (complete vacuum) forms in the medium. This very high vacuum tends to explode any dirt off the surface of the moving wall and is the principle behind ultrasonic cleaning. The phenomenon is referred to as "cavitation" and it is very important in the medical uses of ultrasonics for diagnostic purposes that the level of the ultrasound be kept well below where this will occur.

For ordinary sound in air, the velocities of the moving surfaces are always very, very small compared to the speed of sound.

3.1.2 Analogy With Water Waves

Perhaps the most familiar example we have in nature of the propagation of a local disturbance is the propagation of a water wave such as that formed by a falling stone. A water wave that would be analogous to the sound pulse from a moving wall would be one generated by a moving vertical wall in water. If the motion of water piled up in front of such a wall was similar to the motion of air, the wall would build up water in front of it as shown in fig. 4.9.

Figure 4.9 Diagram representing a water wave creating by a moving vertical wall in water if water behaved in a fashion exactly analogous to the elastic movement of air.

When the wall stopped moving, a "wave" would propagate away from the wall in a fashion similar to that of the sound pulse (fig. 4.10).

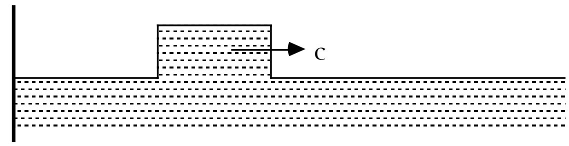


Figure 4.10 Diagram representing the water wave, starting in figure 4.9, after the wall has stopped and the wave has moved away from the wall. Again, the movement of water is assumed to be exactly analogous to that of the elastic movement of air.

The analogy with the sound pulse is not perfect however. Water in front of the wall does not move like air because it cannot be compressed; it has to move vertically along the wall. (Here we are not considering sound waves in water but ordinary water waves on the surface.) The shape of the wave that would actually be produced is shown in fig. 4.11.

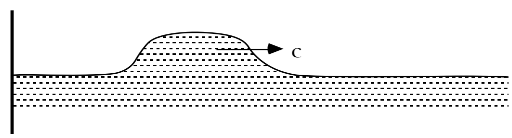


Figure 4.11 Diagram representing the actual water wave that would be created by a movement of a vertical wall in water. Again, the movement of water is assumed to be exactly analogous to that of the elastic movement of air.

As the water wave propagates away from the wall, it loses even this initial shape by flattening out (fig. 4.12).

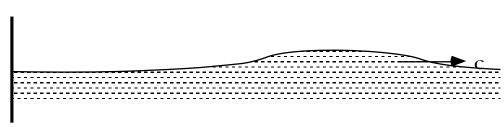


Figure 4.12 Diagram representing the water wave of fig. 4.11 after it has moved further to the right.

For example the initial steep water wave caused by an underwater movement, such as from an earthquake or a volcanic explosion, will spread out as it propagates and, after crossing an ocean, will appear at the shore as just a gradual rising of the water level similar to that of a tide. Such phenomena are therefore referred to as "tidal waves".

The reason for this behavior of water is that the physics of water flow from one elevation to another is different from that of the elastic air flow from a region of overpressure to a region of underpressure. (In technical terms, there is "shear flow" involved). This can also occur in air flow from an overpressure region if the overpressures are high such as in a dynamite blast or a lightning flash. The sound close to a lightning flash is a very sharp loud crack while the sound of a single distant flash is the familiar rolling thunder.

For the overpressures that occur at sound levels that do not damage the ears, the flow will be the elastic flow that leads to preservation of the sharp boundaries between the pressure zones and hence the preservation of the original sharp nature of the pressure disturbance as it passes an observer. An initial sudden movement therefore produces a sharp pulse of sound at a distance. The speed of propagation of sound through a medium is therefore one of the most important properties of sound.

4.1.3 Speed of Propagation of Sound in Air

From the physics of gases and their motions, the velocity of sound in a gas can be shown to be given by the equation

$$c = \sqrt{\frac{\gamma p_o}{\rho}}$$

where c is the velocity of sound, g is the ratio of specific heat of the gas at constant pressure to the specific heat of the gas at constant volume, p_o is the normal (quiet) pressure of the gas and r is its density.

The factor γp_o comes from the compressibility of a gas. The γ factor itself comes into the equation because sound going through the air is a so-called adiabatic process. By this it is meant that the passage of the sound wave through the air is so fast that heat produced by the overpressure cannot leave the gas it is in and is returned as energy to that same gas when the pressure wave has passed. This essentially raises the pressure that is necessary to compress a gas by a given amount. The value of γ for air is 1.40372.

The factor p_o is just the value of normal atmospheric pressure (101,325 Pascals). The factor ρ ιs the air density which at 20 degrees C is 1.205 kgm per cubic meter.

From the equation it can be seen that the velocity of sound increases with pressure or "springiness" but decreases with density. The connection between pressure and density in normal air is such that when the pressure is increased, the density increases in strict proportion; air under twice as much pressure being twice as dense. Thus air under different pressures has the same sound velocity. However, helium at the same pressure as air has a sound velocity considerably greater because of its lighter density.

When all of these figures are put into the equation, one gets a theoretical sound velocity for air at 20 degrees C and normal atmospheric pressure of

$$c = 343.6 \text{ m/s}$$

The actual figure for the velocity of sound in dry air at 20 degrees C and normal atmospheric pressure (101,325 Pascals) for a 1000 Hz sound wave is;

$$c = 343.562 \text{ m/s}$$

The agreement between the theoretical and observed values for the velocity of sound indicate the high degree to which the physics of sound is understood. However, like any medium for sound, air is not perfectly elastic. Air viscosity and non-adiabaticity at high overpressures can lead to deviations in its sound velocity. For very large waves (such as those near the vicinity of an explosion) the velocity can be considerably higher but for sound waves in the sound power range that do not damage human ears, the velocity is constant to within a few parts per million for all sound power levels.

Also the variation of sound velocity with frequency is very small. Over the useful range of frequencies for human hearing, the values are given in table 4.1.

Frequency (Hz)	Deviation of c from value at 1000 Hz (parts per million)
100	-30
200	-10
400	-3
1250	0
4000	+5
10,000	+10

Table 4.1 Deviation of the velocity of sound from the value at 1000 Hz in normal air at various frequencies.

The physical model of sound passing through air being that of a wave passing through a purely elastic medium of uniform density is therefore very good. The significant variations in the velocity of sound come from varying physical properties of the air. For example, the variation with humidity of the air is given in table 4.2.

Humidity (%)	Deviation of c for 1000 Hz Tone (parts per million)
0	0
10	73
20	415
30	775
40	1136
50	1500
60	1860
70	2230
80	2590
90	2960
100	3320 (About 0.3 %)

Table 4.2 Deviation of the velocity of sound from the value at 1000 Hz in normal air at various frequencies.

Taking into account that people with well developed musical abilities can detect a 0.1% to 0.2% change in frequency and that the frequency of many wind instruments depends directly on the velocity of sound in the air contained in them, we see that this humidity factor is on the verge of being noticeable.

At first it might seem strange that humidity would increase the velocity of sound since humid air is slightly heavier than dry air. However, the γ factor for humid air is higher than for dry air and this is the predominate cause of the velocity change for humid air.

The most significant factor in the variation of sound velocity in ordinary acoustic environments is the variation with temperature. This is because for a constant pressure the density of a gas decreases as the temperature increases. The density of an ideal gas is inversely proportional to the absolute temperature of the gas. (To get the absolute temperature of a gas, you add 273 to the temperature in degrees C). Thus the velocity of sound in a gas will be proportional to the square root of this temperature. For example, the absolute temperature of a gas at 20 degrees C is 293K. At 0 degrees C it is 273K. The ratio of the velocity of sound at 0C to that at 20C is therefore;

$$\frac{c_{0^o}}{c_{20^o}} = \sqrt{\frac{273}{293}} = 0.965$$

corresponding to a 3 1/2 % drop in velocity from 20C to 0C. In a wind instrument this would correspond to about a quarter tone drop in frequency, an easily detectable change for a musician. Many musical instruments have mechanisms for adjusting for such frequency changes but fixed instruments such as pipe organs in a church suffer severe tuning problems if the room temperature is not correct.

4.1.4 The Connection Between Pressure and Air Velocity in a Sound Wave

From the diagrams of section 4.1.1 it is clear that there is a relationship between the velocity of the moving wall and the pressure that builds up in the air in front of it. The pressure generated in the air was in proportion to the velocity of the wall generating it. What was not so clear is that the same relationship exists for the air in the sound pressure pulse itself. So show this, diagrams representing three successive instants in the pulse propagation are shown together in fig. 4.13. In this diagram it can be seen that the only air that is moving is the air in the region of overpressure. The rest of the air is stationary.

The same phenomenon occurs for the vacuum wave shown in fig. 4.14 except that in the region of underpressure the air moves backwards towards the wall.

The physics that gives the equation for the speed of propagation of sound also gives the relationship between sound pressure and air velocity. The two are strictly proportional to each other according to the equation

$$pressure = \frac{\gamma p_o}{c} \text{ x } velocity = 413 \text{ x } velocity$$

(for dry air at normal room conditions)

As an example, consider a wall which moved forward at 1 m/s (about a normal walking speed) but only for 1 ms so that its total movement was 1 mm. This, of course, would create a pressure pulse that lasted 1 ms and, travelling at 344 m/s, would cover a region of 34.4 cm in the air in front of the wall. From the above equation, the overpressure in this pulse would be 413 Pa. A wall moving backwards at the same speed would create a vacuum pulse of –413 Pa

4.1.5 Power in a Sound Wave

The relationship between air velocity and sound pressure is another very important property of sound. It means that there is real power or transport of energy associated with sound propagation. This is because power is the product of a force and a velocity. If a force exerted on an object does not move that object, then there is no work done and therefore no power involved. Also, if an object moves but requires no force to keep it moving, again no work is done and no power is involved. Only if a force is associated with a motion is there work being done and the power is then the rate at which this work is being done. This rate is simply the force times the velocity of movement being caused by the force. In the case of the sound propagation from a moving wall, the power involved in the generation of the sound is the force required to move the wall multiplied by the wall velocity.

The force required to move the wall is that required to overcome the effect of the overpressure (or underpressure) of the air on the wall. If the wall is being moved to the right, then a force to the right is required to overcome the overpressure. If the wall is being moved toward the left, then a force to the left is required to overcome the underpressure of the air. In either case, the force is in the direction of the velocity and work must be done to sustain the motion.

The force per square meter of wall is therefore just the air overpressure or underpressure. The power involved per square meter of wall is therefore the air pressure multiplied by the wall velocity.

The concept of power per square meter is very important in acoustics. It is defined as Intensity and by international agreement has the symbol *I* reserved for it. In the simple example of the moving wall, the wall is doing work on the air if front of it with an intensity equal to the pressure of the air times the wall velocity.

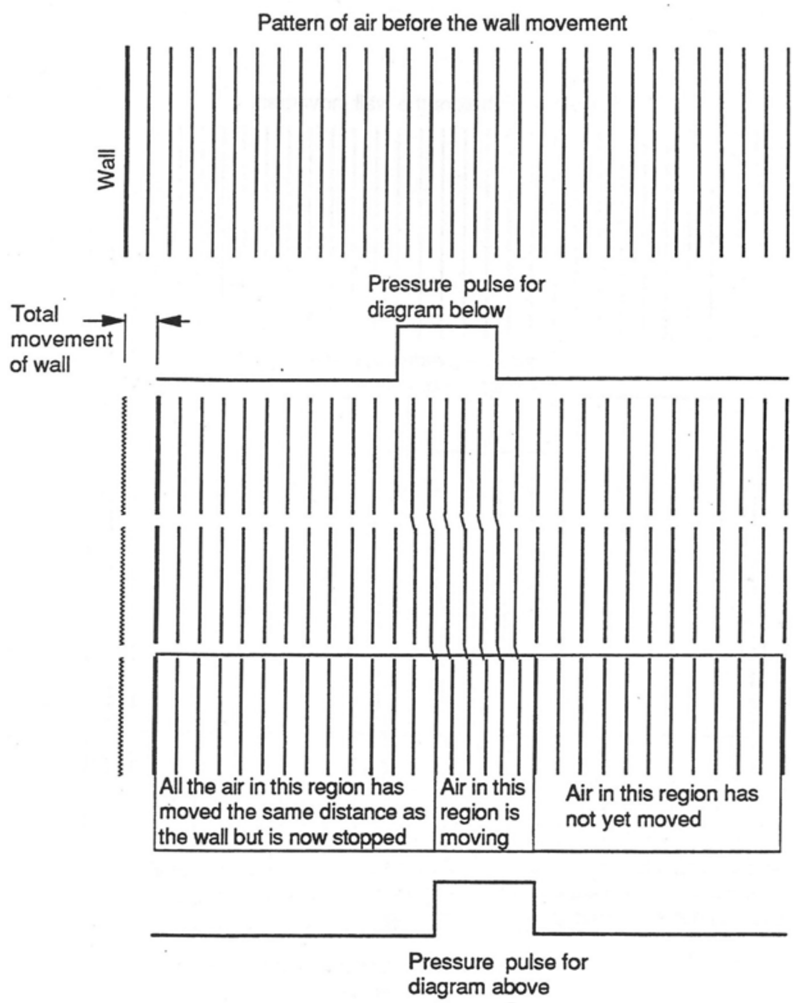


Figure 4.13 Diagram representing the movement of air for a pressure pulse moving to the right as a result of a wall movement on the far left. The only air which is actually moving at any instant in time is the air in the overpressure region. In this case, the air movement is in the direction of the wave propagation.

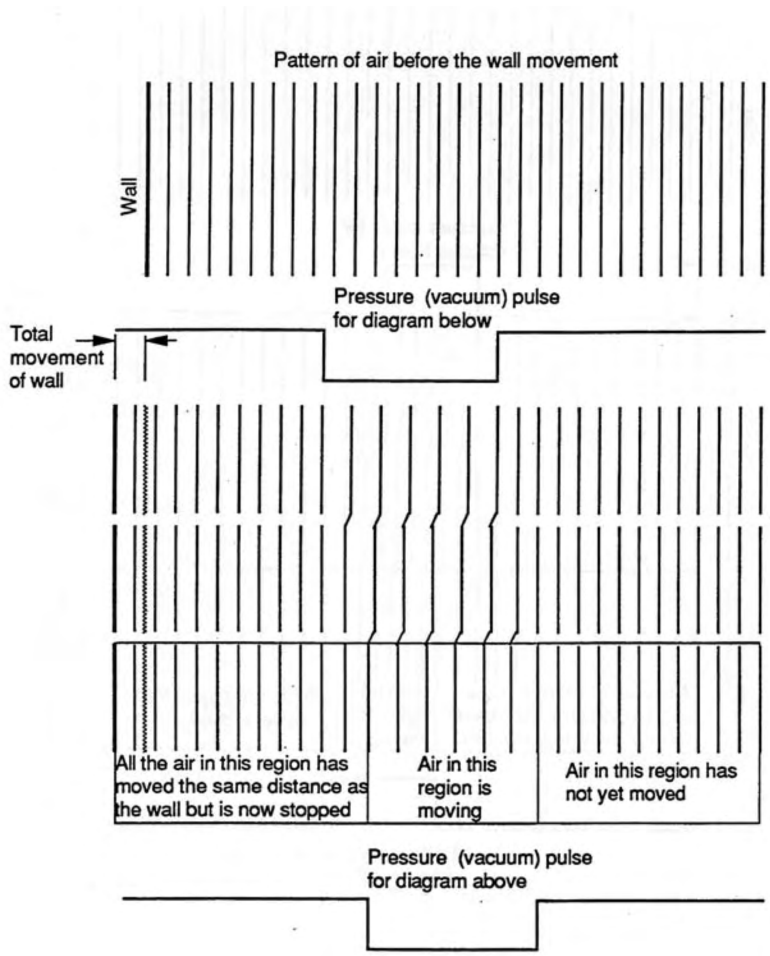


Figure 4.14 Diagram representing the movement of air for a vacuum pulse moving to the right as a result of a wall movement on the far left. The only air which is actually moving at any instant in time is the air in the underpressure region. In this case, the air movement is in the direction opposite to that of the wave propagation.

What happens to the energy that the wall puts into the air in front of it? From conservation of energy in an elastic medium such as air, the energy must go somewhere. Looking carefully at Figures 4.13 and 4.14 should show that the air in the pressure pulse is in fact moving forward (or backward) at the same velocity as the wall which caused the disturbance in the first place. The air at the leading edge of the pressure pulse is therefore doing work on the air in front of it in exactly the same fashion as if it were the wall itself. The effect of the wall movement is therefore being propagated through the air as real power or intensity. If the disturbance ever comes to another wall, it is capable of exerting pressure on that wall. If that wall moves with the air then work will be done on that wall with the same intensity as the original moving wall did work on the air.

This is why the concept of intensity of a sound wave is so important. It represents the amount of power that can be extracted from a sound wave per square meter of the wave-front surface. Because of the relationship between pressure and air velocity in a sound wave there is a simple equation linking sound intensity and sound pressure;

$$\begin{aligned} \textit{Intensity} &= \textit{Pressure} \times \textit{air velocity} \\ &= \textit{Pressure} \times \frac{c}{\gamma p_o} \text{ x } \textit{Pressure} \end{aligned}$$

or in symbolic form

$$I = \frac{c}{\gamma p_o} p^2$$

For normal air in a room at 20C, in SI units (Watts, meters and Pascals) this equation becomes

$$I = 2.412 \text{ x } 10^{-3} p^2$$

Continuing with the example of a wall moving at 1mm in 1ms to produce a sound pulse with a pressure of 413 Pascals, the intensity of this sound pulse would be 411 Watts/meter^2.

4.1.6 The Decibel Scale of Sound Intensity

The human ear intercepts about 1 cm^2 of a sound wave surface and seems to need only about 1 x 10^{-16} Watts to create a detectable motion in its cochlea. The sound intensity at the "Threshold of Sound" is therefore about 10^{-12} Watts per square meter. This is used then as the reference level from which sound intensity is expressed. It is usually referred to symbolically as I_o.

The intensity range over which the human ear can usefully respond to sound is enormous. At the point where pain and physical damage of the ear sets in the intensity is a trillion times that at threshold or about 1 Watt per square meter. Because of this large range and because of other factors such as the approximately logarithmic response of the human nervous system (see Roederer in the reference reading list), the intensity of a sound is usually expressed in a logarithmic scale called the "Decibel Scale". This scale is defined by the equation

$$\text{dB} = 10 \log_{10}\left(\frac{I}{I_o}\right)$$

where dB refers to the decibel level, $\log_{10}$ is the logarithm to base 10, I_o is the reference level specified above and I is the intensity of the sound, both in Watts per square meter. (I_o is 10^{-12} as specified above.)

The translation of a sound intensity level in dB to actual watts per square meter is just the inverse of this equation or

$$I = I_o \text{ x } 10^{\left(\frac{\text{dB}}{10}\right)}$$

From the equation relating intensity and sound pressure, the sound pressure itself can also be determined for a given dB level. The result is the table below linking dB, intensities and sound pressures (with some representative sounds).

dB	I Watt/m^2	p Pascals	
0	10^{-12}	2.036 x 10^{-5}	Threshold of sound for humans
10	10^{-11}	6.438 x 10^{-5}	Falling pin
20	10^{-10}	2.036 x 10^{-4}	Whisper at 1 m
30	10^{-9}	6.438 x 10^{-4}	ppp in music (very quiet hall)
40	10^{-8}	2.036 x 10^{-3}	pp (average modern empty hall)
50	10^{-7}	6.438 x 10^{-3}	p (interior of expensive car)
60	10^{-6}	2.036 x 10^{-2}	mf (subdued conversation)
70	10^{-5}	6.438 x 10^{-2}	f (City traffic)
80	10^{-4}	0.2036	ff (Limit allowed in factories)
90	10^{-3}	0.6438	fff (Steel-railed subway train)
100	10^{-2}	2.036	Loudest sound that can be tolerated for short periods.
110	10^{-1}	6.438	Typical rock concert
120	1	20.36	Threshold of pain
140	10	200.36	Jet engine at 30 m

Table 4.3 dB, intensities and sound pressures for representative sounds. Note that Pascals increase by 10 for each 20 dB increase and by √10 for a 10 dB increase.

In using the decibel scale it is often convenient to use the fact that a factor of two in intensity corresponds to about 3 dB. (The accurate value is of course 10 Log_{10} 2 or 3.01 dB but for the purposes of acoustics, 3 dB is an accurate enough figure.) Thus if sound intensity is doubled in intensity the sound level goes up by 3 dB; if it is halved, it goes down by 3 dB. Another useful fact to remember is that 10 factors of 2 (2^{10} or 1024) correspond closely to a factor or 1000 or 30 dB and therefore 20 factors of 2 correspond closely to 60 dB.

The figures in the above table show the scale of the sound phenomenon. Normal atmospheric pressure is about 100,000 Pa. Sound at a level that causes human beings to experience physical pain in their ears (120 dB) is only about 20 Pa or 1/5000th of this value. Thus sound is indeed a very small perturbation of the pressure of the air. A gentle breeze will put a wind pressure of 100 Pa on your face. A 1 meter climb will cause a pressure decrease due to the extra altitude of about 10 Pa.

Perhaps the most impressive figure is that for the wall which moved only 1mm at the modest velocity of 1 meter per second for 1 ms. The level of the sound pulse would be 146 dB or well beyond the threshold of pain and into the region where the ear can be physically damaged! The ears are indeed very delicate devices.

4.1.7 The RMS Pressure in Sound

In the simple case of the sound pulse created by a wall moving at uniform velocity, the pressure in the sound pulse was constant. In general, sound sources do not move with anywhere near uniform velocity. The sound sources in music have very complex motions. Yet there will be an average rate of transfer of energy to the human ear and hence a certain average decibel level to the sound. For intensity, the averaging is simple; the total energy delivered per square meter is divided by the time it takes to deliver that energy and this gives the average power. The average pressure is not so simple. This is because both negative pressures and positive pressures give sound intensity and the average sound pressure is in fact zero. (By sound pressure is meant the overpressure or underpressure relative to the normal air pressure in a quiet room.)

This problem is gotten around by noting that the average effectiveness of the sound pressure is due to its square. If the sound pressure is doubled, the intensity is quadrupled. The square of a sound pressure is always positive whether the pressure itself is positive or negative. By taking the squares of all the sound pressures and finding the average of these squares (mean square) the average effectiveness of the pressures is obtained. Taking the square root of this average (root-mean-square) then gives the effective average of the sound pressure oscillations.

This value is the one used to express the acoustic pressure of a sound. The above equation should therefore be written for the general case as

$$I = 2.412 \times 10^{-3}\ p_{\text{rms}}^2$$

A special case of importance is the sinusoidal pressure variation that occurs for a pure tone. It can be shown by calculus that the rms average of such a pressure oscillation is $1/\sqrt{2}$ times the peak value of the overpressure (or underpressure). A pure tone of 60 dB and therefore having an rms acoustic pressure of 2.036×10^{-2} Pa would have sound pressure oscillations from $+2.879 \times 10^{-2}$ Pa. to -2.879×10^{-2} Pa.

4.1.8 Energy in a Sound Wave

The final feature of sound to be considered here is its energy. Sound, in essence, represents the transfer of energy through space. The energy comes from the sound source through the air and impinges on the human ear to stir up the cochlea so as to send messages to the brain. While the sound wave is travelling through the air, the energy of the wave is stored in the air. In our simple moving wall example, after the wall stopped moving it put no more energy into the system. It had stopped doing any work. Yet when the wave reached a far wall it was capable of imparting on this wall the work done by the original moving wall. In the period between when the first wall stopped and the second wall received the energy, the energy must have been stored in the sir.

In many considerations of acoustics, it is important to know how much energy is stored in a sound wave. Again, in this introduction there will not be a mathematical derivation of this energy but just a presentation of the facts with a verbal description of how the facts might come to be.

The energy in a sound wave is in two parts; an energy of compression of the medium similar to the energy of compression of a spring and the kinetic energy of motion of the medium. It turns out that in a direct plane sound wave such as being considered here the energy density of each of these two forms is always equal. The total energy density is therefore twice that of either the potential (spring) or the kinetic (motion) part.

The total energy density in a sound wave can be directly related to the rms acoustic pressure or the intensity of the wave. The equation that results is

$$\begin{aligned} E &= \frac{p_{\text{rms}}^2}{\rho c^2} \\ &= 7.013 \times 10^{-6}\, p_{\text{rms}}^2 \\ &= \frac{I}{c} \quad = \quad 2.907 \times 10^{-3}\, I \end{aligned}$$

where E is the energy density of the sound wave in Joule per cubic meter.

Air propagating a sound of 100 dB therefore contains 2.91×10^{-5} Joules of sound energy per cubic meter. The total sound energy in a standard lecture room of 13 x 8 x 3 m at 100 dB would therefore be about 1 thousandth of a joule or about the energy required to lift a gram weight about 10 cm.

Again, the energy involved in sound is very small in normal mechanical terms and ordinarily would not be regarded as of any significance. It is only because of the enormous sensitivity of the aural systems of mammals that it is of significance.

Inverting the equation connecting energy density and sound intensity presents a new way of visualizing this relationship;

$$I = Ec$$

From this equation, the power of a sound wave can be thought of as coming from the rate at which it delivers its energy density to a surface. An energy density of E travelling at a speed c will deliver energy at the rate of $E \times c$ to each square meter of a surface. Thus, although the air itself is not moving at the velocity c, the energy contained in the wave is propagated at this velocity.

4.1.9 Spherical Sound Waves

A plane sound wave from a moving wall may be the simplest possible sound wave but it is not the usual form of direct sound wave that is encountered in a room. A usual sound source is small compared to the dimensions of the room and the sound waves radiate out from the source radially in all directions. Since the sound wave will radiate at the same speed in all directions the wavefront will be spherical, centered about the source.

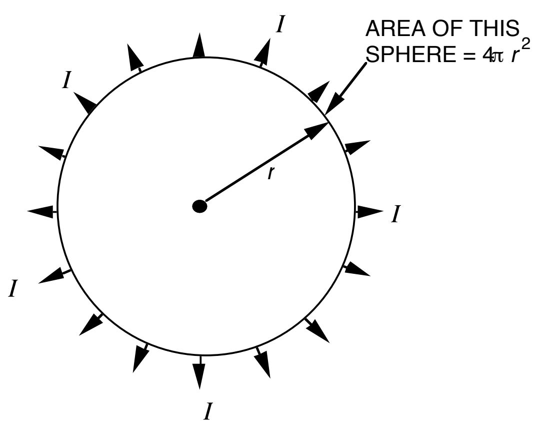


Figure 4.15 The spherical wavefront from a point source. The wavefront moves out in all directions so that at any instant, such as the one shown, all points on the wavefront are the same distance r from the source.

How is the spherical wave different from the plane wave?

The only difference between this type of wave and the plane wave is that the pressure in the sound pulse will diminish with distance from the source. This effect is analogous to the effect of a stone dropped in a quiet pool of water. The ripple will diminish as in moves out in every growing circles about the point of impact of the stone with the water.

For sound waves in open air, the pressure diminishes simply in inverse proportion to the distance from the source. If the pressure is 2 Pascals at 1 meter from the source, it will be 1 Pascal at 2 meters from the source and 0.5 Pascals at 2 meters from the source.

Because the intensity is proportional to the square of the pressure, the intensity will diminish in inverse proportion to the square of the distance from the source. This can be related to the conservation of the total sound power radiating outward from the source. For example, suppose a source is radiating power outward that is falling on the inside of a sphere. The intensity at the sphere will be the power divided by the surface area of the sphere. Now suppose the same power is being radiated outward to a sphere which has twice the radius. The surface area of this sphere will be four times the surface area of the first (area of a sphere is proportional to the square of its radius) and so the intensity will fall to 1/4 th.

Using this reasoning, the actual intensity at a distance from a source emitting spherical waves can be calculated from the total power radiated by the source. As shown in fig. 4.15, the area of a sphere is given by

$$\text{Area} = 4\pi r^2$$

Assuming the radiated sound power is spread uniformly in all directions, the intensity for a power P is given by

$$I = \frac{P}{4\pi r^2}$$

Sound intensity will therefore fall a factor of 100 for every increase of the distance by 10. On the decibel scale, sound intensity falls by 6 dB for each doubling of the distance from the source and by 20 dB for each increase of the distance by a factor of 10.

As an example of the use of the power-intensity formula for a spherical sound wave, consider the intensity at various distances from a one watt source radiating equally in all directions. The table below shows the results.

r (meters)	I ($1/r^2$)	dB ($1/r^2$)	dB (actual)
1	0.080	109	109
2	0.020	103	103
4	0.0050	97	97
10	8×10^{-4}	89	89
30	9×10^{-5}	79	79
100	8×10^{-6}	69	69
300	9×10^{-7}	59	58
1000	8×10^{-8}	49	46
3,000	9×10^{-9}	39	30
10,000	8×10^{-10}	29	-1

Table 4.4 dB and intensities at various distances from a 1 Watt source for an ideal $1/r^2$ dependence and what is actually achieved in open air.

The values calculated from the formula are accurate to within 3 dB for distances up to 1000 meters. However, for greater distances the actual sound level that will be achieved is less than that predicted by the formula. This is because air, like any other real material, is not a perfect transmitter of sound. Some energy is lost due to viscosity in the air movement. This loss is about 3 db per kilometer in clean dry air and the figures in the right hand column of the table include this absorption. However, in normal auditoria the distances for sound propagation, including all the reflections involved in reverberant sound, are seldom more than 1000 m.

What is perhaps astonishing about these figures is the tremendous carrying power of sound. A source of only 1 watt of sound power (remember that an ordinary light bulb will dissipate about 100 watts of electrical power) will produce a deafening sound of about 110 db at 1 meter. If we take a sound level of 70 db as a loud conversational level, then this 1 watt source will interfere with conversation at 100 meters distance. Even perhaps more amazing, if there were such a thing as perfect atmospheric conditions and no other noise, this sound would still be audible at up to almost 10km distance (even taking atmospheric absorption into account). Perhaps those of you who have had the experience of being outdoors far from civilization on a cold winter night on a frozen lake (about as close as you can get to ideal atmospheric conditions for sound) can confirm the tremendous carrying power of sound over great distances. Under such conditions, the sound of somebody talking can be heard over a kilometer away.

In buildings the direct sound wave never travels these great distances. Even in the largest halls, it will reach a surface of the room within 30 m of travel. When it does so, another aspect of the sound wave becomes very important; a great deal of its energy reflects from the surface back into the room. Reflection then becomes another very important aspect of the physics of sound.

4.2 Sound Reflection

4.2.1 Plane Wave Reflection

When a sound wave reaches a surface, that surface normally is almost an immovable object for the air. The moving air in the sound wave therefore builds up against the wall and its internal velocity stops. However, the piling up of the air against the wall due to the momentum associated with the air velocity causes the pressure to momentarily exceed that in the original wave. This extra overpressure generates a backwards wave away from the surface.

To show the dynamics of this, the overpressure region of a sound wave as it reaches an immovable wall is shown in fig. 4.16. At (a), the wavefront has almost reached the immovable wall. At (b) the wave has moved into the wall and the air velocity in the region which has reached the wall has been reduced to zero. Furthermore the air which has already reached the wall has pushed back to kill the velocity of the air in the region which has not yet even reached the wall. While all motion of the air in this piled up region has stopped, the pressure in this region has doubled over that for the incoming pulse.

As this motion progresses, more and more of the wavefront region is piled up against the wall until in fact a momentarily stationary state is reached. This occurs when exactly half the region of compression in the wave has reached the wall (at (c) in fig. 4.16). After this state has been reached, the compressed air now begins to push the uncompressed air on the left backwards. The result is a wave which begins to develop in the direction away from the surface (at (d) in fig. 4.16). Finally, at (d) in fig. 4.16, the fully developed reflected wave appears.

It might be noted that at all times the total energy in the wave during reflection is conserved. For example, at the momentary stationary state the pressure is doubled giving four times the energy density for the pressure part. However, since the kinetic part has vanished and this was originally equal to the pressure part, the overall energy density has only twice that in the original wave. Since the actual extent of the wave at this instant in its reflection is only half that of the original then the total energy (energy density times volume) remains constant. (Those with a physics deviation might like to prove that this is the case at any instant in the reflection.)

The case shown in Fig. 4.16 is for a wave arriving perpendicularly to a surface. If it arrives at an angle, the reflected wavefront takes up an angle with the surface such that the angle of incidence equals the angle of reflection (Fig. 4.17).

An important aspect of this phenomenon is that it is extremely difficult to prevent. For example, with a simple microphone and a high frequency speaker ("tweeter") producing a sharp pulse of sound, it is easy to see on an oscilloscope the reflected sound wave from a thin piece of paper held a meter away from the speaker. It is also easy to see that the reflected wave is strongest when the paper is oriented so that the reflected wave is directed toward the microphone (fig. 4.18).

In order to prevent sound reflection from a surface, the air moving onto the surface must have somewhere to go. Reflection is therefore reduced by providing holes for the air to move into (an open window is, in fact, almost a perfect absorber of sound). However, any real surface, by definition, cannot be made up completely of holes. Carpets, drapes and acoustic absorbing tiles have about as many small holes for the air to move into as it is possible to get in a material which will support itself. The "acoustic absorption coefficient" of such material (fraction of sound energy absorbed from a sound wave falling on the material) is typically from 30% to 80%. The absorption coefficients for normal building materials designed to provide structural strength are much lower. Some typical figures are given in table 4.5.

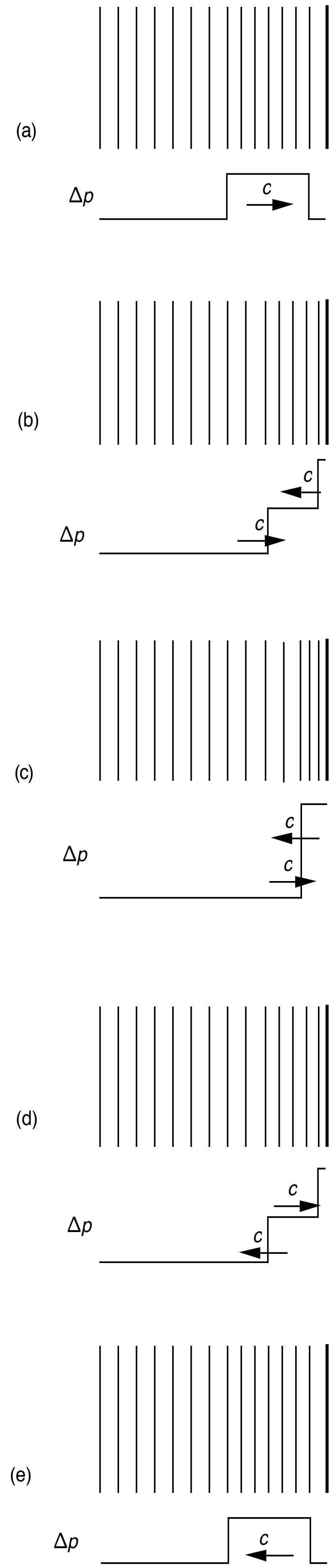


Figure 4.16 The reflection of a sound pulse at a firm boundary.

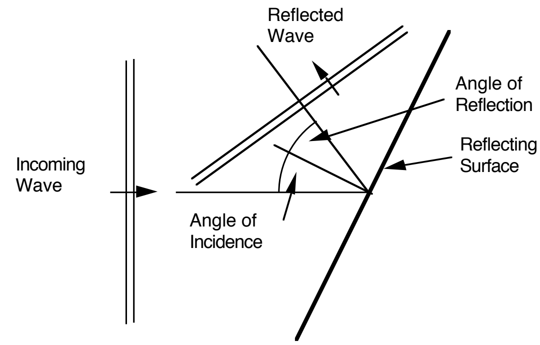


Figure 4.17 The reflection of a sound pulse at a firm boundary.

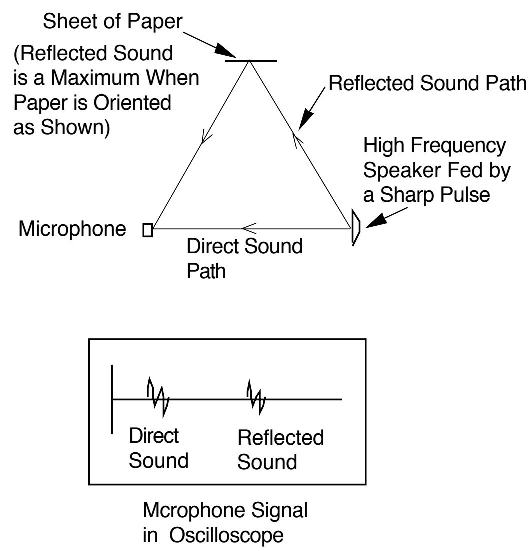


Figure 4.18 The reflection of a sound pulse from a sheet of paper.

Material	Absorption coefficient
Concrete	0.015
Brick wall	0.02
Plaster	0.06
Wood Sheeting	0.10
Carpet	0.20
"Acoustic Tile"	0.80
Open Window	1.0

Table 4.5 Absorption coefficients (fraction of sound energy absorbed upon reflection for various materials.

4.2.1 Spherical Wave Reflection

When sound with a curved wavefront reaches a reflecting surface, there will be a new reflected wavefront which is itself curved. In the special case of the spherical wave radiating from a small source, the reflected wave will be another spherical wave but now appearing to radiate from behind the reflecting surface (see fig. 4.19).

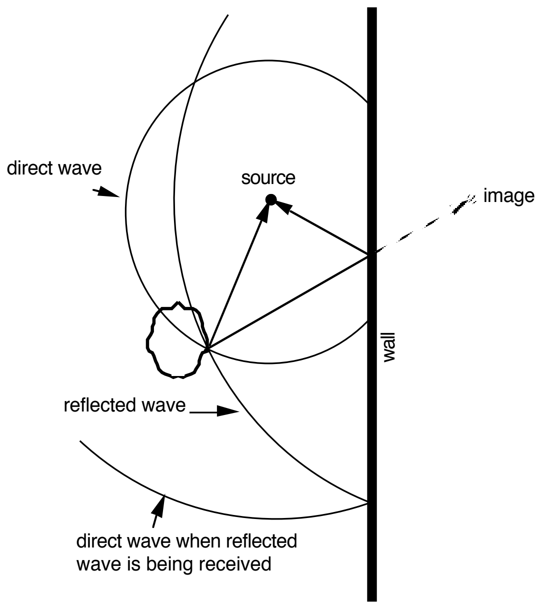


Figure 4.19 The reflection of a spherical wavefront from a wall.

The apparent source of this reflected wave will be a point behind the surface at the "mirror image" of the real source. An observer listening to the sound from the source would, if it were a sharp pulse, hear two waves passing. One would be the direct sound wave from the source and the other would be the reflected sound from the surface. The perceived positions of the sources of these two sounds would be as shown in fig. 4.19

The image source of a reflected sound appears to be a very important source in the human perception of sound in a room. Even though the time difference between the direct sound wave and the reflected sound wave in a typical room would be only a few hundreds of a second, this time difference is used by the brain to tell the distance and direction of the closest wall (the one giving the first reflection). For some reason, this appears to be an important psychological factor in the feeling of well-being of the listener. For example, some modern results in a scientific study of concert hall preference (see chapter 3) have shown that people generally prefer halls in which the first reflection comes from one of the side walls rather than from the ceiling. The manipulation of the first reflections by altering the design of a room is therefore an important aspect of the acoustics of concert hall design.

4.3 Multiple Sound Reflections

So far only the first reflection from one surface has been considered. Normally, all six interior surfaces of a room, the floor, ceiling and four walls, are significant sound reflectors. A direct spherical sound wave will reflect from each of these surfaces, producing in each case another image source. In addition there can be multiple sound reflections such as those in the corner of a room shown in Fig. 4.20.

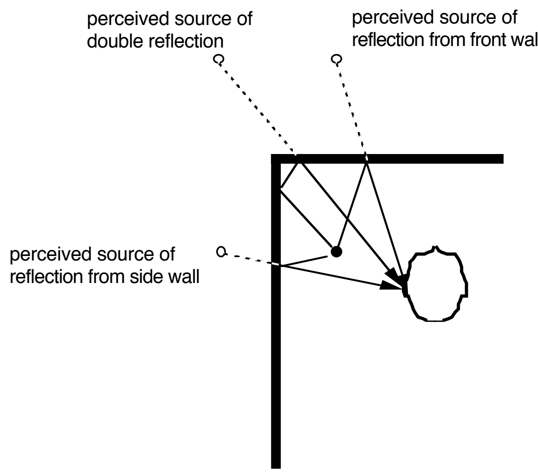


Figure 4.20 Corner reflections of a wave from a point source.

Furthermore, each of these corner image sources have mirror images in the opposite walls (fig. 4.21).

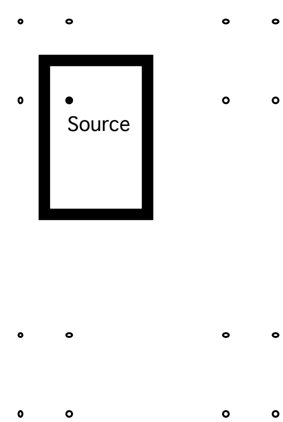


Figure 4.21 Four corner reflections of a wave from a point source.

In turn, all of these images have their images out to very great distances. Because of the high reflection coefficient for sound of normal walls, these perceived images can extend out to very great distances.

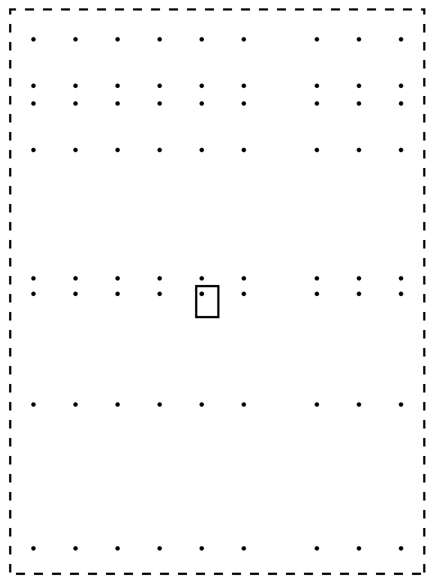


Figure 4.22 Images of multiple reflections of a wave from a point source in a room. The dotted boundary of the room has no real significance but could be thought of indicating that the images at some distance must get too faint to be perceived.

This is similar to being in a room in which the walls are made of mirrors. The optical absorption coefficient of a good mirror is about the same as that of concrete for sound; about 1.5%. To be similar to that for typical sounds in a room, the floor and ceiling would also be mirrors but they would be a little dirty if the ceiling was covered with acoustic tile and the floor with carpeting. Imagine that a bright light was turned on in this room. The result would be a pattern of light in the mirrors which extended to very great distances in all directions, even to some distance above the ceiling and below the floor. The mirrors have provided much more light in the actual room by supplying a host of image sources.

Similarly, the walls of a music room considerably increase the sound falling on a listener's ears by providing a host of images sources in the walls. However, the images sources also provide another effect in sound that would not be perceived in the case of the mirrors. With mirror walls, all the light would disappear as soon as the actual source was turned off. This is because light travels as such a very high speed. (It would actually disappear in about one millionth of a second.) However, because of the much slower velocity of sound, the sound from the distant sound images will take a perceptible time to arrive.

The walls of the room therefore have two important effects on the sound in the room; they increase the amount of perceived sound in the room and they make the sound in the room linger for a perceptible time. This lingering sound is one of the most important sounds perceived in a room and has been given the labels "Room Reverberation" or "Reverberant Sound". This was the sound that was introduced in chapter 2 as the sound which makes many home recordings seem as if the recording microphone was kept inside a barrel.

4.4 Reverberant Sound

Again, one of the most important aspects of the reverberant sound in a room is how long it appears to last. This section is an introduction to the principles by which it is possible to estimate how long a sound will last in a room by knowing the room geometry. No attempt will be made here to prove the equations used as that is not the purpose of these notes. Rather, there will be an attempt to give some idea of where the equations come from.

First it is perhaps worthwhile to get some impression of how long a reverberant sound appears to last in a typical room. This can be done by the "hand-clap" technique. By simply giving a sharp clap of sound into a room and then listening carefully to the response, we can get some idea of the reverberation time. In fact an experienced observer can get a very good estimate of this reverberation time by such a simple test. In a typical lecture room the reverberation time will be something shorter than one second, usually about 3/4 of a second.

The first attempt to calculate the reverberant time of a room was made by Sabine in dealing with the acoustic problems of Fogg Hall at Harvard, opened at the turn of the century. By today's standards, the measuring apparatus that he had to guide him in his theories were extremely insensitive but the results of his work are still regarded as the basics of room acoustics. With his success in diagnosing the problems with Fogg Hall and in suggesting improvements, acoustics became to be regarded as a science.

The essential features of Sabine's approach was to assume that the sound energy in a room was uniformly divided throughout the room and to consider the effect a particular sound absorbing surface would have on this energy. Sabine's reasoned that a sound absorbing surface will remove a fraction of the sound energy falling on it according to the equation

$$\begin{array}{l}\textit{Rate of absoprtion}\\ \textit{of sound energy}\end{array} = \alpha_{abs} \times \begin{array}{l}\textit{Rate at which energy}\\ \textit{falls on surface}\end{array}$$

where α_{abs} is the sound absorption coefficient for the surface (i.e. the fraction of sound energy absorbed in reflection from the surface). In the language of mathematics (i.e. calculus), this equation becomes

$$\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{abs} = -\alpha_{abs} \times \textit{Rate at which energy falls on surface}$$

The rate at which sound energy falls on a surface is proportional to the sound energy density in the room. This sound energy density in the room is itself proportional to the total sound energy in the room. If the sound energy is uniformly distributed throughout the room, the sound energy density is just the total energy divided by the room volume. The rate at which energy falls on a surface is therefore proportional to the total sound energy in the room and therefore, from the above equation, the rate of loss of sound energy in the room is proportional to the sound energy itself;

$$\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{abs} = -\,Constant \times E$$

This is an example of an extremely important equation in physics, engineering and biology; it is an equation in which a quantity is changing at a rate which is proportional to the quantity itself. Perhaps the most familiar example of this type of equation is the growth of money in a saving account. Suppose, for example, that 1000 dollars were put into an account which gave 8% interest annually. The money in this account in the following five years would be as in the middle column of table 4.6.

Year	Money (at 8% interest per year)	Money (at 8% loss per year)
0	1000.00	1000.00
1	1080.00	920.00
2	1166.40	846.40
3	1259.71	778.69
4	1360.49	716.39
5	1469.33	659.08

Table 4.6 Value of money under compound interest or loss. Note that the increase in the 5th year is 8% of 1360.49, and that the decrease in the fifth year of a loss is 8% of 716.39, not 8% of 1000.00.

A related phenomenon occurs when the money in put into a system which loses 8% per year. The resulting figures for this situation is shown in the right hand column of table 4.6.

The behavior of money in such a saving account is not strictly analogous to the situation where the rate of change of a commodity is continuously proportional to the commodity itself. The interest or loss loss was calculated at the end of the year and then added to or subtracted from the account. The growth or decay of the money would only be exactly analogous to the continuous change in a commodity if the interest were "compounded continuously". The results of such a continuous compounding as shown in table 4.7.

Note that the money increases or decreases at a slightly greater rate then under the annual compounding.

But how were these numbers calculated? How does a bank calculate the amount you have in an account when they offer you "continuous compound interest"? One way would be to have a computer calculate the interest every second and add it to your account. The amount would be so small that certainly the increase in the money in the account would be continuous (no discrete jumps). This would be possible with modern computers but it would be a waste of the computer time (computers also cost money to employ them). The computation is done much more simply using the "exponential function";

$$M = M_o \times \mathrm{e}^{(Interest\ Rate\ \infty\ time)}$$

where M is the value of the money at any time and M_o is the initial value of the money. For example, the value of 1000 dollars after 4 years at 8% compounded continuously would be

$$M_{4\mathrm{yrs}} = 1000 \times \mathrm{e}^{(0.08\ \infty\ 4)} = 1000 \times \mathrm{e}^{0.32} = 1377.13$$

Year	Money (at 8% interest per year)	Money (at 8% loss per year)
0	1000.00	1000.00
1	1083.29	923.12
2	1173.51	852.14
3	1271.25	786.63
4	1377.13	726.15
5	1491.82	670.32

Table 4.7 Value of money under continuously compounded interest or loss.

The exponential function describes any continuous growth or decay which is at a rate in proportion to the quantity itself. This is because the unique feature of the exponential function which is that it is its own derivative (see any elementary calculus book).

$$\frac{\mathrm{d}\mathrm{e}^x}{\mathrm{d}x} = \mathrm{e}^x$$

The derivative of a general exponential function

$$y = a\mathrm{e}^{bx}$$

is then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = ab\,\mathrm{e}^{bx}$$

or

$$\frac{\mathrm{d}y}{\mathrm{d}x} = by$$

The exponential function is therefore the solution of any equation of the type

$$\frac{\mathrm{d}y}{\mathrm{d}x} = by$$

In the case of a decay of a quantity, the exponent b in the exponential becomes negative. For example, again in the case of money when there is a constant loss rate

$$M = M_o \times e^{-(Loss\ Rate\ \infty\ time)}$$

The form of this negative exponential function is of particular interest here. It is shown in Fig. 4.23

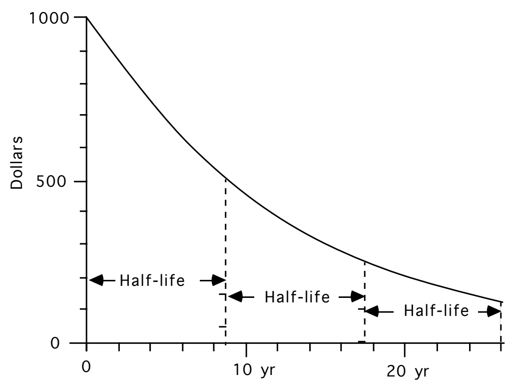


Figure 4.23 The exponential decay function.

The particularly interesting feature of this function is that the fractional decrease is the same for any fixed time interval, no matter when that time interval occurs. For example, the original 1000.00 is halved in the first 8.88 years, about 8 years, 8 months, and it is halved again, to 250.00, in another 8.66 years.

There is, in fact, a very simple mathematical connection between the fractional rate of decay of a quantity and its "half life";

$$t_{1/2} = \frac{0.693}{\text{Decay Constant}}$$

where the "Decay Constant" is the fractional rate of decay. For an loss rate of 8% per year compounded continuously, the Decay Constant is 0.08 per year.

The factor 0.693 is just the natural logarithm of 2. This comes into the relationship because the time at which a quantity is halved is that when

$$e^{-\text{ Decay Constant } \infty\ t} = 0.5$$

$$\text{Decay Constant} \times t = -\text{Ln } 0.5 = \text{Ln } 2$$

Therefore

$$t\ (= t_{1/2}) = \frac{\text{Ln } 2}{\text{Decay Constant}}$$

Another characteristic time often related to the decay constant of a quantity is the "Relaxation Time" defined as the time for a fall to 1/e and generally written as τ. This is quite simply related to the decay constant;

$$\tau = \frac{\text{Ln e}}{\text{Decay Constant}} = \frac{1}{\text{Decay Constant}}$$

Sabine's reasoning therefore led to the conclusion that the sound in a room would decay exponentially. What this means is that it would decay with a certain characteristic half-life. The range over which a human can detect sounds in a short time interval is about 60 dB. A sound of 100 dB which is suddenly decreased in intensity will appear to disappear at about 40 dB. If given some time of quiet the ears will "open up", very much like eyes become dark adapted when a bright light is turned off. It is only then that the ears will be able to detect sounds below 40 dB.

A factor of 2 is about 3 dB and so a loss of 60 dB will take about 20 half-lifes. The reverberant sound in a room should therefore appear to have a duration of 20 half-lives. Sabine therefore defined "Reverberation Time" T by the formula

$$T = 20 \times t_{1/2} = 20 \times \frac{0.693}{\text{Decay Constant}}$$

What remained in the analysis was to obtain the decay constant of the sound in a room from the geometrical properties of the room.

As already pointed out, the rate at which a given absorber remove sound energy from a room is related to the rate at which the sound energy in a room falls on the absorber. This is the difficult mathematical part of the analysis. The sound energy of the room is moving in all directions at the velocity of sound (344 m/s in a normal room). This means that for any given volume of sound energy in the room, half the sound will be moving to one side and half to the opposite. In particular, half the sound energy in the volume immediately in front of an absorber will be half moving toward the absorber and the other half moving away from it. Therefore, half the sound energy does not get to the absorber at all (Fig 4.24).

Also, the sound which is moving toward the absorber is generally travelling at some angle to the absorber. This sound will not see the full effective area of the absorber but only the part perpendicular to its particular direction (Fig. 4.25).

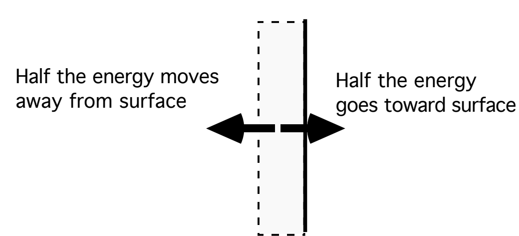


Figure 4.24 The movement of sound energy in front of an absorber. The particular case shown is for sound moving exactly perpendicular to the wall.

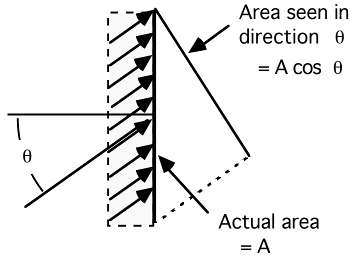


Figure 4.25 The area seen by a sound wave travelling at an angle to the normal of an absorbing surface. The area seen is the actual area of the surface multiplied by the cosine of the angle of the direction with the normal to the surface.

Using integral calculus, it is possible to show that the average of all these projections for all the possible directions at which sound can reach the surface is just half the surface area. (See almost any engineering text-book on room acoustics.)

The rate at which sound energy flows in a room is the energy density times the sound velocity. If one half of one half of the energy in front of an absorber actually flows onto that absorber, then the overall flow rate of sound energy onto the absorber is given by

$$\textit{Rate at which sound energy falls on a surface} = \textit{Energy Density} \times c \times \frac{\textit{Area of surface}}{4}$$

$$= \frac{E_{total}}{V} \times c \times \frac{\textit{Area of surface}}{4}$$

where E_{total} is the total sound energy in the room and V is the room volume.

The rate at which this surface takes sound energy out of the room is therefore given by

$$\left(\frac{dE}{dt}\right)_{abs} = -\alpha_{abs} \times \textit{Rate at which energy falls on surface}$$

$$= -\alpha_{abs} \times \frac{E_{total}}{V} \times c \times \frac{\textit{Area of surface}}{4}$$

The α_{abs} term and the area of the surface can be multiplied together to get the effective area of a particular absorbing surface. Furthermore, this can be done with all the absorbing surfaces in the room and the result added up to get an overall effective absorbing area for the room. The equation for the total rate of energy loss in the room then becomes;

$$\frac{dE_{total}}{dt} = -\frac{E_{total}}{V} \times c \times \frac{A_{eff}}{4}$$

Rearranging this equation gives the equation for an exponential decay

$$\frac{dE_{total}}{dt} = -\frac{A_{eff}}{4V} \times c \times E_{total}$$
$$= -\textit{Decay Constant} \times E_{total}$$

where

$$\textit{Decay Constant} = \frac{A_{eff}}{4V} \times c$$

The reverberation time of a room is therefore given by

$$T = 20 \times t_{1/2} = 20 \times \frac{0.693}{\textit{Decay Constant}}$$

or

$$T = 20 \times \frac{0.693 \infty 4}{c} \times \frac{V}{A_{eff}}$$

For c in m/s, i.e. 344, this equation becomes the standard form used for simple calculations of the reverberant sound in a room;

$$T = 0.165 \frac{V}{A_{eff}}$$

(In some older texts, particularly from the USA, the constant in front is calculated for V and A_{eff} in feet. Since the quotient V/A_{eff} has the units of feet, then the constant in the reverberation time equation becomes 0.165 divided by the number of feet in a meter or 0.165/3.28 = 0.05 (about).

As an example of the use of this formula, consider a typical rectangular lecture room 10 meters wide by 8 meters deep by 3 meters from floor to ceiling with a back wall being a window wall and the front wall the blackboard wall. Suppose the side walls are brick of absorption coefficient 0.02. The ceiling would usually be acoustic tile but not the most absorbing kind and usually of absorption coefficient about 0.5. Assume the floors and window wall have a typical absorption coefficient of 0.05.

The volume of the room is 240 m^3. Calculating the effective absorption area of the walls, floor and ceiling gives

Front wall	$10 \times 3 \times 0.02$	=	0.6	m^2
Window wall	$10 \times 3 \times 0.05$	=	1.5	
Side walls	$2 \times 8 \times 3 \times 0.02$	=	0.96	
Floor	$10 \times 8 \times 0.05$	=	4	
Ceiling	$10 \times 8 \times 0.5$	=	40	
	Total	=	47	m^2

Table 4.8 Calculation of effective absorbing area of the walls of a typical room.

To this we should add the absorption area represented by the people in the room and the chairs and tables.

Average values taken for people in indoor clothing is about 0.5 m^2. The average for the type of hard chairs and tables that are in lecture rooms would be about 0.03 m^2 per seating unit. The total effective absorbing area of the people and chairs for 25 people and 60 chairs be $25 \times 0.5 + 60 \times 0.03 = 14$ m^2.

The total absorption area for the room is therefore about 61 m^2. The reverberation time of the room should therefore be about

$$T = 0.165 \times \frac{240}{61} = 0.66 \text{ seconds}$$

This value will be typical of a modest size lecture room.

The reverberation time calculated from such crude assumptions about a room may not have much worth as it stands. However, if the actual reverberation time is known from measurements on sound in the room, then the formula can be used in reverse to calculate the effective absorbing area in the room and then one can estimate how much absorbing material should be added (or taken out) to get the reverberation time that is desired. This is essentially what Sabine did. He measured the actual reverberation time of Fogg Hall and then prescribed the materials needed to bring the reverberation time down to acceptable levels. In practice, one of the common uses of the formula is to calculate the effect on the reverberation time of a hall (which usually can only be measured when it is unoccupied) of having the hall filled with people.

There are several other useful things that come out of this analysis. One is that we can now turn the problem around and use the reverberation time of a room to estimate what the sound level will be in a room when we put a given sound power source in the room. Since in the steady state condition, the rate of sound energy loss by absorption will be equal to the rate at which sound energy is coming out of the source, we have

$$\begin{aligned} &\textit{Sound Power} \ (= N \text{ in Meyer}) \\ &= \textit{Rate of energy loss} \\ &= \textit{Decay constant} \times E_{total} \end{aligned}$$

This allows the calculation of the energy density in a room;

$$\textit{Energy density} = \frac{E_{total}}{V} = \frac{N}{\textit{Decay constant}}$$

The decay constant is related to the reverberation time by

$$\textit{Decay constant} = \frac{0.693}{t_{1/2}} = \frac{20 \infty 0.693}{T}$$

The energy density is therefore given by

$$\textit{Energy density} = \frac{T}{13.8} \times \frac{N}{V}$$

The energy density is related to the sound intensity in a room by $I = c \times \textit{Energy density}$. This gives a simple formula for the reverberant sound intensity in a room with a sound source of N watts;

$$I_{reverb.} = \textit{Energy density} \times c = \frac{c\,T\,N}{13.8\,V}$$

For a normal room and c in m/s, this formula becomes

$$I_{reverb.} = 24\frac{TN}{V}$$

Applying this formula to the lecture room example with a sound source of 1 watt power gives an intensity of

$$I_{reverb.} = 24 \times \frac{0.66 \infty 1}{240} = 0.066 \text{ Watt/m}^2 = 108 \text{ dB}$$

Again, a perhaps surprising result illustrating the extreme sensitivity of the ears. A 1 watt source of sound power (about the power involved in one person breathing) in a lecture room for 50 people would create a reverberant sound level which would drown out any other normal source of sound and almost be painful to the ears. If exposed to such a sound for a relatively short period of time, temporary hearing loss would occur and if exposed for a longer period the hearing loss would become permanent.

Finally, the analysis can give a formula for the room radius of a point source in the room. Since the direct sound level at distance r from a source is given by

$$I_{direct} = \frac{N}{4\pi r^2} = 24\frac{TN}{V}$$

Solving for r gives the room radius;

$$r_{room} = 0.058\sqrt{\frac{V}{T}}$$

Using the values for the lecture room example above gives a room radius of 1.1 m

Appendix

The wave equation is generally presented in the form

$$\frac{\partial^2\phi}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}$$

This is the form that applies to a wave travelling in the direction x. If the wave can also travel in the direction y, the equation becomes

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}$$

and if the wave can travel in all three dimensions;

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}$$

This final form of the equation is often written in the short-hand form

$$\nabla^2\phi = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}$$

The variable ϕ can be anything that is governed by this equation. In the case of sound it can be pressure, velocity or position of a small part of the medium. The meaning of the equation is easiest to visualize when ϕ is the position. Then the meaning is essentially that the force (left hand side of the equation) to produce an acceleration (right hand side of the equation) of any point in the medium is proportional to the acceleration ($F = ma$).

In the proportionality constant $1/c^2$, c is the velocity with which a wave will propagate. If, when the equations of motion for a variable are derived and they are seen to be of the form of the wave equation with a positive constant in front of the second partial derivative with respect to time, then wave propagation is possible in the variable and the velocity of that wave propagation will be the reciprocal of the square root of the constant term. This was the result of Maxwell's analysis in the mid 19th century of the equations governing electric and magnetic fields in a vacuum. From this analysis, Maxwell was able to prove that electromagnetic waves could exist and that they would have a velocity which was close to that which was then known for the velocity of light. Thus the wave equation solved the mystery of what is light. It is an electromagnetic wave.

Exercises and Discussion Points

1. a) A lightning flash is followed 2.6 second later by a fairly sharp thunder clap. How far away was the lightning bolt?

   b) As the thunder storm moves away, distant lightning bolts no longer produce sharp claps of sound but rolling thunder. Why?

2. A rifle bullet can hit you before you hear the sound of the gun which fired the bullet. How can this happen?

3. There is a curved wall of rock near Sante Fe, New Mexico, which a person can stand in front of at a designated spot and hear quite distinct repetitions of anything spoken by the person or any people nearby at about the same level as they were originally spoken. What is the nature of the curve of the rock? If the echo comes 0.7 seconds after the original, what is the distance of the wall from the speaker?

4. A broad flat surface moves forward 0.4 mm in 2 ms at uniform speed and then stops.

   a) What is the extent of the pressure zone propagated away from the wall?

   b) What is the duration of the pulse of sound heard by an observer near the wall?

   c) What will be the pressure level of the sound pulse?

   d) What will be the decibel level of the sound pulse?

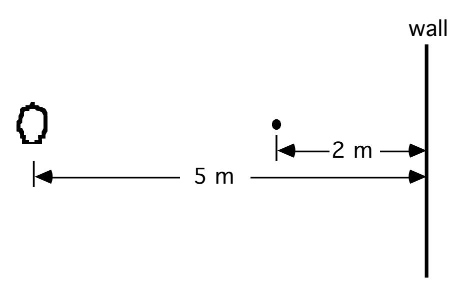
   e) Repeat for a wall which moves at 0.1 mm at 10 cm/sec

5. An oscillating wall is moving at a peak velocity of 1 mm per second. What is the connection between the oscillating velocity and the oscillating pressure in front of the wall? When in the walls motion is the pressure a maximum and when is it zero? What is the peak pressure and what is its rms value? What is the intensity of the sound propagating from the wall?

6. What is the average sound energy density in the air in front of the wall in problem 5?

7. A point source is radiating 0.1 Watt of sound uniformly in all directions into a room. What is the intensity of the direct sound from the source (watt/m^2 and dB) at 2 m from the source? What is the intensity (watt/m^2 and dB) at 5 m from the source?

8. A point source of sound located 2 m from a large reflecting wall is radiating sound equally in all directions. An observer is located 5 m from the wall in line with the source;



a) What will be the time difference between the arrival of the direct and the reflected sound?

b) What will be the relative intensity levels of the direct and reflected sounds (dB difference) if the wall has an acoustic absorption coefficient of 0.02?

c) What will be the relative intensity levels of the direct and reflected sounds (dB difference) if the wall has an acoustic absorption coefficient of 0.5?

9. A listener is sitting directly in line with center stage at a distance of 15 m from a performer. The side walls of the hall are parallel and separated by 18 m.

a) What will be the time difference between the arrival of the direct sound from the performer and the first reflection from the side walls?

b) If the side walls have an acoustic absorption coefficient of 0.05, what will be the relative intensity of the direct sound and the first sound reflection?

10. What are the energy storages in a gas associates with sound waves? Which fundamental parameters of a gas determine the velocity of these waves? Which parameter variation causes the most significant changes in this velocity in ordinary life and how does this variation show in the equation

$$c = \sqrt{\frac{\gamma p_o}{\rho}}$$

11. The variation of the velocity of sound with temperature T can be expressed as;

$$c_{T^o} = c_{0^o}\sqrt{\frac{273 + T}{273}}$$

where c_{0^o} is the velocity of sound at zero degrees and T is the temperature in Celcius. Explain where this equation comes from. (Do not attempt to derive it but merely explain its origin.)

12. Explain the importance in the human perception of sound of the fact that in the human range of hearing, all frequencies of sound have the same velocity.

13. Explain the meaning of the equations;

$$\frac{E_{total}}{V} = \frac{p^2}{\rho c^2} = 7.013 \times 10^{-6} p^2$$

$$I = c\frac{E_{total}}{V} = \frac{p^2}{\rho c} = 2.412 \times 10^{-3}\ p^2$$

14. Describe the experience of being in a room with very sound reflective walls in terms of the analogy of being in a room in which all four walls , floor and ceiling are mirrors. Use the principle of "imaging" in optics to describe the perceived sound sources. Point out the difference that the ability of the human perception system to perceive time intervals in sound makes to the perception of the sound compared to the perception of the light (where no such ability exists because of the extremely high velocity of light waves).

15. How many reflections from a surface of absorption coefficient 2% would be required to reduce the intensity of a sound wave by 60 dB. (Assume no reduction in intensity because of distance travelled by the wave). If the average distance between the reflecting surfaces was 10 meters, how long would this take?

16. Why did Sabine take 60 db as the drop in sound level at which reverberant sound is no longer of any importance?

Sketch in a graph the way reverberant sound intensity will decay with time. How many times does the sound level have to halve to equal a 60 dB drop? How many times does the sound pressure have to halve to equal a 60 dB drop? What is the connection between these two numbers? If the amplitude of a sound as recorded on an oscilloscope halves in 120 milliseconds, what is the room reverberation time?

17. State the basic assumptions made by Sabine in his analysis of reverberant sound in a room. In the equations

$$T = \frac{4}{c}\ \ln(10^6)\frac{V}{A_{eff}} = 0.165\frac{V}{A_{eff}}$$

explain the significance of the terms and where they come from in Sabine's analysis.

18. Define room reverberation time, energy density and room radius as they appear in the following equations

$$T = 0.165\frac{V}{A_{eff}}$$

$$\frac{E_{total}}{V} = \frac{\mathrm{T\ N}}{13.8\ \mathrm{V}}$$

$$r_{room} = 0.056\sqrt{\frac{V}{T}}$$

Give the meaning of the terms in the equations and some qualitative explanation of where these equations come from.

19. Given the equations;

$$\frac{E_{total}}{V} = \frac{T\,N}{13.8\ V} \ ; \quad I_{dir} = \frac{N}{4\pi r^2} \ ;$$

$$r_{room} = 0.056\sqrt{\frac{V}{T}} \ ; \ c = 344 \text{ m/s}$$

suppose one had a musical instrument creating a sound level of 80 db direct sound at 2 meters.

a) What is the sound power of the instrument assuming that it radiated uniformly in all directions?

b) What would be the reverberant intensity in a room 10m x 15m x 20m with a reverberation time of 1.4 s?

c) What would be the decibel level of reverberant sound in part b)?

d) What would be the room radius for the instrument assuming it radiated uniformly in all directions?

e) What would be the relative decibel level of the direct sound and the reverb sound at 15 m from the instrument (near the back of the room)?

20. If an orchestra consisted of instruments of the following average capabilities in sound level production in a room;

5 instruments of 85 db each
3 instruments of 88 db each
1 instrument of 92 db

what would be the decibel level in the room if all 9 instruments were playing at once?

21. A room 10 meters wide, 15 meters long and 4 meters high is constructed with the following acoustical properties;

Front wall of plaster of acoustical abs. coefficient 5%
Back wall of brick of 2%
Side walls of concrete of 1.5%
Ceiling of acoustical tile of 40%
Floor of hardwood of 5%

What would be the reverberation time with 100 people sitting on 100 chairs in this room? (The effective area of a person is assume to be .5 OWU (open window units) and of the particular type of chair in this room, .12 OWU).

How much curtain material of absorption coefficient 35% would have to be added to the room to drop the reverberation type by 0.25 sec? What would be the effect of carpeting the room with a carpet of 55% absorption coefficient?

Answers;

1. a)895m; 3. b) 120m; 4. a)69 cm b) 2ms c)83 Pa d) 132 dB e) 34.4 cm, 1ms, 41.3 Pa, 126 dB; 5. In phase, when wall is at center of its motion, when wall is at maximum of its motion, 0.4 Pa, 0.28 Pa, 83 dB; 6. 5.5 x 10^{--7} J/$m^{3;}$ 7. 2 x 10^{-3} W/$m^{2,}$ 93 dB, 3.2 x 10^{-4} W/$m^{2,}$ 85 dB; 8. a)11.6 ms b)7.4 dB c) 10.4 dB; 9. a)24.5 ms b)4.1 dB; 15. 684, 20 s; 16. 20,10, I α $p^2$1.2 s; 19. a)5 mWatt, b)5.8 x 10^{-5} W/m^2, c) 78 dB, d) 2.6 m, e) -15 dB; 20. 97 dB ; 21. 0.72 s, about 200 m^2, about the same as the curtains.

CHAPTER 5

THE SOUND OF A VIBRATING SURFACE

A point source which radiates uniformly in all directions may be the simplest type of sound source that one can have in a room but it does not resemble very closely the typical musical sound source that one will have in a room. Such a source will not usually radiate uniformly in all directions.

This effect is not so noticeable in everyday life. For example, a loudspeaker which is rotated as it is sounding a musical note, (see fig. 5.1) will not present a noticeable variation as it rotates, even when the note is of fairly high frequency. However, a microphone placed near the speaker while these notes are being sounded will record a distinct variation in amplitude of sound, particularly for the higher frequencies.

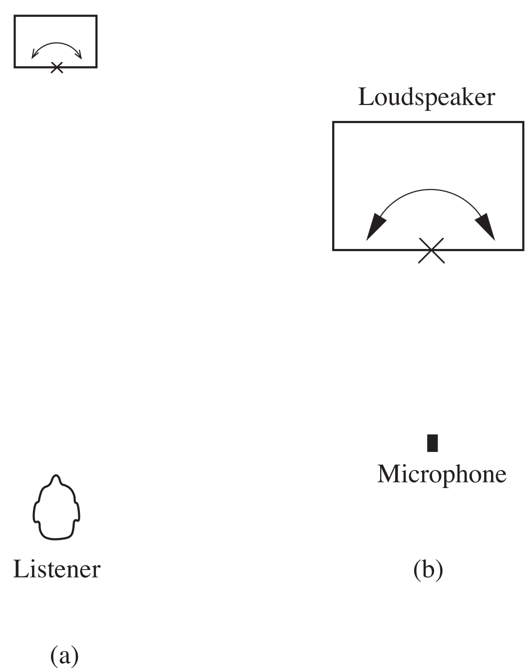


Figure 5.1 The effect of rotating a loudspeaker in a room while it is sounding a note. A human being seated as shown in (a) will usually not notice a significant difference as the speaker rotates but the pick-up on the microphone near the speaker (b) will be quite noticeably different, particularly for high frequency notes.

Most musical sound sources behave in this fashion. In fact a loudspeaker is designed so as to minimize the variation of sound level with direction from the speaker. The reason that the effect was not as noticeable for the far listener is again because of the room reverberation. The greater part of the sound intensity of the note at the listener is due to the reverberant sound and only a small part of it is due to the direct sound from the speaker and, in the case of a sustained note, the ear has no way to discriminate between the direct and the reverberant sound. Since the level of reverberant sound in a room usually depends very little on the original direction of the sound from the speaker the level of sound heard by the listener changes very little as the speaker rotates.

This presents another problem for the recording engineer. The sound picked up by a microphone close to a source will have variations in intensity for different frequencies depending on the direction of the microphone relative to the source. This can be demonstrated by making a simple recording with a microphone close to a speaker's head. By moving the microphone around the speaker's head while the recording is being made, a quite noticeable change can be detected from when the microphone is in front of the speaker to when it is behind. When the microphone is behind, the speech is much harder to interpret than when the microphone is in front (see fig 5.2).

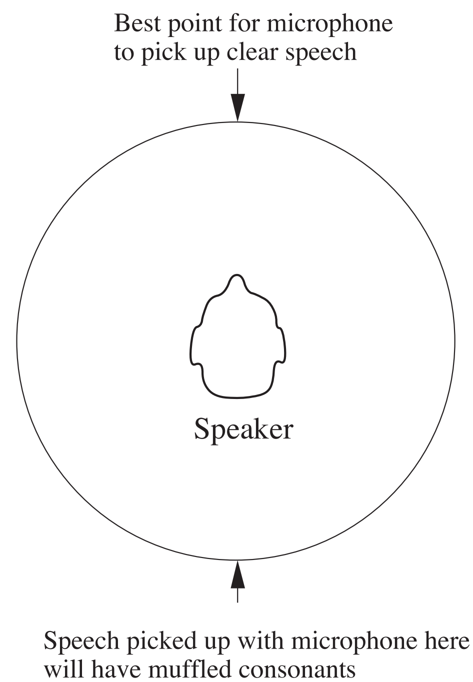


Figure 5.2 The effect of rotating a recording microphone around a persons head while the person is speaking. The speech will be much less intelligible when the microphone is behind the person's head than when it is in front.

The reason for this is that the high frequency components of a persons voice, those which contain the information distinguishing the consonants, radiate predominantly forward while the low frequency components are more uniformly radiated in all directions. The speech picked up behind the persons head will therefore contain too much of the vowel sounds compared to the consonants. In everyday life, this is why a polite person directly faces a listener.

The effect is not so noticeable in a typical room when one is farther from the sound source. Again, this is because most of the sound power reaching a person in such a room will be reverberant sound which will contain all of the frequencies anyway. What is affected is the direct sound which is used primarily to determine the direction and the nature of the source. Since we generally have good visual clues in this regard, the missing clues in the direct sound are not so noticeable. A perhaps too familiar example is that of a professor who lectures facing the blackboard.

However, while it is more difficult to understand a person speaking with his or her back to you, it is still usually possible. When the sound is recorded and played back in another setting however, the visual clues are removed and the deficiencies of the high frequency components of the sound become much more apparent.

The same phenomenon can produce a recorded sound which is much harsher than the natural sound of a musical instrument if the microphone is placed in a position which picks up more of the high frequency components than would normally be heard by a listener. A particular example is the trumpet which radiates its very high frequency components directly forward. When a player is playing a note with a lot of these high frequencies, the trumpet is normally directed upward, away from any particular listener in the audience. If a recording microphone happens to be placed in line with the trumpet when such a note is being played, quite an unnatural effect will be recorded.

The frequency at which a pronounced directional pattern starts to show for a sound source is that for which the wavelength of the sound has approached the source dimensions. This is a common feature of sound sources. It is very important for a recording engineer to know the actual directional pattern of a musical instrument in its different registers (i.e for its various frequency components).

5.1 Polar Diagrams

5.1.1 Decibel Plots

The way directional patterns are usually described for the sound from a particular musical instrument. or even for loudspeakers themselves, is to draw a polar diagram representing the intensity in any direction. These are diagrams on so-called "polar" graph paper with a circular coordinate system. An example of such a diagram for a typical sound source is fig. 5.3 obtained from the data shown in Table 5.1.

Such data are obtained by having the source operate in an anechoic chamber (a room with no perceptible sound reflections from the walls) and measuring the sound intensity in different directions at a chosen fixed distance from the source. The distance chosen is usually limited by the dimensions of the anechoic chamber and the dimensions of the source. A common distance is about 2 meters but for smaller anechoic chambers it will be often 1 meter. Because the difference in the pattern for different frequencies of sound is very important, the measurements will be performed at different frequencies by either having the source produce a pure tone, as in the case of loudspeaker testing, or as in the case when testing musical instruments, having the measuring instrument separate the sound received into frequency components.

TABLE 5.1

Angle (deg)	dB	Int. (Watt /m^2) $\times 10^9$	Press. (Pa) $\times 10^4$	Angle (deg)	dB	Int. (Watt /m^2) $\times 10^9$	Press. (Pa) $\times 10^4$
0	69	7940	574	180	30	1	6.4
10	70	10000	644	190	37	5	14.4
20	69	7940	574	200	39	8	18.1
30	67	5000	455	210	40	10	20.3
40	63	2000	288	220	38	6	16.2
50	57	500	144	230	36	4	12.8
60	50	100	64	240	30	1	6.4
70	30	1	6.4	250	52	158	81
80	42	16	25.6	260	58	630	162
90	48	63	51	270	60	1000	204
100	50	100	64	280	58	630	162
110	49	79	57	290	52	158	81
120	46	40	41	300	30	1	6.4
130	30	30	6.4	310	55	316	114
140	37	5	14.4	320	63	2000	288
150	38	6	16.2	330	66	4000	406
160	40	10	20.3	340	67	5000	455
170	38	6	16.2	350	68	6300	511

Measurements taken at 2 m from a sound source in an anechoic chamber. Average intensity = 1.50×10^{-6} Watt/m^2 corresponding to an acoustic pressure of 2.50×10^{-2} Pa.

Generally, what is measured and plotted are decibel levels such as in Fig. 5.3. One advantage of using decibel levels is that they allow a much larger range of sound levels to be represented on the graph. For example, the data in fig. 5.3 cover a range of 40 dB. This corresponds to a pressure range of a factor of 100. Plotting the information using a straight pressure scale would compress all the information in the lower 20 dB range into the inner 10% of the pressure scale (see fig. 5.4).

Another advantage of the decibel plot of intensity is that it more closely resembles the human sensation of the relative intensity levels in different directions (see any good text on psychoacoustics such as Roederer.)

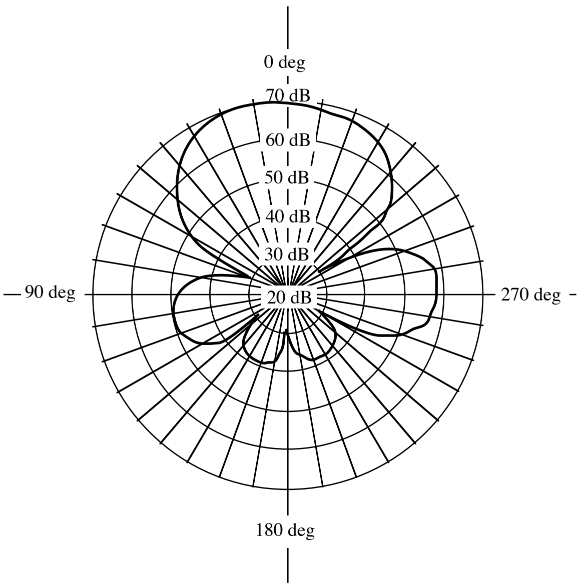


Figure 5.3 A polar diagram representing the variation of intensity with direction for a typical sound source. What is plotted on this graph is the actual intensity in decibels measured in different directions at a fixed radius from the center of the sound source.

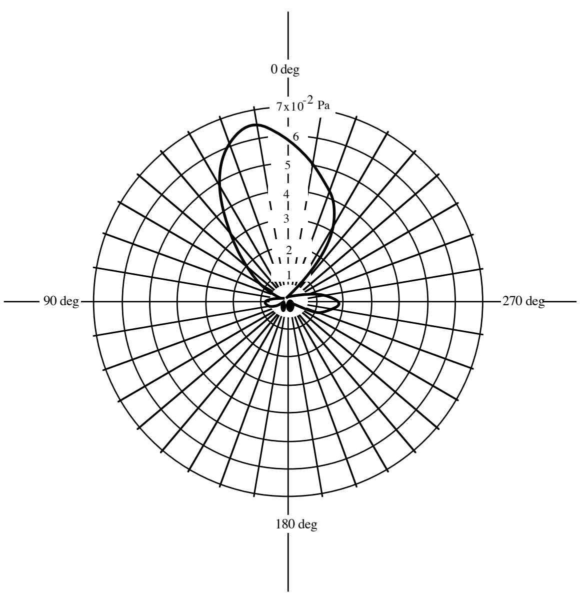


Figure 5.4 Polar diagrams representing the variation of sound pressure with direction for a typical sound source. The decibel plot shows much more of the information in the low intensity region.

5.1.2 Directivity Plots

Sometimes the directional characteristics of a source may be shown in sound pressure levels. Imagine the distance from the source at which the sound would have a particular pressure level. As one walked around the source, this distance would be greater for directions in which the sound propagation was favoured and would be less for directions which were not favoured.

A polar diagram representing the actual sound pressure in any direction can be easily converted into such a plot by noting that the pressure in direct sound increases in inverse proportional to the distance from the source. If the pressure at a given distance in a particular direction is, say, 1/10th of what it is at the same distance in a standard reference direction, usually taken as 0^o, then one would have to move in to 1/10th the given distance to get the same reference level pressure. Thus the diagram of fig. 5.4 representing the pressures can be reinterpreted as representing the relative distance from the source at which there will be a specified pressure level (see fig. 5.5).

One way of expressing this effect quantitatively is to plot the ratio of the distance on the graph in a given direction with the distance at which the same intensity would be received from a source of the same power that radiated uniformly in all directions (i.e an "isotropic" source) with the same total power. To obtain this diagram one has to have the total power radiated by the source. A figure good enough for acoustics work can be obtained by noting the intensity for about every 10 degree interval, calculating this in watts and taking the average for the complete circle. This gives the average intensity from which can be calculated the pressure level for this average intensity.

The average intensity of the polar plot we have been considering is 1.5×10^{-6} Watts /m^2, and that would be the intensity at 2 meters for an isotropic source of the same power as the actual source. The acoustic pressure at this intensity is 0.0250 Pa. For the actual source which gave 0.0650 Pa at 10^o, the ratio is then 2.60. The resulting plot of the ratio of the distance from the actual source to the distance for the same intensity from an isotropic source is shown in fig. 5.6.

Such a plot is called a "directivity" plot and is of interest in recording engineering because of its connection to the room radius for a musical instrument. The room radius (which is the distance from a source at which the direct sound intensity is the same as the reverberant sound intensity) will move in and out in proportion to the way the line on the directivity graph moves in and out. If it is important to be within the room radius in order to favor picking up the direct sound from the instrument, then the room radius for a given direction of the microphone relative to the instrument can be determined by using the directivity plot and the equation connecting the room radius with the room volume and reverberation time

$$r_{room} = 0.058\, \Gamma_{st} \sqrt{\frac{V}{T}}$$

where Γ_{st} is the directivity as determined from the directivity plot.

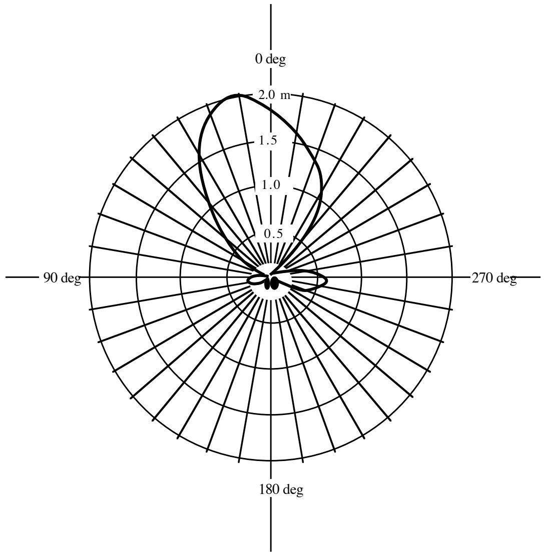


Figure 5.5 A polar diagram representing the variation of distance at which a given sound level will be received in various directions. It is obtained from the pressure polar diagram just by changing the radial scale from pressure to distance.

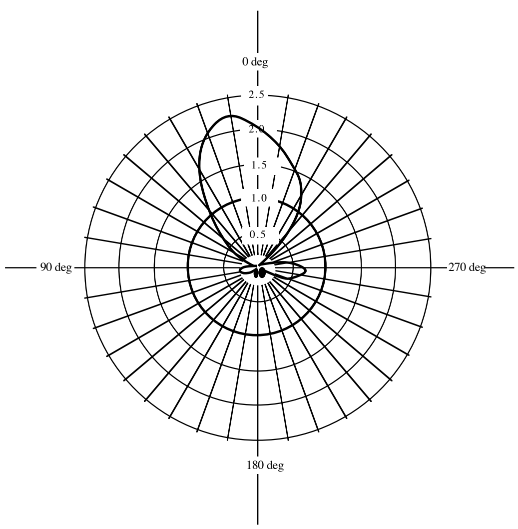


Figure 5.6 A polar diagram representing the ratio of distance at which a given sound level will be received in various directions to the distance for the same intensity for an isotropic source. This is sometimes called a "Directivity Diagram". The circle for the isotropic and the real sources providing the same intensity is shown heavy for reference.

5.2 Radiation Patterns of Some Standard Surfaces.

The directional pattern of sound can be calculated for any vibrating surface by using the methods that will be described in Chapter 8. This has been done for a variety of simple shapes and the results are available in standard works on acoustics (see for example, Olsen). There are two shapes which are of particular interest in music; that of a plane circle vibrating perpendicular to its surface and that of a segment of a cylinder which is vibrating in a mode in which the surface is expanding and contracting along radial lines. The results for the calculation of these two surfaces will be presented here.

5.2.1 The Radiation Pattern from a Vibrating Circular Surface

The radiation pattern for a vibrating circular surface is of great importance in many branches of physics and engineering. This is because it approximates the radiation of light or any other electromagnetic wave through a circular aperture. Its importance in music is that it approximates the radiation pattern of many musical instruments in which the sound leaves the instrument through a circular opening. The best examples of this are the brass instruments in which the sound radiates from the instrument through the bell. The sound pattern from such instruments is similar to that of a vibrating circle (piston) in an infinite baffle as shown in fig. 5.7 (The reason for including the infinite baffle surrounding the circle will be explained later.)

Plots of the radiation patterns of a circular vibrating surface at different frequencies of vibration are shown in Fig. 5.7.

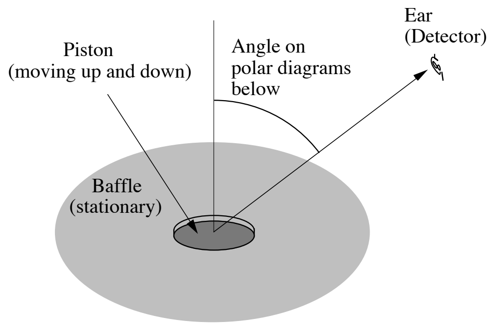
Note again the general characteristic; the lower frequency tones have much less directional characteristics than do the higher frequency tones. The particular frequency at which the wavelength is 1/1.22 times the diameter of the source is important in that it is the lowest frequency at which a distinct node can occur (no sound whatever radiated in a particular direction). At this particular frequency, no sound is radiated in the direction parallel to the surface. As the frequency is raised above this level, the nodal line moves forward until at a wavelength equal to 1/4 the surface diameter, it is at an angle of only 18^o with the forward direction from the surface. Meanwhile, three new nodal lines have moved into the graph at 33.8^o and at 54.3^o.

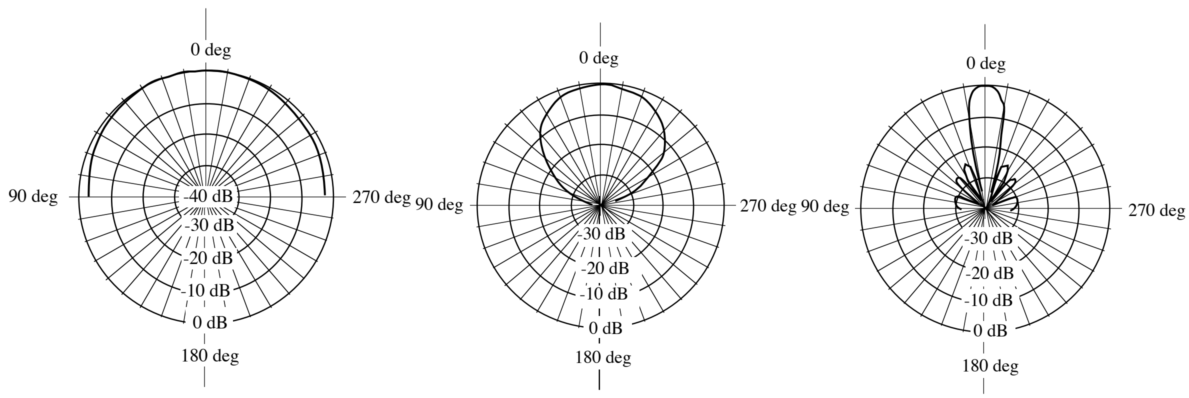
These patterns are, of course, symmetrical about the axis of the vibrating disk, the nodal "lines" actually being the intersection of a nodal cone with the plane of view of the diagram. (see fig. 5.8.)

The actual intensity of radiation from a disk of radius R in a direction θ compared to that in the direction perpendicular to the surface is given by the formula

$$I = I_o \left(\frac{2\, J_1\left(\frac{2\pi R}{\lambda} \sin \theta\right)}{\frac{2\pi R}{\lambda} \sin \theta} \right)^2$$

where J_1 is the Bessel function of order 1 and I_o is the intensity in the direction perpendicular to the surface.





$\lambda = 2 \times$ diameter $\lambda = 1.22 \times$ diameter $\lambda = 0.25 \times$ diameter

Figure 5.7 Polar diagrams representing the directional patterns of the radiation from a circular surface for different frequencies of vibration of that surface. Pattern on the left is for wavelength = 2 times surface diameter, in center 1/1.22 times and on right 1/4 times the surface diameter. Zero degrees is forward, perpendicular to the surface.

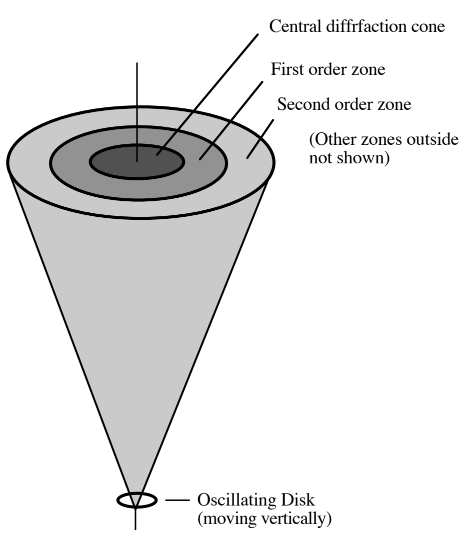


Figure 5.8 Diagram representing the directional patterns in three dimensional space of the radiation from a circular surface for a frequencies of vibration of that surface with a wavelength equal to 1/4 of the surface diameter.

The fact that a vibrating surface can produce no radiation whatsoever in a particular direction is perhaps a puzzling feature of wave propagation, but it is one of the characteristic features of wave propagation. If one has such a phenomena, (no propagation in special directions) then this is regarded as clear evidence that the phenomena is associated with wave propagation. The discovery that light through a hole exhibited such a characteristic was regarded as proof that light was a wave propagation. It took many more years of scientific research before it was discovered that light was in fact an electromagnetic wave.

5.2.2 The Radiation Pattern from a Vibrating Cylindrical Segment

In some music instruments, a significant portion of the sound radiates from a curved surface. The most important example is the sound from the classical stringed instruments such as the violin, viola, cello and bass. In certain frequency ranges, a great deal of the sound radiates from the vibration of the front and back surfaces of the instruments. The sound from such a source can be roughly approximated by that from a cylindrical segment vibrating along radial lines (see fig. 5.9).

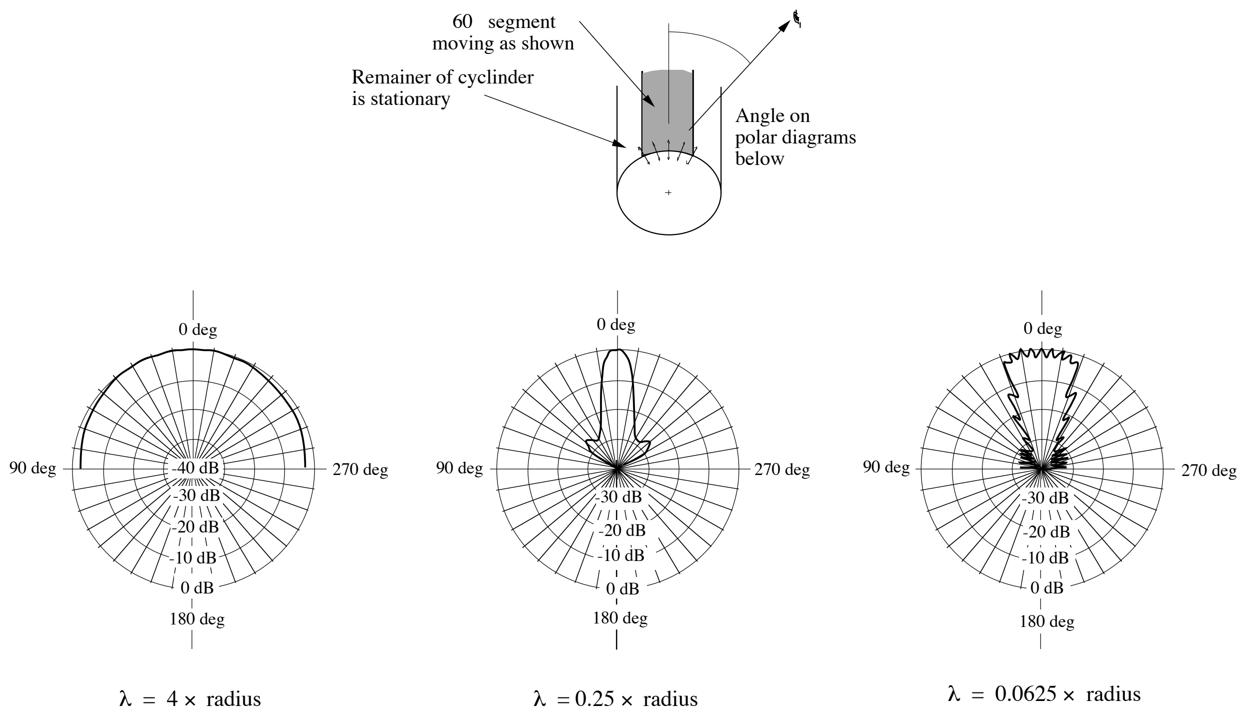


Figure 5.9 The pattern of radiation from a vibrating cylindrical segment where the vibration is along radial lines to the surface. The patterns are for a 60^{o} vibrating segment. All points on the segment move in and out along the radial lines perpendicular to the surface.

The interesting feature of the radiation from such a surface is that the pattern is broad for both very low and very high frequencies. At low frequencies, as for any source, the pattern extends uniformly in all directions. At high frequencies, the pattern approaches the 60 angle subtended by the surface. Note however, the rather complicated pattern for intermediate wavelengths.

5.2.3 The Radiation Pattern of Musical Instruments

The radiation patterns of the simplest possible surfaces discussed here are themselves rather complicated. Acoustic musical instruments are generally made up of many sources of sound in any one instrument. Furthermore, the frequencies of sound radiated by the instruments are generally of wavelengths near some dimension on the instrument; in other words at frequencies where the directional properties are most pronounced. It is to be expected than that the radiation pattern of musical sounds from musical instruments will be extremely complicated, requiring immense computing power to accurately predict. This is indeed the case. However, the actual radiation patterns are of great importance to the recording engineer and so they have been extensively measured for the classical instruments. One of the most extensive collections in one source is that in chapters 4 and 7 of Meyer.

Exercises and Discussion Topics

1. Explain the meaning of the "Directivity Factor" as used by Meyer to express the directional characteristics of musical instruments. Why would it appear in the equation for room radius of a musical instrument as follows;

$$r_{room} = 0.058\, \Gamma_{st} \sqrt{\frac{V}{T}}$$

2. Suppose one measured 107 db as the sound intensity 1.5m in a particular direction from a 0.5 Watt sound source in an anechoic chamber.

   a) What would be the distance from a 0.5 Watt isotropic source at which the sound level would be 107 dB?

   b) What is the directivity of the source in the direction measured?

3. What is the connection between the directivity factor and the polar contour plot for equal intensity of sound? What extra piece of information is needed to get directivity factors from such a contour plot?

4. A sound source is being probed in an anechoic chamber. In moving the sound level meter around the source, the following values are obtained for the distance at which the sound level is 90 dB.

Angle (deg.)	Dist. (cm)	Angle (deg.)	Dist. (cm)	Angle (deg.)	Dist. (cm)
0	164	120	80	240	64
10	153	130	72	250	73
20	140	140	70	260	86
30	132	150	71	270	96
40	99	160	75	280	102
50	90	170	78	290	106
60	88	180	80	300	108
70	88	190	78	310	116
80	92	200	76	320	131
90	92	210	70	330	134
100	90	220	64	340	142
110	86	230	62	350	162

Plot this data on a polar graph. (Copies of polar graph paper may be obtained from the university book-store). Draw a smooth curve through the points. If the distance from an isotropic source of the same power was 104 cm, what would be the directivities at 0, 90, 180, 270 degrees?

5. Measuring the sound intensity of a 1000 Hz tone one meter from a source in an anechoic chamber gave the following numbers;

Angle (deg.)	Dist. (cm)	Angle (deg.)	Dist. (cm)	Angle (deg.)	Dist. (cm)
0	100	120	75	240	74
10	99	130	74	250	72
20	97	140	73	260	70
30	93	150	72	270	65
40	87	160	70	280	66
50	82	170	68	290	68
60	75	180	65	300	71
70	72	190	70	310	78
80	70	200	72	320	86
90	73	210	73	330	93
100	74	220	74	340	97
110	75	230	74	350	99

Plot this data on a polar graph and draw a smooth curve through the points.

6. What is the connection between the polar graph of dB level for direct sound from a musical instrument and the contour line for a uniform intensity level around the speaker? What is the connection between decibel changes for different directions at a uniform distance from a source and the movement of the contour line about the source for a uniform intensity level?

7. Using the fact that the direct sound from a source varies with distance according to the equation

$$I = I_o \left(\frac{r_o}{r}\right)^2$$

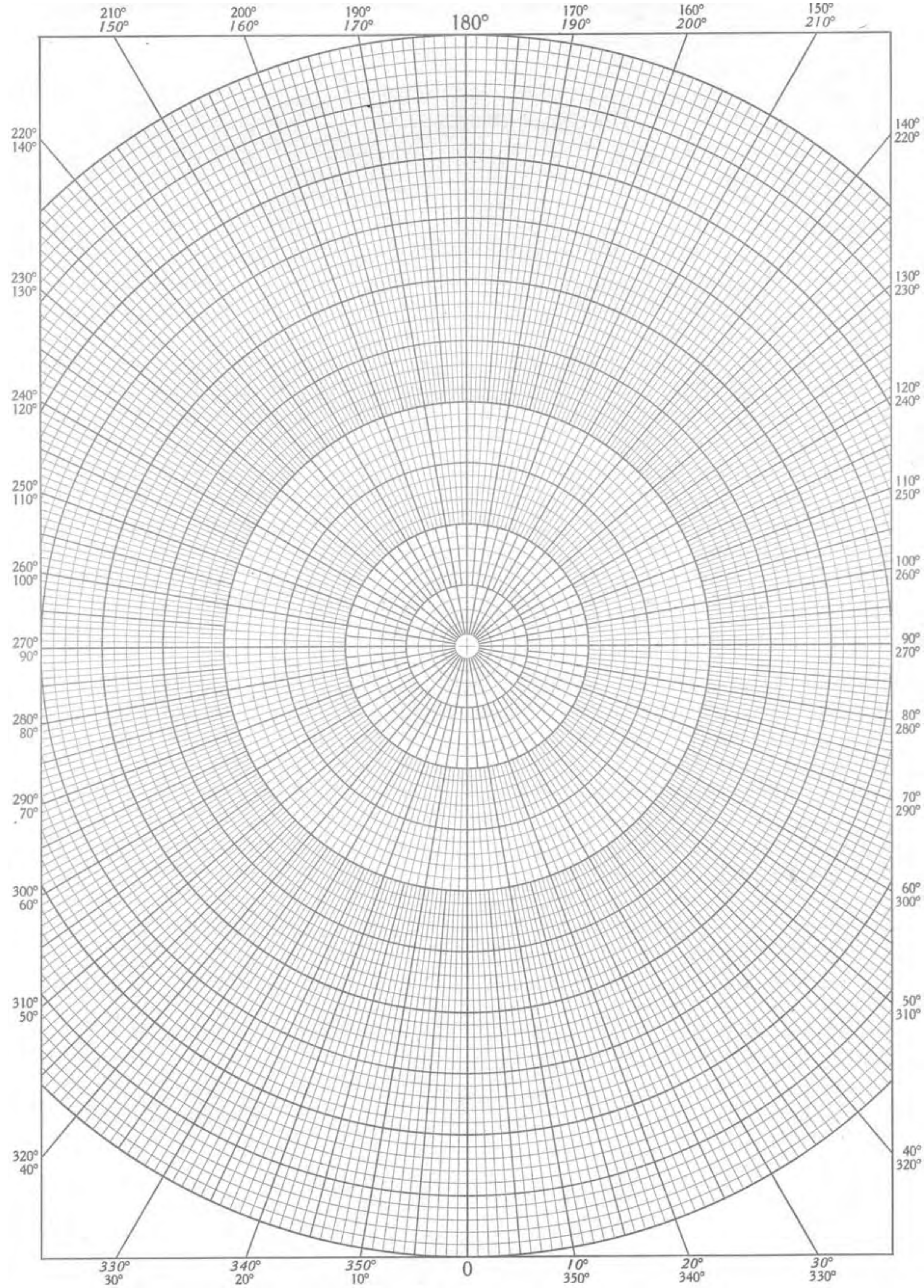
convert the data of problem 4 to obtain a polar graph of the decibel level at 1 m.

8. Convert the data of problem 5 to obtain a contour plot for 100 dB.

9. Discuss the relative usefulness of the polar graph of dB levels at some reference distance versus the directivity factor diagram (or a contour diagram from which it is derived).

Answers

2. 0.89m, 1.69; 4. 1.58, 0.88, 0.77, 0.92; 7. An example: at 0°, 94.3 dB; 8. Examples: 0°, 1m; 180°, 1.8 cm; 50°, 12.6 cm.


210°
150°
200°
160°
190°
170°
180°
170°
190°
160°
200°
150°
210°
220°
140°
140°
220°
230°
130°
130°
230°
240°
120°
120°
240°
250°
110°
110°
250°
260°
100°
100°
260°
270°
90°
90°
270°
280°
80°
80°
280°
290°
70°
70°
290°
300°
60°
60°
300°
310°
50°
50°
310°
320°
40°
40°
320°
330°
30°
340°
20°
350°
10°
0
10°
350°
20°
340°
30°
330°

CHAPTER 6

SOUND DIRECTION AND RELATIVE PHASE

This chapter introduces the physics of why sound waves travel in particular directions and why the sound patterns from relatively simple sources can be complicated.

To start, consider the direction of sound from the simplest possible source; a sphere which has a surface which is expanding and contracting in an oscillatory fashion at a regular rate;

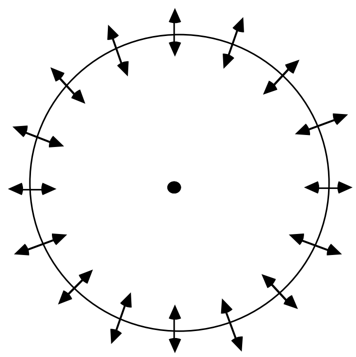


Figure 6.1 Schematic diagram of the surface motion of a sphere that becomes an isotropic source of sound. All points on the surface move in and out at the same amplitude and frequency and in phase. This mode of oscillation of a sphere is sometimes referred to as the "breathing mode"

Such a source pushes air directly outwards from its center and this air pushes against air which is just outwards from it and so on ad infinitum. Because of the symmetry, it is easy to agree that such a source will radiate sound equally in all directions. To an observer outside the sphere, the sound would appear to come from a point at the center of the sphere. Such a source would therefore be equivalent to the simple point source considered in Chapter 4.

However, what if the source is more complicated than this extremely simple point source. All musical instruments are more complicated than point sources and therefore more complicated sources must be studied to understand how musical instruments actually radiate sound. How are such sources to be studied?

6.1 The Superposition of Sources

The basic principle by which complicated wave sources can be studied was invented by Huygen in the 17th century. Any source which is very small compared to the wavelength of the sound that it radiates will radiate isotropically as if from a point at its center. The surface of a complicated source can therefore considered to be made up of many small pieces with each piece radiating as a point. The trick then is to add up the effects of all these little point sources radiating at the same time. This principle is called "superpostion of sources".

As a start to how small isotropic sources add, consider four small spherical sources each radiating spherical waves when operated on their own. Such sources can be created by putting small speakers (about 4 cm diameter or less) in small glass jars and sealing the mounting joint with modeling clay. (The result is a kind of "acoustic suspension" speaker in which the air in the glass jar behind the speaker is compressed and expanded by the speaker cone motion but does not radiate sound into the room. This will be shown to be an important consideration later.)

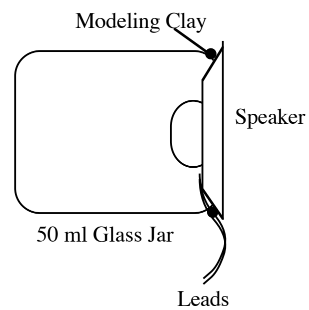


Figure 6.2 Speaker mount for a small source that is isotropic to about 2500 Hz

By moving the microphone around one of these speakers operated alone, it is possible to check that it is in fact radiating sound almost equally in all directions.

A particularly simple arrangement of four of these speakers is to have them mounted in a line equally separated by about 9 cm and each emitting a 2000 Hz tone, the speakers all being fed from the same source and all being connected together so that each speaker is exactly in phase with each other and each producing the same sound power. By placing the microphone close to each speaker in turn, it is possible to check that each speaker is indeed putting out sound of about the same amplitude and in phase.

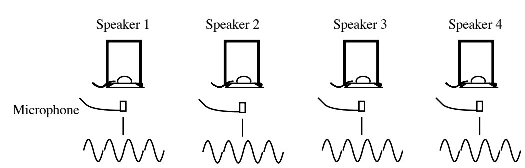


Figure 6.3 Speaker set-up. The center line of each speaker is separated by 9 cm from its neighbor(s).

Now suppose the same microphone is used to probe the sound level at much greater distances from the speaker combination. Here it is important to remember that we are trying to probe the direct sound from the speakers. For this the microphone must be well within the room radius for the sound. For 2000 Hz the room radius will be about 2 m. However, the microphone must also be far enough from the speakers to receive sound almost equally from all four. It would seem that a distance of about one meter would be most appropriate.

Placing the microphone at this distance from the speakers on a line directly in front of the speakers results in a fairly good signal being received. Placing it in line to the side of the speakers, we see hardly any significant signal.

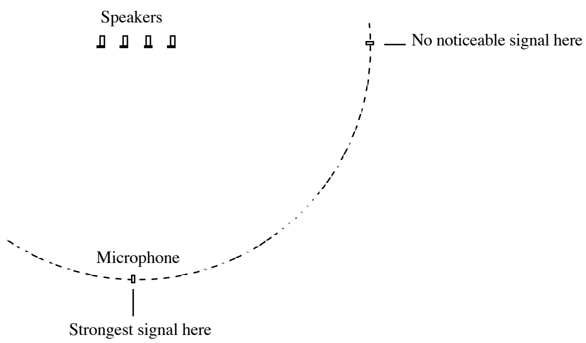


Figure 6.4 Probing the sound field at 1 m around the speakers, all speakers in phase.

This is perhaps surprising. What happened to the sound travelling sideways from the speakers?

A further puzzle arises when the speakers are connected so that speakers 2 and 4 in the array are in opposite phase to speakers 1 and 3 (fig. 6.5).

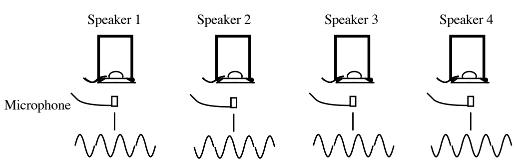


Figure 6.5 Speakers 2 and 4 reversed in phase

Again it can be checked that the speakers are in opposite phase by probing with the microphone close to each speaker.

Probing at one meter, it will now be seen that there is a strong sound to the side of the speakers. On the other hand, there is now no significant sound to the front of the speakers (fig. 6.6).

The direction of radiation of sound from this array of sources has been completely altered by merely changing the relative phase of the sources. When they were all in phase, they radiated in a direction perpendicular to the line of the array. When they were phased so that there was a delay of half a cycle between speakers, they radiated in a line along the line of the array. What is going on?

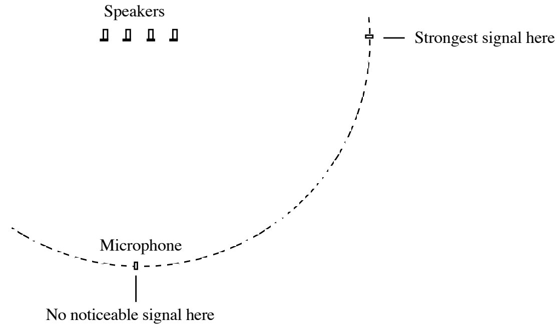


Figure 6.6 Probing the sound field at 1 m around the speakers when speakers 2 and 4 are antiphased.

To understand this phenomenon, one has to consider how sources of different phases add. In the simple case set up here, the sources are either in phase or 180º out of phase. (More complicated phase differences will be considered later.) Such sources either simply add or cancel (see fig. 6.7).

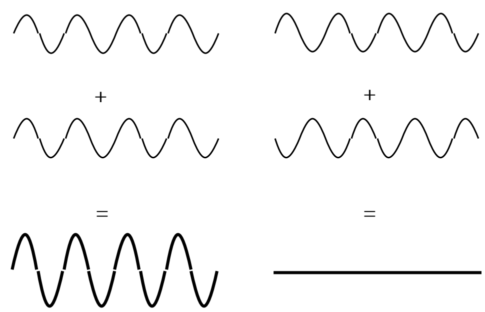


Figure 6.7 Addition and cancellation of oscillations

Now consider the phases with which the sounds from the speakers arrive at some point in space around the speakers. Take the simplest case first; that of a point directly in front of the speakers when the speakers are all connected in phase (fig. 6.4). At this point the sounds from each of the speakers will arrive in phase. They will therefore all add together giving a good sound signal.

However, at the point to the right of the array the sounds from the speakers will not all arrive at the same phase. This is because the sounds from the farther speakers will be delayed. In fact, the dimensions have been set up and the frequency deliberately chosen so that the speakers are one half-wavelength apart. Therefore, at a point to the right of the array the sound from speaker 3 in the array will be one half oscillation behind the sound of speaker 4, the sound of speaker 2 will be one half oscillation behind 3 and the sound of speaker 1 will be one-half oscillation behind speaker 2.

Speaker 3 will threfore cancel the sound of speaker 4 and speaker 1 will cancel the sound of speaker 2. This results in no sound in the direction of the line of the array.

Consider now the case when speakers 2 and 4 were delayed in phase by 180° (fig. 6.6). Now when the sound of speaker 3 arrives at a point to the right of the array the delay due to its extra distance has just allowed it to match up with the output of speaker 4 which had been already delayed by 180° at its start. Similarly for speakers 1 and 2. All four speakers therefore reinforce each other along a line to the right.

However, for points directly in front of the array, the sounds from the four speakers will arrive with their original phase differences. tjhis is because they are all delayed by the same amount because their sounds had to travel the same distance through air. The sounds will therefore cancel.

In summary, the direction of the sound propagating from this array of sources can be switched simply by changing the relative phase of the sources making up the array. All that has to be done to get the sound to swing from going forward to going to the right is to change the relative phases of the sources.

In the example shown, the direction of propagation would swing by exactly 90°. Can the direction be aimed it at somthing between straight forward or to the side?

6.2 Directing Sound to Any Desired Direction

To see that it is possible to direct sound to intermediate angles, consider the result if speaker 2 was only 90° in phase behind speaker 1, speaker 3 90° behind speaker 2 and speaker 4 only 90° behind speaker 3. Then the sound would propagate in the direction 30° to the right (see fig. 6.8).

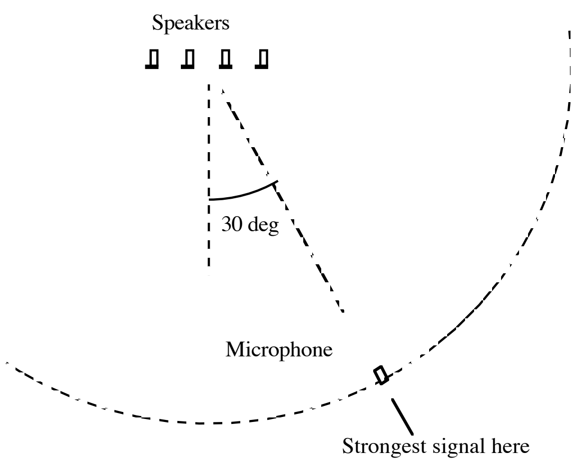


Figure 6.8 Each speaker is out of phase by 90° to its neighbors; speaker 2 90° behind 1, speaker 3 90° behind 2 and speaker 4 90° behind 3.

To understand this, consider sound propagating from the various speakers in this direction (fig. 6.9). The sound from speaker 1 would have travelled exactly the right distance to have delayed its phase enough to be in step with the sound starting from speaker 2 (i.e it has travelled one-quarter wavelength giving a 90° phase delay). The same thing happens when these sounds travel the distance from speaker 2 to speaker 3 and from speaker 3 to speaker 4; they all are delayed by the extra distances they travel so as to be in step with the more delayed sources.

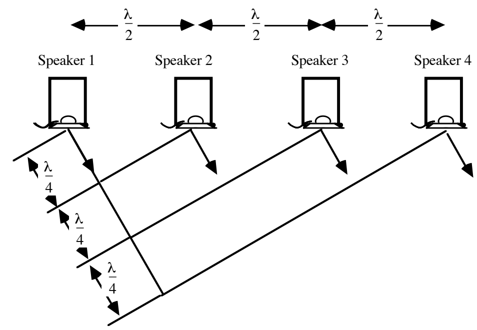


Figure 6.9 The geometry of wave propagation at 30 degrees to the normal to an array of four sources separated by half a wave-length..

It is not necessary to understand the mathematics of this to understand the principle. However, for those readers that are interested, the angle at which the sources will propagate is given by

$$\alpha = \arcsin\left(\frac{\Delta\phi}{d}\right)$$

where α is the angle between the line of propagation and the normal to the line of the sources, $\Delta\phi$ is the phase delay from one speaker to the next in fractions of a cycle and d is the distance from one speaker to the next in fractions of the wavelength of the sound. In the case considered, the time delay from one speaker to the next would be one quarter of a cycle and the speaker separation would be one-half a wavelength giving an angle whose sine is 0.5. That angle is of course 30°. Taking one more case as an example, if the phase delay were 0.4 cycles and the speaker separation 1.2 wavelengths, the angle would be that whose sine is 0.3. That angle would be 17.5°.

Thus there is a way of setting up any direction we wish for wave propagation from an array of sources. All that is required is to be able to adjust the relative phases of the sources. This is a technique commonly used for radio transmission were the phase delay between sources (the array of antennae commonly making up the radiating system) is easily controlled electronically. In this way radio waves can be beamed to the areas of greatest population density from a given fixed array of antennae. Furthermore, if the p[opulation densities change the direction can be easily changed to accomodate this by adjusting the phases of the sources. Another example in modern technology is the scanning of radar beams where the scanning is not done by rotating radiators, as can be often seen on many ocean going vessels, but by electronically switching phase delays between small sources. (See the February 1985 issue of Scientific American). In this way the scanning times are not limited to the mechanical speeds with which you can spin the radiator but only to the speed with which you can electronically switch the phases. This can be very fast.

6.3 Directional Microphones

Another example of the use of relative phases to determine the direction of a wave is in the highly directional microphone. This device consists of a line of small omnidirectional microphones all connected so as to add to the input of the same amplifier (fig. 6.10). However, before being added to the amplifier input, the outputs of these microphones all go through delays, each microphone having its own delay. This delay is set to be as close as possible to the time it would take for sound to pass from one microphone to the next if it were travelling straight along the line of the array. For microphones 10 cm apart, this time delay from one microphone to the next, the microphone farthest upstream being the most delayed, would be 0.29 ms.

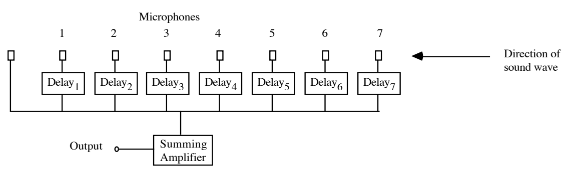


Figure 6.10 Schematic of a highly directional microphone made up of a set of small omnidirectional microphones and electronic delay circuits for each microphone. The outputs of the individual microphones will only add coherently for sound coming in the line of the microphones. For a separation of 10 cm between microphones, the delays would be set to increase by 0.29 ms from one to the next.

In such a system, only waves travelling exactly along the direction of the array will cause the outputs of all the microphones to arrive at the amplifier at the same time and therefore reinforce each other.

In some older versions of directional microphones the delays are achieved by using hollow tubes to bring the sound to a region where the pressures are added before being registered by a single microphone.

6.4 Phase Relationships and Wave Direction

Thus wave direction is fundamentally connected to phase relationships within the wave. For those with a previous introduction to the physics of waves, this is perhaps familiar. Consider for example any wave travelling through a medium. It is perhaps best to visualize a water wave travelling along the surface of the ocean on which there is a line of floating objects. Fig. 6.11 represents a view looking down on such a water wave with crests moving in various directions relative to a horizontal line of objects. Each of these objects will be oscillating up and down as the wave passes them. However, there will usually be a phase delay from one object to the next due to the time it takes for the wave to travel from one to the other.

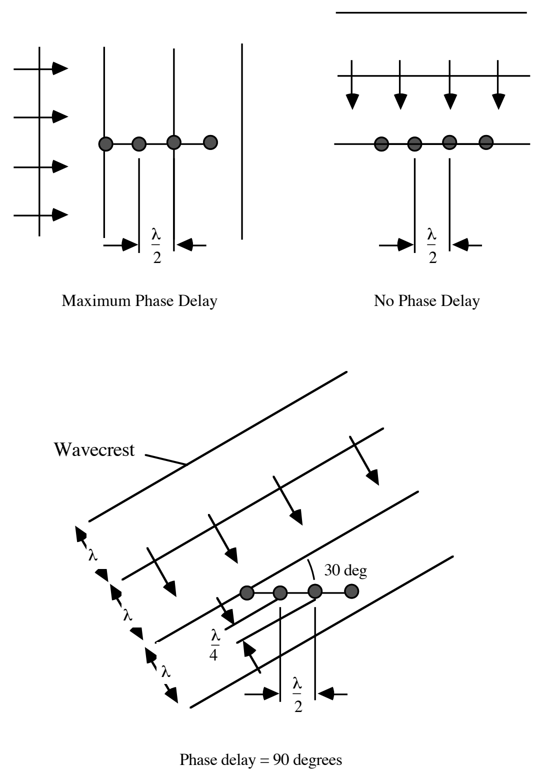


Figure 6.11 Diagram representing four regularly spaced pieces of flotsam in a regularly spaced series of water waves. The angle between the line of the pieces and the direction of the wave is shown for three case; in line with the wave direction, perpendicular to the wave direction and at 60 degrees to the wave direction. The phase delay in the oscillation of the pieces of flotsam will depend on the wave direction as shown.

Thus wave propagation can be considered as resulting in a specific phase relationship between objects experiencing that wave at different points in space. In this chapter, the point has been made that waves are propagated in specific directions if the correct phase relationships are established between many small sources that are producing the wave.

Here then is the basis of Huygen's idea about wave propagation. The two phenomena of wave motion producing phase relationships at points in its medium and points in the medium producing wave motion are in fact just two different ways of looking at the same thing; wave motion. Thus at any time a wave in space can be decomposed into a multitude of small sources at all points in the medium. These sources would then have, of course, the phase relationships produced by the wave in the medium at that time.

One is left with a conundrum; which is the cause and which is the effect? Like for most things in physics, this is a meaningless question; the cause and the effect are the same thing. They are just different ways of looking at a system. We then come to a general statement regarding sound propagation from an array of sources:

> **The intensity of a wave disturbance and the direction of its propagation at any point in space is determined by the strengths and relative phases of each of the elementary sources making up the total source of the wave.**

The elements making up the source of the wave can even be the oscillating segments of a medium that has been excited by a wave travelling through it. This is, in effect, how waves propagate.

In the next chapter we will look at how to use this principle to treat more general problems in wave propagation.

Exercises and Discussion Topics

1. Explain why 4 isotropic sound sources emitting a pure tone in phase and set in a straight line will give no sound in the direction of that line if the sources are equally spaced at half a wavelength for the tone.
2. Explain why the same sources in problem 1 will radiate strongly along the line but not at all perpendicular to the line when they are alternately phased.
3. Why is there no sound radiated at 30 degrees to the perpendicular in problem 1?
4. Explain the principle of the highly direction microphone based on delayed pickup from a group of isotropic microphones arranged in a line. Is the directional property of this microphone very dependent on the frequency of the sound being picked up? Why?
5. How are the relative phases of the oscillations in a medium through which a wave is passing related to the direction of the wave? Explain how this relationship can be used to set up a wave travelling in a desired direction.

CHAPTER 7

SOUND WAVE DIFFRACTION

The previous chapter dealt with the importance of the relative phase of combined sources. In this chapter a simple technique will be introduced with which you can conceptually deal with many sources of different phases. The full mathematical treatment of this subject is called Fresnel integration because it was first used by Fresnel in the study of the diffractive properties of light. While no doubt the most powerful way to treat the subject, here such a mathematical treatment would be inappropriate and could in fact hide the essential principles involved.

What will be introduced here is a graphical approach based on the concept of phasors as elemental sources of sound. As an introduction, consider again the simple four speaker array of sound sources that was studied in chapter 6 (see fig. 7.1).

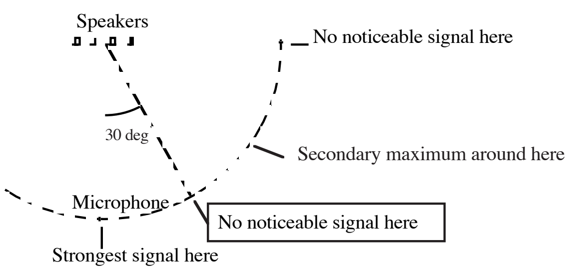


Figure 7.1 Probing the sound field at 1 m around the speakers, all speakers in phase.

Again all the four speakers are in phase but now make a more careful probing of the sound field around the speakers. It will be noticed that there is no sound in a direction 30° to the straight forward direction!

By moving the microphone a bit beyond this point, it will be seen that the sound level grows again after 30°, reaching a maximum somewhere around 45° before it falls to zero again at 90°.

What has happened to cause this cancellation of sound at 30°? One simple way to look at this problem is to recognize that at 30° the sounds from speakers #1 and #3 will arrive at 180° relative to each other (#1 is one half-wavelength behind #3 and the same for speakers #2 and #4). Therefore speakers #1 and #3 cancel as do speakers #2 and #4. No sound is therefore propagated in this direction.

Are there any other regions where such cancellation can occur? How can one determine how much sound is propagated at the angle of 45° compared to straight forward? Indeed, how can one calculate the relative sound level that would result from this speaker combination in any arbitrary direction that we wish to know about? The simple way of only looking for the angles at which sounds reinforce or cancel does not answer such questions and such questions must be answered if one is to understand the directional properties of sound sources

A more powerful way of looking at this general sort of problem is through the concept of phasors.

7.1 The Treatment of Oscillations as Phasors

A phasor is a vector way of representing the amplitude and phase of an oscillation. The length of the phasor is the amplitude of the oscillation and its direction on the paper on which it is drawn is determined by the phase angle of the oscillation. By convention, an oscillation of zero phase angle is drawn as a phasor that points to the right. Oscillations with positive phase angle have rotations counterclockwise (as do positive angles in trigonometry) and oscillations with negative phase angles are shown as phasors rotated clockwise. Representative phasors are shown in Fig. 7.2.

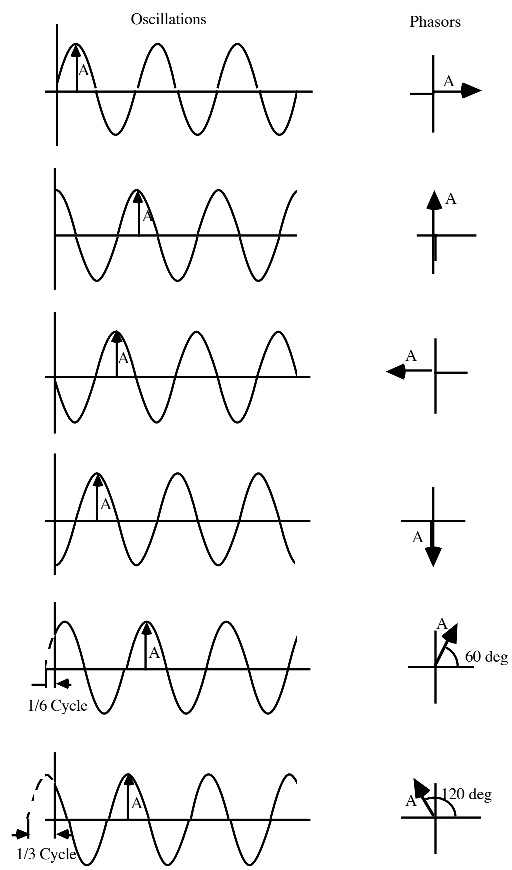


Figure 7.2 Representative phasors of oscillatory motion

Look at them carefully until you understand exactly the relationship between the arrows and the waves. One way of visualizing the connection is to notice that the intercept of the waveform of the oscillation with the *y* axis is the same as the projection of the phasor on the *y* axis.

The great advantage of the phasor way of looking at oscillations is that now one can add oscillations of different amplitudes and phases by simply adding the phasors. This is of course vector addition which can be done graphically by drawing the vectors head to toe as shown Fig. 7.3.

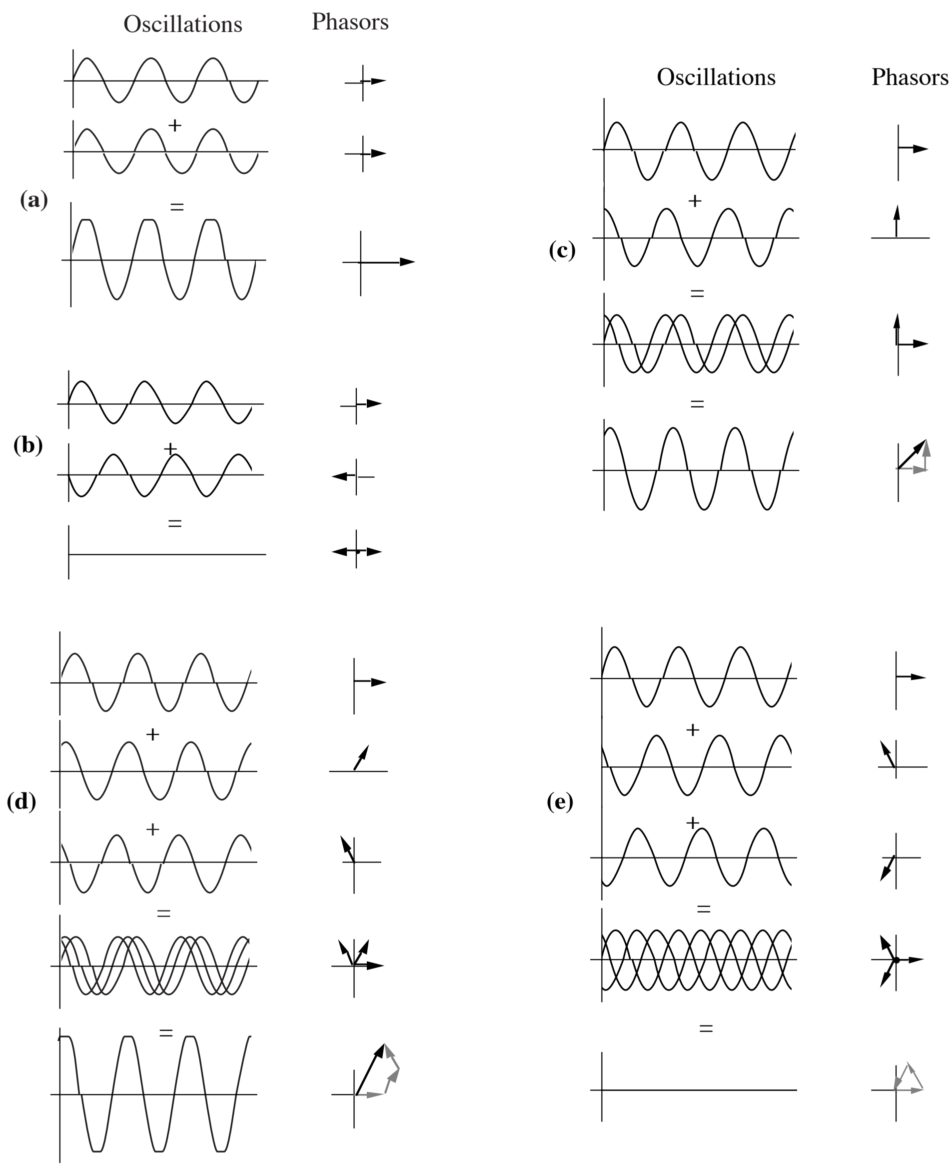


Figure 7.3 Addition of some representative oscillations using the phasor method. (a) is a very simple case; that of adding two equal oscillations that are in phase with each other. The result is an oscillation of the same phase and twice the amplitude. (b) is that of two equal amplitude oscillations that are out of phase with each other. The result is, of course, zero. (c) is two oscillations which are only 90° apart in phase. The graphs of the oscillations tell us only that the result of adding these two oscillations is in fact another oscillation of the same periodicity as the two originals. The phasor addition tells us that it will have an amplitude 1.414 or √2 of the two components and a phase of 45 degrees. (d) is an even more complicated situation where it can be seen that three oscillations 60° apart in phase can be added to give an oscillation which is twice the amplitude of any one of the components and has the phase of the middle oscillation. The final example is the case of three oscillations separated in phase by 120°. Here it can be seen that the result of the addition is zero. Thus oscillations can add to zero total amplitude even when there are no oscillations exactly out of phase with each other. All that is required is that the vector sum of the phasors representing the individual oscillations be zero.

7.2 Phasor Treatment of a Four Speaker System

We now have another way of understanding the cancellation of the sound of the four speaker system at 30°. Consider first a point directly in front of the speakers. It can be seen that the phasors of the oscillations of the sound arriving at this point from the four individual speakers all line up (fig. 7.4)

Now consider what happens as we move toward the right with our observation point. The phasor of speaker 4 moves counterclockwise because its phase is moving forward as speaker 4 becomes closer to the point of observation while that of speaker 1 moves clockwise due to its greater phase delay. The result of the phasor addition is as shown in the diagram. Notice that the resultant phasor is now shorter than the one for the central point.

As we move farther, we would have a phasor pattern that curls more and more. At a point in the direction of 30° to the central axis, the phase change from one speaker to the next will 90° and the phasor diagram will appear as shown. The resultant phasor of this diagram is of course zero; adding the phasors head to toe simply brings one back to the start of the phasor diagram.

Now consider what happens as we move even farther to the right. Now the phase angle between the arriving oscillations from the speakers is greater than 90°. The resultant phasor diagram is as shown and now the resultant phasor is no longer zero! In fact it is starting to grow again. At the point shown, the angle between the phasors is 120°. At this point we have a phasor diagram where the resultant phasor is the phasor left over after the first three have cancelled each other.

Doing the vector addition carefully gives the rise to a maximum at an angle of about 47° with then a falling in net amplitude until we have reached 90°. Here the individual phasors are 180° apart and merely run back and forth along each other when they are added. Because there are an even number of sources, the result is zero.

The resultant diagram of the radiation pattern of the four source system is as shown. The pattern will be repeated, of course, in the region on the diagram above the sources. There are then six radiation lobes from this source combination; two major ones front and back and four minor ones roughly pointing at 45° to the main lobes.

A multi-lobe pattern is characteristic of multiple source systems when the sources are separated by significant fractions of a wavelength.

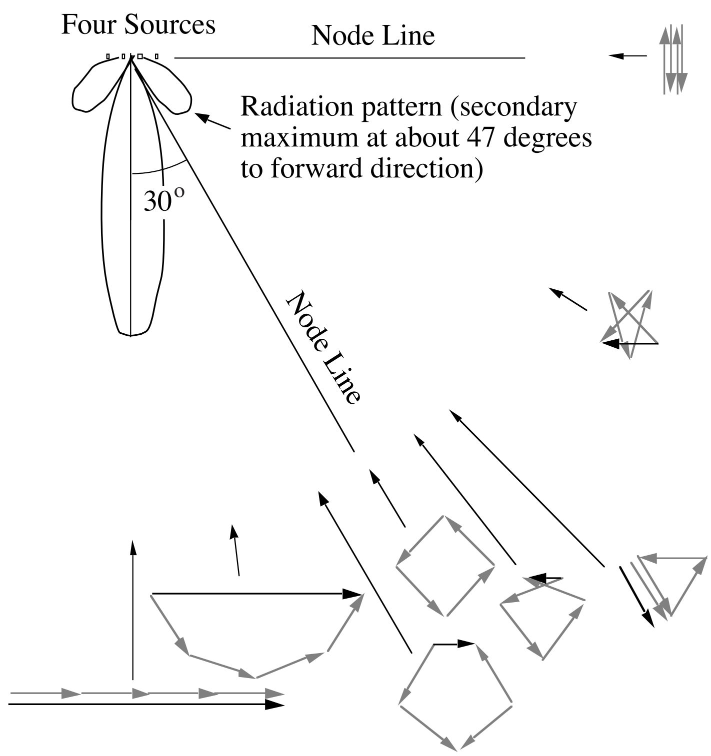


Figure 7.4 The phasors for the sound arriving at various observation points around a four source system, the four sources all operating in phase and at the same amplitude. The individual phasors are added in order from the source farthest to the left to the source farther to the right. At a very large distance where all the speakers are the same distance from the source, they will exactly line up but for closer distances such as we have here there will be the slight bending of the direction of the vectors due to the noticeable extra distance of speakers 1 and 4 from the central point. The resultant phasor would still be practically the algebraic sum of the four speaker oscillations even for oscillations arriving at points fairly close to the speakers.

7.3 Phasor Treatment of a Dipole Radiator

Having seen how phasors can be used to sum the oscillations from four sources, return now to an even simpler case; that of two equally strong sources physically close to each other (separated by a small fraction of a wavelength) but 180° out of phase to each other. This is a very important type of source in wave theory and is called the "dipole radiator".

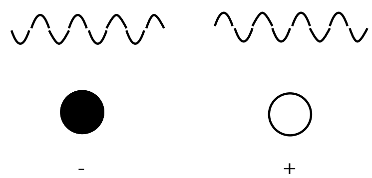


Figure 7.5 A dipole radiator. Both elements of the dipole radiator are of equal strength but of opposite phase.

A common example of such a radiator is an individual electroacoustic speaker that is not in an enclosure.

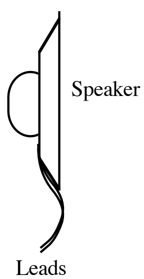


Figure 7.6 A bare electroacoustic speaker as a dipole radiator. With a forward motion of the speaker cone, he air expanding from the front of the speaker would be a positive source and the air being sucked into the back of the speaker would be a negative source. Because of the thickness of the speaker, there will be a separation of these two sources.

A microphone moved around very close to such a speaker will show that the sound is not radiated uniformly in all directions. No sound is radiated to the side and the sound radiated to the rear of the speaker is out of phase to the sound radiated toward the front.

This is not difficult to understand. The sound is produced by the movement of a speaker cone. When that cone is moving forward to produce a positive pressure buildup in the front of the speaker, it is creating a vacuum at the rear of the speaker. This vacuum propagates backwards as a wave that is 180° out of phase with the wave propagated from the front.

Such a radiator will have a very distinct radiation pattern which can be understood by the way the waves from the two sources add together at various points in space around the radiator. Consider for example a point directly to the side of such a dipole. At this point the waves from the two sources making up the dipole will arrive with their original phase differences intact. They will therefore exactly cancel.

For points to the front or back of the dipole (i.e in the direction of the line of the dipole) the source nearer will be advanced in phase relative to the source which is farther. This means that the waves will now not exactly cancel and that some wave action will appear. For points that are not exactly in front of or to the rear of the dipole the phase delay will be less and so the cancellation will be greater.

The phasor diagrams showing this is detail for the dipole radiator are shown in fig. 7.7.

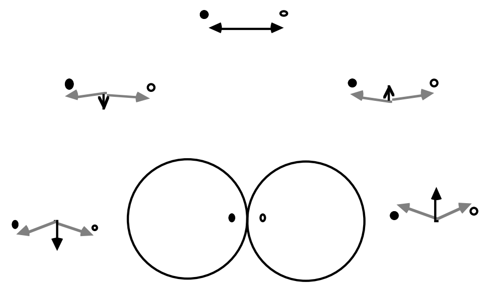


Figure 7.7 Phasor diagram of the two sources in a dipole radiator.

To the side the phasors cancel and therefore there is no sound oscillation. However to the front and back, because of the different distances of the two elements of the dipole the phasors are no longer back to back; the phasor of the farther element is turned clockwise and that of the nearer element is turned counter-clockwise. The result is a small vector sum representing the sound amplitude that will arrive at such points. It is easy to see that this vector sum will be the greatest when the rotation of these phasors is the greatest. That will be along the line of the dipole at points both front and back. The radiation pattern of the dipole is thus easily explained in detail and the actual amplitudes at any point in space can be easily calculated if one wished to do so. The result is shown in Fig. 8.6 where it is seen that the characteristic radiation pattern of sound pressure around a dipole is in fact two circles back to back, the sound in the forward circle being 180° in phase relative to the sound in the backward circle.

The phasor treatment of the dipole radiator also explains a very important feature of the open loudspeaker cone as a dipole radiator; that it lacks bass even in the forward and backward directions of the speaker. This can easily be seen when it is realized that the phase difference between the arrival of the two sounds from the two elements of the dipole depends on the wavelength of the sound. Longer wavelengths have less phase delay for the same distance travelled than do shorter wavelengths. Thus the phasors of the two elements of the dipole line up more closely to each other for the bass; the lower the frequency, the more cancellation of the phasors and the less sound that is radiated.

7.4 Slit Diffraction

The purpose of introducing phasors in this chapter is to use them to understand diffraction, one of the most important ideas in physics and of tremendous importance to all wave phenomena, acoustics included. Diffraction is the reason for the complicated radiation patterns for even a simple circular surface introduced in chapter 5.

Again, we will start by considering the simplest possible example, in this case a wave falling on a slit opening. This could be a sound wave arriving at a gap in a doorway, a light wave arriving at a slit between two razor blades or a water wave reaching a gap in a breaker in a harbor. For simplicity we will have the wave arriving at the opening with its crests parallel to the plane of the opening (ig. 7.8).

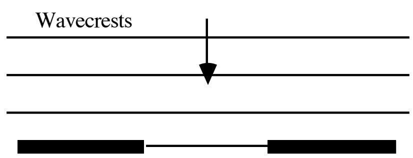


Figure 7.8 Wavecrests arriving at a slit opening in a wall. The slit is regarded as being long into the plane of the view. The slit width is shown. The waves are arriving with their crests lined up with the wall

Consider in detail how this wave will pass through the opening. A first simple-minded approach would perhaps lead to the conclusion that it goes through the opening and spreads out the other side. We might guess that the distribution of the wave action on the other side would not be uniform; there would be more wave action in the original direction of the wave and less to the side but we would be quite prepared to believe that some wave action would show up in any direction from the opening that was not in the original direction of the wave.

In fact this will not be the case. If the wavelength of the waves is less than the slit width, then the wave will enter the region after the opening <u>avoiding certain directions</u>!

This was the puzzle investigated by Fresnel and others that led to the invention of very small phasors and the so-called Fresnel integral calculus, which is a vector calculus method for adding many small phasors. They were trying to solve the puzzle of light which had been shown by Young to behave in this fashion when it went through a very fine slit.

The phasor approach to this problem starts with Huygens principle, the wave at the opening is thought of as many little sources very, very close together in a line along the source opening (fig. 7.9).

Figure 7.9 The wavecrest at the slit opening is theorized to be made up of a line of very small individual sources, all of the same strength and all in phase.

Each of these sources is in phase. Taking up some point away from this slit on the far side of the opening and directly in line with the original wave, all of the little phasors representing these little sources will be in line and add up like little vectors in line (see fig. 7.10)

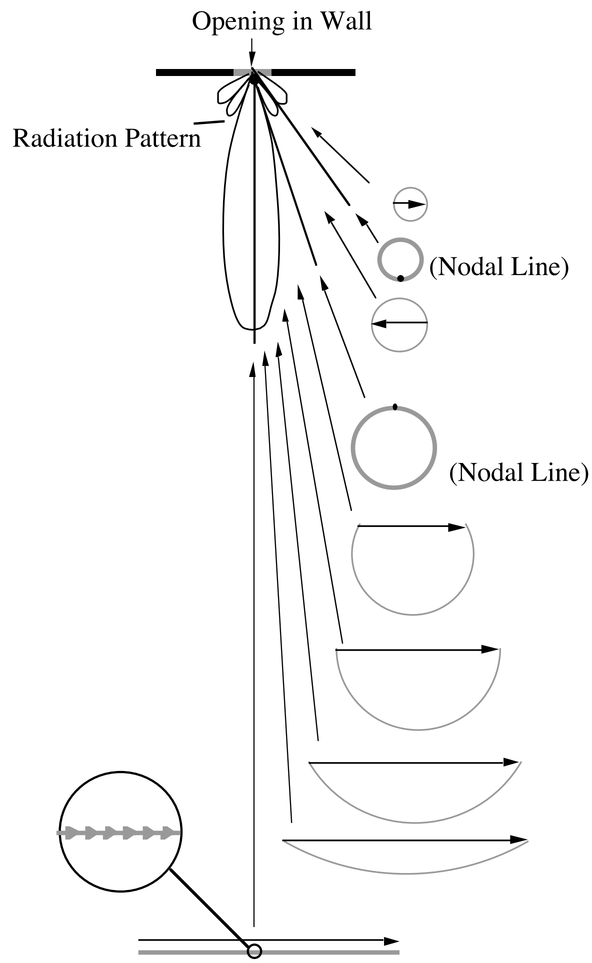


Figure 7.10 The phasor diagrams for the multitude of sources assumed to be in the slit opening. The individual small phasors for each small part of the slit are all assumed to be of equal length. With change in position in the slit, the phasors will rotate by an amount equal to the phase delay associated with that position.

Now consider what happens to this phasor diagram as we move away from the central direction. Suppose that we move to the left. The phasors corresponding to sources to the right of the slit will now rotate counterclockwise because these sources are closer to the point of observation. The phasor diagram starts to curl as shown. The resultant phasor for the total wave action arriving at the point (A) is therefore a little shorter.

Now consider what happens as we go farther and farther. The phasor diagram curls more and more until it is in fact a complete circle! At this point the resultant phasor is zero!.In which direction does this occur? Looking carefully at the phasor diagram it can be seen that the individual little phasors must change their direction by exactly 360^o from beginning to end. This means that the phase difference between the arrival of the oscillations from the two extremes of the opening (the two edges) is one complete oscillation or 360^o. The two edges of the opening must then be one wavelength different in distance from the point of observation.

Fig. 7.11 shows the geometry of this critical situation.

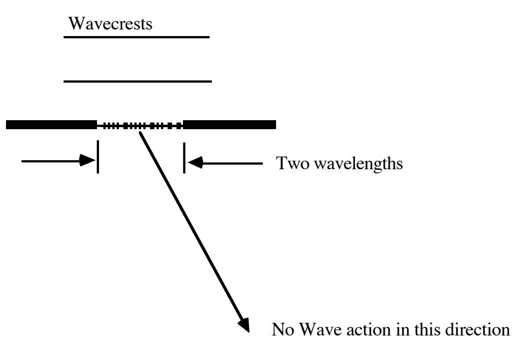


Figure 7.11 If the slit is twice as wide as the wavelength, there will be no wave action along a line at 30^o to the direction of the original wave.

The angle for this critical direction in space is given by a very simple formula;

$$\sin\,\theta = \frac{\lambda}{w}$$

where θ is the angle, λ is the wavelength of the wave and w is the width of the slit opening.

Waves of wavelength say 10 cm going through an opening 20 cm wide will show no wave action whatsoever in a direction 30^o to the direct line from the opening.

An enlarged view of the diffraction pattern of the wave passing through a slit is shown again in fig. 7.12.

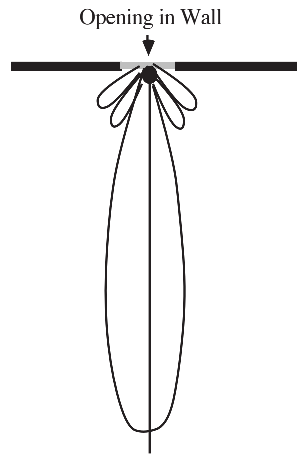


Figure 7.12 Enlarged view of the polar graph of intensity of a wave after passing through a slit.

The width of the central lobe is very dependent on the width of the opening through which the wave passes. The wider the opening the smaller the distance we have to move away from the central axis to get a one wavelength difference in distance of the edges of the opening. This means that there is broad dispersal of the wave on the other side of the slit only if the opening is about a wavelength or so. If the opening is many wavelengths wide, the wave on the other side will have a very narrow angular divergence after it passes through the slit. It will cast itself into a beam.

This is the answer to the puzzle of light. Light is a wave action with a very short wavelength. Ordinary openings that would be barely visible to the naked eye would still be gigantic compared to this wavelength. Openings that would show broad dispersal of transmitted light would be so small that not enough light to see would get through. It is only with modern lasers that the diffraction properties of light can be easily demonstrated.

Of what importance is this to acoustics? One item of importance is the way sound will reflect from or radiate from a surface. Such a source of sound produces waves very much like those that come through the opening we have just considered. The sound does not radiate in a smooth pattern in all directions but radiates in preferential lobes governed by the wavelength of the sound being radiated and the size and shape of the radiator. Thus the directional properties of musical instruments are determined by diffraction theory and it is only through this theory that the complicated radiation patterns of musical instruments and loudspeakers can be understood.

7.5 Diffraction Through a Circular Opening

As an example, return to the vibrating circular surface. Imagine the surface to be divided into many little regions the effects of which are then added together as in the case of the slit. Here the segments can be lines made up of points which are all equidistant from the observer (fig. 7.13).

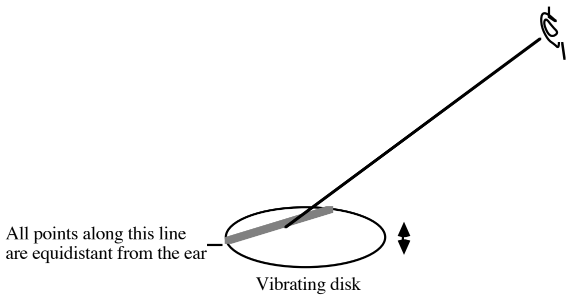


Figure 7.13 The elemental phasor source for the vibrating disk.

Unlike the case for the slit, the strengths of these elemental sources are not now all of the same; the strength will depend on the length of the line on the disk. The phasor diagrams will now look something like that shown in fig. 7.14.

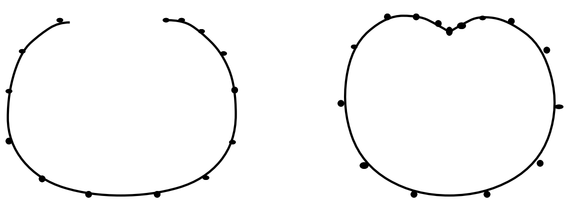


Figure 7.14 The phasor diagram for the wave propagating from a vibrating disk in two distinct directions. The diagram on the left is for a direction in which the far edge of the disk is one wavelength more distant than the near edge of the disk. The phasors at the extremes are very short because the lengths of the disk segments for these phasors are very short.The diagram on the right is for a direction where the far edge of the disk is 1.22 wavelengths more distant than the near edge of the disk. The phasors for this direction do lead to complete cancellation.

The result is that the phasor diagram does not give zero for this direction, as it did for the case of the slit where each phasor had the same length. This does not mean, however, that there is no point at which such a cancellation can occur; it occurs at a slightly greater phase difference between the two edges.

There is then a nodal cone about the axis of the disk containing a central maximum lobe of radiation, similar in pattern to that of Fig. 8.10 for the slit, the difference being that the nodal line comes at an angle θ given by

$$\sin\theta = 1.22\frac{\lambda}{d}$$

where d is the diameter of the disk. (Compare this formula with that for the slit; $\sin\theta = \lambda/d$.)

The radiation patterns of a vibrating disk that were introduced in chapter 5 are therefore explained fully by phasor addition applied with Huygens principle of wave propagation. Exactly the same diffraction pattern as for a vibrating disk will therefore occur for a wave falling on a circular opening. Again, this is because the wavefront in the circular opening can be thought of as a set of many small oscillators, all in phase.

7.6 Some Consequences of Wave Diffraction

This chapter will conclude with pointing out another remarkable feature of wave diffraction. The elemental sources that make up the wave in a circular opening not only propagate forward in the direction of the wave but also backward; they are assumed to be isotropic. At points on the wave side of the opening, these elemental sources can therefore be heard (fig. 7.15).

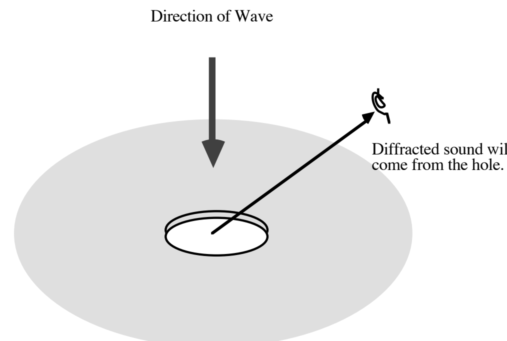


Figure 7.15 A hole will diffract a wave back into the room from which the wave was originally travelling. A sound can therefore be perceived as coming from the hole which is, of course , a perfect absorber!

The effect of these elemental sources on the wave side of the opening are therefore the same as on the far side; oscillations are diffracted backward in a lobe pattern just as they are propagated forward.

What does this do to our concept of what happens to sound in a room with absorbers and openings such as doors and windows. Up to now we have taken the point of view that sound which falls on a perfect absorber or goes out a window (which physically is the same thing) is lost from the room. Now diffraction theory is telling us that sound bounces even from such absorbers and scatters back into the room. In fact it would seem that half the sound in the opening would bounce backward and half continue to go forward. The simple theory of room reverberation assumed that the absorber merely removed the sound energy that fell on it. Was an error made in this assumption?

The mathematics to prove it is not appropriate for this material but rest assured. The sound energy arriving at the opening does go out the opening and disappear from the room. Where then does the sound energy scattered back from the opening come from?

The sound energy does scatter back in the way diffraction theory would indicate. The energy involved in this backscattered wave is robbed from the energy in the wave falling on the opening in the vicinity of the opening. Thus not only the energy actually falling on the slit is lost from the wave but energy of an equal amount is scattered back from the wave in directions determined by diffraction theory. Thus absorbers not only take energy out of the sound wave in the room but also scatter an equal amount of the remaining energy about the room as well. This backscattered energy does not represent a loss in sound energy in the room since it stays in the room. Therefore it does not affect the elementary theory of room reverberation as developed by Sabine.

However, the backscattered sound from an absorber has profound implications in how the absorbers should be placed. Because people seem to like the sound waves bouncing around a room to be scattered so as the appear to come from all directions, having the absorbing surfaces broken up into pieces of dimensions about equal to the wavelengths of the important components in the reverberant sound in a room can give a much more pleasant effect than having all the absorbing done by one flat wall.

As a finish to the subject of diffraction, one of the difficulties with understanding modern physics will be pointed out. In modern physics, particles (electrons, protons, neutrons etc. i.e. the fundamental building blocks of the world as we know it) are regarded as wave packets travelling through space. Thus they can exhibit the bizarre properties we have been talking about here for sound waves. In principle, all objects are collections of waves and therefore, in principle, an object arriving at an opening (such as a baseball about to enter a living room window) should break into parts some of which will bounce back from the opening. This is of course not what happens.

The particle "waves" that we talk about in modern physics are "waves of probability" and what modern physics is really trying to tell us is that there is a probability that when a baseball arrives at an open window it will bounce back from the window. Modern physics gets around the problem of explaining this to any normal person by stating that, of course the wavelengths of the probability are so extremely short that this would never be witnessed in real life (just as light diffraction is never seen in ordinary life because of its short wavelength).

While the problem is of no consequence in ordinary matters, it does pose some extreme philosophical problems which are not a subject for these notes but which have occupied philosophers and modern physicists quite a lot since these principles were discovered in the early 1920's.

Exercises and Discussion Topics

1. What would be the phasor representation of 3 equal intensity pure tones of the same frequency arriving at a point in space 120 degrees apart in phase? What would be the net result of the arrival of these three tones? What would be the result if they arrived 60 degrees apart in phase? What would be the result if they all arrived in phase?

2. What is a "dipole radiator" in acoustics? What are its directional properties? What is the relative phase of the front and back waves? Why is a bare loudspeaker, at frequencies for which the wavelength is much longer than the speaker dimensions, essentially a dipole radiator?

3. Why does a loudspeaker in an infinite baffle become a monopole radiator? What are the directional properties of a monopole radiator? Why is the loudspeaker baffle (or enclosure) so important at the lower frequencies?

4. State the principal assumptions of wave diffraction theory. What are the consequences in the extremes of source size being very small compared to the wavelength and of a flat source which is very large compared to the wavelength? Qualitatively, what sort of things happen when the source size and wavelengths are comparable?

5. From the assumptions of Huygens principle and diffraction theory, explain why an opening in a reflecting surface (such as a window in a concrete wall) will disturb the pattern of the reflected wave.

CHAPTER 8

THE RADIATION PATTERNS OF MUSICAL INSTRUMENTS

This material is from that of Meyer and Olsen on the radiation patterns of musical instruments and other sound sources. The material in Meyer is to be found in chapters 4 and 7. The material in Olsen is to be found in chapter 4, section 4.12 (beginning at page 100) and in chapter 6, section 6.5 (beginning at page 231).

At the present time, this material has not been transcribed and no permission has been obtained for copying the material from these sources. The reader is strongly encouraged to read the original material for factual information on the subject of the directional patterns of sound from musical instruments.

Exercises and Discussion Topics

1. Select one of the following categories of musical instruments; strings, horns or winds, and discuss the predominate directional characteristics of the direct sound from that class of instruments. Present the common features of the members of that class and the change of the features from instrument to instrument within that class.

2. What are the directional characteristics of the human voice and how do they relate to the ease of perception of speech with different orientation of the speaker relative to the listener?

CHAPTER 9

THE FREQUENCY SPECTRUM OF SOUNDS

The diffraction of sound waves around a source and objects in front of it leads to complex directional patterns to a sound wave coming from that source. At the same distance from a source the sound will generally have much different intensities in different directions. Furthermore, this pattern of directionality will be different for different sound frequencies. As a general rule, low frequencies will not have much directionality, i.e. they will radiate with about the same intensity in all directions, even towards the back of the source. On the other hand, high frequencies will radiate in quite specific directions with many directions in which the sound will be very weak.

Diffraction of sound therefore means that the timbre of a note, i.e. the relative amounts of fundamental and higher harmonics that give a note the characteristics that distinguish an A on a trumpet from an A on a violin, will appear to be different in different directions. This is generally an acceptable situation for the normal seats in a music theater because only a small fraction of the total sound in a sustained note is the direct sound where these directional features occur. The majority of the sound intensity used to discern the timbre of a note is in fact the reverberant sound and in a good auditorium this sound will eventually reach the listener with a good balance in the various frequencies no matter in what direction it left the source.

However, the majority of the sound that should be picked up in a good recording will be the direct sound itself. The directional features giving the sound a particular timbre at the location of the microphone can then be very important.

9.1 Frequency Components of the Human Voice

As a specific example, consider the human voice. A spectrogram of a typical sound such as a sustained vowel "ee" sung at a particular note will be seen in a spectrum analyzer to be particularly rich in high frequency components. A representative spectrum for the "ee" vowel sound is shown in fig. 9.1.

While the fundamental frequency, 110 Hz, of the spectrum shown in fig. 9.1 designates the pitch of the tone it is seen to by far not the strongest component in the spectrum, this being typically in the range of about 440 or 550 Hz for the "ee" sound. Also, note that there are strong components in the 3000 Hz range and even quite a noticeable amount in the 6000 Hz range.

These "clumps" of components are referred to as "formants" in speech and music and the reason for their existence will be an important part of the second half of this course. For now the important point is that they exist and are very important in perceived quality of a musical note. For instance, the difference between an "e" and an "ah" vowel sound is that in the "ah", the higher frequency formants peak at different frequencies and are much weaker (fig. 9.2).

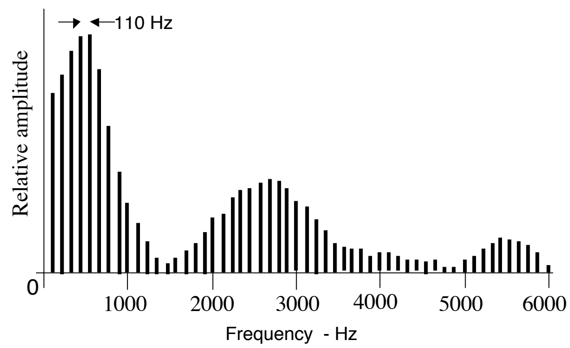


Figure 9.1 A typical spectrogram of a vowel "ee" sung at A_{110} by a male singer.

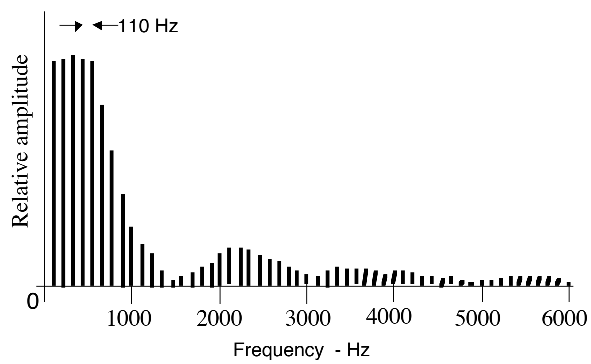


Figure 9.2 A typical spectrogram of a vowel "ah" sung at A_{110}.

As an example consider what an incorrectly placed microphone can do to the quality of an "e" sound of a singer. The spectrum with the microphone in front of the face will be as shown in fig. 9.1 but the spectrum with the same microphone behind the head will have most of the high frequency components missing.

Practically all musical sounds of any significance will have such complex frequency spectra and will have their timbres severely affected if the directional characteristics of the various components are not taken into account in a recording. The actual frequency spectrum of a sound source therefore is important when considering the directional characteristics of that source.

As in acquiring knowledge of any complex phenomenon, an understanding of its basic causes in usually a help. In the case of musical acoustics this, of course, implies understanding the physics of how the spectra arise. Again, one starts any such physics study with the simplest cases.

9.2 The Frequency Spectrum of Some Simple Tones

9.2.1 A Pure Tone

The very simplest musical tone that can be created is that of a pure tone resulting from a sinusoidal (simple harmonic) oscillation of air pressure. Such a tone is created by the oscillation of a simple musical device such a tuning fork. The variation of pressure with time and the resultant spectrum is shown in fig. 4.3.

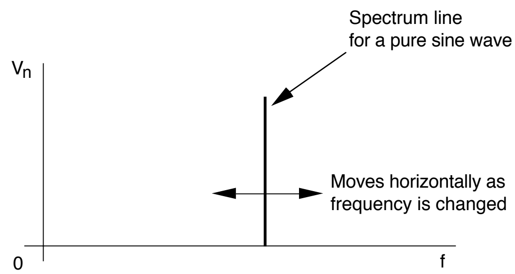


Figure 9.3 The waveform and spectrogram of a pure tone.

As the tone is changed in loudness it can be seen that the height of the single line rises and falls. As it is changed in pitch, it is seen that the line moves sideways, left for lower pitch and right for higher pitch.

9.2.2 A Square Wave form Oscillation

The next simplest musical tone that can be created is that of a pressure which is alternating between an overpressure of a certain amount and an underpressure of the same amount. Such a sound would be created by a wall moving back and forth at constant velocity but abrupt changes in direction, producing first a pressure wave and then a vacuum wave as described in Chapter 4. The variation of pressure with time and the resultant spectrum for a repetition rate of 500 Hz is shown in fig. 9.4.

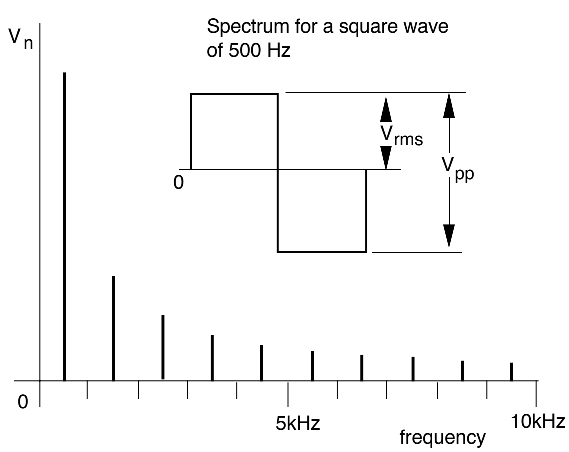


Figure 9.4 The waveform and spectrogram of a "square" 500 Hz tone.

It is seen that this tone does have the fundamental as its strongest component but contains also all the odd harmonics up to beyond the range of the spectrogram. It is these higher harmonics that give the "square" tone it much harsher quality compared to the pure tone.

In these notes the numbering of the harmonics will start at 1 for the fundamental, i.e. the fundamental is the first of the harmonic series making up a tone. A "harmonic" is not to be confused with the term "overtone". A musical term more in keeping with the term "harmonic" is "partial". Thus the fundamental is one of the partials of a musical tone.

Sometimes, but not always, an overtone or a partial can be a harmonic. This is because a harmonic is rigorously defined to be an integer multiple (i.e. exactly 1, 2, 3 or etc, times the fundamental frequency) while an overtone or a partial is sometimes not so exactly related. There is much confusion in the literature, even in the Harvard Dictionary of Music, about this point and it will be discussed more thoroughly later in the course.

9.2.3 A Triangular Waveform Oscillation

The last simple musical tone that will be presented here is that of a pressure which is swinging between an overpressure and an underpressure in a fashion that presents a triangular waveform (fig. 9.5). The variation of pressure with time and the resultant spectrum for a repetition rate of 500 Hz is shown.

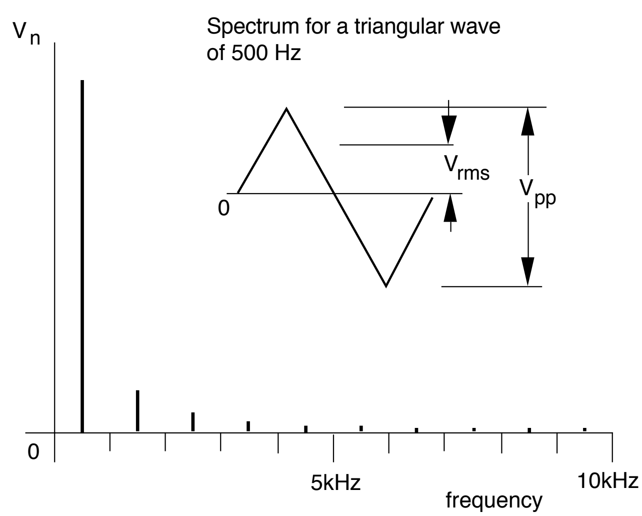


Figure 9.5 The waveform and spectrogram of a "triangular" 500 Hz tone.

It is seen that this tone, like the "square" tone, contains all the odd harmonics but that the higher ones are much weaker. This is the reason that the tone is not as harsh.

Why do these tones have these odd harmonics and only these odd harmonics? To gain some understanding of this phenomenon consider the simple addition of a few low order odd harmonics of a tone.

9.3 The Synthesis of Some Simple Tones

9.3.1 A Square Tone

First add to a fundamental a third harmonic at one third the amplitude of the fundamental, both starting at the same zero phase angle. The result is as shown in fig. 9.6.

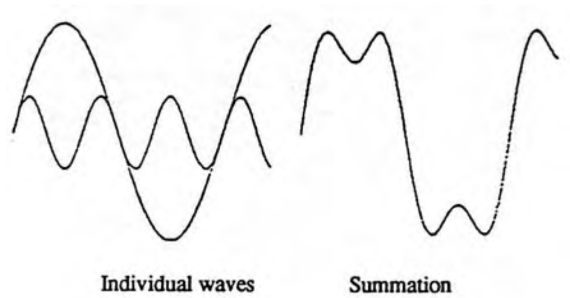


Figure 9.6 Pure sinusoid plus 1/3 amplitude 3rd harmonic.

Suppose now we add yet another wave, the fifth harmonic at one fifth the amplitude. The result is as shown in fig. 9.7.

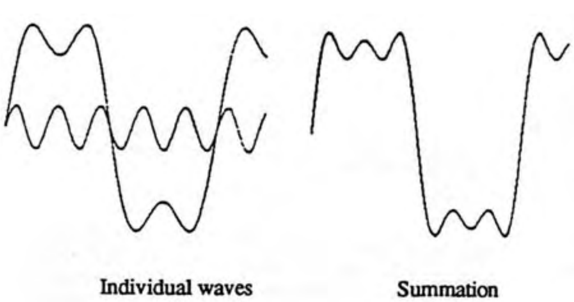


Figure 9.7 Pure sinusoid plus 3rd and 5th harmonic.

Continuing with some seventh harmonic at one seventh the amplitude, and finally with some ninth harmonic at one ninth the amplitude, the results are as shown in fig. 9.8 and 9.9.

Listening to the sounds of these waveforms, you would easily discern the presence of each harmonic as it is added. Aurally, at least, they seem to stay separate even as they are added into the system.

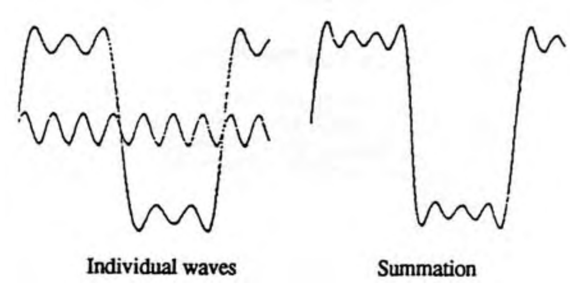


Figure 9.8 Pure sinusoid plus 3rd, 5th and 7th harmonic.

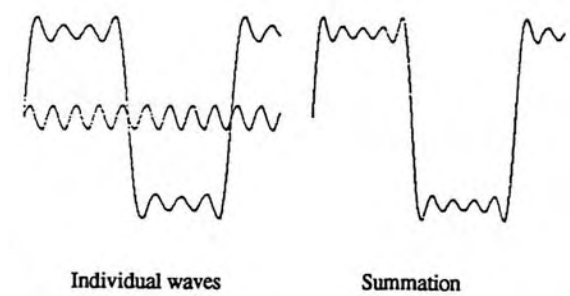


Figure 9.9 Pure sinusoid plus 3rd, 5th, 7th and 9th harmonic.

But also you would notice that the sound of all the harmonics simultaneously is approaching that of the square wave. Furthermore, the shape of the waveform is approaching a square wave.

From this it would appear that the sound of the tone and the picture of the waveform will become closer to that of the square wave if more harmonics are added in this fashion. The rule used is only odd harmonics added in phase at a amplitude relative to the fundamental given by the reciprocal of the harmonic number.

This can be confirmed graphically if you have a computer. The result of adding up to the 21st harmonic is shown in fig. 9.10.

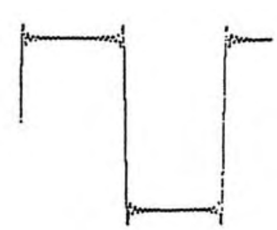


Figure 9.10 The result of adding to the 21st harmonic.

9.3.2 A Triangular Tone Plus Others

It is also possible to do the same sort of thing for a triangular waveform. Again, this wave can be simulated by adding together carefully selected harmonics of carefully selected amplitudes and phases (see fig. 9.11)

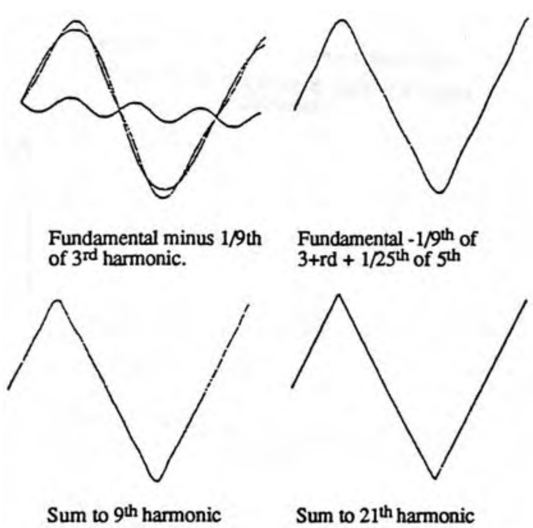


Figure 9.11 The result of adding to the 21st harmonic to create a triangular waveform.

Note now that the waveform approaches the triangular much more quickly and with much less amplitude for the harmonics than for the square waveform. This is related to the triangular waveform sounding less harsh than the square waveform.

It should not be surprising that this can be done for any waveform. Note the result of adding together another selected group of the first 9 harmonics in fig. 9.12. This group of harmonics approaches what is called the "sawtooth" waveform. Here all the harmonics had to be used, including the even ones.

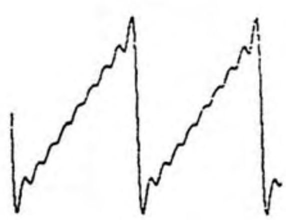


Figure 9.12 The result of adding to the 9th harmonic to create a sawtooth waveform.

The conclusion is that any repetitive waveform can be synthesized by adding together pure tones; a fundamental and selections of its harmonics. The reason that only the harmonics can be used to create these waveforms is that for the overall waveform to be repetitive at the fundamental frequency all the components must start again at the restart of a fundamental cycle. Only those oscillations that have completed an integer number of whole cycles in this fundamental period can do this.

This result has a powerful corollary: Any repetitive waveform created by any means can be decomposed into pure tone components, or can be thought of as being made up of those pure tone components. For example, consider the square waveform, which is often generated in electronic circuitry by simply having a switch going from being connected to a positive voltage to being connected to a negative voltage. No harmonics are actually used to generate the resulting "square" waveform but, nonetheless, the waveform does contain all these harmonics.

This can be demonstrated by adding to a square waveform a pure sinusoidal wave of phase opposite to that of the square wave so that it will cancel any fundamental that is in it (fig. 9.13).

Figure 9.13 The result of adding a fundamental component out of phase to a square wave.

Note what happens as the amplitude of the added pure tone is increased. The fundamental tone of the square wave is heard to disappear at a very definite amplitude of the superimposed pure tone. The resultant waveform at this condition is what is shown in fig. 9.13.

This exercise can be repeated for a removal of the 3rd harmonic from the square wave. When listening carefully to the resulting tone as complete removal of the harmonic is approached, the initial presence of these components stands out.

The effect of switching these cancelling sinusoidal waveforms on and off can be clearly seen in the spectrograms of the sounds (fig. 9.14)

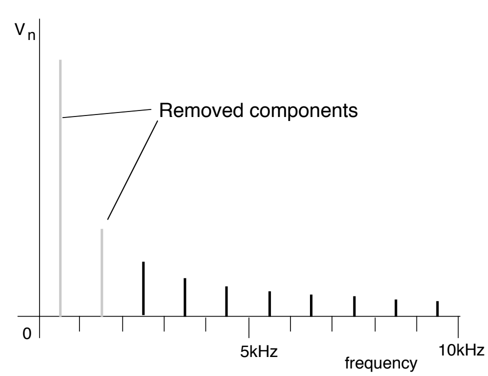


Figure 9.14 The spectrogram resulting from the addition of the cancelling waveforms for the fundamental and the 3rd harmonic to a square wave.

The superimposed pure tones are exactly canceling the pure tone components of the same frequency in the fundamental, leaving only the higher harmonics in the square waveform to be heard. Thus the square wave does indeed contain a pure fundamental component even though the mechanism that produced it used no such tone.

You might even note that the fundamental component of the square wave has a larger amplitude than the square wave itself. At first it might be puzzling that the square wave contains a fundamental which is higher than it. However, notice again what happened in the initial combining of the harmonics to simulate the square wave. The primary effect of adding the third harmonic was in fact to bring down the top of fundamental. The effect of the remaining harmonics appeared to be to make the top flatter and to make the sides of the waveform progressively steeper. The origin of the higher harmonics appears to be the sharp rise and the sharp corners of the waveform.

9.4 The Frequency Spectrum of Sharp Pulses

In all of the simple examples discussed so far, the higher harmonics making up a tone were much weaker than the fundamental. However there is another conceptually simple tone for which this is not the case and which is one of the most important tones in musical acoustics. This is the sound from a series of sharp pulses. As a start consider a waveform that is not very different from the square waveform that has already been considered, only the low pressure period compared to the high pressure period has been reduced (fig. 5.15)

Figure 9.15 A asymmetrical square pulse.

Playing such a pulse through a loudspeaker at the same repetition rate of as a square waveform or a triangular waveform will produce a sound quality that is much harsher than either. Furthermore, the quality gets harsher still as the duration of the pulse is shortened.

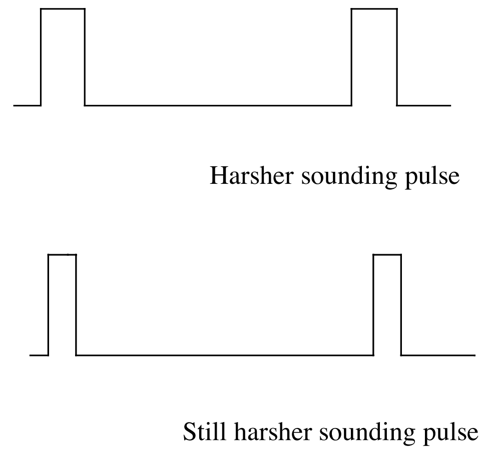


Figure 9.16 Progressively harsher asymmetrical square waveforms for sound pulses.

The harshness of the pulse seems to increase continually until the pulse duration is only about 0.05 ms, after which further reduction just seems to diminish the sound level of the pulse without changing its timbre.

Using the Fourier analyzer to determine the frequency spectrum of such repetitive pulse shows why this should happen. Starting with a pulse which is exactly one-half as long as its repetition period, i.e.of a square waveform, we see that the frequency spectrum of the square wave is reproduced. However, slowly decreasing the pulse duration shows that at the particular point when the duration is one-third of the repetition period every third harmonic disappears (fig. 9.17).

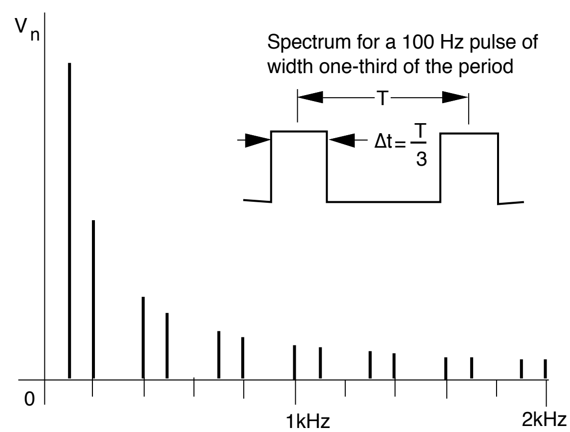


Figure 9.17 Spectrogram for a square pulse which has duration of 1/3rd of a cycle period.

Decreasing the pulse duration further shows that at a duration of one-fifth of the repetition period, only every fifth harmonic disappears (fig. 9.18).

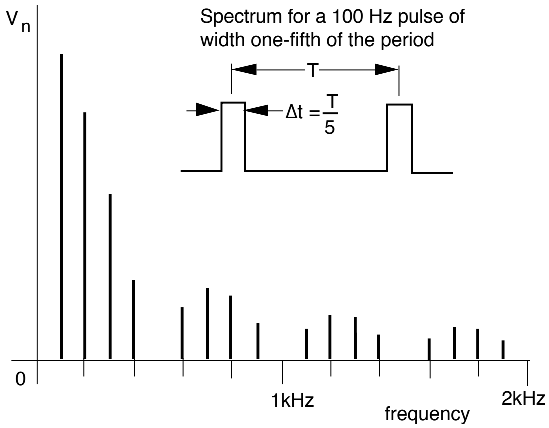


Figure 9.18 Spectrogram for a square pulse which has duration of 1/5th of a cycle period.

Finally, at a pulse duration which is only one-tenth of the repetition period, every tenth harmonic disappears (fig. 9.19).

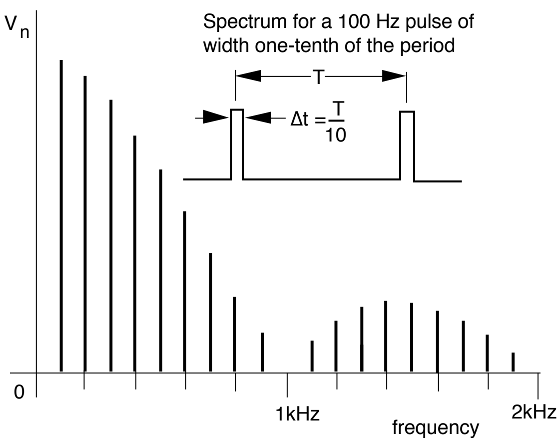


Figure 9.19 Spectrogram for a square pulse which has duration of 1/10th of a cycle period.

These spectra show that as the duration of the pulse narrows, more and more harmonics are developed. Also it is seen that there is a particularly simple relationship between the duration of the pulse and the lowest frequency which is cancelled. Referring to this lowest frequency as a "band-pass" Δf gives

$$\boxed{\Delta f = \frac{1}{\Delta t}}$$

where Δt is the duration of the pulse.

This is the reason for the timbre of the pulse becoming harsher as it is made progressively narrower. For durations of only 0.1 ms, the first null is at 10kHz. For durations shorter than that, the first null frequency goes even higher but into a range which is not possible for the simple speakers to reproduce or, even if they did, for you to hear very loudly. Durations of less then 0.1ms therefore do not result in an increase in timbre but in throwing more power into high frequencies which cannot be heard. The sounds therefore appear just to get weaker.

At first it is usually very hard to understand the bandwidth theorem at first. However, it is extremely important in physics and electrical engineering. One aspect of it can be shown be lowering the repetition rate while maintaining a constant pulse duration. For example fig. 9.20 shows what happens as the rate of repetition of a 0.4 ms pulse is lowered from 500 Hz to 250 Hz;

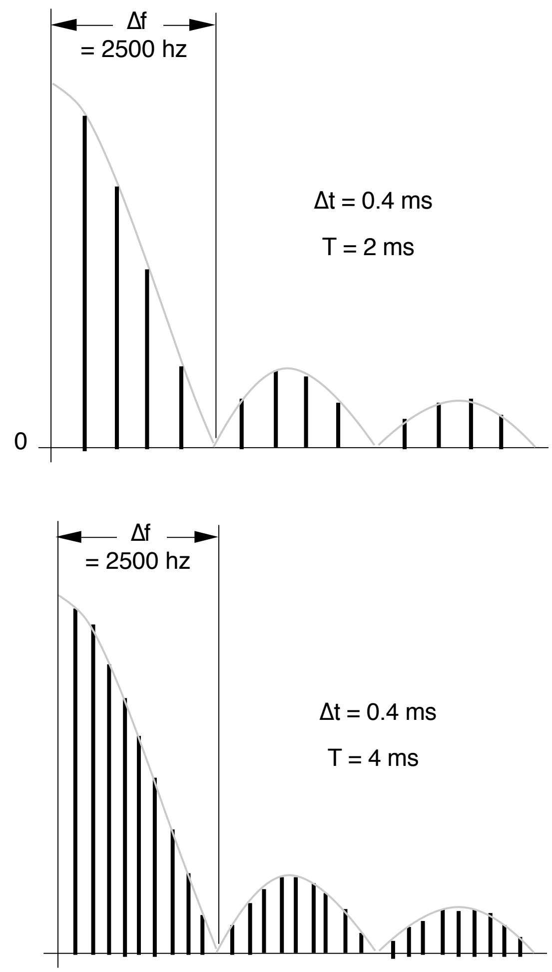


Figure 9.20 Spectrogram for a square pulse of duration 0.4 ms at two repetition rates; 500 Hz (top) and 250 Hz (bottom).

The shape of the spectrum has not changed from the 500Hz rate but there a twice as many spectrum lines in the same frequency interval. By lowering the repetition rate, the spectrum has become richer!

A little thought might make this seem somewhat plausible. The fundamental is now reduced to 250 Hz and so there are more harmonics possible in any given frequency range. What is still not perhaps so plausible is that the shape of the spectrum has nothing to do with the basic repetition rate of the pulse but only its duration!

This phenomenon is very important in human speech. Both male and female voices have about the same spectrum shapes for the different vowel sounds in a language. However, these vowel sounds are made up of harmonics of a relatively low fundamental. They come about because the basic sound production is from sharp puffs of air injected into the vocal tract through the tightly stretched muscles forming the larynx and it is the frequency of these puffs which is the fundamental.

However, because of basic physiological differences in the structure of the throat the frequency of these puffs is much lower for men than for women. For women it is typically about 220Hz, for men only about 140Hz. (Yet another case where women are faster than men.)

Typical spectra for the vowel sound "ee" are shown in fig. 9.21 for both male and female voices.

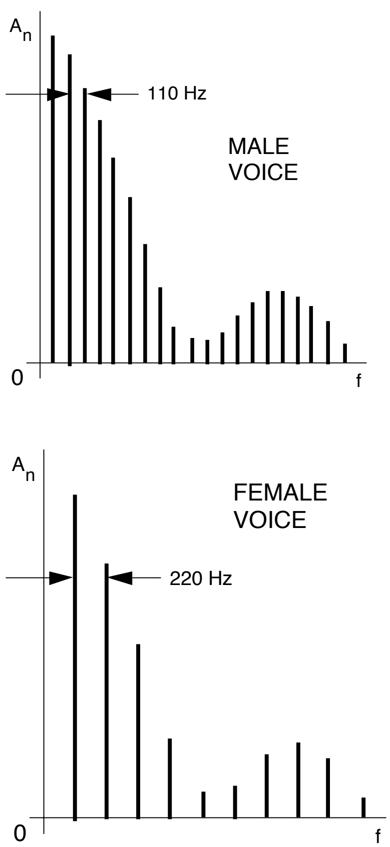


Figure 9.21 Spectrogram for a typical male voice (top) and a typical female voice (bottom) for an "ee" sound.

The result of the slower repetition rate of the male is that the spectrum of the vowels sounds for a male voice is much richer than that for a female voice. Some people will say that it is not so much richer as noisier. In any case, it makes the male voice distinctly different in quality to the female voice and provides, for learning children, a different perceptual problem. For most children brought up in a normal home environment, this poses no difficulty since the father will be screaming at a child nearly as often as the mother. However, in special cases such as the training of deaf children that have been given the benefit of hearing through powerful hearing aids in special situations, there can be a problem. This comes about because most of the hearing experiences with human voices for such a child will that of a female voice, either of the therapist, who is usually female, or the mother, who is usually the one who has given up her job to train the child. This tends to make the child unreceptive to male voices since they are different to what is being intensively experienced in the training sessions. Many people in the field of auditory rehabilitation for children are concerned about the low number of males entering the profession and would like to encourage more to get involved.

To return to the pure (i.e more useless) physics of this phenomenon, consider what happens as the repetition rate of the pulse further lowers. At about 25 Hz, the sound would appear to undergo an abrupt change in character from a continuous tone (or noise) to a discernable series of clicks. However, it still has the same spectral shape that it did at 500Hz. This is because it still has a duration of 0.4ms

Going all the way down to 3 Hz, would give quite clearly distinct pulses 3 times a second. Yet the spectrum shape is still the same, except that now it is impossible to see any distinct lines because they are separated by only 3Hz (fig. 9.22).

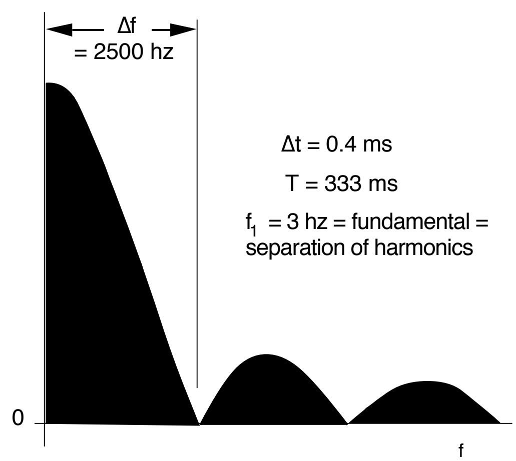


Figure 9.22 Spectrogram for a square pulse of duration 0.4 ms but at a repetition rate of only 3 Hz..

Yet there is a distinct pitch to this pulse. For example, if the pulse duration is changed from 0.4 ms to 1.0 ms, the pitch of the pulse would appear to lower. If the duration were changed to 0.1 ms, its pitch would appear to go higher. There would be an apparent change in pitch without a change in the fundamental frequency (3 Hz). This is because the spread of frequencies up to the first null depends only on the duration of the pulse and is given by the bandwidth theorem.

This leads to some difficult philosophical questions. The first is what happens when the repetition rate drops to zero; i.e when there is just one pulse and it is never repeated? Fourier analysis seems to be telling us that the shape of the frequency spectrum of such a tone burst or pulse will be the same as it is when the repetition rate is 500 Hz!

To get even deeper into this philosophical quagmire, consider what happens when we make such a burst infinitely narrow in time. The result will be a Δf of infinity. Thus we have the result that an infinitely narrow single tone burst of pulse contains all frequencies at equal amounts.

The mathematics which proves that this must be the case is not apprropriate for this course. It is called the Fourier transform and is related, of course, to the mathematics of Fourier analysis. Here, an attempt will be made to show how the phenomenon comes about by adding up a series of sine waves made up of all the harmonics of a fundamental but now with all the components of the same amplitude. They will all start at their maxima. (In terms of phase this means that they are all starting at a phase angle of 90°).

The individual waves are shown for the fundamental (500 Hz) and for the second harmonic (1000 Hz) in Fig. 9.23.

Notice that they both start at their crests but that the 1000 Hz has an extra crest in between the two shown for the fundamental and therefore tends to cancel the through of the fundamental at this point when they are added together.

Now consider the result of adding the third harmonic (1500 Hz) as shown in Fig. 9.24. Notice that again the oscillations all add at the crests of the fundamental but that elsewhere there seems to be no coherent summing. The resulting oscillation between the fundamental crests is even less than if we had the 1500 Hz tone alone.

Fig. 9.25 shows the result of adding the rest of the harmonics up to 4500 Hz. That is as far as the simple device used for this addition will allow, but you can imagine what the result of more additions would be; the peak that has developed would get higher and narrower. What has happened as more and more components are added is that reinforcement only occurs at the crests of the fundamental. In between these crests, the waves tend to cancel each other out. The result of adding all these higher frequency components onto the fundamental is to create a rather sharp spike at the fundamental frequency.

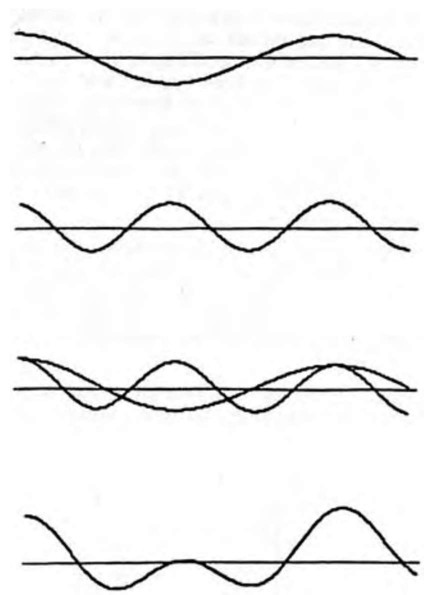


Figure 9.23 The result of adding a 500 Hz tone and a 1000 Hz tone, both of the same amplitude and both of the same starting phase (90°).The top graph is for the 500 Hz tone, the second for the 1000 Hz tone and the third shows both tones simultaneously. The result of the addition is shown in the bottom graph.

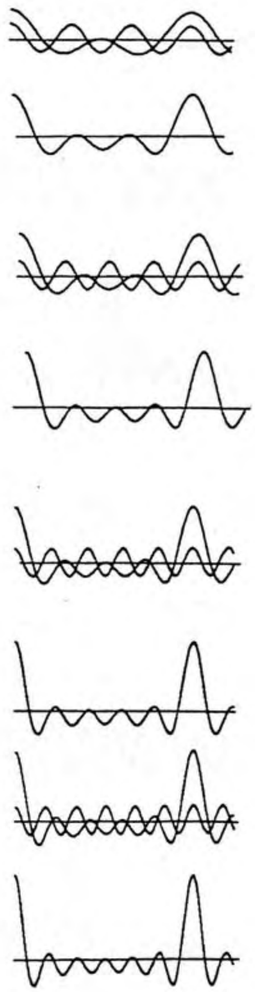


Figure 9.24 The result of adding the 1500, 2000, 2500 and 3000 Hz tones at the same amplitudes and phases as the components in Fig. 9.23. The top graph, left, shows the result of in fig. 9.23 with the third order harmonic (1500 Hz) overlaid. The next graph below it shows the result of adding these first three harmonics. The next graph below that one has the fourth harmonic overlaid and the graph below it the result of adding this fourth component. This continues to the lowest graph showing the result for all six harmonics.

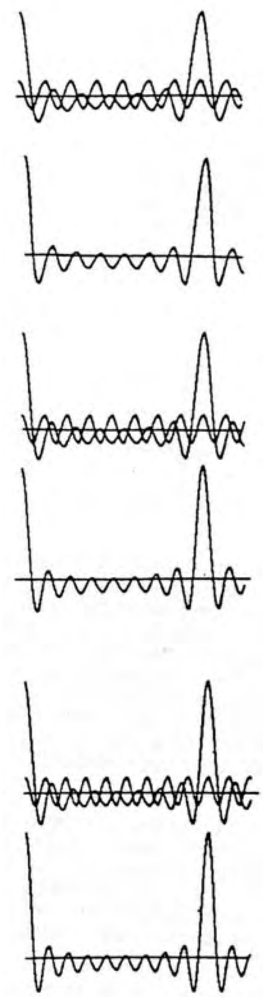


Figure 9.25 The result of adding 3500, 4000 and 4500 Hz tones at the same amplitudes and phases as the components in Fig. 9.1. The upper graph shows the result of Fig. 9.2 and the seventh order harmonic. The lowest graph shows the result of the addition of all of the first nine harmonics of 500 Hz.

Fig. 9.26 shows the result of a calculation of the sum of 100 harmonics of 500 Hz (up to a frequency of 50,000 Hz).

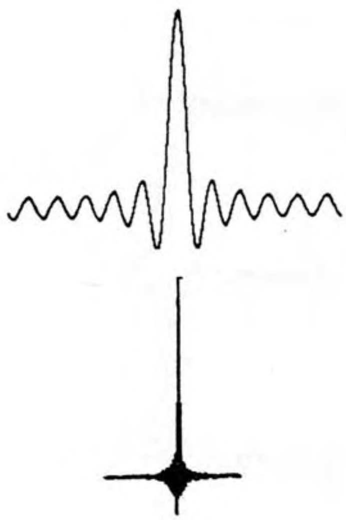


Figure 9.26 The result of adding up harmonics, all of the same amplitude and starting phase (90°). The upper graph shows the result of the first 9 harmonics as in fig. 9.25 but with the phases shifted to locate the peak in the center of the diagram. The graph below it shows the result of adding up the first 100 harmonics using the same phase relationships. The components in the graph on the right are only about 1/12th as large as they are for the graph on the left.

Thus a very sharp pulse can be created by adding together at the proper phase a large number of harmonics, all of equal amplitude. Again, this can be turned around. If a sharp pulse has been produced by any means whatsoever, then all these harmonics, with the correct phase relationships, have also automatically been produced.

Now notice what happens if the phase relationship governing the harmonics in the spike is altered? This can be demonstrated by showing the result of adding up harmonics from 500 Hz to 4500 Hz but now with a different phase relationship. Fig. 9.27 shows the result of adding harmonics which all start at zero phase angle. It is seen that different shaped pulses are produced than when the phase started at 90°. Here the pulses are "bipolar" meaning that they make sharp excursions to both above and below the axis, the negative excursion occurring first if the starting phases are 0° and the positive excursion occurring first if the starting phases are 180°.

It is perhaps not too surprizing that the upside down version of the original spike can be obtained by starting all the oscillations with a phase of 270° (see Fig. 9.28).

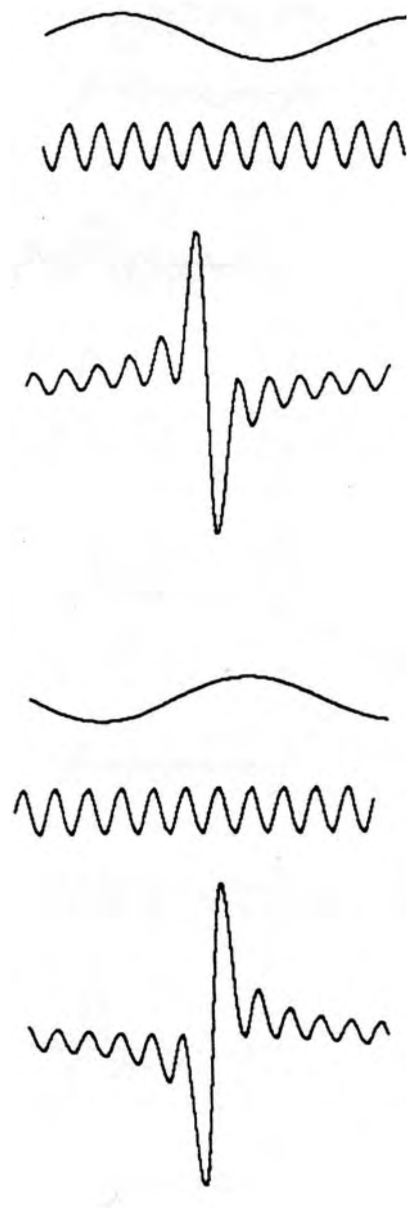


Figure 9.27 The result of adding up the first 9 harmonics, all of the same amplitude and starting phase. The graph on the top shows the result if the starting phase is 0°. The graph on the bottom shows the result if the starting phase is 180°.

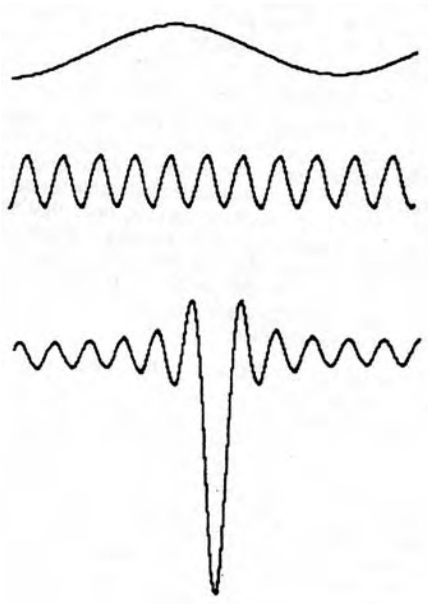


Figure 9.28 The result of adding up the first 9 harmonics, all of the same amplitude and starting phase, in this case 270°.

9.5 The Connection Between Time and Frequency Spectra

The important point from the previous section his that each different phase relationship between an infinite set of harmonics creates a different sort of spike or "transient". Now again turn this thought around. Each different spike has its own phase relationship between the harmonic components making up that spike. Therefore, if recording apparatus does not faithfully reproduce that phase relationship in its output, the original quality of the a sharp sound pulse may well be lost. Rather the sharp transient will appear to be just a noise burst.

Thus the preservation of phase relationships in the components of a sound is very important in preserving the quality of the attack of the sound and, as pointed out in the early part of this course, the attack of a sound is a very important part for our recognition of the nature and the direction of the source. This was not only important for the survival of our ancestors but is apparently also very important in our appreciation and enjoyment of music.

Perhaps the logic of all of this may become a little clearer by looking at the frequency-time relationships in two extremes; that of a steady pure tone that last forever and that of an infinitely sharp pulse. The first case is shown in the top half of fig. 9.29 and the second in the bottom half.

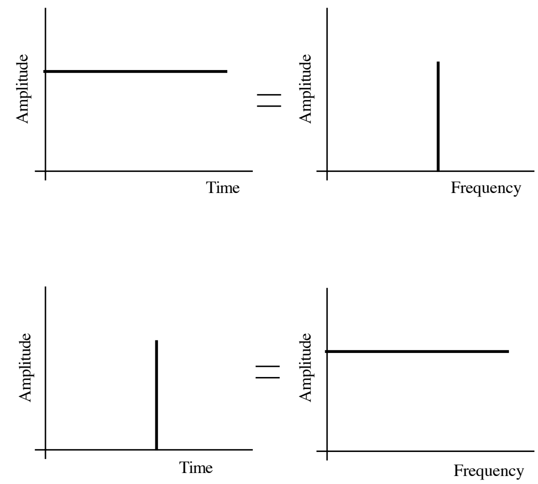


Figure 9.29 Diagrams representing the two extremes of types of oscillations; the infinitely stable single frequency pure tone (top diagrams) and the infinitely sharp spike (bottom diagrams). Note the inverse relationships in the diagrams; the frequency spectrum of one has the same form as the time spectrum of the other and vice versa.

Thus frequency and time have changed roles in the diagrams.

This is the beginning of an understanding of Fourier analysis and the Fourier Transform. There are two equivalent ways to look at any oscillation; in the time domain of actually following the motion and in the frequency domain of perceiving the frequency spectrum. Those of you who have taken Physics 224 should be quite familiar with this concept since the frequency spectrum is what the human brain perceives in discerning sound and the subset of sounds called music.

Thus any oscillation can be viewed in these two equivalent ways. One is called the "Fourier transform" of the other. If one of the forms is known, then the other can be obtained by the mathematics called the Fourier transform.

Again, I would like to include something about philosophical implications. As pointed out in chapter 8, when it was found that particles were collections of waves in space there were enormous philosophical difficulty. One of these is that an event such as a collision between particles must be capable of being described by a set of frequencies. If then it were indeed a single event, then it must be made up of all the frequencies. Furthermore, for that event to occur, all of these frequencies had to have the right set of phases at the start so as to line themselves up at the instant in time the event occurred. How did these phase relationships get set up? What are the governing rules of nature in this regard? Rationalizing these questions with our observations of nature was one of the most difficult problems that mankind has ever had to cope with in philosophy.

However, to return to more mundane matters, the frequency spectra that we have been dealing with are rather special; the harmonics of a fundamental. To look at another feature of the "Fourier" transform, consider yet another type of addition of pure tones; that leading to beats.

9.6 The Addition of a Bundle of Close Frequencies

As always, one starts with the simplest possible case; a pure tone. As a specific example to focus on, take one with a frequency of 3000 Hz. Such a tone is shown in the top graph of Fig. 9.30.

Suppose as a second example, two tones, close in frequency, are added together. Fig 9.30 show the result for a tone of 2800 Hz and another of equal intensity at 3200 Hz. The result is the familiar beat pattern at a beat frequency of 400 Hz also shown on the figure.

Now go to something that you probably never encountered in any previous physics course that you had to suffer through; that of the combination of three frequencies which are close together. This is also shown in the diagram. By keeping the separation at 400 Hz and having the proper balance between the center frequency and the two side frequencies (in this case 2 to 1), you can get a very distinct beat pattern, even more distinct than for the simpler two frequency case.

Here the beat frequency is the same as for the two frequency case but the null is of longer duration.

By adjusting the frequencies so that the gap is halved from what it was, then you will get a beat frequency which is only half of what it was. (See bottom diagram of the set).

Again, there are two ways of looking at the phenomena; the amplitude in time and the frequency spectrum.

However, now lets go to another stage. Take a five frequency beat pattern. How I get the right relative intensities for this beat pattern is not important here. The important thing is that there is a set of these frequencies and amplitudes that, when programmed into a microcomputer results in these graphs. Notice that for a spectrum of five frequencies of the relative amplitudes shown, a beat pattern occurs with a very distinct hollow empty region in between the beats. Notice that the beat frequency is now only half again what it was in the previous diagram.

Finally on the same diagram, is included what happens when nine components of the relative amplitudes shown in the frequency spectrum are summed. Here the frequency separation is only 50 Hz. Notice that now the beat pattern repeats itself only every 20 milliseconds. Also notice how broad the quiet region in between the beats has become.

But now notice a rather important effect. If the amplitude balance is kept the same as before, but the frequency interval is broadened to to 100 Hz, the beats are brought back to being only 10 ms apart but the duration of the oscillation bunch is also shorter.

There is thus an intimate connection between the shape of the frequency spectrum and the shape of the amplitudes within a beat. Note that they are both the "Napolean's hat" shape which in mathematics is called Gaussian. However, more important, note that the widths are interconnected. The broader I make the frequency spectrum, the narrower I make the amplitude duration in the beat. Conversely, the narrower I make the frequency spectrum, the broader I make the amplitude duration.

In the lower part of the diagram, is shown the result of adding 32 oscillations of the correct amplitudes. The result is a beat which has not noticeably changed its amplitude in time, but for which the second beat has moved cleanly off the page. All that was done to achieve this was introduce to introduce a set of frequencies that more completely filled the region of the spectrum making up the beats. If I continued to do this the result would be to move the second beat farther and farther away in time. Again, going to the logical extreme, all the frequencies in the spectrum were filled in, the second beat would never occur!

9.7 The Fourier Transform

Putting together this whole mess of frequencies would be practically impossible. However, it is certainly possible to turn an amplitude on and off in the manner that was the result of the addition of the 32 component frequencies in Fig. 9.29. If we do that once and only once, then we must have automatically produced the associated frequency spectrum.

This is the principle of the Fourier transform. Any modulation of the amplitude of an oscillation perturbs the frequency of that oscillation; it renders it from the infinitely narrow spike that defines a continuous oscillation at constant frequency, to a spectrum of frequencies. To obtain what the frequency spectrum actually is we do the mathematics of the Fourier transform.

The Fourier transform of two extremes has already been considered. The Fourier transform of the continuous steady oscillation is a sharp vertical line. The Fourier transform of a sharp spike in time is the flat frequency spectrum of all frequencies (fig. 9.30).

The Fourier transform of other amplitude variations (and the inverse Fourier transform going from a frequency spectrum to an amplitude variation) is more complicated and requires the mathematics of the Fourier transform to be carried out. Luckily, even for electrical engineers, there are now commercial devices which will do the Fourier transform by computation, and will do it very rapidly. Also with the advent of cheap computing power, these devices are becoming more and more reasonable in cost. They are called "Fast Fourier Transform" or FFT devices.

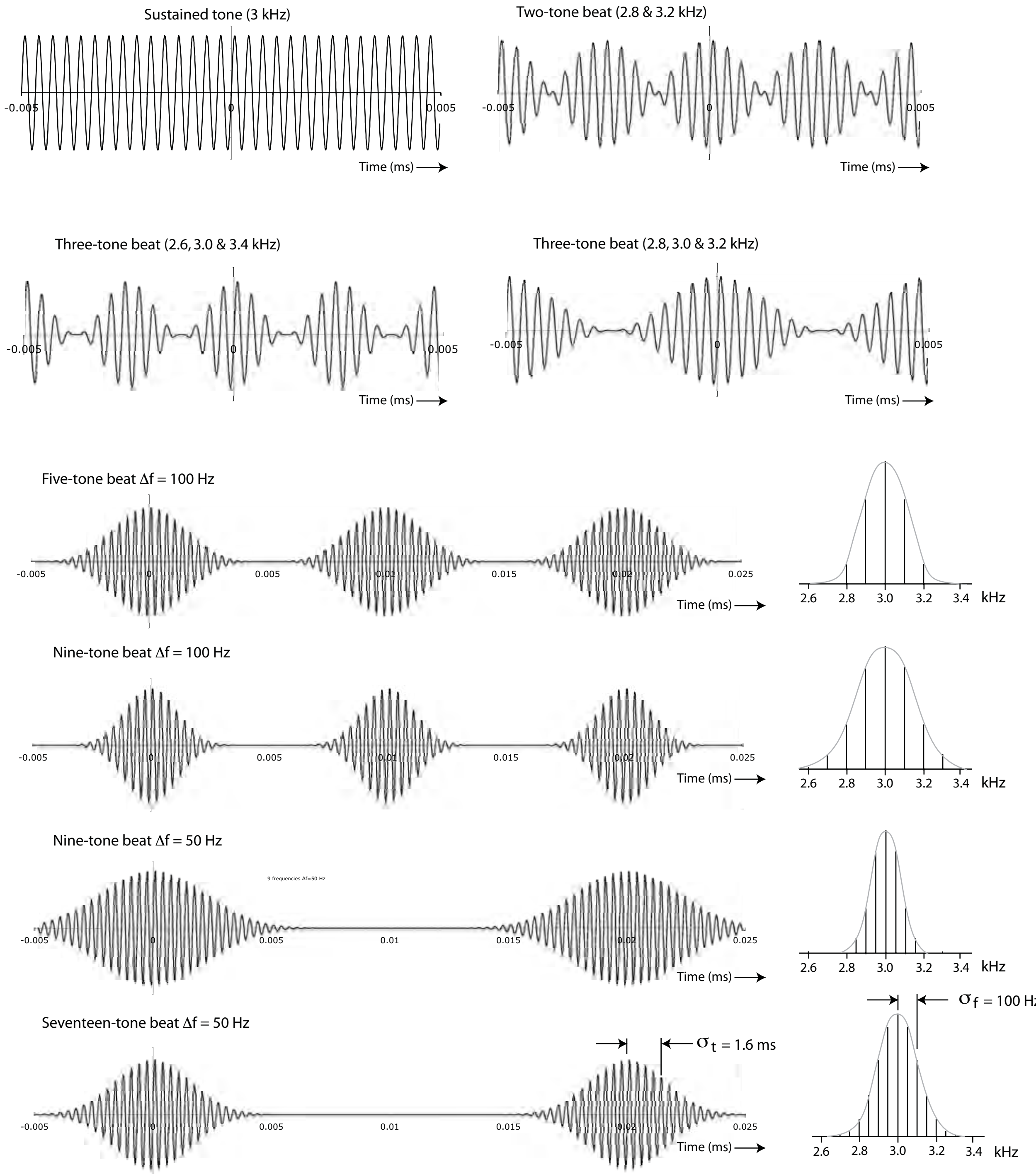


Figure 9.30 Diagrams showing the result of adding pure tones very close together in frequency.

About all the mathematics that I will give here is that the Fourier transform of the so-called "Gaussian" shape previously referred to is another Gaussian shape with an intimate connection between the two shapes. That connection is the width of the two shapes. As already pointed out, the wider one shape is, the narrower is the other. The actual connection is that if one takes what is called the "σ" (sigma) of the curves (this turns out to be how far you have to go away from the maximum to fall to about .605 of the maximum) then the product of the two sigmas will be $1/2\pi$;

$$\sigma_{\text{time}}\,\sigma_{\text{frequency}} = \frac{1}{2\pi}$$

The sigma of a gaussian is often referred to as the uncertainty in the value at the center of the guassian. The above equation is therefore often quoted and "the product of the uncertainties in frequency and time is given by $1/2\pi$".

This is another version of the bandwidth theorem. If the manner in which you turn an oscillation on and off can be approximated by a Gaussian, then you can immediately estimate by how much you have broadened the frequency by this act of turning it on and off. The figures that you get for the diagrams in Fig 9.30 are shown and, considering the rough way the numbers were estimated, show a remarkable agreement with the figure of $1/2\pi$.

Again forgive some more fundamental physics in closing. Modern physics has shown that energy is always quantized in units of planks constant times frequency.

$$E = hf$$

Thus energy is intimately related to the frequency of oscillations. If that is the case then the only way we can have a very definite energy is to wait forever. If we try to do anything to the energy of a system, we are turning energy on and off and therefore changing this frequency. The uncertainty in energy and the uncertainty in time will therefore be related by the famous Heisenberg uncertainty principle of modern physics

$$\sigma_{\text{energy}}\,\sigma_{\text{time}} = \frac{h}{2\pi}$$

Another point that is important for these notes is illustrated by the following thought experiment. The notes of a piano and all their overtones constitute a set of oscillations that cover a bandwidth of about 4000 Hz with a sigma of perhaps about 2000 Hz. Such a collection of frequencies would correspond to an amplitude of sound being turned on and off with a sigma of $1/2\pi$ divided by 2000 or about 80 microseconds. In other word, the Fourier transform of the sounds of the piano is a sound pulse which lasts 80 millionths of a second!

Yet everybody knows that if one bangs down on all of the keys of a piano at once, one gets more then 60 millionths of a second of sound. What has happened here? Is the Fourier transform just a mathematical abstraction that has little relevance in the real world.?

What has happened is that we forgot about the phase. Sure the Fourier transform gives us the frequency spectrum resulting from a given amplitude variation with time but it also gives us the exact phase relationships between all these frequency components. If we do not preserve these phase relationships, we will not get the right amplitude variation with time when we do the inverse Fourier transform back from the frequency spectrum. The frequency spectrum of an 80 microsecond sound pulse <u>will be</u> roughly the same as all the notes on a piano played at once but the phase relationships will be quite different. The phase relationships in the sound pulse will be tightly governed by the rules introduced in the beginning of this chapter. The phase relationships between all the frequencies in the piano will be random. Again, this illustrates the importance in preserving the phase relationships of all the frequencies in any recording. If they are not, then one can turn a sharp transient into a sustained noise such as the roar of a piano keyboard with all its keys pressed at once.

Exercises and Discussion Topics

1. How can two very different oscillations have the same frequency spectrum of sound?
2. Explain why preserving the relative phasing of the frequency components of a sound is important in reproducing the quality of a sound, particularly the attack of musical notes.
3. What is the difference between the attack of percussion musical instruments and that of musical instruments that give sustained notes? By diagrams of amplitude versus time, illustrate the attack of a few representative musical instruments.
4. What is the Fourier transform? Where does the concept of phase come into this transform? Qualitatively, what is the connection between the duration of a tone and the frequency spread of that tone? Explain the importance of the Fourier transform in understanding sound perception by humans.

CHAPTER 10

THE ORIGINS OF MUSICAL SOUNDS

The perception of all types of sound is a very important part of our everyday lives. Yet for human beings (and perhaps for some mammals and birds), there is a very special class of sounds called musical sounds which seem to be perceived as quite different from the others. For many people, the perception of these sounds is so enjoyable that they will engage in a great deal of playful creation of such sounds or even just in playful exercise of the perception of such sounds made by other people.

Just as the playful exercise of perception and motor skills in sport heightens these skills, so does the playful exercise of performing and listening to music heighten the perception skills related to that type of sound. This means that people who engage in such exercises can become highly critical of musical sounds. Any artificial creation of musical sounds, or any reproduction of "natural" musical sounds, by electronic devices, must be very faithful to the nature of these sounds. A knowledge of the nature of sounds that are generally regarded as musical is therefore important to musicians and recording engineers.

What makes a particular kind of sound musical is a very difficult question. It seems that music is somehow connected with very short interval timing mechanisms within the nervous system, the same sort of timing mechanisms that are associated with the perception of sound direction. Whatever the neural mechanism involved, the result is that sounds that contain tones which have frequencies related by simple numbers will generally be regarded as musical. The phenomenon is a large part of the subject of psychoacoustics and is regarded as outside the subject material of these notes. Here it will be assumed that certain sounds are indeed musical and present a somewhat technical description of how they originate in so-called "acoustic" arise (as opposed to "electronic" instruments).

Since antiquity, people have made music with practically any implement that could be handled. Again as an introduction, the simplest possible example, and probably the oldest in human history, will be considered; that of the sound produced by blowing into the neck of a bottle. Such an action can produce a very pure musical tone, the pitch of which depends on simply on how much empty space there is in the bottle and the size of it's neck. Although known since antiquity, this device was first analyzed scientifically by Helmholtz in the mid 19th century and is now called the Helmholtz Oscillator.

10.1 The Helmholtz Oscillator

If one blows across the opening of a short-necked bottle with a capacity of about 340 ml, such as the old standard Canadian beer bottle - the "stubby", which has a neck of 16 mm inside diameter and about 2.5 cm long, a pure tone of about 220 Hz (the A below middle C) can be easily obtained.

The explanation of the source of this sound is fairly simple. Suppose, as a starting point, that there was an underpressure of air in a cavity to which there was a tube connected to the outside as in fig. 10.1.

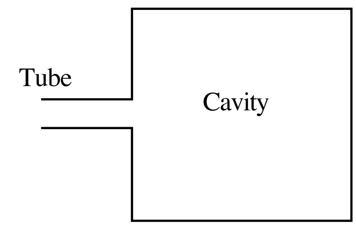


Figure 10.1 A schematic diagram of the Helmholtz oscillator. It consists simply of a cavity of undefined shape but a definite volume connected to the outside by a tube with a definite cross-sectional area and length.

Air in the tube would be pulled into the cavity. However, this air has mass and therefore does not move immediately into the cavity. The underpressure of the cavity has to act on it for a period of time to build up a flow.

After the flow has built up there comes a point where the underpressure of the cavity will have been relieved and there will be no more tendency for air to be pushed into the cavity. However, this does not mean that no more air will flow into the cavity. Rather, it is the flow rate that will no longer increase. The air that is in the tube will continue to flow due to its own momentum and now as this air moves into the cavity it will create an overpressure. It is as this overpressure builds up that the flow will finally stop.

At this point the cavity has an overpressure. Air will now start to flow out of the cavity, creating exactly the reverse of the case when the cavity had an underpressure. Eventually the cavity returns to the negative pressure it had at the start of the sequence.

Considering the energy of the system, the elastic energy of the overpressure (or underpressure) in the container is oscillating with the kinetic energy of the air in the tube. When the elastic energy is at a maximum (i.e. the pressure in the container is at a maximum overpressure or a maximum underpressure) the kinetic energy in the tube is zero because there is no air flow at this point. Similarly, when the elastic energy is zero (i.e. there is no overpressure or underpressure) the kinetic energy is at a maximum because the air in the tube is then either flowing into or out of the container at a maximum rate.

This analysis, of course, neglects energy losses in the system. One such loss is that due to viscous drag in the air flow through the tube. A more important loss, from the point of view of the Helmholtz resonator as a sound source, is that due to the air flowing out of the hole of the tube having to push away the air which is already outside the tube. This causes sound energy to be radiated away from the opening.

A thin walled container with a hole is also a Helmholtz oscillator. This is because, from the aerodynamics of the flow of air through a hole, it is equivalent to a short pipe. If the hole is circular, it can be shown to be equivalent to a pipe of length equal to about 1.7 times the radius of the hole.

Helmholtz analyzed this system in a way analogous to that of an oscillating mass on a spring (fig. 10.2).

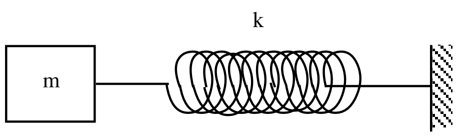


Figure 10.2 A schematic diagram of the elementary mechanical oscillator. It consists simply of an object of undefined shape but a definite mass connected to a spring of a definite spring constant *k*.

In this system the oscillation is described by the two equations for the force in the system. One of these equations is that for the spring force;

$$F = kx \qquad (10.1)$$

where k is the "spring constant" representing the stiffness of the spring and x is the stretch of the spring.

The other equation is that for the force on the accelerating mass;

$$F = ma \qquad (10.2)$$

where m is the oscillating mass and a is the acceleration.

Anyone who has taken an elementary course in physics will have seen the resulting equation for the oscillation frequency of this system;

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \qquad 10.3)$$

Thus the stiffer the spring, the higher the frequency and the heavier the mass the lower the frequency.

In the Helmholtz oscillator, the overpressure of the cavity is analogous to the spring force. A stretch of the spring is analogous to a volume of air which has been pulled out of the cavity. The relationship between the pressure in the cavity and the amount of air put into it can be shown to be as in (10.4) (see appendix).

$$p = \frac{Q}{C_A} \qquad (10.4)$$

In this equation Q is the volume of extra air put into the cavity (in cubic meters) and C_A is a quantity called the acoustic capacitance, This acoustic capacitance is related to the volume V of the cavity by the equation;

$$C_A = \frac{V}{\rho c^2} \qquad (10.5)$$

where ρ is the density of the air and c is the velocity of a sound wave in that air. It can be seen that the acoustic capacitance is an inverse concept to the spring constant of a mechanical spring. A large acoustic capacitance means a large volume flow to reach a given pressure. A large spring constant means a small stretch to get a large force from the spring.

(In dealing with loudspeaker design problems, a concept which is the inverse of the spring constant, and hence more analogous to acoustic capacitance, is often used. This is the "mechanical compliance" which is the ratio of stretch of a spring to the force the spring develops. Thus a large compliance means that a given force will produce a large stretch to a spring.)

The quantity in the Helmholtz oscillator analogous to the acceleration of the mass in (10.2) is the rate of change of the flow of air through the tube. The quantity analogous to the mass being accelerated is the so-called "inertance" of the air in the tube. The equation analogous to (10.2) is then

$$p = M \times \textit{rate of change of flow} \qquad (10.6)$$

where M is the inertance and again p is the pressure, in this case the overpressure of the cavity which is causing the flow rate to change.

It can be shown (again see the appendix to this chapter) that the inertance of the air in a tube is the mass of the air which is moving divided by the square of the cross-sectional area of the tube through which it is moving;

$$M = \frac{\textit{mass of moving air}}{(\textit{cross sectional area})^2} \qquad (10.7)$$

For a tube of length L and radius R, the inertance is

$$M = \frac{\rho\,(L + 1.7 \times R)}{\pi R^2} \qquad (10.8)$$

In this equation, πR^2 is, of course, the cross-sectional area of the tube and $L + 1.7 \times R$ is the effective length of the air moving through the tube. The term $1.7 \times R$ is the length of tube that has already been pointed out to be equivalent to a circular hole. It comes from the air in the vicinity of the openings which has to move some distance from the opening before its velocity slows to insignificance.

Completing the analogy is the equation for the frequency of a Helmholtz resonator;

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{C_A M}} \qquad (10.9)$$

$$\left(f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \right)$$

Putting in the expressions for C_A and M, gives a working equation which predicts the oscillation frequency of of a Helmholtz oscillator;

$$f = \frac{cR}{2\sqrt{\pi V\,(L + 1.7 \times R)}} \qquad (10.10)$$

For an internal volume of 340 ml, and a neck of length 2.5 cm and internal diameter 16 mm, the predicted frequency of the air oscillation is 214 Hz.

The Helmholtz oscillator is an important part of many musical instruments, most notably the violin family and the guitar. The curly "F" holes in the body of the violin and the single circular hole in the guitar are designed to produce this type of oscillation of the air in the body of the instrument. Its importance here is that it is the very simplest example of how air can be made to oscillate at a specific frequency by an interplay of the energies associated with its pressure and its velocity.

There are, however, many other modes in which air can be made to oscillate, even in the simple case of an empty bottle. If one blows very hard on the type of bottle described here, one gets a shrill note at about 1500 Hz. By using compressed air one could get an even higher, more shrill note at about 2800 Hz. The types of motion of the air in the bottle that give rise to these oscillations will now be considered.

10.2 Standing Waves on a String

Here there will be what at first might appear to be a digression; a discussion of the modes of vibration of a stretched string. This is also one of the musical instruments of antiquity and could be discussed on its own merits. It is introduced here however as a means to understand the analogous modes of vibration of air in a container.

The modes of vibration of a stretched string can be easily seem by vibrating the string at the frequencies of these modes. For a string which is stretched and tied down at the ends they will look like those shown in fig. 10.3. The frequencies at which these vibrations take place are successively integer multiples of the frequency of the lowest mode (the top mode of the figure). Thus the frequency of the second mode is twice that of the lowest, the frequency of the third mode is three times that of the lowest and so on.

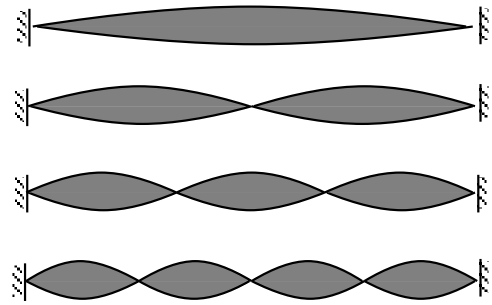


Figure 10.3 The lowest four modes of vibration of a stretched string. The higher modes are similar but with successively more nodes (points of no vibration).

It is possible to have a heavy string which is stretched and not tied down but attached to very light long strings so that the ends of the heavy string are quite free to move. The modes of vibration will then be as shown in fig. 10.4.

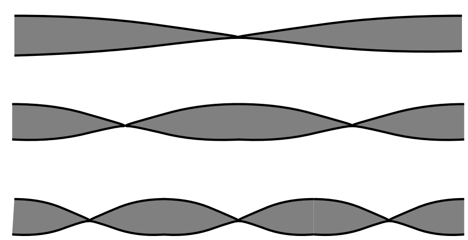


Fig. 10.4 The lowest three modes of vibration of a stretched string free to move at each end. The higher modes are similar but with successively more nodes.

These modes, for the same string under the same tension as that in figure 10.3 will have the same frequencies.

The modes if one of the ends is tied down and the other is free to move are as shown in fig, 10.5.

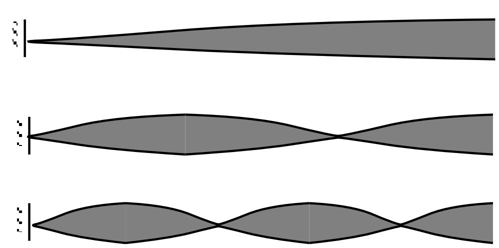


Fig. 10.5 The lowest three modes of vibration of a stretched string free to move at one end but tied down at the other. The higher modes are similar but with successively more nodes.

In this case the frequencies of the modes will be odd integer multiples of the lowest frequency mode, the frequency of this lowest mode for the same string under the same tension as in the previous cases being only one-half the frequency of the previous lowest modes.

The detailed motion in these modes of vibration are shown in Fig. 10.6. This diagram shows the motion for one half-cycle of the third mode of the stretched string tied down at both ends. It can be seen that adjacent maxima on either side of a node are always out of phase with each other.

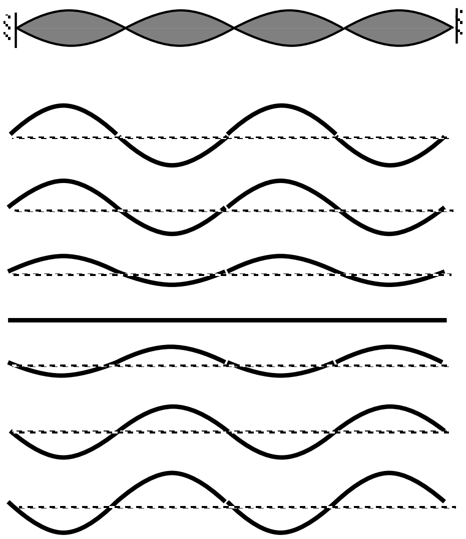


Figure 10.6 The detailed motion of the string for the 4th mode of vibration in Fig. 10.3

The modes of vibration of a string shown in fig. 3 to 5 are often referred to as "standing waves". This is because they have the appearance of waves that do not move but oscillate in a standing pattern, the peaks becoming troughs and back to peaks and the troughs becoming peaks and back to troughs.

A very important property of these standing waves is that they all have distinct frequencies. If their motions are coupled to the surrounding air then there will be distinct frequencies of sound waves propagated through this air. These distinct frequencies will be heard as tones and if the frequencies of these tones are related by integer multiples then they will normally be regarded as resulting in a musical sound. Since the modes of vibration of a stretched string are integer related (they are in the sequence 1:2:3:4... or 1:3:5:7....) then coupling the motion of a string to the surrounding air results in a musical instrument. The stretched string is indeed on of the first musical instruments invented by humans.

To understand the modes of vibration of a stretched string one must understand the physical principles governing the frequency of the vibration. The general treatment of this subject which is applicable to all types of systems involves the solution of the wave equation in bound systems and will be found in an advanced text-books in acoustics. What will be presented here is an elementary introduction using pictures that should illustrate more clearly how these standing waves arise.

A standing wave can be produced by two interfering waves travelling in opposite directions (see fig. 10.7)

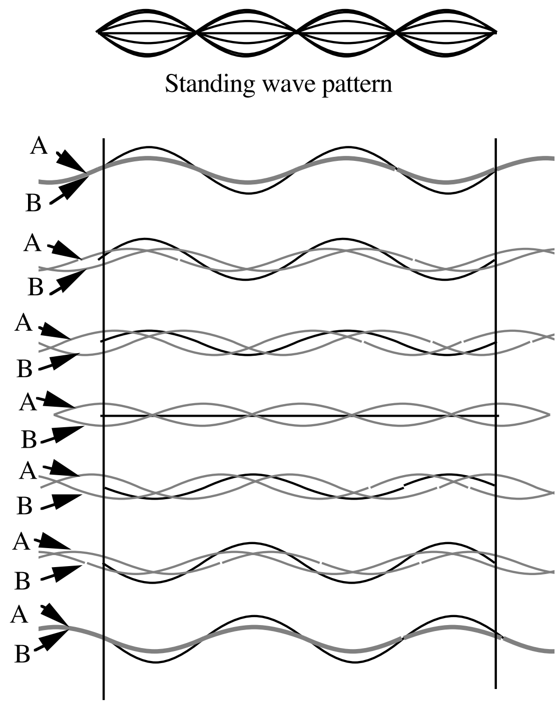


Figure 10.7 The production of a standing wave by two equal amplitude travelling waves travelling in opposite directions. The standing wave produced is shown in the top diagram. The successive pictures below it show how this standing wave is produced by movements of the two travelling waves. The first picture of this sequence is for when the two waves overlap, producing a result which is twice that of either. The fourth picture is for when they have moved just the right amount to cancel. The final picture is for when they have moved so that they reinforce each other in a waveform upside-down to that of the first picture.

It can be seen by the motions in this picture that the time for one half-cycle of the standing wave is the time for motion of the individual travelling waves through one-half cycle each. Thus the frequency of the standing wave is just the frequency of the individual travelling waves from which it is formed.

The frequency of a travelling wave is directly related to its wavelength by the simple equation $f = c/\lambda$ where c is its velocity. Thus we can get the frequency of a standing wave from the wavelength of the travelling waves that would make it. It is easily seen that this wavelength is twice the distance between adjacen nodes in the standing wave.

The frequencies of the standing waves on a stretchec string of length L held down at both ends, or free a both ends, are therefore

$$f = \frac{c}{2L}, \frac{2c}{2L}, \frac{3c}{2L}, \frac{4c}{2L}, \text{ etc.} \qquad (10.11)$$

where c is the velocity of travelling waves on the string The frequencies for a string of length L tied down a one end and free at the other are

$$f = \frac{c}{4L}, \frac{3c}{4L}, \frac{5c}{4L}, \frac{7c}{4L}, \text{ etc.} \qquad (10.12)$$

As pointed out at the beginning of this section, it migh appear as a digression from the subject of the seconc mode of vibration of the air in a bottle. What will be introduced now is the analogy between the transverse motion of a stretched string and the motion of air in a tube.

10.3 Standing Waves in Air in a Tube

The motion of air under the influence of a sound wave has already been shown in chapter 3. It is a velocity of the air in the direction of the wave motion which occurs in connection with the pressure in the wave. For the case of a pure simple harmonic motion the motion will be as shown in Fig. 10.8.

From this diagram it can be seen that the pressure and the velocity are in phase. For a wave travelling to the left, the velocity diagram would be inverted resulting in the pressure and the velocity being 180° out of phase if the velocity is still regarded as being positive toward the right.

Thus the pressure and velocity diagrams for sound in open air look very much like the shape of travelling waves on a stretched string which has no boundaries. To complete the analogy of standing waves on a string to the same sort of wave pattern in a column of air, consider what happens to the sound waves when there are reflecting boundaries to the wave motion similar to that of a stretched string which is tied down at both ends. As an example, consider sound waves in a closed pipe. The pattern of vibration analogous to the first mode of the stretched string is shown in fig. 10.9.

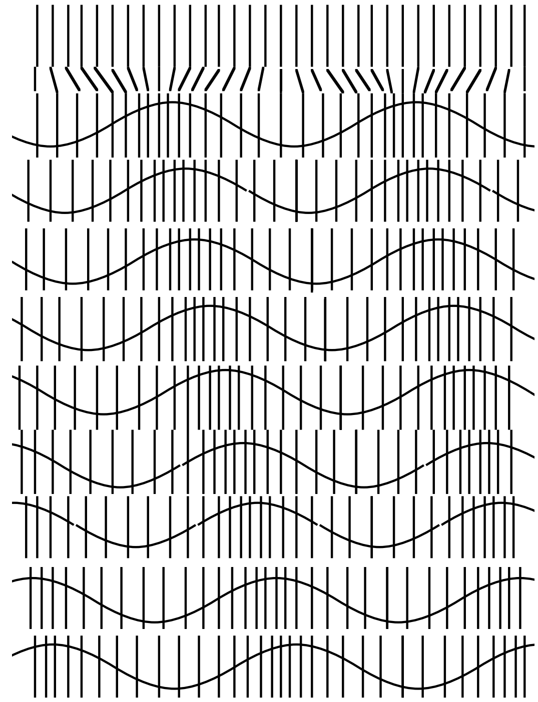


Figure 10.8 The motions involved in a sound wave in open air. The regions of compression can be clearly seen and these regions progress towards the right for the successive pictures. The detailed motions of the air can be noted by comparing one picture with the next. A vertical line representing a particular region of air can be seen to be merely oscillating back and forth in the direction of the wave motion. (This can be most clearly seen for the lines at the extremes of the pictures.) The heavy lines represent the pressure on a vertical scale and it can be seen that these progress to the right in the successive pictures. These same lines can also be seen to represent the velocity of the lines representing the regions of air; where there is a maximum concentration of the lines and therefore a maximum pressure, there is maximum velocity to the right. Where there is a minimum concentration there is maximum air velocity to the left (in an algebraic sense minimum velocity to the right).

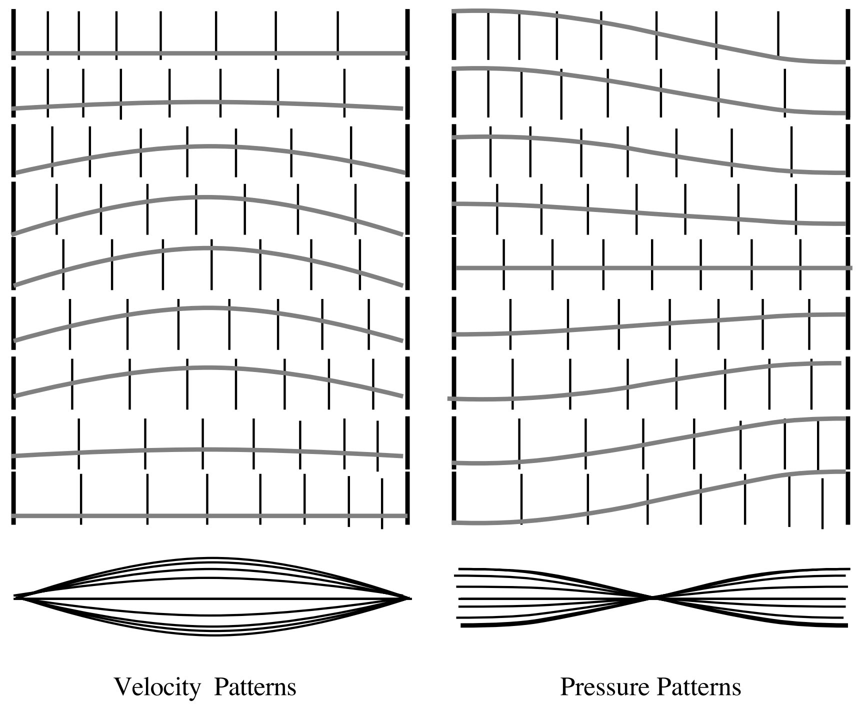


Figure 10.9 The motions involved for one half-cycle of the first standing wave mode of vibration of the air in a closed pipe. The vertical lines represent the positions of the air at the instant in time represented by the diagrams. Successive diagrams downward represent successive instants in time. The diagrams are repeated side by side so that one set can be used to indicate the velocity patterns and the other to indicate the pressure patterns (both shown as shaded lines on the diagrams). The diagrams on the bottom give an overview of the velocity and pressure patterns for a complete cycle.

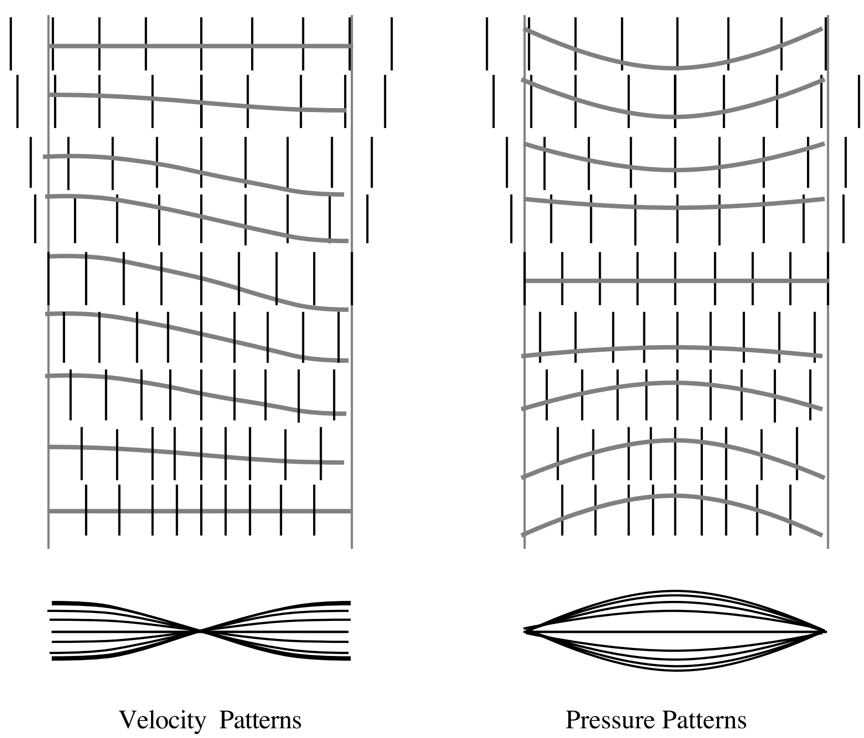
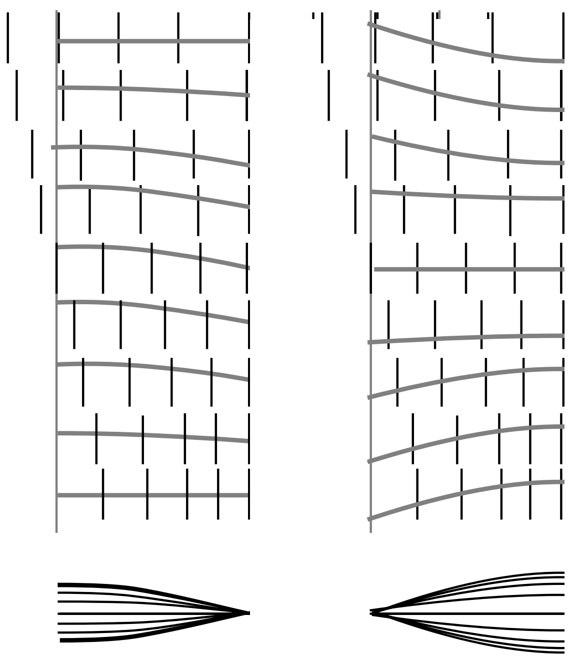


Figure 10.10 The motions involved for one half-cycle of the first standing wave mode of vibration of the air in an open pipe. The diagrams are laid out as in fig. 10.9.

Two aspects of these patterns are immediately obvious. The pattern for the velocity oscillation is analogous to the oscillation pattern for a stretched string tied down at both ends but the pattern for the pressure oscillation is analogous to the oscillation pattern for a stretched string which is free to move at both ends. This reversal of the patterns for pressure and velocity also appears in the case of an open tube, the diagrams for which are shown in fig. 10.10.

Finally, the motions of air in a pipe which is closed at the right end but open at the other is shown in fig. 10.11. Here it can be seen that the pattern is the same as that of the left half of the patterns in Fig. 10.10. A mirror image of this pattern would occur for a pipe closed at the left end and open at the right; a pattern that would be just the right hand half of the patterns of Fig. 10.10. This comes about because the oscillation of the air in the lowest mode of the open pipe involves no motion at its center.



Velocity Patterns Pressure Patterns

Figure 10.11 The motions in one half-cycle of the first standing wave mode of vibration of the air in a pipe which is open at the left but closed at the right end.

The modes of vibration of the air in a pipe are therefore very analogous to those of a stretched string and the simple formulae 10.11 and 10.12 can be used to calculate their frequencies. The old standard Canadian "Stubby" beer bottle approximates a tube closed at both ends and of length about 13 cm. The frequency of the lowest standing wave mode of the air in this bottle should therefore be about

$$f = \frac{c}{2L} = \frac{340}{0.26} = 1308 \text{ Hz} \qquad (10.13)$$

This explains the shrill higher note that can be obtained by blowing very hard across the neck of such a bottle. The still higher note that could be obtained by compressed air from a nozzle would be the second standing wave mode of the air in this bottle.

Thus air in an enclosure can vibrate in modes other then just the simple oscillation of air in and out of a hole. From the simple analysis given here where the modes of oscillation of air in a pipe are taken as analogous to those of a stretched string, it is apparent that, in fact, there are practically an infinite number of such other modes.

However, the analogy between the motion of a stretched string and that of air in an enclosure does not show the full richness of the modes of vibration of the air in an enclosure. This is because air in an enclosure is a three dimensional system while the stretched string is a one-dimensional system. To consider what this does to the possibilities of air motion in an enclosure, first consider the modes of vibration of two-dimensional systems.

10.4 The Modes of Vibration of Surfaces

Vibrating surfaces are also one of the most primitive form of devices for producing music. They are the basic parts of drums and bells and are essential components of many more modern instruments such as the violin. The modes of vibration of a surface therefore determine many of the characteristics of the sound from musical instruments.

Again, we start with the simplest possible example, in this case that of a flat circular surface. The modes of vibration of such a surface will approximate those of the stretched membrane which typically forms the head of a drum.

As in the rest of these notes, only a diagrammatical description will be given. The full mathematical description of the modes of vibration of a circular surface is rather complicated and is not even given in some excellent advanced text-books on acoustics but is left as an exercise in advanced mathematical physics.

> The solution of the wave equation for two dimensional surfaces with circular symmetry turns out to involve Bessel functions which have some of the properties and some of the appearance of trigonometric functions which are, of course, the solutions of the one-dimensional system such as the stretched string. The modes of vibration of a circular surface will therefore have a resemblance to those of the stretched string but there will be features that have no analogy in the string.

The lowest mode of vibration of a flat circular surface that is held down around its edge appears very similar to the lowest mode of a stretched string tied down at its ends. This is a mode in which the center of the surface undergoes the maximum motion (see fig. 10.12).

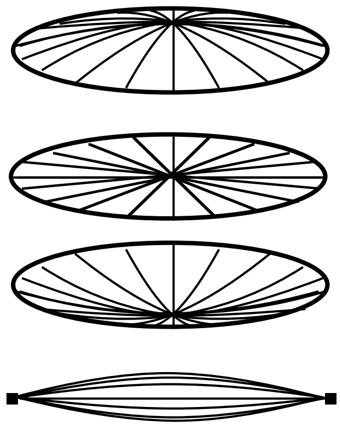


Figure 10.12 The motions involved for one half-cycle of the lowest mode of vibration of a circular plate held down around its rim.

The next lowest mode of this type is one in which the center region of the plate moves in the opposite direction to that of the outer region (shown in fig. 10.13). This mode has some of the characteristics of the third mode of the stretched string except that the distance between the nodal points is not uniform. Also, the amplitude of motion of the outer region is not as great as that of the central region.

There is, nonetheless a mode which is analogous to the second mode of the string. It has the appearance shown in fig. 10.14. This mode of vibration is, however, of a different class of symmetry than the other two already introduced; it is asymmetrical about a particular diagonal of the plate, the nodal line, whereas the others are symmetrical about any diagonal.

This opens up a whole new dimension for the modes of vibration. There can be modes which have two perpendicular nodal lines, with adjacent regions of the plate vibrating out of phase but opposite regions vibrating in phase as shown in fig. 10.15.

Furthermore, the two classes of modes can be combined, the lowest member of this combined class being one in which there is a nodal circle similar to that for the symmetrical mode but also a diagonal nodal line. Fig. 10.16 shows the pattern for two of these combined modes.

All the modes of the circular plate then form a two-dimensional array as shown in Fig. 10.17.

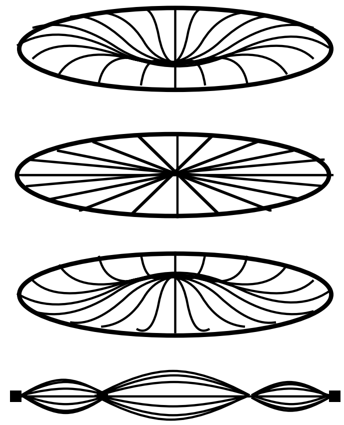


Figure 10.13 The motions involved for one half-cycle of the second symmetrical mode of vibration of a circular plate held down around its rim.

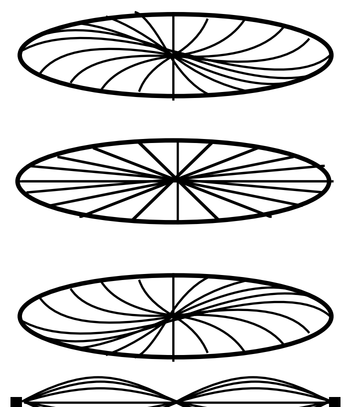


Figure 10.14 The motions involved for one half-cycle of the first asymmetrical mode of vibration of a circular plate held down around its rim.

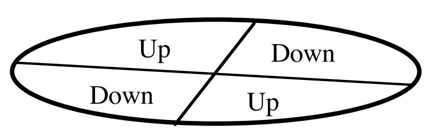


Figure 10.15 The pattern of the second mode of vibration with diagonal nodal lines for a circular plate held down around its rim.

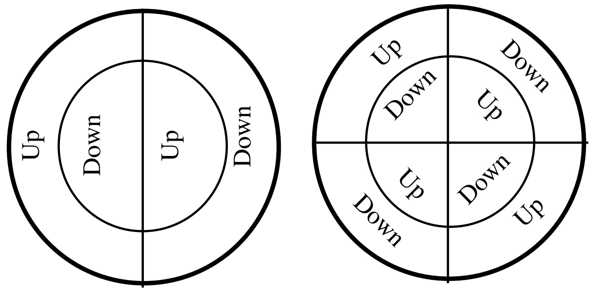


Figure 10.16 The patterns of two of the modes of vibration with diagonal and circular nodal lines for a circular plate held down around its rim.

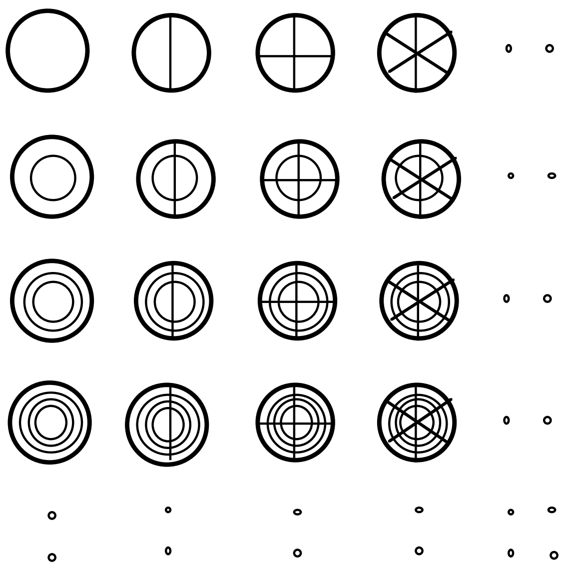


Fig. 10.17 The patterns of the higher modes of vibration of a circular plate held down around its rim. Only the first 16 are shown of the set which extends to infinity in each of the symmetry directions of the diagram.

An equivalent but more restricted set of modes also exist for a circular plate which is not tied down at its edge. In this case the symmetries of the motion and conservation of momentum during the vibration require that there be at least one nodal circle and an even number of diagonal nodal lines.

An important characteristic of all of these modes is that, unlike for the stretched string, there is no simple numerical relationship to their frequencies. The tones which result when these oscillations are coupled to air will therefore not be of much musical use. To get musical tones out of a vibrating plate, it will generally have to be modified so as to have variable thickness or variable mass loading over its surface so that some of the important lower modes of vibration will actually have frequencies which are close to being related by simple numbers. A bell, despite its shape, is essentially a vibrating surface, the lower modes of which have their frequencies tuned to musical intervals by adding metal to or shaving metal from the surface at particular points. This operation is one that involves a high degree of experience and craftsmanship. Another similar surface tuning operation of importance in music is that of the wood panels that make up the top and bottom of a violin.

The mathematical treatment of the vibration of surfaces in typical musical instruments would be enormously complicated and of very little use in setting up the correct modes and their frequencies, this setting up being normally accomplished by trial and error based on a great deal of experience. The reason for introducing the concepts here is that they should lead to a better understanding of the phenomena of vibrations in musical instruments and give a new perspective from which to view the wonderful experience of the production and perception of musical sounds.

Another reason for introducing the modes of vibration of surfaces is to get a better understanding of the cause of the complexity of the direction patterns of sound radiated from a typical musical instrument. The complexity of the radiation pattern from a surface that is vibrating as a whole has already been pointed out in Chapter 5. In a typical musical instrument for which the vibration of its surfaces is an important part of the generation of a musical sound, the surfaces are not vibrating as a whole but in the various modes that have been introduced here. It is to be expected that the radiation patterns from these modes are even more complicated than those shown in Chapter 5.

A final reason for introducing the modes of vibration of a surface is that it can lead to a better understanding of the modes of vibration of air in an enclosure. These modes are another important source of sounds from musical instruments and, as well, are important in the generation of sound in a room. Air in an enclosure is essentially a three-dimensional system and so to understand its modes of vibration , one has to extend the consideration of modes of vibration to three-dimensions.

10.5 The Modes of Vibration of Air in an Enclosure

The modes of vibration of three dimensional systems such as air in an enclosure form an important branch of study in physics and engineering. Consequently they have been extensively analyzed for a variety of geometries. Again, as for the vibrating surface, the simplest possible system to visualize is the one with the greatest symmetry; the circular disk for a surface and the sphere for a volume.

It should not be surprizing that the modes of vibration of air in a spherical enclosure have similarities to those of a vibrating disk. The simplest class of modes have spherical symmetry. These are modes in which the air at any point in the system moves radially in and out from the center (see fig. 10.18). There is a difference from the two dimensional case in that all the modes of this class have a node at the center. (The air at the center cannot move in any direction without violating the spherical symmetry).

Analogous to the modes of the vibrating disk are the modes of vibration of air in the sphere which have symmetry about a diametrical plane. It is perhaps not surprizing that now, however, there is a third type of symmetry possible; that about an axis of the sphere. This opens up another class of modes involving circular motion about such a symmetry axis.

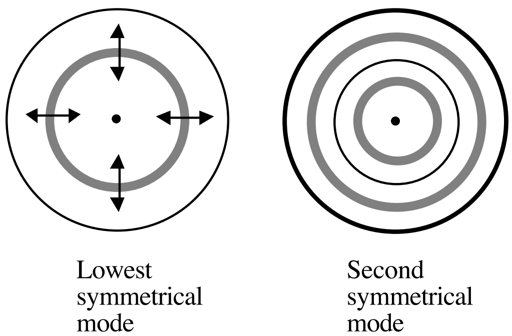


Figure 10.18 The patterns of the first two spherically symmetrical modes of vibration of air in a spherical enclosure. All modes have a velocity node at the center with all modes after the lowest having intermediate spheres on which then is a velocity node.

The modes of vibration of a spherically symmetric system have been analyzed very thoroughly in modern physics because of their importance in understanding the dynamics of the electron cloud forming an atom. The solutions for such a system are expressed in terms of "spherical harmonics". The dynamics of an electron around an atomic nucleus are not the same as those of air in an enclosure but the form of the solutions are very analogous. The various modes of the system are specified by n, ℓ and m "quantum numbers". Here "n" specifies the complexity of the vibration pattern and may be thought of as number of nodal surfaces, axis or points in the system plus one. (The electron cloud in the lowest mode has no nodes.) ℓ may be thought of as the degree of complexity of the pattern about a particular plane, the spherically symmetric modes having $\ell = 0$. m may be thought of as the degree of complexity around a particular axis.

The modes of vibration of air in a sphere have elegant symmetries which are of relevance to modern physics but they are not of much relevance to musical acoustics. This is because spherical enclosures are not often used in music, either in instruments or in auditoria. Of more importance here are the modes of vibration of air in a rectangular box such as a typical room.

While a box may be more complicated to visualize than a sphere, the motions of air in a box are, in fact, easier to visualize than those in a sphere. The lowest modes are particularly easy to visualize (see fig. 10.19). They are just motions parallel to the walls. For example, the lowest frequency mode is just motion back and forth along the longest dimension of the room (usually regarded as the room length);

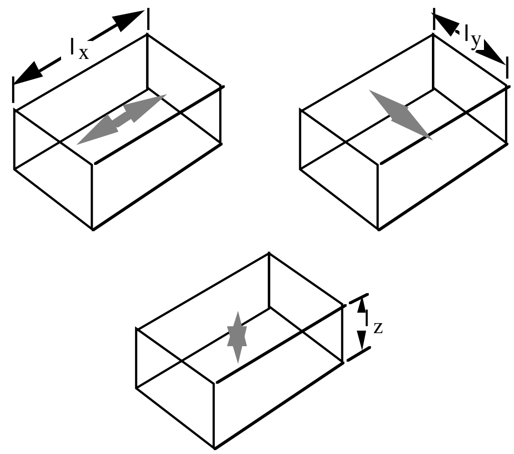


Figure 10.19 The motion of air in the three simplest modes of vibration in a rectangular enclosure.

Since the dynamics of this motion are the same as for those in a closed pipe, the frequencies of this motion is simply $c/2L_x$ where L_x is the room length.

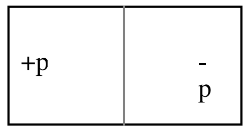
The next lowest mode is motion along the middle dimension of the room (usually the room width). The frequency of this oscillation will, of course, be $c/2L_y$ where L_y is the room width. There will also be, of course, a vibration mode in which the air moves along the shortest dimension of the room (usually the room height). The frequency of this oscillation will be $c/2L_z$ where L_z is the room height.

The type of pressure pattern in the room during these modes of oscillation is shown in Fig. 10.20. Note that there is a pressure node in the middle of the room and that therefore this type of air oscillation will not shown a pronounced effect on a typical pressure sensitive microphone placed in the center of the room.

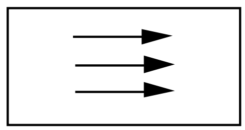
There will, of course, be higher modes along each of the three dimensions, the frequencies of these higher modes along any dimension being just integer multiples of the lowest mode in that dimension.

There are also higher modes with symmetries in two dimensions. The lowest mode combining motion in the x and y direction is of this type shown in Fig. 10.21

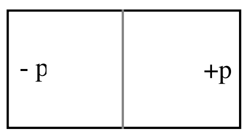
The motions for the lowest mode combining all three directions is shown in Fig. 10.22. Again, it can be seen that there is a three dimensional pattern of symmetries in the modes of oscillation of the air in the room, corresponding to the three dimensional nature of the motion.



Maximum pressure on theleft (no motion)

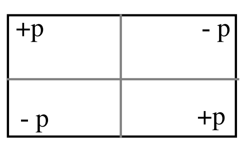


Mazimum motion to the right (no pressure)

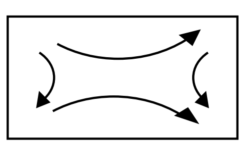


Maximum pressure on the right (no motion)

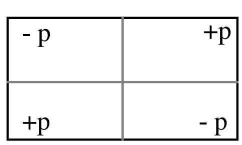
Figure 10.20 The motion of air for one half of a cycle of the lowest mode of vibration in a rectangular enclosure.



Maximum pressure (no motion)

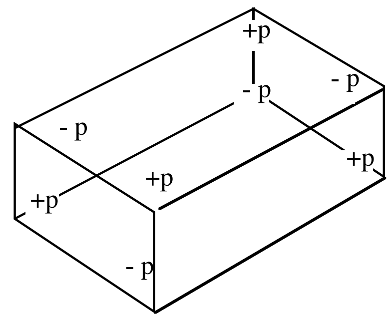


Mazimum motion (no pressure)

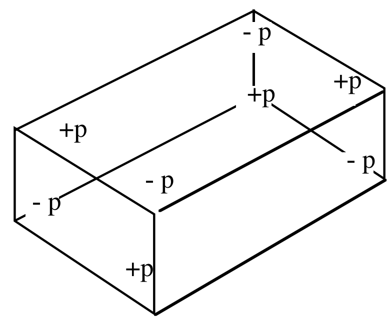


Maximum pressure (no motion)

Figure 10.21 The motion of air for one half of a cycle of the lowest crossed mode of vibration in a rectangular enclosure.



Pressure at start of half cycle (pressure maximum, no motion)



Pressure at end of half cycle (pressure maximum, no motion)

Fig. 10.22 The motion of air for one half of a cycle of the lowest mode of vibration in a rectangular enclosure encompassing all three directions of motion.

The frequencies of all the possible modes of vibration of air in a rectangular enclosure is given by a rather simple formula

$$f = \frac{c}{2}\left[\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2\right]^{\frac{1}{2}} \qquad (10.14)$$

where n_x, n_y and n_z refer to the number of pressure nodal lines in the x, y and z motions respectively. For example, the lowest mode of vibration in the system in the x direction would have $n_x = 1$ but n_y and $n_z = 0$. This mode would be referred to as the (1,0,0) mode. The lowest node that combines motion in the x and y direction would have n_x and $n_y = 1$ but $n_z = 0$ and would be referred to as the (1,1,0) mode. The lowest mode that would combine all three would be the (1,1,1) mode.

For the normal room containing the sounds produced by a musical instrument, the frequencies of the sounds will correspond to high modes of the room vibrations. As an example, consider a typical lecture room 13 m long, 8 m wide and 3 m high in which an instrument is playing a note of 250 Hz. The modes of air motion along the length of the room with frequencies nearest this note are (19,0,0) and (20,0,0) at 248.5 and

261.5 Hz respectively (assuming the velocity of sound to be 340 m/s). The modes in the other directions are the (0,11,0), (0,12,0,), (0,0,4) and (0,0,5) modes with frequencies of 233.8, 255, 226.7 and 283.3 Hz respectively.

An important feature of these modes is that for a normal room there are usually very many of them in the range of frequencies important to the pitch of a musical note. For example, in the room just considered there would be 69 modes between 245 and 255 Hz. This means that, in principle, there will be some mode or modes very near the actual frequency of the tone and which can be excited by the tone from the instrument. This gives a completely new way to consider reverberant sound in a room; it is made up of sound in the various modes of air vibration in the room.

However, many of the modes of air oscillation involve motion between the floor and the ceiling. In fact, of the 69 modes between 245 and 255 Hz, only 13 involve only motion parallel to the floor and ceiling. If the sound absorption of the room is concentrated on the floor and ceiling, then the only modes that will be easily driven by the musical instrument will be from these 13 modes. Thus there is a possibility that notes of different frequencies will not be evenly enhanced by room air vibrations.

Again the need is seen for the sound absorption to be scattered uniformly throughout the room so that all modes of room air vibration can take part evenly in the sound. By extending the number of possible room air vibrations in the given frequency interval of 245 to 255 Hz from 13 to 69, there will be a much more even distribution of the sound energy over the possible modes.

Another way to introduce more modes of vibration into a given frequency interval would be to raise the height of the room. Taking a room of 12 × 8 × 4.8 m, which has about the same seating capacity as the above "lecture" room but now with the recommended dimension ratios for music of 4:3:1.6, one gets 122 modes of air vibration in the frequency range of 245 to 255 Hz. Such a room, providing the sound absorption surfaces are scattered throughout the room, should provide a much more even distribution of the reverberant sound than the lecture room.

In can be shown that the number of modes of air vibration in a given small frequency interval at high values of n is proportional to the square of the center frequency of that interval and the volume of the room. The formula is

$$\Delta N = \frac{4\pi}{c^3} f^2 \, \Delta f \, V \qquad (10.15)$$

where ΔN is the number of mode in the frequency interval Δf, f is the center frequency of the interval and V is the room volume. This result is an important one in physics. Its importance in acoustics is that it shows the importance of room volume in obtaining a rich spectrum of modes in room reverberation. It also shows that the problem of having adequate numbers of vibration modes occurs mostly for the lower frequencies.

Another aspect of the modes of air vibration in a room that is important to the recording engineer is that they all have pressure nodes somewhere in the room. Placing a pressure sensitive microphone anywhere in the room can produce a distortion of the spectrum of the room modes caused by some of the nodes near a particular note having pressure nodes at the position of the microphone. Placing the microphone very near a wall will eliminate this possibility for many of the modes but the only sure place to avoid nodes is to place the microphone in a corner of the room. The richest spectrum of room modes in the reverberant sound will therefore be obtained by such a microphone placement.

10.6 Some General Aspects of Standing Waves; The Concept of Normal Modes

In these notes, standing waves have been introduced as special, separate ways that systems can oscillate. How are such modes related to the general oscillation of systems at any frequency?

To begin this subject, consider a simple mechanical system made up of two identical pendula lightly coupled by a thin rod. Such a system can be easily constructed as shown in fig. 10.23 from two 200g masses, two pieces of string and a soda straw.

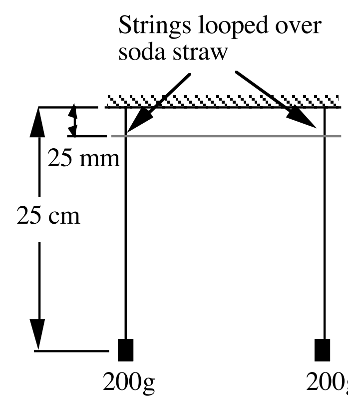


Figure 10.23 A simple coupled pendulum made of two 200g masses, two pieces of string and a soda straw. The strings are looped over the soda straw at about the position shown, the distance of the soda straw from the top support being the same for each string. The soda straw transfers the motion of one of the pendula to the other.

If one of these pendula is pulled to one side and released while the other is not disturbed, the initial motion will be just an oscillation of the moved pendulum. However, after about 20 seconds, the motion of this pendulum will have stopped and the other pendulum will have picked up the motion of the first. After another 20 seconds, the motion will transfer back to the first.

This curious behavior is at first a little difficult to understand. However, drawing a graph of the motions will give a clue as to some underlying cause. Such a graph is drawn in fig. 10.24.

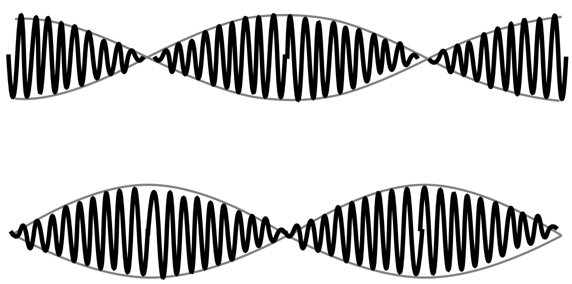


Figure 10.24 The motions of the two pendula in Fig. 10.23 when one of the pendula is given an initial displacement and released. The top diagram refers to the pendulum given the initial displacement and the bottom diagram to the other.

It is apparent that there is a beat phenomenon in the motion of the two pendula. A beat occurs when there are two equal amplitude simple harmonic motions of slightly different frequency. This means that there are two simple harmonic motions that are "beating" in this system.

The two simple harmonic motions that are in this system are fairly easy to set up. All that is required for one of them is to displace both pendula the same amount before releasing them. The resulting motion will be a constant swinging of both pendula, in phase, at a frequency of about 1 Hz. The second motion can be set up by displacing each pendulum the same amount but in opposite directions before releasing. The resultant motion will again be steady simple harmonic motion of both pendula but this time out of phase and of frequency about 1.05 Hz.

It is relatively easy to see, in passing, how these two frequencies come about. For the symmetrical motion, the length of the pendulum motion is the full length of the strings. For the asymmetrical motion, the length of the pendulum motion is the length of the strings from the soda straw to the masses.

The beat frequencies of these two types of simple harmonic motion would be 0.05 Hz or 20 seconds per beat. What this implies is that the initial displacement of only one of the masses is equivalent to putting equal amounts of these two simple harmonic motions into each mass and having them beat together.

That this is so can be seen by combining equal amounts of an initial amplitude of the two simple harmonic motions (see fig. 10.25).

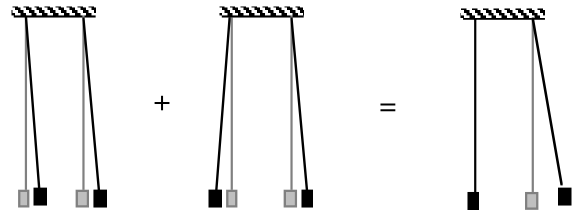


Figure 10.25 The input of equal amounts of the two pure simple harmonic motions of the two pendula in Fig. 10.21. It can be seen that this addition is equivalent to an initial displacement of only one of them.

This introduces one of the most important concepts in physics. The combined motions of the pendula that resulted in steady simple harmonic motions for each are called the "normal modes" of the system. Any initial set of displacements of the masses can be decomposed into initial amplitudes and phases of these two fundamental normal modes.

The importance of this analysis is that, at any time after the initial set-up, the condition of the system is the result of the superposition of these two modes. In other words, after the set-up and release, <u>the two normal modes behaved completely independently of each other</u>.

This, in fact, is why these modes are called "normal". The word "normal" here does not have the meaning of "usual" or "expected" but in the mathematical meaning of "perpendicular" as in the normal to a plane. Normal coordinates, such as the x, y and z coordinates in the usual perpendicular coordinate system, are mathematically independent. (In everyday terms, walking so as to change only your x and y coordinates doesn't chance your height z.)

Many mechanical systems have normal modes of oscillation, sometimes hard to identify. An amusing example is the Wilberforce Pendulum made up of a single mass with extended arms and a single spring. By pulling down the mass and releasing it, a curious motion is set up in which the initial up and down motion of the mass translates itself completely into a spinning motion of the mass and back again (see fig. 10.26)

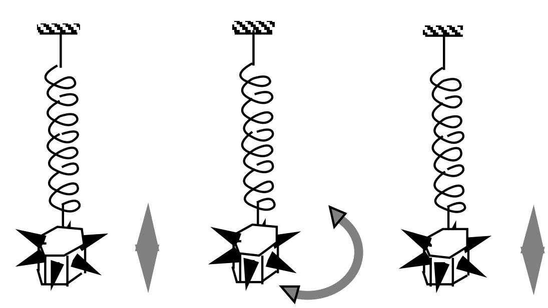


Figure 10.24 The motions of the Wilberforce Pendulum.

Again, there is obviously some sort of beating motion. Here it takes some experimentation to find that the normal modes of this system are a certain amount of clockwise twist, looking down, with a downward displacement and another with the same amount of counterclockwise twist associated with the same downward motion. To get the amusing beats, these normal modes are tuned by adjusting the lengths of the arms on the mass until they are very close in frequency. A straight downward pull with no twisting is then equivalent to putting in equal amounts of the normal mode with a clockwise twist and the normal modes with the counterclockwise twist. (A twist to the right and a twist to the left at the same time is no twist at all.)

There is another important aspect of normal modes. The number of normal modes in a system is equal to the number of degrees of freedom of all the elements of the system. In the case of the Wilberforce pendulum there are two degrees of freedom for the one mass; an up and down motion and a twisting motion. Hence there are only two normal modes of vibration. For the coupled pendula there are two masses. If the pendula are allowed to only move in the plane of the diagram of Fig. 10.21, then there is only one degree of freedom for each pendulum and so the number of normal modes is again two.

A string, a plate or a sphere is made up of practically an infinite number of individual masses, i.e. atoms. There are therefore, in principle, almost infinite numbers of normal modes of oscillation. The standing waves that have been considered in this chapter are, in fact, the normal modes of oscillations of these systems.

The fact that the standing waves in continuous systems are normal modes of these systems means that these modes are independent. Energy put into these normal modes at a start-up of a system will therefore stay locked into each normal mode without transferring from one to the other. Each normal mode can be expected to have its own rate of energy dissipation and so some of the modes will die more quickly then the others.

This explains the behavior of many struck musical instruments such as the piano or the guitar. The initial displacement of the string by an impact will correspond to a certain mixture of normal modes of that string. Each of these normal modes then decays at its own rate, in general the higher frequency components most quickly. The energy put into the room by these modes as they decay are the "partials" of the musical tone the instrument is producing. The tone of these instruments therefore starts with a very rich mixture of partials in which the fundamental is relatively weak and decays rather quickly into predominantly the fundamental.

It should also be noted that in such struck instruments the partials have the frequencies of the normal modes. These frequencies are only harmonics of the fundamental for the ideal string. In real strings such as those of a piano or a guitar there is a finite thickness and the system becomes a complex three dimensional one. The frequencies of the higher modes are then not exactly harmonics of the fundamental and so the musical partials of the tones produced are not exactly harmonics. This gives the tones a special characteristic; close enough to being harmonics to be musically related but far enough off to be more interesting than pure harmonics.

Exercises and Discussion Topics

1. Describe the action of a Helmholtz resonator through one cycle. In your description of the action, answer the following questions;

   a) In what part of the resonator does the air velocity play a dominant role and in what part does the pressure play a dominant role?

   b) What is the analogy to the oscillation of a mass on a spring?

   c) What are the two forms of stored energy in the system and how do each behave in time through the oscillation?

2. What would be the acoustic capacitance, the acoustic inertance and the resonating frequency of the lowest frequency oscillation of a 4 liter wine jug with a 3 cm long neck of 2 cm diameter?

3. Suppose that all dimensions of a bottle are doubled.

   a) By what factor is the acoustic capacitance of the bottle changed?

   b) By what factor is the acoustic inertance of the neck changed?

   c) By what factor is the Helmholtz resonating frequency changed? How is this related to your answers for parts a) and b)?

4. Describe the motion of air as a sound wave with a pure tone is passing through it. What is the relative phase of the pressure oscillation and the velocity oscillation? What changes between pressure and velocity when a wave goes in the opposite direction? Distinguish clearly between the actual air velocity and the wave velocity of the sound.

5. Describe the motion of air in the lowest two modes of vibration in a closed pipe (use sketches as much as possible to save words in your descriptions.) What is the relative phase of the pressure oscillation and the velocity oscillation in different parts of the pipe? What changes between pressure and velocity as one looks at different parts of the pipe?

6. Describe the general features of the normal modes of vibration of two and three dimensional systems and why one would expect the radiation patterns of these modes from typical musical instruments to be very complex.

7. Discuss room reverberation as being made up of the normal modes of vibration of a room and their decay after the sound source has been turned off.

8. What would be lowest frequency normal mode in a room which was 10 meters wide by 15 meters long by 4 meters high? What would be the next five frequencies?

   b) Where should you place a high impedance microphone to pick up all the normal modes of room resonance that may be in a reverberant sound? Why? What type of modes would you miss by placing a microphone in the exact center of the room?

9. What are the two normal modes of the simple two coupled pendula oscillator? How is it that these modes can be said to each be normal? How is the simple movement to one side and release of one of the pendula described in terms of these two normal modes?

10. In what sense are the normal modes of oscillation of a system "normal"? Of what significance are the normal modes in describing any arbitrary oscillation of a system due to some driving force. What importance do the normal modes have in how the vibration energy of a system is dissipated?

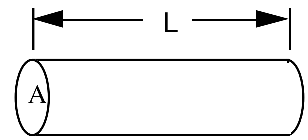
11. Define harmonics, musical partials and normal modes of oscillation. Point out clearly how any one of these may not be equivalent to another and discuss their importance in music in general.

Answers

2) 2.82 x 10^{-8} m^4s^2/kg, 179.6 kg/m^4, 70 Hz; 3) a) x8,b) x1/2, c) 1/2 = $\sqrt{1/(8 \times 1/2)}$; 8) (1,0,0) 11.33Hz, (0,1,0) 17 Hz, (1,1,0) 20.43 Hz, (2,0,0) 22.66 Hz (2,1,0) 28.33 Hz

APPENDIX

The following is a short derivation of the equations of the Helmholtz oscillator. Consider a cylinder of gas of density ρ , length L and cross-sectional area *A*.



The mass of this gas will be the density times the volume;

$$m = \rho L A \qquad (A10.1)$$

The force on this gas by a pressure on one of its end faces is just *p* times *A*, giving an equation;

$$F = ma \ ; \ pA = \rho L A\, a \qquad (A10.2)$$

or

$$p = \rho L\, a \qquad (A10.3)$$

The acceleration in this equation can be translated into a rate of change of volume flow rate by noting that the volume flow rate is the velocity of the air multiplied by its cross-sectional area. Thus the rate of change of this flow rate is the cross-sectional area multiplied by the rate of change of velocity. The rate of change of velocity is , of course, the acceleration. Thus

$$A\, a = \textit{Rate of change of flow} \qquad (A10.4)$$

Rewriting the equation in terms of *p*, *M* and rate of change of flow gives

$$p = \rho L \times \frac{\textit{Rate of change of flow}}{A} \qquad (A10.5)$$

On the other hand

$$p = M \times \textit{Rate of change of flow} \qquad (A10.6)$$

Therefore

$$M = \frac{\rho L}{A} \qquad (A10.7)$$

To get the other equation for the Helmholtz oscillator (that dealing with the overpressure for a given flow input to a container), use the equation for the adiabatic compressibility of a gas;

$$pV^{\gamma} = Constant \qquad \text{(A10.8)}$$

From this, using simple calculus it can be shown that

$$\frac{\mathrm{d}p}{p_o} = -\gamma \frac{dV}{V} \qquad \text{(A10.9)}$$

$$\mathrm{d}p = -\, p_o\, \gamma \frac{\mathrm{d}V}{V} \qquad \text{(A10.10)}$$

where p_o is the normal atmospheric pressure on the gas in the container and dp is the excess pressure associated with the change in volume dV. This is the pressure that would be required to reduce a volume of gas V by dV which would be equivalent to bringing an outside volume of gas dV and putting it into a fixed container of volume V. Expressing this quantity of gas as Q gives;

$$\mathrm{d}p = \frac{p_o\, \gamma}{V} Q \qquad \text{(A10.11)}$$

The pressure involved in the motion of the air is the dp of this equation. Putting this as the sound pressure p gives

$$p = \frac{p_o\, \gamma}{V} Q \qquad \text{(A10.12)}$$

On the other hand,

$$p = \frac{Q}{C_A} \qquad \text{(A10.13)}$$

Therefore

$$C_A = \frac{V}{\gamma p_o} \qquad \text{(A10.14)}$$

Using the relationship giving the velocity of sound in air;

$$c = \sqrt{\frac{\gamma p_o}{\rho}} \qquad \text{(A10.15)}$$

gives the final form for the acoustic capacitance of a volume of air;

$$C_A = \frac{V}{\rho c^2} \qquad \text{(A10.16)}$$

CHAPTER 11

THE GENERATION OF MUSICAL SOUNDS

The normal modes of vibration in a musical instrument are the origins of musical sounds in that instrument. The frequencies of these normal modes and their relative strengths determine the timbre of the notes produced by the instrument and their patterns determine the radiative properties of the sound from the instrument.

However, a musical instrument left to itself is quiet; it's normal modes of vibration merely represent the possibility of producing music. Music is generated when this instrument is manipulated so as to generate a particular set of it's normal modes. The notes must grow from nothing to a level sufficient to be heard as component parts of the music.

The nature of the growth of the sound for notes from a particular instrument, a property called the "attack" of the notes, is a very important characteristic of the instrument. It is particularly important in identifying a particular instrument as the source. Demonstration recordings in which the initial 50 ms of the notes have been removed show the importance of preserving the attack; in such recordings the instruments are almost unidentifiable. Also, since the normal modes of oscillation in a typical musical instrument grow to full power in less time than it takes for the room reverberation modes to grow to full power, most of information concerning how the modes develop is in the direct sound and the first few room reflections.

This is yet another reason why faithful reproduction of the direct sound is very important in a recording. How the normal modes of an instrument develop is therefore an important consideration in recording engineering.

The methods of exciting the normal modes of a system can be usefully divided into three basic classes; resonance, impulse and feedback. These will each be considered in turn with an overview at the end of how they all interrelate in musical instruments.

11.1 The Excitation of Normal Modes by Resonance.

One of the simplest ways to excite a normal mode of vibration is by resonance. This is the process by which a system with a normal mode of a particular frequency is excited by vibrating the system with an oscillating external force at or near that frequency. This is an important phenomenon in many musical instruments, an example being the string instruments in which the Helmholtz oscillator formed by the box of the instrument and its holes is excited in resonance by the strings of the instrument.

To examine the basics of excitation by this mechanism, consider again the Helmholtz oscillator formed by the old-fashioned Canadian beer-bottle, the "stubby". This oscillator can be resonated by attaching a small loudspeaker to its bottom with modeling clay and the resonance can be observed by a small capacitance microphone lowered into the bottle (see fig. 11.1). When the output of an amplifier for a pure tone is fed into the small speaker in this set-up and the output of the microphone is displayed on an oscilloscope a pronounced increase in the height of the displayed waveform is seen when the tone generator hits the frequency of the Helmholtz oscillation. For the standard "stubby" beer bottle, this frequency will be 216 Hz.

Above this frequency the waveform sharply diminishes but as the frequency is increased there will also be pronounced peaks at 1496, 2860, 3380, 3750 and 3950 Hz. These frequencies, together with the Helmholtz frequency, are the normal mode frequencies of this system.

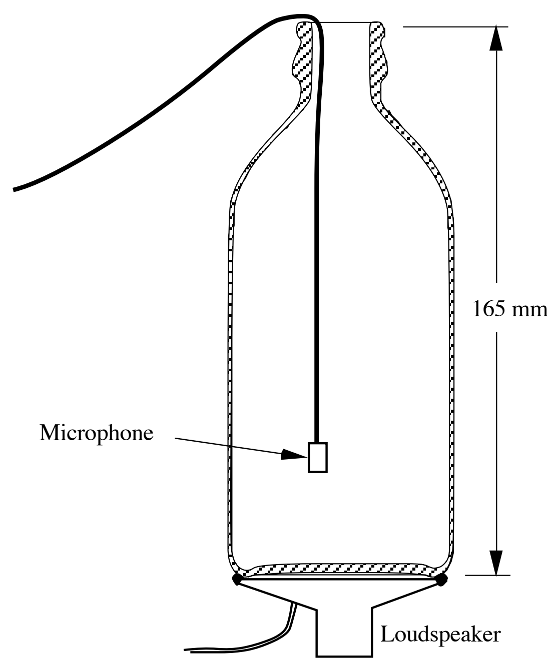


Figure 11.1 A system for investigating the normal modes of air in a bottle by resonance. The microphone output is led to a preamplifier the output of which is displayed on an oscilloscope.

In addition to being used to detect the response of the system to different frequencies of a driving force, the microphone can be used to investigate the vibration pattern of a particular mode at its resonance. Thus at resonance on the Helmholtz frequency the microphone shows practically the same response as it is moved anywhere throughout the interior of the bottle. This is to be expected for the Helmholtz oscillation since it is an assumption of the simple model of the oscillation presented in Chapter 10 that the air entering the bottle through the neck builds up a uniform pressure inside the bottle.

Another important feature of resonance can be seen by noting the sound pressure inside the bottle compared to that immediately outside it at the opening. In the set-up used here the sound just outside the bottle at resonance cannot even be detected by the oscilloscope, indicating that only a small fraction of the energy stored in the resonance is radiated away from the system in any one oscillation of the system. This can be seen more dramatically by observing the sound pressure inside the bottle when the Helmholtz oscillation is exciting by blowing across the neck. When there is only a gentle sound outside the bottle due to this oscillation, the sound pressure inside the bottle is seen to overload the microphone. Since this microphone-preamplifier combination is designed to operate at up to about 120 dB, this means that the sound pressure inside the bottle as a result of the gentle blowing is ear-shattering.

However, at the resonance at 1496 Hz the microphone shows a maximum response at the bottom and at the top of the bottle but almost no response at all when it is exactly half-way in the bottle. Looking at the waveform carefully with the oscilloscope triggered on a synchronization pulse from the signal generator, it can be seen that the pressure at the bottom of the bottle is 180^{o} out of phase with that at the top of the bottle. This is the response to be expected for a "half-wave" resonance in the bottle since a capacitance microphone responds to the sound pressure rather than the sound air velocity and the sound pressure has a node in the center of a closed pipe for the half-wave. The probing of the resonance with the microphone therefore proves that the mode at 1496 is the "half-wave" mode and that the hole in the bottle is not large enough for the bottle to be regarded as an "open-ended" pipe.

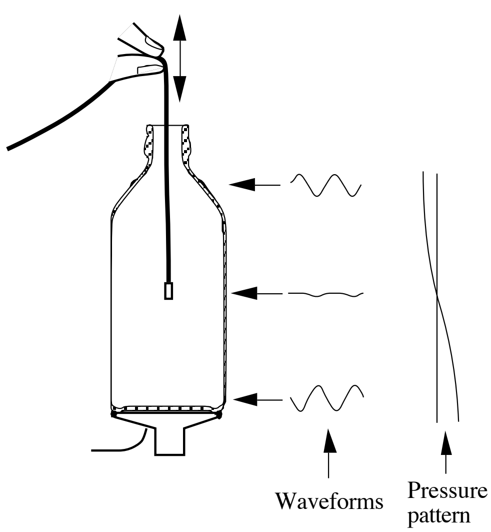


Figure 11.2 The pattern of pressure oscillations in the bottle shown in Fig. 11.1 when the 1496 Hz mode is excited.

It is perhaps interesting to note that the half-wave resonance for a closed pipe of 14 cm length (approximately the length of the interior of the bottle) would be about 1230 Hz, indicating that the model of the bottle as a closed pipe is only very approximate.

It is also perhaps interesting to note that while the higher normal modes are being excited by resonance, the simultaneous excitation of the Helmholtz oscillation by the background noise in the room can be seen as a 216 Hz ripple on the higher frequency mode. This again shows the independent character of the normal modes; the excitation of one by resonance has no effect on the possibility of exciting another at the same time.

For the resonance at 2860 there is seen to be nulls in the pressure at two depths in the bottle, indicating a "full wave" mode. For the resonance at 3380 Hz there is seen to be a null in the response anywhere along the axis of the bottle but a maximum response at any depth along the inside of the wall indicating a cross mode of oscillation.

While excitation by resonance is a very good way to detect the normal modes of a system and to determine their frequencies and vibration patterns, what is of interest here is how a normal mode grows when from excited from an initially quiet state by resonance.

11.2 The Growth of Normal Modes when Excited by Resonance.

A familiar example of resonance which occurs on a time scale such that the growth can be easily observed is that of an adult pushing a child in a swing by a sequence of small pushes in synchronism with the child's motion. Gentle pushes will slowly build up an amplitude of swinging to a level determined by the adult as being appropriate; the stronger the pushes, the faster the swinging motion develops and the higher the eventual degree of motion.

In the case of resonance sound in music, the rate of growth of the resonance is important. The advantage of using an oscilloscope to probe the sound level of a normal mode is that the display is so fast that the growth of the resonance oscillation can still be observed. In the case of the Helmholtz resonance of the system shown in Fig. 11.1, it will be seen to be that shown in Fig. 11.3. This picture is obtained by triggering the oscilloscope display on the start of the signal in the bottle.

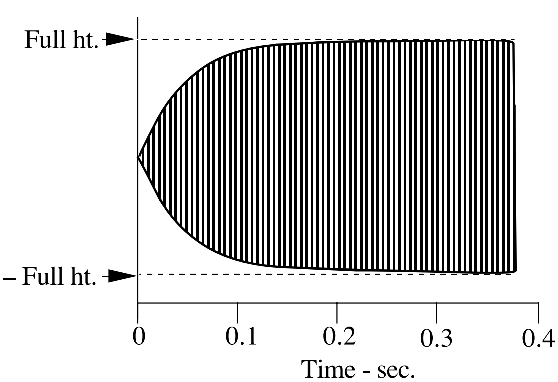


Figure 11.3 The growth of the Helmholtz oscillation by resonance in the bottle of Fig. 11.1.

This is seen to be like the saturation curve for the growth of room reverberation studied in Chapter 4. If indeed it is similar to the case for room reverberation then there will be an exponential decay with a particular half-life after the source of the sound has been turned off. This can be checked by timing a shut -off of the tone generator with the oscilloscope display so that the decay of the oscillation can be seen (see Fig. 11.4).

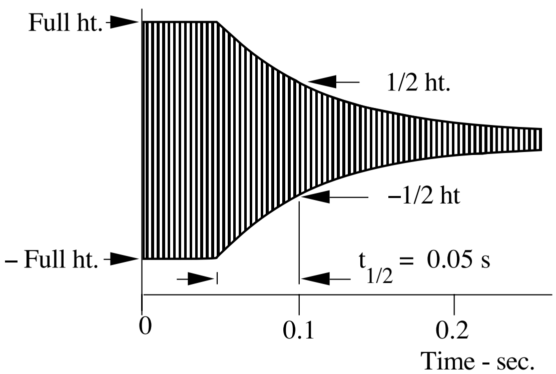


Figure 11.4 The decay of a Helmholtz oscillation excited by resonance in the bottle shown in Fig. 11.1 after the tone generator has been shut off.

It appears that the decay is indeed exponential with a "half-life" of about 75 ms. The half-live for such decays does not depend on the sound level before the decay starts. This in turn means that the time it takes for a sound to reach a certain fraction of its final level will <u>not</u> depend on the strength of the excitation. All that a higher excitation power accomplishes is a higher excitation level.

11.3 The Oscillation Amplitude of Normal Modes when Excited by Resonance.

The actual amplitude that is reached in the excitation of a normal mode of oscillation by an oscillating driving force with a frequency near the frequency of that mode is an important subject in engineering. In engineering the object is often opposite to that of music, the purpose being to prevent an oscillation from occurring rather than to deliberately try to generate one. For example, a wind-created oscillating driving force resonating with a normal mode of torsional oscillation destroyed the famous Tacoma suspension bridge in Washington state shortly after it was opened to the public. A long freight train, such as in western Canada, can develop dangerous longitudinal waves if an inexperienced engineer applies an oscillating engine force. For this reason, the physics of driven oscillators can be found in any intermediate text in mechanics for physics or engineering.

Again, the purpose of these notes is not to present the underlying mathematics of the phenomenon but to give the results in as understandable a form as possible for someone being introduced to the phenomenon. (An outline of the mathematics is given in the appendix to this chapter.)

Also, as in the rest of these notes, the starting point is the simplest possible example. Here that example is a mass on a spring, a system that has only one mode of oscillation. Consider what happens as this system is shaken by an oscillating force of constant amplitude but at a successively increasing frequency. To take a concrete example, a mass hung on a vertical spring can be given such a force by moving the point of support up and down by hand at a fixed amount at various frequencies (see Fig. 11.5).

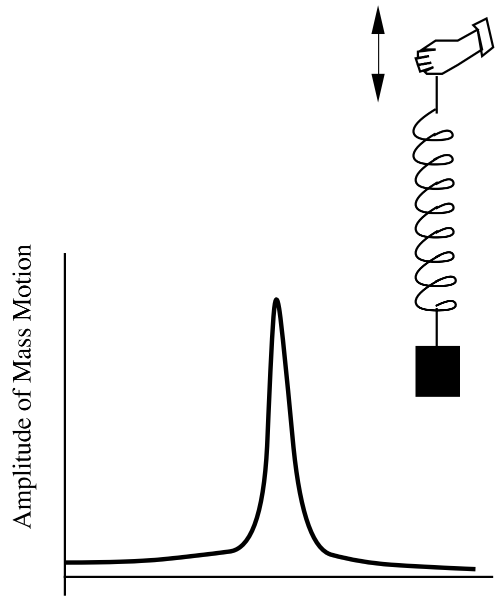


Fig 11.5 A set-up which will apply a constant amplitude force of varying frequency to a simple oscillating system. The hand is moved up and down with the same amplitude of motion at different frequencies. Since the force applied to a spring is proportional to its stretch, the applied force due to the hand motion is also of constant amplitude. The resonance curve that would result is shown on the right. For a typical mass on a spring, the hand movement to get the mass to move up and down through 20 cm has to be only about one millimeter.

An important result of the physics of such a system is that, if the change of the frequency is carried out very slowly, there will be a steady vibration of the mass <u>at the frequency of oscillation of the force</u> (not the natural frequency of the oscillator itself.) Also, the amplitude of this steady oscillation will change with the frequency of the force; at very low and very high frequencies it will be small but at the natural frequency of the oscillator it will be very large.

When a system with a natural oscillation is vibrated at the frequency of that natural oscillation, the system and the driving force are said to be in "resonance" and the resulting vibration of the system is at a maximum for that force. The actual amplitude at resonance will, of course, depend on the magnitude of the oscillating force but it will also depend on the frictional forces resisting the motion. The higher the friction, the less will be the amplitude of vibration at resonance. In the example of the child on a swing, if the supports for the swing ropes are rusty, it will take a larger oscillating force to get a desired amplitude of swing.

One of the most common types of frictional force encountered in oscillators is that in which the frictional force is proportional to the speed. This is the type of frictional force involved in movement through air or water at moderate speeds and is usually termed "viscosity". It is also the type of force encountered by an electrical charge moving through a typical conductor. For such a frictional force, the amplitude of an oscillator at resonance will be proportional to the "resistance" of the system where this resistance is defined as

$$r = \frac{Force}{Velocity} \qquad (11.1)$$

In the case of electricity, the resistance is defined as the voltage divided by the electrical current. Thus electrical voltage is analogous to mechanical force and electrical current is analogous to mechanical velocity.

Turning to the specific example of a mass on a spring, if the mass is suspended in air a very small oscillating force will cause a very large motion at resonance. In fact, it will be very difficult to achieve a steady state oscillation at resonance; the oscillation will continue to grow until it is so violent that the limits of allowable motion of the mass will be reached. If, however, the mass is immersed in water, the motion at resonance will be much more restrained and if it is immersed in thick oil it will be still more restrained. Typical behaviors at resonance will be as shown in Fig. 11.6.

The mathematical relationship of the amplitude of a driven oscillator to the dynamics of its motion is given in the appendix. Here only some of the features of the results will be described.

11.4 The *Q* of Oscillators

A common method for expressing the amplitude gain of an oscillator at resonance is by it's "Q" value where *Q* refers to the "Quality" of an oscillator. There are several equivalent ways to define *Q*. One of the most intuitive is that of the amplitude of the oscillation at resonance for a given amplitude of driving force.

11.4.1 Relationship of Q to Amplitude at Resonance

Consider an oscillation which when driven by an oscillating force at very low frequency compared to its natural oscillation frequency reaches an amplitude of oscillation of A_{low}. If an oscillating force of the same amplitude but at the resonant frequency results in an amplitude A_o then Q may be defined as

$$Q = \frac{A_o}{A_{low}} \qquad (11.2)$$

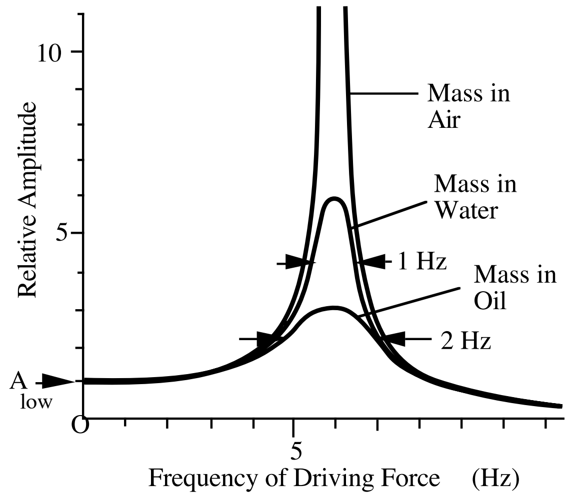


Figure 11.6 The steady-state amplitude of vibration of a mass on a spring at different driving force frequencies for different environments of the mass. The top curve is for air, the middle for water and the lower for thick oil. A_{low} is the amplitude of oscillation that is achieved for very low driving frequencies. The curve for the mass in air would reach a peak of about 100 A_{low} and would have a separation of only 0.06 Hz between the low and high frequency points that resulted in an amplitude of 70 A_{low}.

For the examples in Fig. 11.6, the *Q* for the mass in oil would be about 3, for the mass in water about 6 and for the mass about 100. The *Q* of oscillating systems involved in the mechanical generation of musical tones will generally be in the range of 10 to several hundred.

11.4.2 Relationship of Q to Width of the Resonance Curve

There is, however, an equivalent definition of *Q* which generally allows an easier experimental determination of its value. This is related to measurements taken only near the resonance itself. It turns out to be of particular importance when there is more than one mode of oscillation of a system where the amplitude at low driving frequencies cannot be related to any one particular mode.

This definition of *Q* is related to the "width" of the resonance curve. It can be seen from the diagrams in Fig. 11.6 that for heights on either side of the resonance that are at the same fraction of the resonance height, the frequency gap is greater for the lower *Q* systems. This effect can be expressed quantitatively by noting the frequency interval between two points on a horizontal line drawn through the curves at some arbitrary fraction of the peak height and referring to this gap as the resonance "width".

The actual fraction of the peak height taken for defining *Q* by a resonance width is $1/\sqrt{2}$ or about 70%. For the curves of Fig. 11.6 the resonance widths so defined are about 2 Hz for the mass in oil, 1 Hz for the mass in water and 0.06 Hz for the mass in air.

For the cases of the mass in oil and in air the Q can be determined by noting the relative heights of the resonance and the amplitude at low frequency (A_{low}). These are seen to be 3 and 6 respectively. For these cases, at least, Q is related to the resonance widths by the simple formula

$$Q = \frac{f_o}{\Delta f} \qquad (11.3)$$

where f_o is the resonance frequency (6 Hz) and Δf is the resonance width.

This is, in fact, the accepted definition of Q in terms of resonance width. The particular fraction of resonance amplitude taken to define the resonance width comes from energy considerations. The energy of an oscillation is reduced to 1/2 when the amplitude is reduced to 1/√2. The "A = 1√2" points are therefore the "E = 1/2" points. If the oscillation is connected with the generation of a sound, then the intensity will usually be directly related to the energy of the oscillator and so the "E = 1/2" frequencies will become the "*minus* 3 db" frequencies. In electrical engineering the frequency interval between the two "*minus* 3 db" points in the response of a resonating system is often called the "*Bandwidth*" of the oscillator .

The determining of the Q of an oscillator by measuring the resonance frequency and the frequency interval between the "1/√2" points is equivalent to the determination by measuring the amplitude at resonance and the amplitude at low frequencies. This removes the necessity of making sure that the amplitude of the driving force on an oscillator is the same for very low frequencies and for resonance (usually a difficult task for such a large frequency range) and is therefore usually a much more convenient method for determining the Q of an oscillator. Also, as pointed out above, it can be used when there is more than one mode of oscillation of a system.

As an example of the use of resonance width to determine the Q of a system consider again the Helmholtz oscillator of Fig. 11.1. Careful tuning of the tone generator shows that the minus 3 dB points are at 214 and 218 Hz. From this and (11.3) the Q of the oscillation is 54.

The effect the height of the resonance for a given driving force when the Q of the resonance is altered can be shown by placing a small sliver of thin cloth across the opening of the bottle. The effect of this is to lower the resonance frequency a little to 213 Hz but, more significantly, to lower the height at resonance by about 50%. In addition, the minus 3 dB points are now seen to be 209 and 217 Hz for a resonance width of 8 Hz. The Q of the oscillation is therefore 27, or about half that for the unimpeded opening. Thus it appears that for a given driving force the amplitude at resonance is proportional to the Q of the resonance.

11.4.3 Relationship of Q to Rate of Energy Loss of an Oscillator

It may be seen from the above that the Q of an oscillator is rather closely related to energy loss. The motion of the air through and around the small sliver of cloth in the previous example takes more energy per cycle than when the cloth was not there. It turns out that the same Q as defined by the ratio of amplitude at resonance to the amplitude at low frequencies or the ratio of the resonance frequency to the bandwidth can be given yet another equivalent definition;

$$\boxed{Q = 2\pi \times \frac{\textit{Energy stored in oscillator}}{\textit{Energy lost in one cycle}}} \qquad (11.4)$$

While this form of the definition of Q does not have the same direct and intuitive connection as the others to the nature of the resonance curve, it does have a more direct connection to the basic dynamics of the oscillatory motion and is therefore usually regarded as the "fundamental" definition of Q. An idea of its importance may be obtained by noting that the power (P) put into an oscillation is given by

$$P = \textit{Energy loss per cycle} \times \textit{frequency} \qquad (11.5)$$

If the system is at resonance, then the frequency is f_o and by rearranging the equation for Q one gets the energy stored in the system as

$$E = \frac{Q}{2\pi f_o} \times P \qquad (11.6)$$

The energy stored in an oscillator at resonance is therefore proportional to the Q and the power and is inversely proportional to the frequency of the resonance.

The importance of this use of Q can perhaps most easily be seen in the decay of an oscillation once all driving forces have been removed. Such a decay is the exponential one pictured in Fig. 11.4. It can be shown that the half-life of an exponential decay of an oscillator is directly connected to the Q of the oscillator. It turns out that the time constant (time for amplitude to decay to $1/e$ of it's initial value) is $Q/\pi f_o$ and this is related to the half-life by

$$t_{1/2\ Amplitude} = 0.693\ \tau_{Amplitude} = \frac{0.693}{\pi f_o} \times Q \qquad (11.7)$$

(For a derivation of this equation using differential calculus see the Appendix to this chapter.)

Thus the half life of the Helmholtz resonator with a Q of 54 should be 56 ms. This is indeed about the value that was observed for the decay shown in Fig. 11.4.

The half-life for the energy will be half that of the amplitude. This is because the energy is proportional to the square of the amplitude and so the energy goes through two half-lives (to 1/4) while the amplitude has gone through just one. The equation for the energy half-life and lift-time is therefore;

$$t_{1/2\ Energy} = 0.693\ \tau_{Energy} = \frac{0.693}{2\pi f_o} \times Q \qquad (11.8)$$

This is perhaps the most intuitive idea of all for the concept of the Q of the oscillator; it is directly related to the half-life of an oscillation once all driving forces have been removed.

The other factor in the half-life of an oscillation is the frequency. It can be seen that for two oscillations of the same Q, the half-life of the higher frequency oscillation will be less than that of the lower frequency one. This is, in fact, a common property of musical instruments; higher frequency oscillations tend to die away more quickly than lower frequency ones.

11.4.4 Relationship of Q to the Growth of an Oscillation

The energy concept applied to the Q of an oscillator also explains the growth of an oscillation which is being driven by resonance. If an oscillating system is driven by an oscillating force at resonance, the energy of the oscillation of the system will be seen to grow as shown in Fig. 11.7

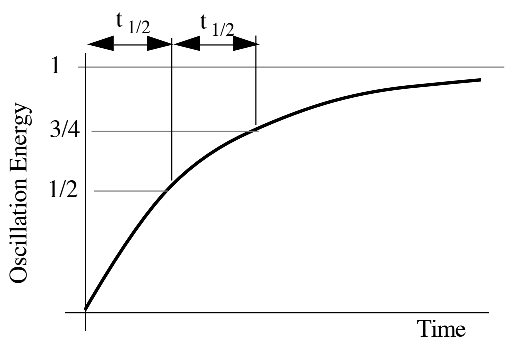


Figure 11.7 The growth of a resonant oscillation due to a constant amplitude driving force. The curve will be an inverted exponential curve, called the "saturation" curve. Here half-life refers to the time it takes for a halving of the difference from the final saturation level.

Again, this type of growth was considered in Chapter 3 in the growth of room reverberation. Repeating the ideas in that chapter, it can be seen that the curve is just an up-side down version of the exponential decay curve where the exponential decay is now the decay of the difference between the oscillation energy and its final value. This curve is another very common one in science and is often referred to as a "saturation" curve. The concept of half-life here refers to the time it takes for the difference from the final "saturation" value to decrease by half.

The half-life, and hence the "time constant", of the saturation curve for the energy of a driven oscillator is therefore the same as that for the exponential energy decay curve that results when the driving force is turned off. (This is proven using differential calculus in the Appendix to this chapter.) Thus

$$t_{1/2\ \text{Energy (Saturation)}} = 0.693\ \tau_{\text{Energy (Saturation)}}$$

$$= \frac{0.693}{2\pi fo} \times Q \qquad (11.14)$$

The time it takes for a driven oscillator to reach 1/2 of its final energy is important in acoustics. It is, of course, the time it takes for the sound to reach to within 3 db of its final level. For a Q of about 250 and a frequency of 250 Hz, it can be seen that this will be about 0.1 seconds.

It is important to note that the time it takes for a sound to reach its −3db point is proportional to the Q of an oscillator. Oscillators with very high Q's at low frequencies are therefore not very useful as musical instruments because of the perceptible time it would take for their sounds to develop.

It might seem strange that high Q, or high quality oscillators take longer for their sounds to develop than do low Q oscillators. This is because, for the same input power, they develop much more energy than do the low quality oscillators and so it takes longer for this energy to develop.

11.4.5 The Growth of an Oscillation Driven off-Resonance

In many cases in instruments an oscillation created in one part of an instrument may drive a normal mode but not necessarily exactly in resonance. An example which will be discussed later in this chapter is when a fundamental vibration with a strong set of harmonics (such as the vibration of the reed of an oboe) has a harmonic which is close to, but not exactly that of one of the higher modes of the instrument. The final steady-state of such a system has already been discussed; it is an oscillation with a steady amplitude, the amplitude being reduced from that at a resonance following the sort of resonance curves shown in Fig. 11.6. However, what is the growth pattern of the sound of such a driven system?

An outline of the mathematics is given in the appendix and only a few simple statements about the result will be given here. The amplitude of a simple oscillator driven off-resonance will grow as shown in Fig. 11.8.

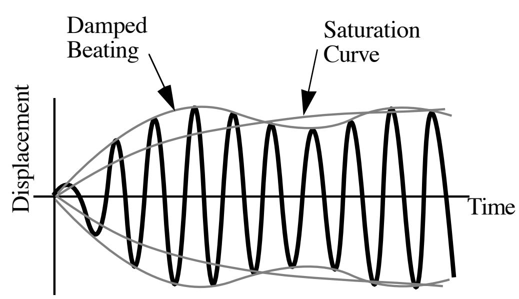


Figure 11.8 The growth of oscillation in a system driven off resonance by a constant amplitude driving force. The curve will be the saturation curve for the oscillation leading to the final steady-state oscillation at the driving force frequency but modulated in a beating pattern with its natural oscillation frequency. The beating dies away with the same time-constant as that for the decay of amplitude of the oscillator when it is left alone after it has been excited.

The curve will be the saturation curve for the oscillation leading to the final steady-state oscillation at the driving force frequency but modulated by a beating with its natural oscillation frequency. The interval between amplitude maxima will therefore be the reciprocal of the difference between the driving frequency and the natural frequency of the oscillator. For example, if the driving frequency is 252 Hz and the natural frequency is 250 Hz, then the interval between beat maxima will be 1/2 sec.

The beating pattern dies away with the same time-constant as that for the decay of amplitude of the oscillator when it is left alone after it has been excited ($\tau = Q/2\pi f_o$). If the 250 Hz system being "resonated" has a Q of 1000 then the decay time constant will be about 0.6 seconds.

Again one can see the undesirability as musical instruments of systems with very high Q's. Such systems could have decay times of the beating from slightly off-resonance excitation which would extend to seconds. Such a wavering, uncertain development of a musical tone would not be desirable.

11.5 Exciting a Multi-Mode System by Resonance

The concepts of resonance and Q of an oscillation are easily extended to more complicated systems with more than one degree of freedom of motion. Here the concept of normal modes, introduced in the previous chapter, becomes very important. A normal mode of oscillation of a system is one in which all elements of the system are oscillating at the same frequency and in phase (or 180^o out of phase, which is in phase with negative amplitude). Therefore it is possible for a driving force applied at one point in the system to drive all the elements in the system in resonance. The individual elements of the system may all have different amplitudes of motion but for any one normal mode of the system will have a definite amplitude for any given application of the driving force. Plotting the response of the system for different driving force frequencies will therefore give a resonance at each of the normal mode frequencies.

A specific example of such a system would be a short section of pipe as shown in Fig. 11.9 closed at the bottom end and open at the top and driven by a small speaker placed at the bottom. For a pipe with the dimensions shown, resonances will occur at 178, 450 and 750 Hz. These resonances will be on the three lowest normal modes of the system, corresponding in the ideal open-closed pipe model to the "1/4 wave" fundamental, and the next two modes at three and five times the frequency frequency. That they are these types of modes can be seen by lowering the microphone into the pipe and noting the changes in the sound pressure on the microphone for each of the modes. It will be seen that when the speaker is emitting a tone of 178 Hz, the sound pressure will increase considerably as the microphone enters the pipe and will reach a maximum at the very bottom of the pipe. For the 450 Hz tone, the pressure will increase as the microphone enters the pipe but will then decrease and come to a sharp minimum with the microphone about 22 cm from the bottom. Thus this mode exhibits the characteristics of the "3/4 wave" resonance of an ideal pipe.

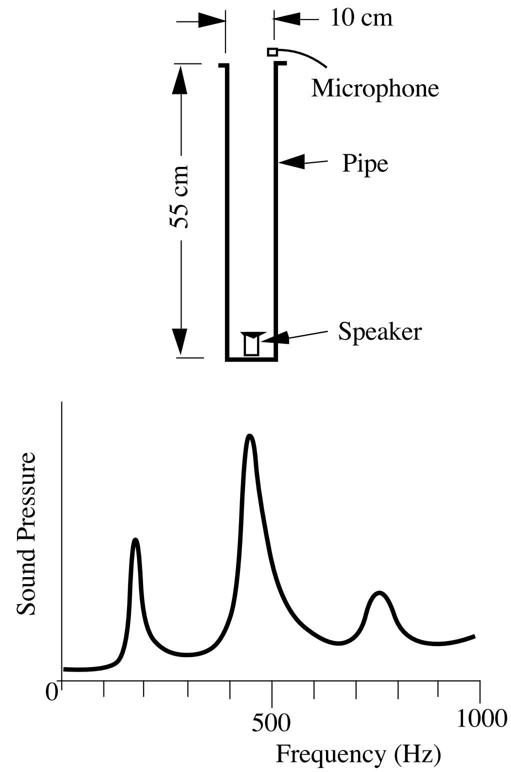


Figure 11.9 The response to a driving force of a typical system with a number of normal modes at various frequencies. The response is measured in terms of the sound pressure registered by the microphone placed near the open end of the pipe for the frequencies of pure tones played in the speaker. For the system shown, the resonances occur at 178, 450 and 750 Hz.

As pointed out in the previous chapter, each of these normal modes is independent. This also means that they will not only have distinct frequencies but also distinct Q values, the Q values being determined by how effectively the motions are coupled to energy dissipating processes. As an example, in the piano there appear to be two modes of vibration involving motions parallel to the sounding board and perpendicular to the sounding board. The motion parallel to the sounding board does not transfer energy as quickly to the board, and hence into the room, as does the motion perpendicular to the sounding board. The independent beating sound in each of the various harmonics that is characteristic of a piano tone is related to this phenomenon (see "The Coupled Motions of Piano Strings", G. Weinreich, Scientific American 240 (1), 118-127)

The results for the pipe in Fig. 11.9 show that it is not very well approximated by the model of a simple 1/4 wave linear system; its mode frequencies are not in the simple ratio 1:3:5 etc. This is because the open end of the pipe is not a simple termination of the pipe at no sound pressure. There is indeed sound pressure at the open end of the pipe or no sound would be radiated from the pipe into the room.

The detailed mathematical solution in three dimensions of even this simple geometry is very complicated and usually not worth-while. The normal modes of systems are determined empirically by a scan of the resonances with a variable frequency driving force.

11.6 The Excitation of Normal Modes by Impulse

11.6.1 The Excitation of Normal Modes by a Single Impulse

In the previous chapter, normal modes of oscillation of a system were introduced as properties of the system which could be set up at the beginning by a proper release of a system. In the case of the two coupled pendula, releasing the pendula with identical displacements to the one side set up one of the modes and releasing them with equal but opposite displacements set up the other. Giving only one an initial displacement resulted in both normal modes being introduced by equal amounts simultaneously and it was stated that any initial condition of the system was equivalent to a particular combination of the two modes. Here this concept will be extended to much more general systems.

As examples of more general systems, consider a long heavy rod, a rectangular plate and a circular disk (see Fig. 11.10). All can be suspended by threads so that they can be relatively free to vibrate in their respective normal modes.

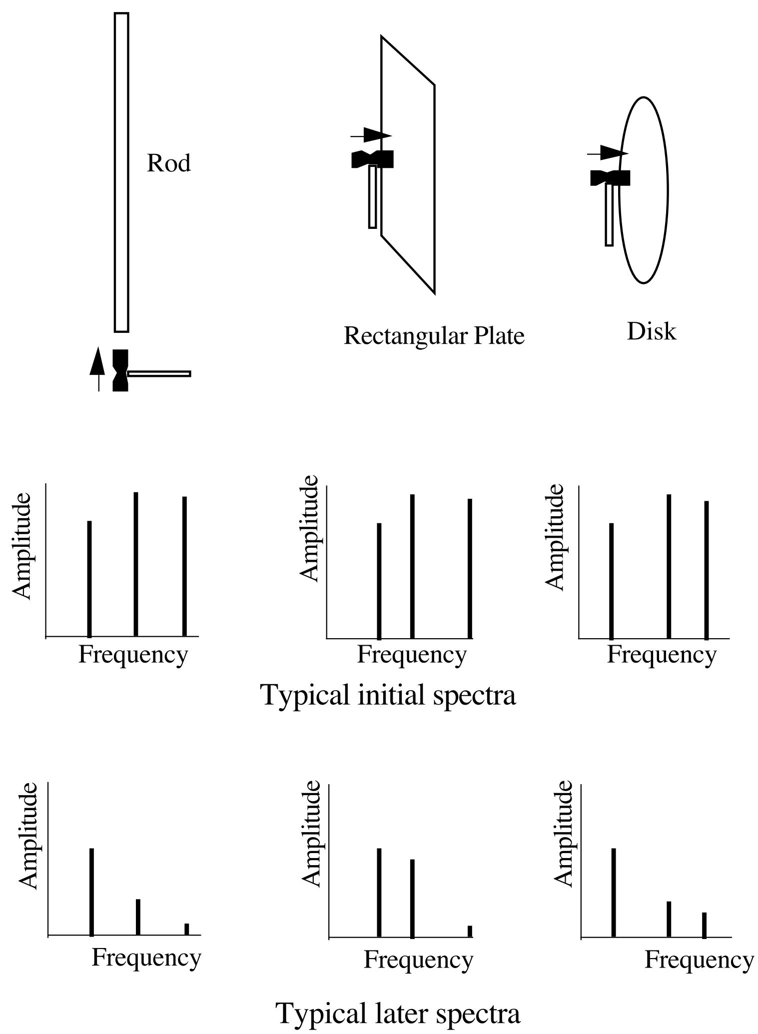


Figure 11.10 Three systems with normal modes that can be investigated by delivering impulses. The diagrams below are representative of the spectra that will be obtained by a spectrum analysis of the sounds that will be picked up by a microphone close to the struck objects.

By using a microphone and a spectrum analyzer to look at the sound produced by the normal modes it is possible to see the spectrum of normal modes produced by any impulse to the systems. If a hammer is used to give the rod a sharp blow on its end the normal modes can be observed in the spectrum. They seem close to being harmonically related and result in a somewhat high-pitched but musical tone. The normal modes induced by the impulse will also all decay independently, the higher ones decaying generally more quickly then the lower. This tendency of higher modes to decay faster than the lower modes has already been discussed in terms of the Q's of the modes.

The important point to note here is that, by hitting the rod in different places, the system can be started with different amounts of each of the modes. This difference would be seen on the spectrum analyzer and would be heard as a difference in timbre of the struck sound. The same phenomenon can be observed with the other two systems, where the sound is much more like noise because of the lack of any discernible musical relationship between the various modes. Hitting the plates at different points will produce different spectra and hence different timbres of sounds.

However, there would also be quite a noticeable difference in the timbre and the relative amount of the normal modes when the steel hammer is substituted by the heel of a shoe. Now the timbre of the tone would be much softer, high frequency normal modes being not nearly as strong in the spectrum.

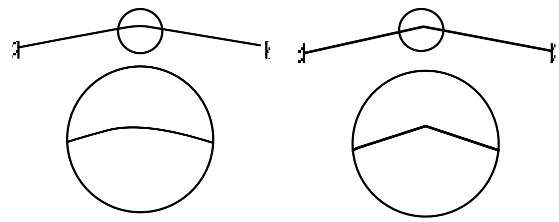
The point here is that different impulses to the system correspond to putting in different mixtures of the normal modes of the system. The impulse itself can be thought of as being made up of a particular recipe of normal modes.

The Fourier transform (see Chapter 9) is of relevance here. By using such a transform any transient motion was shown to be equivalent to a spectrum of pure oscillations. Similarly, any given geometrical state of a system can be arrived at by adding up enough of the right values of normal modes of oscillation with the correct amplitudes and phases. The concepts are essentially identical.

The significant points to be brought out here are the qualitative connections between the type of impulse to a system and the timbre or frequency spectrum of the partials (normal modes of oscillation) produced. In general, the sharper the impulse (the shorter the duration of the impulse), the more high frequency modes that are excited. Again this can easily be understood in terms of the Fourier transform analogy; the shorter the duration of a transient sound, the more high frequency components it has.

However, there is another factor in determining the content of high frequency partials in an impulse; that of the degree of geometrical distortion created by an impulse. Consider, for example, a string which is pulled aside by a soft finger stroke or a small hard plectrum (see Fig. 11.11).

The qualitative difference in the timbre of the notes produced by these two initial displacements is fairly obvious; the displacement by the finger would have much smaller high frequency components then the displacement by the plectrum.



Soft finger pluck Hard plectrum pluck

Figure 11.11 The displacement of a stretched string by a smooth finger and by a hard plectrum. The larger circles under the smaller ones are enlarged views of the center region of the string.

As a further illustrative example, consider the piano, one of the most familiar examples of a musical system in which the normal modes of oscillation are fired up by an impulse. The timbre of the tone produced is directly related to the content of high frequency normal modes in the impulse. This impulse can be altered by either changing the speed of the hammer on contact with the piano string or by changing the material and curvature of the hammer itself. Raising the speed of the hammer by striking the key harder therefore not only raises the level of the sound by putting more energy into the piano string vibration but also changes the timbre by introducing more of the higher frequency normal modes of vibration. This makes the piano intrinsically different from the harpsichord where the string is pulled aside and released by a plectrum in a given geometrical displacement which does not depend very much on how hard the key is struck. It gives the piano player the ability to drastically alter the timbre of the tone by the way the keys are "stroked".

The other way the timbre of a piano note can be changed is by changing the material and geometry of the striking head. For good pianos, the material in the head is of extreme importance in achieving the quality of the tone. How this can be changed is obvious to anyone who has heard a "honky-tonk" piano in which thumbtacks are placed in the heads at the point where they contact the strings.Such piano have the "tinny" sounds associated with the presence of a lot of high frequency partials in the initial attack of the notes.

To repeat, there is obviously an underlying similarity between the Fourier transform and the spectrum of normal modes of oscillation. The concept that any disturbance of a system can be thought of as a spectrum of normal modes of that system is essentially a more powerful version of the Fourier transform. The Fourier transform deals specifically with the way some single variable, such as pressure, changes with respect to another variable such as time. The concept that any state of a system is made up of normal modes of oscillation of that system is a more generalized concept in which the variable describing a system, (again it could be the pressure throughout a system) is allowed to vary with more then one variable. An example for the pressure in a system would be variations, not only in time, but in the *x*, *y* and *z* coordinates throughout the system.

11.6.2 The Excitation of Normal Modes by Successive Impulses

An important class of normal mode excitation in music is that of excitation by a regular succession of impulses. An example of this is the excitation of the normal modes of the vocal tract by successive impulses from the larynx in the production of vowel sounds in human speech. For example, the vowel sound "ee" will have spectra as shown in Fig. 11.12 for male and female voices.

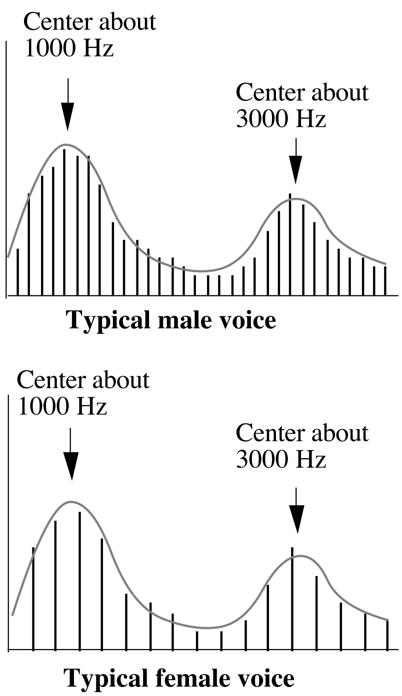


Figure 11.12 Spectra of the vowel sound "ee" for typical male and female voices. The vertical scale here would be db, with a range of about 30 dB for the spectra shown. The two peaks in the spectra are due to the relatively low *Q* normal modes of the air in the vocal tract when it is shaped so as to make the "ee" sound. The harmonic interval is the fundamental frequency of the two types of voices (about 140 for men and about 280 for women).

The spectra shown here exhibit the response of the air in the vocal tract to the input from the larynx. This input is a series of sharp puffs of air which break through the larynx when it is held taut and pressure is applied to it from the lungs. The pressure pulses in the vocal tract just above the larynx by these puffs is shown schematically in Fig 11.13.

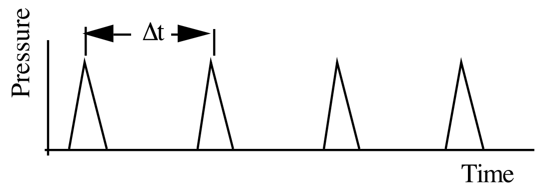


Figure 11.13 The pressure applied to the base of the vocal tract just above the larynx by the puffs of air that come through the larynx during speech. The period Δt between puffs is about 60 ms for men and about 30 ms for women.

Because of the sharpness of the pressure pulses into the vocal tract they are rich in harmonics. The normal modes of air in the vocal tract are resonated by these harmonics to produce the clusters of harmonics seen around the frequencies 1000 Hz and 3000 Hz for the vowel "ee". In speech and music, these clusters of harmonics in a sound spectrum due to resonance of normal modes are called "formants".

The formants are resonances of the normal modes of the by the harmonics of the basic repetitive pulsed input to the system. This phenomenon can also be demonstrated for the short stub of pipe open at one end as shown in Fig. 11.9. When the speaker in this system is driven by sharp pulses at a steady repetition frequency, the modes that have a frequency which is an integral multiple of the pulse frequency will be excited (see Fig. 11.14).

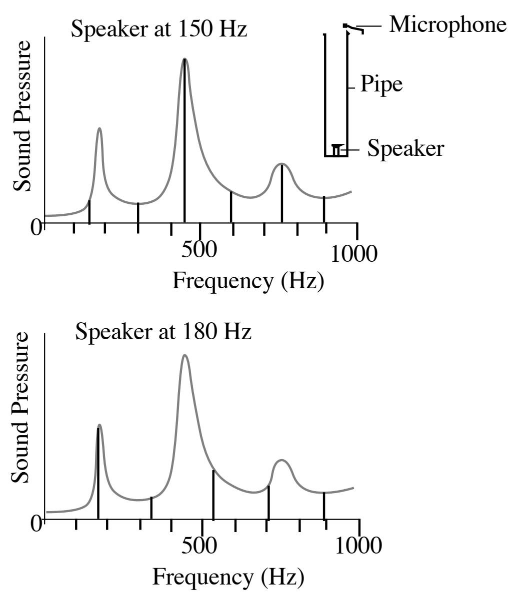


Figure 11.14 The response of the closed-open pipe to pulsed input from a speaker at its bottom. The upper spectrum is for a pulse rate of 150 Hz. Here the 450 and 750 harmonics of the pulser are enhanced due to resonance with the normal modes while the fundamental is, in fact, considerably reduced. The lower spectrum is the result when the pulser is tuned to the fundamental mode of the pipe (178 Hz). The harmonics of this pulse do not significantly excite either of the other two modes.

It can be seen from Fig. 11.14 that the fundamental of a tone can often be one of the weakest components of a musical sound if that fundamental is associated with harmonics which resonate strongly with higher modes of the system.

11.7 The Excitation of Normal Modes by Feedback

So far, the generation of musical sounds in perhaps the largest class of musical instruments has been ignored; that is musical instruments which give sustained tones. The generation of such sustained tones by resonance avoids the issue. How were the frequencies of the driving agent generated in the first place?

To understand the principles of the generation of sustained tones in musical instruments it is necessary to understand one of the most important phenomena in nature; that of the control of systems by feedback.

11.7.1 The Concept of Feedback

Feedback is a term which seems to have recently crept into popular language so it should not be strange. Professors ask for feedback from their students in a course. Customers give feedback to businesses so that the businesses, in principle, can give better service.

"Feedback" as a term was first used by electrical engineers in their development of control systems in the 1930's. The relevance of feedback to controlling systems is fairly easy to understand. If a system deviates from a desired state, then knowledge of that deviation is a very important part of any control system. The knowledge of the degree of deviation can then be used to adjust some input to the system so that the deviation is corrected. For example, if the professor is fed back information that what he is saying is too mathematical for the students, then he should try to lower the level of the mathematics. In another example, if a house is too warm, then the heating should be lowered (the input heat is too high).

In automatic control systems, the input to correct a deviation from the desired state of a system is automatically derived from information about the deviation. This process forms what is called a feedback loop (see Fig. 11.15).

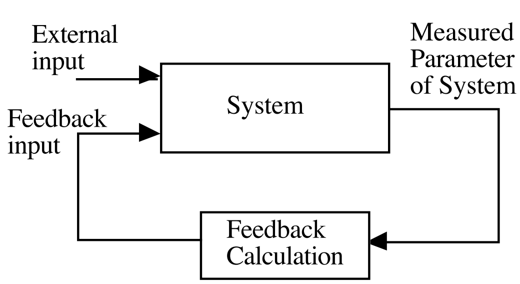


Figure 11.15 A schematic of an elemental control system. Knowledge of the output is used to generate a "feed back" signal which adds to the external input to determine the overall input to the system.

The type of feedback necessary for control is called negative feedback. This does not mean that the information fed back is always negative but that it is used to create an input which we know will cause the system to move in the opposite way to the deviation that was measured. Thus in the case of heating a house, a positive deviation of the house temperature from a set temperature would be used by the feedback system to reduce the heating; a negative deviation would cause the input heat to be raised. The actions due to the feedback are those which will cause the negative of the deviation that was observed. Such negative feedback control is the essence of automatic system control. It is used extensively by biological organisms to adapt to environmental changes. For example, when you sense a drop in external temperature, (usually, and most reliably by temperature sensors in the back of the neck) the body uses this information to turn up the metabolic rate so as to maintain the correct body temperature.

Electrical engineers developed the theory of feedback control so that they could design automatic control systems for electrical and mechanical devices. This theory is now finding many applications in the life and social sciences from biology to economics.

11.7.2 Oscillations in Fed-back Systems

To understand some of the consequences of negative feedback, consider one of the mysterious things that can happen. This phenomenon is familiar to anyone who has had to set up a public address systems or loudspeakers to enhance the sound of a live concert. The apparatus involved is simply a microphone connected to an amplifier which puts the amplified output of the microphone into a loudspeaker.

It is easy to see that this makes the elements of a system with feedback. The output of the microphone is fed back to the microphone by the amplifier and loudspeaker which put sound out into the room to fall back on the microphone. This is a system with a feedback loop (see Fig. 11.16).

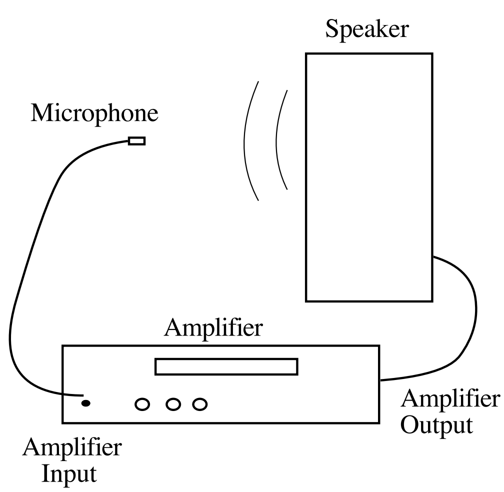


Figure 11.16 A feedback loop created by a microphone, amplifier and loudspeaker

When such a system is turned on and the gain of the amplifier turned up, one often gets an earsplitting howl. To prevent this howl the gain of the amplifier must be turned down or the microphone must be shielded from the loudspeaker output. Having the microphone close to the loudspeaker is almost certain to produce the howl.

This phenomenon is usually explained by pointing out that a sound out of the loudspeaker goes back into the microphone and this sound is then reamplified by the amplifier so as to put even more sound onto the microphone and so the system explodes into a very loud howl.

However, this sort of system would appear to be positive feedback. An output from the microphone is amplified by the loudspeaker and fed back into the microphone to produce more output, not less. Such positive feedback will always produce explosive results. It is similar to the feedback situation in a stick of dynamite. A little explosion in a small piece of the dynamite (which is actually hard to explode and requires a very small very sharp explosion by what is called a "cap" to get it going) will cause more dynamite around it to explode and this in turn causes even more dynamite to explode until the whole stick is rapidly consumed. In the positive feedback system of the microphone, amplifier and loudspeaker, the system rapidly goes to the maximum output that the amplifier can give to the loudspeaker.

But is this system always a positive feedback system? Not necessarily. If the speaker is incorrectly "phased" so that a positive output of the microphone to the amplifier causes the speaker cone to actually pull in, then the feedback is negative; a positive pressure at the microphone causing the loudspeaker cone to produce a vacuum which will tend to cancel the positive pressure. In a system so phased and with adequate bass response, this can actually be seen by observing the motion of a bare loudspeaker cone as the microphone is pushed rapidly towards it. The speaker cone will be observed to recoil away from the microphone.

Yet even such a system, when the gain is turned up, will go into a fierce "feedback" howl. How is this possible in a system which is under negative feedback, the essential element for the control of systems?

Before going into the answer to this puzzle, consider another phenomenon. If the connections of the loudspeaker are changed so that they are correctly phased (the loudspeaker cone moves out with increased pressure on the microphone) there is then positive feedback in the system. The result as the gain is turned up would again be a feedback oscillation but now of a very different frequency than the case for the negative feedback. Such oscillations when the microphone is very close to the speaker will usually be a very low bass rumble whereas the feedback oscillations with negative feedback for the same microphone placement will be of a shrill howl.

That the frequency of the feedback oscillation is different for the cases of positive and negative feedback is a clue to the origin of both. In an oscillating system driven by feedback, the input to the system and the feedback must be in phase for them to add up to an explosive situation. For positive feedback, this can occur at very low frequencies. For negative feedback, it can only occur when the delay in the propagation of sound from the speaker to the microphone is one-half cycle or an odd integer number of half-cycles.

There will be many high frequencies at which this can happen. The system picked out one of these from the general background noise that the microphone is picking up anyway and went wild with it.

What about the oscillation when the system was fed back positively? Why did the speaker cone not just move over as far towards the microphone or as far away from the microphone as it could and just stay there?

The answer here is that the steady state condition of the speaker cone pulled to its limits either in or out and being held there presents no feedback to the microphone. A propagated pressure from the loudspeaker to the microphone requires that the speaker be moving. This means that there has to be an oscillation for positive feedback in this system. Again, the system picks out some frequency for which this feedback is most effective and goes wild with that frequency. Again, all sorts of frequencies are possible (a microphone placed close to the tweeter of a speaker system with positive feedback will give a very shrill howl and probably burn out the tweeter in less than a second). What is required of such a feed-back oscillation is that there be close to a number of whole cycles of the oscillation in the overall delay of the feedback.

Thus it seems that any system with feedback is capable of breaking into oscillations, whether it is fed back negatively or positively. Does this mean that we should not try to control any system by feedback?

The answer is of course no. We can control systems by negative feedback if we understand what we are doing and are careful doing it. This is why electrical engineers found that they had to develop the theory of feedback control to use it effectively. This theory of feedback control tells us not only how to control a system while limiting oscillations to tolerable amounts (it turns out that all practical control systems will have some residual oscillation which is called "hunting") but also what one has to do if one wants to deliberately generate an oscillation of a particular frequency by feedback.

Control theory will not be developed here; what will be presented is merely an illustration of the principles. As a start to the subject, consider an example which may have more relevance to your personal life than any mechanical, electrical or musical system. It is a system clearly under negative feedback "control" and yet which undergoes disturbing oscillations.

The example starts with what happened when the Russians launched Sputnik in 1957. This event offered all the information needed by the U.S to conclude that it's scientific and engineering capabilities were badly in need of upgrading. The specter of a Russian satellite orbiting over the United States every 90 minutes or so while their rockets were blowing up on their launching pads was very powerful feedback of a deficient situation. The problem was almost immediately traced to a science and engineering training system that was allowed to fall into decay after the end of the Second World War 12 years before, while the Russians obviously put a lot of effort in highly disciplined science and engineering training programs.

The required response was obvious. More scientists and engineers had to be trained and so a lot of money and resources were put into upgrading science education in the United States.

Unfortunately, immediate results did not show. The Russians advanced to orbiting animals and then men and the American rockets were still not getting off the ground with any significant payload. The pressure on the scientists and the engineers kept growing. Stories of demands for highly trained scientists and engineers were widely circulated and any smart high school student automatically chose science and engineering for university study.

Finally, of course, results did come. The massive capabilities of the United States for action began to show and the United States space program took off to become one of the great achievements of mankind.

But then what happened?

When the spectacular achievements made it obvious that America had regained its superiority, the massive political support required for the immensely costly space program began to dwindle and spending was cut back. Soon stories began to appear about PhD's in electrical engineering having to take jobs as garage mechanics. It didn't take long to discourage students from undertaking the rigors of the demanding science and engineering programs. Smart students were then going into political science so that they could cope with the apparent revolution of social priorities.

However, the pipeline was full of engineers and scientists in training. Even with the supply at the input end drastically reduced, the output was still there trying to get whatever jobs they could get. Most went into high school teaching or other jobs which made only marginal use of their skills. Finally of course, the supply dried up. What then do you expect started to happen?

Right! Stories of Japanese industry killing American industry because of their superior engineering and technology began to surface and it became harder to get into electrical engineering and computer science than it is into medicine at many of the better schools in North America.

The system clearly has negative feedback control but also it has a very disturbing oscillation. How can this oscillation be controlled?

Clearly one way to control it is to not apply any feedback. One of the reasons in seems that the United States goes through such strong oscillations is the strength of the feedback due to its massive communications systems. Reduce the strength of this feedback and the oscillations should die down. Reduce the gain of the amplifier and the feedback howl will disappear. However, the danger with this method of control is that the feedback may not be strong enough to prevent a very undesired deviation of the state from the desired one. Some rigidly planned economies with not enough feedback effect would seem to fall into this category and the result is the tragedies that is witnessed in the early 1990's in such systems.

A better response would be to take into account the basic cause of the oscillation. In the case of science and engineering students, this cause was the time it takes to train an engineer or a scientist. This is about 7 to 10 years, and the fundamental reason it seems for the approximately 15 year oscillation in the system.What can you do about this?

Here one must understand the necessary relationship between a driving force and the velocity of an oscillation if one is to have the oscillation gain or lose energy. A force in phase with a velocity will increase the energy of an oscillation while a force 180^0 out of phase with the velocity will decrease the energy of an oscillation. Thus to kill the oscillation of the attendance of engineers in universities, the greatest discouragement to enrollment should not come when the number of unemployed engineers is the greatest but when the number coming out of the system per year is the greatest (i.e. the "velocity" of engineers out of the system is the greatest). Put in very personal terms, if you want to be sure of a job when you graduate, go the opposite way that everyone else is going!

In other words, one should be responding to the rate of change of the supply rather than the supply itself! The force should be opposite to the velocity of the displacement of a system rather than the displacement.

Now the relationship before the driving force and the oscillation to be driven should be coming obvious. To get an oscillation going by positive feedback, that positive feedback has to have a phase angle which is leading or lagging the velocity oscillation by no more than 90^0. If it leads or lags by more than 90^0, then the component of force "in phase" with the velocity is actually negative and the oscillation will be damped by the feedback.

Perhaps a good example to illustrate the principle is that of pumping up a swing by yourself. If you think about how you do this you might note that for maximum effectiveness you put your effort in the forward motion when you are moving forward at maximum velocity at the bottom of the swing. On the return, you put your maximum backward effort when you are at the bottom of your swing moving backwards at your maximum rate.

Likewise, to kill your oscillation so that you can safely step out of the swing, you apply these forces in reverse. This takes some training. In general, one is inclined to try to kill an oscillation by applying a maximum reverse effort when the oscillation is at its maximum, rather then when it is at its maximum velocity. In the example of the oscillations of long trains on the prairies, it takes a great deal of training in simulators to get engineers to increase engine power when the locomotive is moving backwards at maximum relative velocity to the train rather than when it has in fact moved backwards as far as it will go.

Perhaps it is useful to note here that applying a force opposite to the displacement rather than opposite to the velocity is, in fact, the condition that leads to oscillations in simple systems such as a mass on a spring.

To conclude, manipulating an oscillation in the way that we want involves creating an effective feedback force which has the correct phase relationship with the velocity of the oscillation, taking into account the delay in the effect of the feedback force due to any inherent delays in the system.

This works well for systems in which there is only one mode of oscillation of any importance. However, in a system with an infinity of possible normal modes of oscillation such as air in a tube and with enough gain, there is bound to be some oscillation which the system picks up with the right feedback for the driving force to have the correct phase relationship with the velocity so as to send all the energy of the system into that mode of oscillation. This is the principle behind practically all "acoustic" musical instruments with a sustained tone, whether the sustained tone is achieved by bowing, scraping, blowing or any other action.

11.7.3 The Factors Leading to the Selection of a Particular Mode for Feedback Oscillation

In the case of the feedback howls produced by a microphone of a public address system placed too close to a speaker, what determines the frequency at which the system will oscillate? In the more important case for the subject matter of these notes, what factors will determine the particular normal mode of vibration of a musical instrument that will be selected by the system for a feed-back generated oscillation?

Return to the example of the simple bottle as a resonator. By blowing gently, the Helmholtz oscillation will be produced. Blowing harder will excite the next lowest normal mode, the half-wave "fundamental" resonance of a tube closed at both ends. Blowing still harder, usually only possible with a compressed air source, would excite even higher modes.

There is another similar example of a system with normal modes of oscillation that can be excited by feedback; the whirling tube music maker that was popular among physicists and other children in the early 80's. It is a hollow flexible tube about a meter long, open at both ends and with an internal diameter of about 2.5 cm. When it is twirled around, one can get a variety of fairly musical tones, the frequency of the tone depending on the speed with which the tube is twirled. At low speed the tone frequency is low, at high speed the tone frequency is high.

A remarkable thing about the tones generated in these two instruments is the purity of their spectra. They are heard as clean flute-like tones which on a spectrum analyzer show as only one line. Why is it that of all the modes of vibration possible in those systems only one is fired up at any one time? In both cases, the sound is produced by air rushing over the ends of the openings of the systems exciting the normal modes of the systems. As the speed of the air increases, the noise spectrum shifts to have its center at higher and higher frequencies. Why in the example of the twirling tube, does the tone progress sharply from the excitation of one of the modes to the excitation of the next higher mode? Why isn't there a gradual change with the lower frequency mode getting weaker and the higher frequency mode getting stronger as the frequency spectrum of the noise of the air rushing over the end of tube rises with the speed?

To get an insight into the physics of this phenomenon, consider again the open pipe. If the small speaker at the bottom of this tube is powered by an amplifier driven by the microphone anywhere in the tube, then one has a feedback situation which can lead to an oscillation.

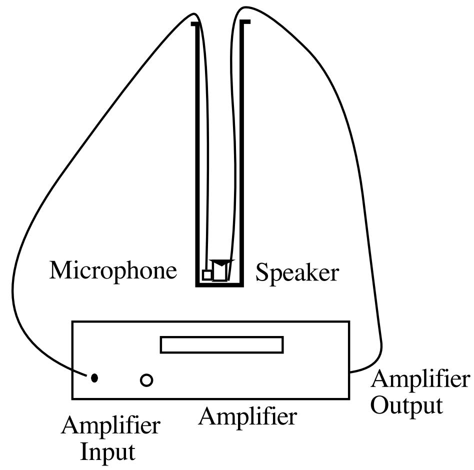


Figure 11.17 The system of Fig 11.9 in a feedback situation.

When the microphone is placed at the bottom of the tube near the speaker, and the gained turned up (very carefully so as not to blow the speaker) a feedback oscillation will break out. The frequency of this feedback oscillation will depend very much on the phasing of the speaker to the microphone output. If it happens to be in phase then the feedback oscillation will most likely be 450 Hz. If it is out of phase, it will likely be a high-pitched squeal of about 2000 Hz.

If now the microphone is drawn out of the tube to near its entrance and the experiment repeated, again there will be feedback oscillations with frequencies dependent on the speaker-microphone phasing. Now the frequency will likely be 450 Hz with the speaker and microphone out of phase and probably about 600 Hz with the in-phase connection.

How are these phenomena to be explained?

The 450 Hz oscillation with the speaker and the microphone both at the bottom of the pipe is perhaps expected. The 450 Hz mode is the one most strongly excited by the speaker in the resonance curve. This mode has a pressure maximum at the bottom of the pipe and, since the microphone is sensitive to the pressure of the sound, this pressure maximum will be fed to the speaker in phase. The speaker and the microphone are therefore in positive feedback on this mode and it is therefore excited.

The next oscillation that is easy to understand is that which occurs when the microphone and speaker are out of phase and the microphone is just inside the open end of the pipe. Here the pressure of the "3/4 wave" mode is out of phase with it's pressure at the bottom of the pipe. When the speaker and microphone being out of phase, puts the feedback back into phase and this mode is again easily fired up.

But where did the weird oscillations at 2000 Hz and 600 Hz come from?

The 2000 Hz oscillation with the microphone and speaker out of phase at the bottom of the pipe must be a cross mode vibration in which the pressures at the microphone and the speaker are out of phase with each other. The origin of the 600 Hz oscillation with the microphone and speaker in phase and the microphone near the pipe opening is not easy to diagnose but it can be noted that it corresponds to a wavelength which is very close to the microphone-speaker separation.

It is possible to set up a feedback oscillation on the fundamental of 178 Hz? It turns out that it is if the microphone is placed at a position inside the tube where it is insensitive to the pressure in the "3/4 wave" mode at 450 Hz. By turning up the gain very carefully to get just the onset of a feedback oscillation, with the microphone and speaker in phase it is possible to get a 178 Hz oscillation started.

The idea to be retained from all this is that the right feedback conditions can allow a system to fire up on one of its normal modes, the mode being fired up depending on the nature of the feedback. This is the basis of tone production in many musical instruments, particularly acoustic instruments producing sustained tones.

In most cases, the automatic feedback due to a normal mode of oscillation in the instrument is also under another kind of feedback control from the player. Thus if the player finds that the tone is about to break into another mode due to an incorrect manipulation of the instrument, then that manipulation can be adjusted so as to prevent this "mode-hop". (A common example of such a breaking into a wrong mode is the "overblown" notes that can be achieved on a recorder.)

11.7.4 The Feedback Process in Air Flow over Surfaces

The feedback mechanism in the system of the speaker in the tube driven by an amplification of the pickup from a microphone is fairly obvious, it was the amplifier connecting the microphone to the loudspeaker. What is the feedback process in the beer bottle and the twirling tube?

The feedback process here is the basis of many of the wind musical instruments. The basic phenomenon involved is that air flowing over a surface creates a vacuum which tends to pull the surface into the air stream (see Fig. 11.18). This puzzling fact was first investigated scientifically by Bournoulli and so the effect bears his name.

If the surface that the air is blowing over can actually move into the air and thereby force it to flow on its other side, then there will now be a vacuum pulling the surface back to its original condition (see Fig. 11.19).

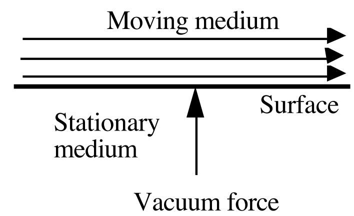


Figure 11.18 The force on a surface over which a medium is flowing on one side but stationary on the other.

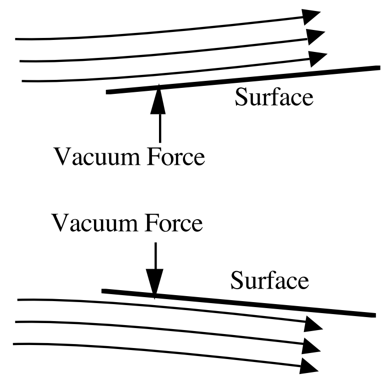


Figure 11.19 The action of a flexible surface angled pointing into a wind. In the top diagram the surface is angled so that the air must move over the top side. The vacuum force this causes moves the surface upward so that it appears as in the lower diagram where the air is forced to move over its bottom surface. The reverses the direction of the vacuum force.

Thus the situation whenever air flows over a moveable surface is unstable and an oscillation can and probably will result. Common examples are the flutter of a flag or a piece of paper in a wind. Another example is the vibration produced by blowing across a blade of grass or a strip of paper stretched directly in front of the lips.

When the surface that the air is blowing over will not move but the air could move to one side or the other, the air motion itself becomes unstable. This type of air flow is called a Venturi oscillation and is the source of noise when air flows over the edge of a surface. The spectrum of the noise that is produced when air blows over the edge of a surface is shown in Fig. 11.20.

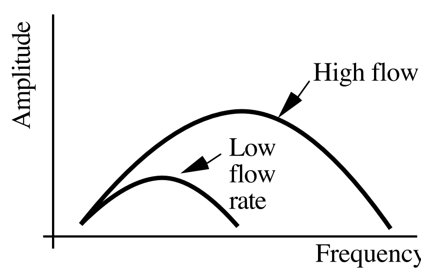


Figure 11.20 The general shape of the spectra of Venturi oscillations for two speeds of air onto the edge of a surface.

At low velocities, the noise spectrum peaks at a low frequency and the peak value is low. At high velocities, the peak is at a higher frequency and the peak value is much greater.

Venturi oscillations about a fixed surface in open air are therefore not musical but have random frequencies and hence appear as noise. However, if a mode of vibration at a definite frequency is excited by such a collection of Venturi oscillations, a coherent Venturi oscillation can be set up by feedback.

As an example of how this can happen, consider the Helmholtz mode of a bottle. Suppose the initial air flow is directed so that it is across the neck and not into the bottle. The result of this will be a vacuum which pulls air out of the bottle. Thus the air flow across the bottle will generate a vacuum in the bottle which will eventually balance the vacuum force of the air flow across the mouth. There will then be an unstable situation in which the air, through a random fluctuation in its flow, can be flipped to flowing into the bottle. This air flow into the bottle will now quickly build up pressure in the bottle until the flow pattern is ready to be forced to the outside again. If this occurs at a time when the Venturi oscillation near the center of the spectrum is about ready to flip to outside flow, this flip will be enhanced by the pressure buildup. The Venturi oscillation at the neck is enhanced by feedback from the Helmholtz oscillation in the bottle. The driving force of the oscillation is the vacuum produced by the flow of air across the mouth of the bottle but the frequency of the oscillation produced is governed by the Helmholtz oscillation. The Venturi oscillation therefore settles on the Helmholtz frequency.

If the air flow is increased so that the natural Venturi oscillation is centered on about 1500 Hz, then the "1/2 wave" normal mode of vibration of the air in the bottle will set up a Venturi oscillation by feedback, taking all the energy from the Helmholtz oscillation.

This effect occurs in all blown instruments such as flutes, whistles, recorders, beer bottles and "twirly tubes". The Bernoulli effect of the vacuum created by any flow provides a feedback mechanism by which a particular normal mode of oscillation of the air in an enclosure can be excited.

11.8 Growth of Feedback Oscillations

Feedback oscillations have a particular pattern of growth, different again from the resonance driven oscillator already described. This is because the power going into the the oscillation is not constant as in the resonance driven oscillator but grows exponentially itself from some small "seed" disturbance. The beginning of a feedback driven oscillation is therefore an exponentially increasing amplitude (see Fig. 11.21)

When the tone has increased in volume, dissipation effects will begin to show and the saturation characteristic of the oscillation will set in. The overall appearance will therefore be as shown.

The initial growth period of a feedback oscillation will depend very much on the strength of the feedback. If it is just sufficient to get the oscillation going, it can be very long. If there is strong feedback, it can be very short. This is why sustained notes played with great force so as produce loud sounds will generally also have much steeper attacks than softer notes.

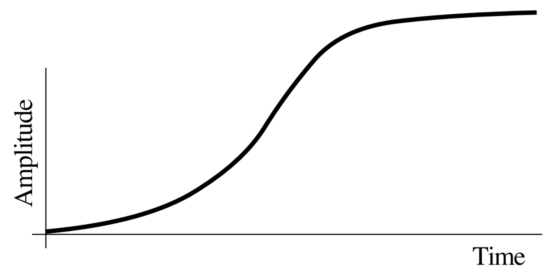


Figure 11.21 The typical growth of a feedback generated tone.

11.9 Normal Modes, Feedback, Resonance and Harmonics

At this stage we are finally at a point where we can take a broad overview of musical instruments and how they produce musical sounds. The principle of percussion instruments have already been discussed. Energy is put into the normal modes of these instruments by an impulse and the timbre of the note produced is determined by the nature of this impulse. What is to be concentrated on here are the principles of the instruments that produce sustained tones by actions such as blowing, spitting, bowing, scraping or stroking.

In general, the starting point is the excitation of a normal mode in the system by a feedback loop. In some cases the complete feedback loop may be difficult to analyze but one can be fairly certain that it is there. For example, in the playing of the trumpet the feedback loop is the production of a pulse of sound by the players lips "spitting" a pulse of air into the mouthpiece of the trumpet. This pulse of air causes a pulse of sound to be propagated down the tube of the trumpet until it comes to the bell where, because of the sudden freedom from the walls of the trumpet tuber, it releases in a spurt. The sudden movement of the air away from the trumpet bell causes a vacuum region to form due to the momentum of the moving air. The pressure pulse that arrived at the bell therefore results in an inverted, or vacuum pulse being reflected back along the trumpet tube.

That means that a vacuum pulse comes back up the tube to slam the players lips shut if, indeed they are still open. This vacuum pulse is then reflected back out the tube by the closed end of the system as still a vacuum pulse. It is then reflected as a pressure pulse by the bell of the trumpet and this pressure pulse returns down the tube to tend to force the players lips open. If the player adjusts mouth air pressure and the lip tension so that the lips are ready to deliver another burst of air after the period taken for this quadruple traversal of the tube length, then the arrival of the input burst reflection as a pressure pulse at the lips will trigger their opening and cause another burst of air to enter the system. This burst of air will occur on top of the pressure pulse that triggered its release and so a positive feedback buildup of the pressure pulse will occur. This positive feedback is felt by the trumpet player as a tingling of the lips that is not there when the same effort is put into blowing a simple "raspberry" (the harsh vibration produced by blowing air through tightly closed lips which have nothing in front of them.)

In the case of the violin, the bow scraping along the string instantaneously forms a glue bond with the string and drags it to one side. At some point this glue bond breaks and the string snaps back to near its equilibrium position whereupon it again forms a glue bond with the hair of the bow. (This is why resin is so important to a violinist).

Meanwhile the pulse formed by the last breakage of the glue bond travels along the string and is reflected by the string boundary. The reflected pulse then comes back along the string. If it arrives at the bow just when the bond between the bow hair and the violin string is about to break again, then there is a positive feedback situation where the reflected pulse helps the tension of the deflected string to break the bond and release the string.

So there is always some feedback mechanism to determine the mode of oscillation that will be excited and the strength to which it is excited.

If this were the only thing that happened, all one would get are the sort of pure tones of the beer bottles and the twirling tubes. That is indeed about all that happens with pure toned instruments such as the flute and the recorder but many musical instruments with sustained tones have very rich timbres with lots of high frequency components. How did these components get into the system?

In general, an oscillation produced by feedback does not result in a pure sinusoidal oscillation of the driving element of the system. In the simplest case of the oscillation of a speaker cone by feedback from a microphone in a room, the motion of the cone is so violent that it is usually just a flipping back and forth from one extreme of its motion to the other. The result is a square waveform oscillation which is rich in harmonics. In the case of a trumpet players lips, the feedback oscillation results in short spurts of air into the bell. These spurts again are very rich in harmonics. Similarly the feedback system in an oboe uses the lowest frequency normal mode of the air in the body of the oboe to snap the two reeds of the oboe open and shut at the frequency of this mode. Such an action puts very sharp pulses of air into the oboe body at this frequency. This action is also very rich in harmonics.

If now there are higher normal modes of oscillation in the musical instrument which match any of the harmonics of the basic feedback oscillation, these harmonics will be enhanced by the resonant action of these modes.

Note that it is the harmonics that are enhanced, not the normal modes of oscillation. The harmonics do not fire up the normal modes themselves. Recall that when a simple oscillator is driven by a force oscillating at a frequency different from the natural oscillation frequency, the oscillator vibrates at the driving frequency, not its natural frequency. Thus the partials of the oboe are indeed harmonics of the fundamental. This is true of all sustained tone instruments in which a feedback oscillation creates a basic fundamental oscillation frequency rich in harmonics which are enhanced by normal modes of the system.

The richness of the tone now depends on how many of the harmonics find matching normal mode frequencies. For instruments such as the violin, the matching is very good since all the useful higher modes of oscillation of the string are very close in frequency to harmonics of the fundamental. In many instruments based on air cavities, only clusters of harmonics will have resonance enhancement by normal modes of the system. One such important musical instrument already discussed is the human voice. The basic oscillation frequency here is the flapping of the vocal chords (more accurately called the larynx) which has a frequency in men of about 150 Hz. However, this sharp pulsing open and shut, like the tone of the oboe, is rich in harmonics. The vocal tract above the larynx (including the nasal cavity) has several important normal mode resonances in the regions of 500 Hz, 1500 Hz and 2500 Hz. At these frequencies, the harmonics of the basic vocal chord frequency will be enhanced. These regions of enhancement are the "formants" of speech and singing. By modifying the shapes of our vocal tract, we can modify the frequencies of the normal modes giving these formants and even change their Q so as to enhance or reduce their effectiveness. This manipulation of our vocal tract while the vocal chords are delivering pulses of air into the vocal tract is what produces the vowel sounds such as "a", "e", "i", "o", "u" and the various diphthongs.

Summarizing then, the production of sustained tones in musical instruments generally involves;

1. A basic oscillation of some driver produced by feedback from a particular normal mode of the system, usually the lowest.
2. The harmonics which may be present in this basic driving oscillation being filtered or (enhanced) into formants by the other normal modes of oscillation of the system.

APPENDIX

Derivation of Q from amplitudes of oscillation

The steady-state amplitude of an oscillating mass m with a resistance r and a spring of spring constant k when driven by an force of amplitude F and frequency ω can be shown to be

$$A = \frac{F}{\sqrt{(\omega r)^2 + (\omega^2 m - k)^2}} \qquad \text{(A11.1)}$$

For very low ω, this becomes simply

$$A_{low} = \frac{F}{k} \qquad \text{(A11.2)}$$

At resonance, $\omega^2 m = k$ and the equation becomes

$$A_o = \frac{F}{\omega r} \tag{A11.3}$$

The definition of Q then gives

$$Q = \frac{A_o}{A_{low}} = \frac{k}{\omega r} = \frac{\omega m}{r} \tag{A11.4}$$

Derivation of Q from Bandwidth

The width of the resonance at the $A/\sqrt{2}$ points is given by solving the equation for A when

$$(\omega r)^2 + (\omega^2 m - k)^2 = 2(\omega_o r)^2 \tag{A11.5}$$

Noting that for narrow resonances, $\omega \approx \omega_o$;

$$(\omega^2 m - k)^2 = (\omega_o r)^2 \tag{A11.6}$$

$$\omega^2 m - k = \pm\omega_o r \tag{A11.7}$$

$$\omega^2 - \frac{k}{m} = \pm\frac{\omega_o r}{m} \;;\; \omega^2 - \omega_o{}^2 = \pm\frac{\omega_o r}{m} \;;$$

$$(\omega - \omega_o) \times (\omega + \omega_o) = \pm\frac{\omega_o r}{m} \tag{A11.8}$$

Noting again that $\omega \approx \omega_o$,

$$\omega - \omega_o = \pm\frac{r}{2m} \tag{A11.9}$$

The frequency interval between the two solutions is therefore

$$\Delta\omega = \frac{r}{m} \tag{A11.10}$$

and the ratio of the resonant frequency to this frequency interval becomes

$$\frac{\omega_o}{\Delta\omega} = \frac{m\omega_o}{r} = Q \tag{A11.11}$$

Since the ratio for the radian frequencies ω is the same as for the cyclic frequencies ($f = \omega/2\pi$)

$$\boxed{Q = \frac{f_o}{\Delta f}} \tag{A11.12}$$

Derivation of Q from Decay Constant

The derivation of this relation follows from elementary differential calculus. The rate of loss of energy is the loss per cycle divided by the time for one cycle, which is of course the period of the oscillation;

$$\frac{dE}{dt} = -\frac{\textit{Loss of energy in one cyle}}{T} \tag{A11.13}$$

From the definition of Q and the fact that the period is the reciprocal of the frequency, this rate of energy loss can be expressed as;

$$\frac{dE}{dt} = -\frac{2\pi f_o}{Q} \times \textit{Energy stored in oscillator} \tag{A11.14}$$

The energy loss dE is from the energy E stored in the oscillator so that the equation becomes the standard for exponential decay;

$$\frac{dE}{dt} = -\frac{2\pi f_o}{Q} E = -\alpha E = -\frac{1}{\tau} E \tag{A11.15}$$

leading to the solution

$$E = E_o\, e^{-\alpha t} = E_o\, e^{-\frac{t}{\tau}} \tag{A11.16}$$

where

$$\tau = \frac{Q}{2\pi f_o} \tag{A11.17}$$

For an oscillator gaining energy from a source while also losing it from its own motion;

$$\frac{dE}{dt} = P - \frac{2\pi f_o}{Q} \times E$$

$$\frac{dE}{dt} = P - \frac{2\pi f_o}{Q} E = P - \alpha E$$

$$= P - \frac{1}{\tau} E \tag{A11.18}$$

leading to the solution

$$E = E_o(1 - e^{-\alpha t}) = E_o(1 - e^{-\frac{t}{\tau}}) \tag{A11.19}$$

where

$$E_o = \frac{QP}{2\pi f_o} \;;\quad \tau = \frac{Q}{2\pi f_o}$$

It is sometimes worthwhile to note that the amplitude of such an oscillation does not follow a simple exponential saturation curve but becomes

$$A = A_o\sqrt{(1 - e^{-\frac{t}{\tau}})}$$

Exercises and Discussion Topics

1. If a system oscillates with a Q of 250,

   a) If the frequency of the oscillation is 400 Hz, what would be the "bandwidth" of a resonance of the oscillation?

   b) What would be the time constant for the sound energy decay (the time for a decay to 1/e of the energy) if the oscillation were excited by giving the system an initial impulse? What would be the half-life? What would be the "reverberation time"?

2. For a sound source which matches in frequency one of the normal modes of a room, sketch the variation with time of the sound amplitude from the moment the sound is turned on to when the sound has died away after the source has been turned off. What is the connection between the growth part of the curve and the reverberation decay part of the curve?

3. Explain how giving an impulse to a system which has many normal modes of oscillation can give different spectra of these normal modes depending on exactly where and how the system is hit. In what way can the spectra be different and in what way will they all be the same?

4. Explain in terms of the normal modes that are excited, why using a plastic plectrum gives a brighter sound to a guitar than does a finger stroke and why bowing a violin near the bridge will give a sound with relatively more high frequency content than bowing it nearer the center of the string.

5. What is "feedback" and why is it important to a musician trying to get a sustained note on a musical instrument with normal modes of vibration?

6. Why will a system which takes part of its output to create its own input generally oscillate even though it may be connected in "negative feedback" so that the output created from the input is of opposite sign? Why is this often a general feature of systems which are to be kept under control by such negative feedback?

7. What determines the frequency of the feedback oscillation that will occur in a system with many pronounced normal modes of oscillation? What will be the nature of that oscillation? What will determine whether the other modes will be of any significance when this mode is fired up?

8. Explain the feedback process that occurs in blowing a note into a bottle or blowing a note in a flute. Why does overblowing a recorder generally produce a discordant note?

9. What was the feedback process that blew down the Tacoma bridge?

10. Explain why the box under a tuning fork is designed to be a resonator with a frequency which resonates with the tuning fork.

11. What is resonance in the technical sense of sound vibration? How is it related to the subject of normal modes of vibration of a system?

12. What generally distinguishes the pattern of frequencies of the normal modes of vibration of a musical instrument compared with that of a non-musical noisemaker? Why are the higher modes of vibration important even if only the lowest or fundamental mode is being fired up by feedback?

13. What are formants in speech and music? Relate them to normal modes of oscillation of systems and resonance of these modes.

CHAPTER 12

THE CHARACTERISTICS OF MUSICAL INSTRUMENTS

The material of this chapter is in large part a combination of the material in the books by Olson and by Meyer, referred to in the handouts at the beginning of the course. To limit the volume of material, it will concentrate on the acoustic instruments used in modern western orchestral music. However, it is hoped that the material allows an understanding of the principles of acoustics that can be applied to the broad range of musical instruments from other cultures.

Olson is particularly good for his detailed descriptions of the musical instruments of an orchestra (Chapter 5) and the overall features of the tones that they produce (Chapter 6). The growth and decay characteristics of musical tones are presented in Fig 6.45 on page 238 of Chapter 6. Meyer has more complete material on the starting transients of musical instruments. The particularly relevant sections of Meyer are the detailed subsections of Chapter 3 dealing with the starting transients of specific instruments in turn.

As to general acoustic properties, modern mechanical (as distinct from electrical) musical instruments can be grouped into broad categories. Meyer groups them into Brass, Woodwind and Strings. Olsen groups them into Strings, Wind and Percussion. In these notes the classification scheme of Olsen will be followed but the order of presentation will be altered so that the instruments with the conceptually simplest operations will be considered first.

12.1 Percussion Instruments

12.1.1 General Features of the Class

As the name implies, these instruments produce musical tones as a result of a hit upon a vibrating system. The direct sound from these instruments is therefore characterized by a sudden onset of a note with a subsequent decay. The maximum loudness of the direct sound is therefore at the very beginning of the note (i.e the very first oscillation cycle of the note). This gives a sharp time signal for the note leading to the predominate use of such instruments in establishing a rhythm or beat to the music.

The principal concern with recording percussion instruments is then to make sure that there is adequate direct sound compared to the reverberant sound. This is because of the importance of the first few milliseconds of sound from a percussion instrument. In a large orchestra, this can be difficult because the percussion instruments usually placed at the back of the orchestra and so will be considerably farther from microphones than will be what are regarded as the more important instruments such as the first violin and, possibly, a piano.

When a percussion instrument is struck, the initial displacement of the vibrating parts by the blow sets up a particular combination of the normal modes of the instrument. The timbre and loudness of the resulting tone is determined by the particular combination of normal modes that are achieved by the blow.

The combination of normal modes that are set up by a blow to a percussion instrument is determined by a variety of factors including the hardness of the hammer, the hardness of the point of the system being struck, the position of the point being struck and, of course, the swiftness of the strike. A particular percussion instrument can therefore sometimes produce a wide variety of sounds, depending on the particular needs of the music.

From the material of the preceding chapter, the following can be considered to be general rules:

1. A strike with a hard object will set up more of the high frequency modes, resulting in a brighter timbre, than a strike with a soft object.
2. A sudden, sharp strike will not only set up higher excitation levels for all the modes but will also favour the higher frequency modes.
3. In many instruments, such as a drum, the excitation level of a particular mode can be varied by varying the point of impact upon the instrument. In general, the closer the impact is to a rigid support of the system, the more the high frequency normal modes will be favoured.
4. Since the normal modes are independent, they will all decay with their own particular time-constants. Some of the modes will have a high Q then their frequency neigbours and therefore tend to last longer. However, since the decay time also varies with the reciprocal of the mode frequency the higher frequency modes will generally decay faster than the low frequency modes. The timbre of a note from a percussion instrument will therefore usually become more mellow as it dies.

The partials in tones from percussion instruments are the normal mode vibrations and these are not necessarily harmonics of the frequency of the lowest mode. If the frequencies of at least one significant higher normal mode is close to being harmonic to the lowest mode, or if only the lowest mode is significant in the sound, then the instrument is referred to as having "definite-pitch". If not, then the instrument is referred to as having "indefinite pitch".

12.1.2 Definite-pitch Percussion Instruments

The definite-pitch musical instruments used in a modern western orchestra include the tuning fork, xylophone, marimba, glockenspiel, celesta, chimes, bells and kettledrum.

(a) Tuning Fork

The tuning fork is a massive bar bent into the form of a "u" so that the two ends can vibrate against each other (see Fig. 12.1).

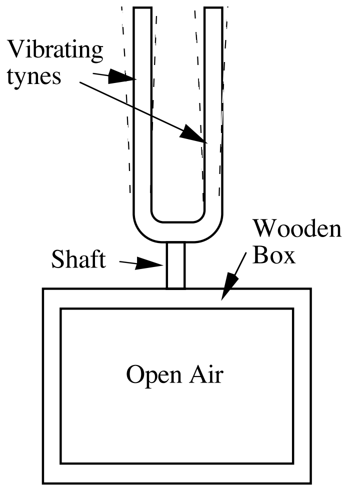


Figure 12.1 A typical music tuning fork.

A rigid support shaft is attached to the bottom of the "u" and fixed to a wooden box. This allows there to be a large amplitude of motion of the ends, and hence a large amount of stored energy, with very little motion of the point of support. The system therefore has a very high Q, the highest of any of the musical instruments. Also, the rigid point of support at the center suppresses all other modes of vibration except those of very high frequency which decay very quickly. A very short time after the instrument is hit, the tone therefore becomes very pure. The instrument is therefore not used as an instrument to actually make music but, as its name implies, to check the tuning of playing instruments.

To most effectively couple the mechanical vibration of the fork to the air in the room, and therefore get the loudest sound into the room, the box is rectangular and open at both ends and designed to have its lowest air mode at the frequency of the tuning fork. This mode will be the "half-wave" mode which will have a pressure maximum at the center of the box. The center of the box therefore becomes a high impedance point for the air resonance. This improves the coupling of the very high-impedance oscillation of the fork to the air in the box (see Chapter 13 on Acoustic Impedance).

(b) Xylophone

The xylophone consists of a number of metal or wooden bars lying horizontally on soft material, the points of support being at nodal points for the lowest transverse mode of oscillation of the bars (see Fig. 12.2). The notes are generated by hitting the bar with a soft hammer and under the bars there is usually a pipe resonator for each bar tuned to the frequency of the lowest mode of the bar.

The design of the xylophone is obviously to enhance the lowest mode of vibration and so give a very pure tone. However, as with any struck instrument, there will be higher frequency modes induced by the impact of the hammer. The effect of the design favouring the lowest mode is therefore to have these higher modes decay very quickly, usually in only a few cycles or several tens of ms. This gives the zylophone its particular form of attack; a short, but fairly soft, transient followed almost immediately by a pure tone.

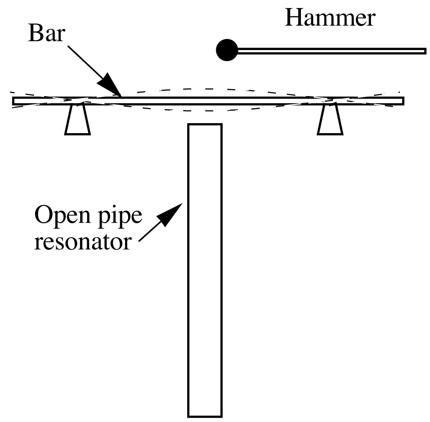


Figure 12.2 A configuration of a xylophone.

Xylophones have frequency ranges of two or four octaves, the usual frequency range being C_3 to E_7 (130.8 to 2637 Hz). The marimba is very similar to the zylophone but with a larger frequency range, F_2 to F_7 (87.3 to 2794 Hz).

Other instruments that are based on a struck vibrating bar are the glockenspiel and the bell lyre. These are smaller than the xylophone typically from C_3 to C_6 (130.8 to 1046 Hz) and do not have resonating air columns for amplification. The bell lyre is the hand-held version used in marching bands.

Another relative of the xylophone is the celesta which is actuated by hammers connected to a keyboard as in a piano. This instrument incorporates mechanical dampers which cause the vibration of a particular bar to cease after the key for that bar is released. The range of the celesta is typically C_4 to C_8 (261.6 to 4186 Hz) .

(c) Chimes and other bells

Chimes, sometimes called "tubular bells", are related to the zylophone in that the basic oscillation is the transverse oscillation of a uniform bar. However, in chimes the "bars" are hollow tubes and they are suspended vertically from their ends. This allows a different set of normal modes to be excited then in the xylophone. In particular it allows significant amounts of the higher frequency modes to be sustained for a short time. Chimes therefore have a richer timbre than the xylophone and have a characteristic more like that of bells. To excite these higher modes of oscillation, the hammer is usually harder than that used in the xylophone class of instruments.

Bells are typically of metal formed into an inverted cup. They are essentially the two-dimensional versions of the tuning fork in that instead of a bar formed into a "u" it is a circular plate formed into a cup. The higher modes of oscillation are therefore analogous to those of a circular plate shown in Chapter 10. The thickness and shape of the metal is formed so that one or two of the higher modes have frequencies which are close to harmonics of the frequency of the lowest mode, i.e. the pitch of the bell. This gives the characteristic tone of a musical bell.

The carillon is a collection of bells in which the pitches are selected to form a musical scale. The individual bells can be excited by various mechanisms, sometimes electromechanical and sometimes in a keyboard arrangement. However, because of the large number of higher modes that are not harmonic to the fundamentals, the simultaneous ringing of many bells can produce a discordant sound.

(d) Kettledrum

The kettledrum, sometimes called the timpani, is formed of a leather skin stretched over a hollow hemispherical bowl. The shape of the bowl and the density and tension of the stretched skin is set so as to give a musical relationship between the frequencies of the various modes of vibration of the system. In this way the kettledrum differs from other drums where such a musical relationship is not usually maintained and the instruments are therefore classified as being of indefinite pitch and of limited musical use.

A mechanism operated by a pedal is used to adjust the tension of the stretched skin very quickly and accurately so that a musical scale can be played with just one instrument. There are two standard sizes of timpani, the smaller producing tones ranging from $B\flat_2$ to F_3 (116.5 to 174.6 Hz) and the larger from F_2 to C_3 (87.3 to 130.8 Hz).

12.1.3 Indefinite-pitch Percussion Instruments

Indefinite-pitch percussion instruments used in modern orchestral works include triangles, drums, tambourines, cymbals, gongs and castenets.

(a) The Triangle

The triangle is a steel bar of varying cross section bent to form a triangle (Fig. 12.3)

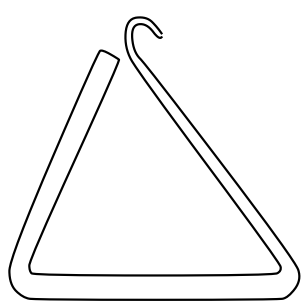


Figure 12.3 The configuration of a musical triangle.

The hammer (called the beater) is also usually of metal. The result of a hit is a complex mixture of transverse, longitudinal and torsional vibrations, giving a sound which is very rich in high frequency content and with a very sharp transient. It is for this reason that the instrument is used primarily to mark rhythm and to call special attention in a musical passage.

(b) Drums

Drums come in roughly three classifications, bass drums, military drums and snare drums. A related instrument is the tambourine. All of these instruments are formed by stretching a leather skin over a structure forming an air cavity but with no particular attention being paid as to whether the various vibration modes of the system are harmonically related.

Bass drums are used to mark time in music and to augment the general output of sound, particularly in low frequencies and in out-door instruments. They have the familiar pill-box shape, identical skins stretched over each end of a short cylinder, and range in size from about 60 cm to over 3 m in diameter.

The military drum is similar to the bass drum except that it is smaller and is formed from a cylinder which has a greater length to diameter ratio than for the bass drum. This gives the instrument more high frequencies components then the bass drum and gives it the characteristic "marching" timbre. Transients in the sound of this drum can be much more pronounced than in the bass drum.

The snare drum is essentially a miniature version of the bass drum but with a set of catgut strings attached to the side of the drum which is not struck by the batons. The strings are designed to be touched by the leather membrane when it vibrates, resulting in a buzzing sound. The tone of a snare drum is brilliant and crisp, indicating a complex transient in the attack of the note.

The tambourine is also a miniature version of the bass drum but with a very short cylinder which is more like a ring than a cylinder and with only one side covered by a stretched membrane. It is essentially a noise-maker with small disks of metal added to the outside of the ring to enhance the effect of the instrument. Its sound will have a very complex waveform with very transient characteristics.

(c) Cymbals and Gongs

A set of cymbals is formed of two disk of brass each with a concave section at the center. The vibration patterns of each disk are therefore very much like that of a circular disk but, unlike in the case of the bell, no attention is paid to having the various modes of vibration harmonically related. The instrument is therefore one of indefinite pitch.

Sometimes a single cymbal is mounted on a drum so that the drummer can play it in conjunction with the drum performance by using a drum-stick. Usually however, they are played in pairs, the sound being made by striking the two disks together.

In any case the result is an extremely complex pattern of vibrations in the disks. In the hand-held version the actual sound pattern radiated into a room can even be further complicated by the relative placement of the two disks by the performer after the note has been struck.

The gong is related to the cymbal in that it is a basically a metal disk struck by a hammer. Compared to a cymbal however, it is usually very large and the hammer is very heavy and has a soft head. This produces a set of very low frequency modes of great power. However, the modes are again not harmonically related and the initial sound from the instrument is often referred to as a loud roar. However, because of the tremendous power of the instrument the sound can last a long time. After a long time only the most powerful low frequency mode is left and so the sound from the instrument gradually progresses from a loud roar to a pure tone.

(d) Castenet

The castenet is essentially a three-dimensional version of a set of cymbals in that it is formed of two approximately hemispherical shells clapped together to form a sound. It therefore also has a great deal of complexity in its modes of vibration.

However, castenets are formed of wood instead of metal and are much smaller and are hand-held. They therefore produce much less sound than do cymbals and have a very different timbre, this timbre not only being due to the properties of wood compared to metal but also being influenced by resonances in the air cavity formed between the two hemi-spherical shells.

Because of their small size, the radiation patterns of all the modes from a castenet are very simple, essentially uniform in all directions.

12.2 String Instruments

12.2.1 General Features of the Class

The vibrating string is the basis of some of the oldest musical instruments. These instruments were very important in the development of music itself. This is because generally humans find a great attraction to sounds that simultaneously contain harmonics of a fundamental tone, i.e. contain parts that are "in harmony". In musical terms they are sounds that contain "partials" that are harmonically related to the fundamental. The availability of instruments that provided such sounds therefore considerably expanded the range of musical possibilities and it was discovered very early that the vibrating string provided such sounds.

The vibrating string is one of a very few, select, systems which has its higher mode frequencies close to being harmonics of the fundamental. It is extremely rare in nature to find an object which has this property, despite the fact that practically all objects will make some sound when struck and many will produce ringing sounds. This is the reason for the vast number of indefinite-pitch type percussion instruments in the music of the human race and the relatively late development of percussion instruments of the definite-pitch type. These instruments typically required eons of development of the human skills necessary to craft the devices into a form where the higher modes of oscillation had frequencies which were multiples of that of the lowest mode.

This was not so for two very simple systems; the stretched string and a long hollow tube. These systems then form the basis of some of oldest musical instruments. Because of the ease of understanding the mechanisms by which a stretched string is put into vibration, this form will be considered first.

(a) The Frequencies of a Stretched String

If some simplifying assumptions are made about a stretched string, the frequencies of the various modes of vibration of that string can be easily calculated. These assumptions are:

1. The string has uniform linear density along its length, i.e. each millimeter of its length will have the same mass.
2. There is a uniform tension along the string.
3. The string has no resistance to bending, i.e. it is perfectly flexible like a loose extremely fine chain.
4. Other than the tension, there is no force on the string aiding or resisting its motion.
5. The string is firmly fixed so that there is no motion whatsoever at its two ends.

If all these conditions are met, then the string will have normal modes of oscillation (the "standing waves of Chapter 10) which are transverse vibrations of frequency

$$f_n = \frac{n}{2L}\sqrt{\frac{T}{m}} \qquad (12.1)$$

when n is an integer designating the mode number (1 being the fundamental or lowest mode), L is the string length (in meters), T is the tension in the string (in Newtons, where 1 Newton is approximately the weight of 100 gm) and m is the linear density of the string in kg per meter (1 kg/m = 10 gm/cm).

From (12.1) it can be seen that, since L,T and m are all constants of the string, the frequencies are proportional to the integer n. They are therefore harmonically related.

Thus if the conditions listed above are all met, then a vibrating string will have harmonically related modes of vibration. Of course, for any real string they will not be met absolutely but for strings that can give vibration frequencies in the range that is of interest in music, they can be easily met to a very high degree. The practical conditions are:

1. The string must be strong and capable of a high tension with very little mass so that the ratio of T over m in (12.1) is large enough for the frequencies to be in the musical range of interest.
2. The string must be uniform over its length.
3. The string must not provide any significant resistance to bending. This condition is met for a string which is very thin compared to its length.
4. The string must have very low frictional forces to its transverse vibration.
5. The points of support at the ends must be sharply defined and very rigid.

In practical wires used for such strings, it is possible to achieve normal mode frequencies for n up to about 8 which are no more than 0.1% different from the harmonics of the lowest mode. This would be typical for a violin string but for a piano where more sound power is required and the string is heavier and therefore thicker, the deviations for the $n = 8$ mode can be as high as 1%. However, deviations from the harmonics at these high modes are usually not of significance because the higher modes decay very quickly. Since it takes some time for the pitch of a tone to be established (see the material on the Fourier transform of Chapter 9) the harmonic relationships of the partials of a tone are not perceived until the higher frequency components have significantly diminished.

However, while the tonal qualities of a vibrating string must have been noted very early by prehistoric man it did not become the basis of musical instruments until the development of systems for coupling the string motion to the surrounding air.

(b) Coupling a Vibrating String to Surrounding Air

The basic method by which the motions of a vibrating string are coupled to air is through a "sounding board". This is a large surface area, usually of wood, to which is firmly attached a structure which defines one end of the vibration patterns of the wires (see Fig. 12.4).

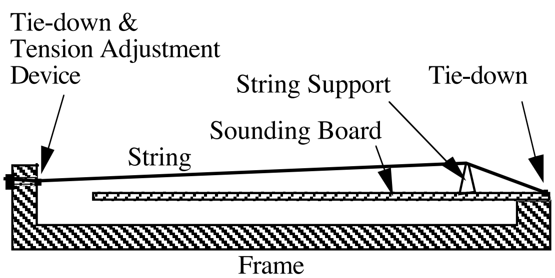


Figure 12.4 A schematic representation of a sounding board system for coupling the motion of a vibrating string to air. The system shown is a simplified version to that used in a piano.

As a consequence of this coupling, the modes of vibration of the sounding board itself come into play. To have a good coupling of the string motion to the board there must be a dense spectrum of normal modes of the board covering the frequencies of the modes of vibration of the string. A large board will normally have such modes. Also, of course, a large board is a very efficient radiator of sound from its vibrating modes (see the Chapter 13 on acoustic impedance).

However, for many stringed instruments, particularly those that are hand held, a large board is undesirable. A smaller, lighter board is then used and accompanied by a box behind the board. Such a box will have, in addition to the vibration modes of the sounding board under the strings, modes of vibration of its back surface and of the air in the box. To use all these modes, the back surface is connected to the sounding board by a wooden post, thereby even further increasing the possible modes of vibration, and the air in the box is coupled to the surrounding air by one or more holes in its surface. (Of course, one of the most prominent air modes of such a box will be the Helmholtz oscillation.)

The radiating pattern of a typical stringed instrument will therefore be that of the various modes of the system which resonate with the string modes.

There are basically three types of stringed instruments, the plucked string, the struck string and the bowed string. When a stringed instrument is plucked or struck, it has many of the characteristics of a definite-pitch percussion instrument. This sub-class of stringed instruments will be discussed first.

12.2.2 Plucked and Struck String Instruments

Examples of ancient plucked string instruments are the lyre, lute and zither. Examples of modern plucked string instruments are the guitar, mandolin, banjo and ukelele, harp and harpsichord. The principle examples of modern struck string instruments are the dulcimer and, of course, the piano, perhaps the most prevalent of all musical instruments in the western world.

The difference between activating a stretched string by plucking or striking is in the manner in which the initial normal modes are excited. In a plucking action, the various levels and phases of the initial modes are those that add up to give the shape of the string just before it is released (see Fig. 11.11). By plucking the string at a point near the center the fundamental and all subsequent odd numbered modes can be favoured. By plucking the string near one of the tie-down points, the higher frequency modes can be favoured. By gripping the string more securely at any particular point before releasing, the level of all the modes can be raised. However, if the string is always plucked at the same point and by the same implement, the timbre of the note will not change a great deal with the level of the note produced.

When as string is struck, the primary effect on the string is to give it an initial velocity rather than an initial displacement. The initial modes that are excited are therefore those that have velocities which add up to the velocity induced by the strike. Thus an impact with greater velocity will tend to favour the higher frequency modes. The timbre of a struck string will therefore noticeably change with the strength of the strike, an effect which is of great importance in the piano where the timbre of the notes can be appreciably altered by the "touch" of the pianist.

The consideration of particular plucked and struck stringed instruments will start with what is usually considered to be the guitar family

(a) The Guitar Family of Instruments

The guitar family of stringed instruments originated with the lyre of ancient Greece. A simplified sketch of its configuration is shown in Fig. 12.5

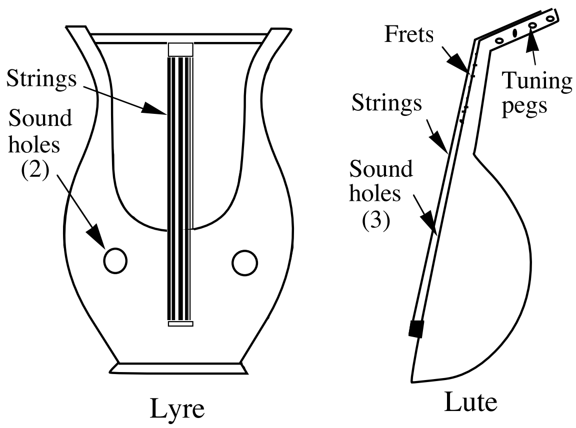


Figure 12.5 Sketches of the lyre and the lute.

The lyre is not only regarded as the forerunner of the guitar family but for all stringed instruments in that it incorporates the essential features of the stringed family. In addition to a structure for supporting the vibrating strings and putting them under the correct tension, it has a sounding board in the form of one side of a box and holes in the box to couple the air vibrations within the box to the surrounding air. The length of a particular string under vibration could be controlled by pressing the string against a fingerboard immediately behind it, very much like in a ,modern violin.

Because of the relatively small cavity of the box, and other features of the instrument, the sound level produced by the lyre was not very great by modern standards. However, because the general noise level in society was considerable lower at the time of the lyre than in modern times, not as much sound power was required in a musical instrument. Remember that this was the age when actors could entertain audiences in large open amphitheaters with, of course, no electrical amplification of their voices.

The lute is regarded as the more immediate forerunner of the guitar in that the box has a larger cavity and the device has frets which the player can use to limit the vibrating lengths of the strings. It was developed about 1000 years ago.

The use of frets instead of the soft tissue of a finger to define the vibrating length of a string allows the higher frequency modes in the string to last considerably longer and gives a brighter tone to the instrument compared to the lyre. Also the structure of the instrument, particularly the large round cavity and larger holes gave more efficient coupling of the string vibrations to the surrounding air and hence a louder sound.

However, the sound level from a lute was still much lower than that from the modern members of the guitar family. Compared to the lute, these instruments have heavier strings and a stronger body allowing the same tone frequencies as in the lute to be generated with much more power and richer timbres.

Of the modern instruments, the mandolin has essentially the geometry of the lute except that it has only four strings, tuned to G_3, D_4, A_4 and E_5 (196, 293.7, 440.0 and 659.3 Hz). The frets are set so that the fundamental vibrations resulting from the string being pressed against two adjacent frets in succession will be one semi-tone apart.

These strings are stretched over a bridge, as in a guitar or violin, giving a much better coupling of the string to the sounding board formed by the flat surface of the box. This gives the instrument much more sound power than lute. The instrument is plucked with a pick or plectrum rather than by the fingers and so the tone is rich in higher frequency modes.

The modern guitar is obviously a further development of the lute in that the resonating box is even larger than for a mandolin and the strings are longer and heavier. Consequently the fundamental frequencies of the strings are much lower than for the mandolin. It has six strings tuned to E_2, A_2, D_3, G_3, B_3 and E_4 (82.4, 110, 146.8, 196, 246.9 and 329.6 Hz respectively) and frets to produce tones at half-tone spacings.

The strings of a guitar can be plucked by using bare fingers or a plectrum. By plucking softly with a finger in the center of the free length a relatively soft pure tone can be produced. By striking with the finger nails near the bridge a tone which is very rich in harmonics can be produced (see Fig. 11.11). Thus a sudden and very great change in timbre can be achieved using the instrument, a feature which is characteristic of the guitar.

There are some other commonly played stringed instruments of the guitar family; the ukulele, the Hawaiian guitar and the banjo. The ukulele is essentially a small version of the ordinary guitar with only four strings. The Hawaiian guitar has a unique arrangement for limiting the vibrating lengths of the strings by using a sliding metal fret. The higher frequencies and the sliding tones that can be produced while a strong high frequency content give this form of the guitar its unique sound. It typically has only four strings.

The banjo is another four-stringed instrument. It produces its unique tones by forgoing a resonating box and using a membrane stretched over a relatively small round rim as a sounding board. There is no covering on the other side of the rim so there is no resonating air cavity.

Because of the structure and shape of the banjo, the high frequency modes of the strings are much more efficiently coupled to the surrounding air than are the low frequency modes. The timbre of the tones produced are therefore rich in high frequencies. The four strings are tuned to C_3, G_3, D_4 and A_4 (130.8, 196, 293.7 and 440 Hz).

(b) The Piano Family of Instruments

The ancestor of the piano can be considered to be the zither. This instrument of antiquity is still played in modern times. It is made up of a set of strings of varying composition stretched horizontally side by side over the top of a flat hollow box. The edge of the box forms a frame over which the strings are stretched and the box itself forms a sounding board. The box has a very large hole to help in coupling the sound in the box to the surrounding air.

The modern instrument consists of 32 strings of which four are located over a fretted board. These strings are used to play the melody while the remaining 28 strings are used for accompaniment. The stings are played by plucking with a ring-type plectrum.

Because of the small size of the zither and the weight of the strings, it is not a very loud instrument, particularly compared to the piano. The range of the strings is also limited. The four melody strings of a modern zither are tuned to C_3, D_3, D_4 and A_4 (130.8, 146.8, 293.7 and 440 Hz) and the accompaniment strings range from C_2 to $A\flat_4$ (65.4 to 415.3 Hz).

The harp can also be considered a forerunner of the piano in that it consists of a set of heavy strings stretched over a massive frame and covering a wide range of fundamental frequencies (from C_1 to G_7, or from 32.7 to 3136 Hz).

In the case of the harp the strings are mounted vertically and are played by plucking or stroking with the fingers. The lower section of the harp, which supports the lower ends of the vibrating strings, is widened to form a sounding board. However, while the massive strings can contain a great deal of vibration energy, the relatively small sounding board, compared to that of a piano, does not couple the string motion very effectively to the surrounding air. Thus the sound of a harp is much gentler than that of a piano but, on the other hand, can last much longer.

Because it is instrument plucked by fingers, the harp is a very mellow instrument with not very much of the high frequency modes being excited.

A more immediate ancestor of the piano is the harpsichord. This instrument has a large number of steel strings stretched over a frame, the strings lying in a horizontal plane. A particular string is plucked by a leather or fibre plectrum which is actuated by the press of a key associated with that string. The shape and arrangements of the keys are that of the grand piano. The range of the keys of a typical harpsichord is from G_1 to F_6 (49 to 1397 Hz).

Because of the mechanical plucking arrangement, the sound of a harpsichord is characterized by having very little variation in intensity and timbre of its notes compared to that of a piano. Also, the overall intensity of the sound produced by the instrument is considerable weaker than that of a piano. While this makes it more attractive to some people as a musical instrument, it makes it of considerably less versatility in musical works.

Another forerunner of the piano is the dulcimer. It resembles a piano without legs and is played by hand-held hammers, one in each hand. More variation in intensity and timbre can therefore be achieved than in the piano. It is also perhaps significant that this was one of the first forms of struck stringed instruments.

As mentioned earlier, the piano is perhaps the most predominant modern musical instrument. There are many good books written on the piano, its mechanism and the sounds it produces. The amount of development which has gone into the modern grand piano has resulted in a very complex and very subtle instrument. The efficiency which which a grand piano can transform human finger power into sound is truly amazing. No attempt will be made here to fully describe this important instrument but only to summarize its important features.

The piano is basically a large number of steel strings of various thickness, densities and construction (some are wires wrapped by other spiral wires to increase the strung mass without diminishing its flexibility) stretched over a heavy steel frame. A very large sound-board is used to couple the vibrations of the strings to the surrounding air. The strings vary in fundamental frequency all the way from A_0 (27.5 Hz) to C_8 (4186 Hz).

The normal modes of the strings are activated by felt hammers connected to keys laid out in the familiar piano key-board arrangement. One key activates two or three strings simultaneously to increase the intensity of the sound. The strings for any particular note can be tuned to give beating effects. Because of the massive strings, the higher frequency modes, as already mentioned, deviate noticeably from being harmonics of the lowest mode. Furthermore, the suspension of the strings is such that the plane of vibration of a particular mode rotates in time, each mode rotating at a different rate. The coupling of the mode motion to the sound-board is much more effective when the vibration is in a plane perpendicular to the board then when it is parallel to the board. This causes a bobbing up and own of the perceived intensity of each the normal modes, each mode bobbing up and down independently. All of this makes the piano tone one of the most subtle sounds in the musical world and, of course, one of the most difficult to tune for optimal performance.

The importance of the piano as a musical instrument also contributes to problems in recording its music. Because it is heard so often by so many people, the subtleties of its tones are appreciated by a great number of people. It is therefore important that a recording be very faithful to its sounds

The major problems with recording a piano's sound stems from its large size and the large size of its radiating surfaces; the sounding board and, in a grand piano, the reflecting lid used to direct the upwardly propagating sound of the sounding board horizontally towards the audience. The size of these surfaces gives even the low frequency components of the sound a very directional characteristic.

However, this directional characteristics of a grand piano can be used to some advantage. Provided a microphone is placed within the main radiating fields of the lid and, possibly, the floor underneath the piano, the room radius will tend to be much larger than for an isotropic source and so the microphone can be safely moved far enough away from the instrument to receive a balanced sound from all of the piano's radiating components.

12.2.3 Bowed String Instruments

The bowed string instruments are the first in this sequence of considerations to represent a sustained tone generator. To generate any sustained tone from a human effort there has to be some sort of feedback effect which causes the energy put into the system by the human to be transformed into energy of oscillation. This is because the frequency of sound oscillations are far beyond that which can be produced directly by human shaking of the instrument.

The basic feedback mechanism at work in bowed string instruments has been presented in sec 11.8 of Chapter 11. It depends critically on the property of resin which gives it a very high static friction and a very low kinetic friction. It other words, resin is sticky. This means that when a resined bow has been pressed against a string, a sort of glue joint forms. As the bow is pressed sideways, perpendicular to the string, the string will accompany the bow for a while until the force required to displace the string breaks the glue joint. The string now flips rapidly back to its rest position and, because of its momentum, a little beyond. It then comes to rest and a new glue joint forms with the resined bow.

This sequence would normally be repeated with varying release times and intensities as the various regions of the bow, with varying amounts of glue etc. pass over the string. However, the first release causes a travelling wave to go along the string to the far end where is is reflected back towards the bowed end. The result is that when it returns to just under the bow, this reflected wave can trigger a break of the glue joint. The pulse caused by this new break will add to the pulse that caused it, resulting in an increase of the disturbance on the string.

Thus it is seen that the basic conditions for a feed-back oscillation are met. The previous pulses cause a new pulse which is added to the old ones and the resultant oscillation grows. It can also be seen that what is achieved is a sawtooth oscillation of a very definite frequency; that of the lowest mode of oscillation of the string.

In chapter 9 it was shown that a saw-tooth oscillation is very rich in all harmonics. If then, there are modes of the system which have frequencies equal to these harmonics, then these modes will resonate at these harmonic frequencies. This is the reason for the richness of the timbre of bowed string instruments.

A common characteristic of any sustained tone generated by feedback is that the perceived sound level in a room will grow in a complex fashion. First there is the starting transients; the initial pulse of sound from the first release of the string from the glue of the resined bow. Then there will be subsequent pulses of relatively uncertain periods (times between pulses) while the reflected pulses on the string are weak and not yet capable of reliably causing a fresh break of the bow-string glue joint when they pass under it. However, as the pulses travelling up and down the string grow, they will settle into a definite pattern and the frequency of the glue joint break will resolve into a definite pitch.

At this point the tone will rapidly grow in intensity until a point is reached where the maximum possible amplitude of oscillation is reached for the particular bowing action used. The sound level from the instrument will then level off as shown in Fig. 12.6.

In most musical instruments producing a sustained tone, the sound reaches to within 3 dB of the sustained level within less than 0.1 seconds. However, by then another phenomenon has set in; the growth of the reverberant sound in the room. In a typical concert hall this will also take about 0.1 second to reach to within 3 dB of its finally sustained level. (For a concert hall with a reverberation time, or time for a 60 dB decrease, of 2 second , the time for a 3 dB change would be 1/20 of that or 0.1 see chapter 4.)

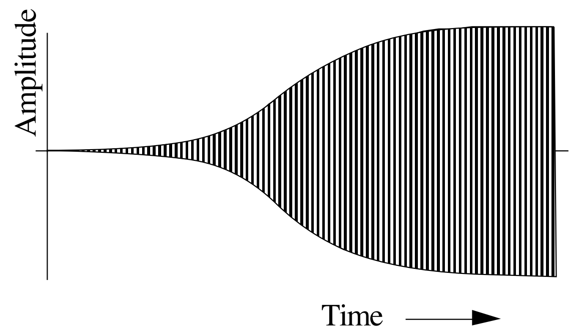


Figure 12.6 The general growth pattern of a feed-back oscillation.

Thus most of the power in the pitch defining components of a tone in the direct sound from an instrument grows at about the same time as the reverberant sound in the room. The pitch and timbre of both the direct and the reverberant sound therefore develop together, becoming much louder than the original starting transient.

However, in a good listening environment, the initial starting transient can be clearly heard in the direct sound because it is not masked by the sustained tone which has not yet grown. This is particularly true for solo instruments where the passage between notes is not masked by sounds from other instruments..

Within these considerations the individual members of the modern bowed string instruments will be described. These are the violin family; the violin itself, the viola, violoncello and the double bass (contrabass).

(a) The Violin

The violin is an instrument with 4 strings tuned to G_3, D_4, A_4 and E_5 (196, 293.7, 440 and 659 Hz). In a general sense its construction is similar to that of the guitar except that it is smaller. However, in the shape of the resonating box, the cross section and the type of wood used for the sounding board and the back plane, and even in the shape of the air holes, much more attention is paid to achieving a broad spectrum of modes of oscillation in the system to give the proper timber to the sound in the room.

The violin is the most common of the bowed string instruments. This is because its frequency range fits into the middle of the musical spectrum and in any well-balanced symphony orchestra has, by far, the most players of any of the instruments in the orchestra. Typically there will be 35 violins, 12 violas, 10 violoncellos and 8 contrabasses.

The modern concert violin, like the grand piano, is a marvel of human craft. There is great complexity and subtlety of its tones and in their development following an initial bowing motion, and these subtleties and complexities are individual to any great performer. Because the violin is heard so often by people who favour classical concerts, as for the piano, the subtleties of its tones are appreciated by a great number of people. It is therefore important that a recording be very faithful to the violin.

However, the violin does not seem to present as many problems in recording as does a piano. This is because the instrument is considerable smaller and its direct sound is therefore not as directional as from the piano. Also, because of the slower growth of the full sound, the ratio of direct sound to reverberant sound is not so important. The starting transient of a violin with softly attacked notes can be as long as 300 ms. For sharply attacked notes the main sound power will have a starting transient of from 30 to 60 ms, depending on the note that is played. This compares with a "starting transient" for the room reverberation of about 100 ms.

However, in a sharply attacked note there will be a much shorter transient for the high frequency components and so the direct sound must still be given prominence, particularly for the lead violinist.

The frequency spectrum of the sounds from a violin, particularly that of the upper strings, is very rich in harmonics. In fact, except for the lowest string, the G string, the strongest partial in a tone is not the fundamental. The frequency spectra of the tones from the higher strings peak at 3 to 4 kHz and have significant components at 10 kHz. Even in the G string, there are frequency components at 8 kHz which are only 25 dB down from the fundamental. This is, of course, the 40th harmonic of the fundamental of that string!

However, while the fundamental and the lower tones do not make up a very large part of the overall sound of a violin, they are very important in establishing the pitch of a tone. They must therefore be faithfully recorded. Since the high frequency components and the low frequency components have vastly different radiation patterns (see Chapter 6), microphone placement for a solo violinist can be a difficult decision to make.

(b) The Viola

The viola is somewhat like a large violin. It is played in the same manner as a violin and has about the same shape. It just has heavier strings and a larger body. The strings are tuned to C_3, G_4, D_4 and A_4 (130.8, 261.6, 293.7 and 440 Hz). Thus it is close to being a violin with all the frequencies reduced by about 30% or about a musical interval of a fifth. It therefore fills in a different part of the musical spectrum, towards the bass compared to the violin. In a way its function relative to the violin is similar to that of the left hand versus the right hand in piano playing; that of a support for the melody of the violin.

However, the viola has a high enough register to have significant musical works composed for it in its own right. The same care taken with recording the violin should then taken with the viola.

While the sound of a viola is very similar to that of a violin transposed down about 30%, taking a recording of a violin and playing it as 70% speed will produce a sound which is noticeable different from a viola. This is because to make the instrument manageable its dimensions are less than is 1.5 times those of a violin. (1.5 is the factor that would bring about a 33 % reduction in all frequencies.) In fact it is only slightly bigger than a violin, being about 66 cm long compared to the violin which is 60 cm long.

The viola is therefore not exactly a scaled up violin and its radiation patterns are slightly different. Also, relative to the violin it has an even higher fraction of its sound power in the higher harmonics compared to the fundamental.

(c) The Violincello and the Contrabass.

The violoncello is again similar to the violin but even larger than the viola. Its strings are tuned to be exactly one octave down from the viola. Because the violoncello (commonly called the "cello") is played between the knees and not on the shoulder, a much larger box can be used. There is a therefore a considerable jump in size, to an overall length of 117 cm, from the viola to the cello.

The directional and tonal characteristics of a cello are therefore very much like those of a violin with the frequencies all scaled downward by an octave and a fifth (to about 33% of those of the violin).

A similar step occurs to the contrabass. The four strings of this instrument are tuned to E_1,, A_1, D_2 and G_2 (41.2, 55, 73.4 and 98 Hz). This is over two octaves down from the violin. The length of the contrabass is 198 cm, about the proper ratio to maintain acoustic proportions with the violin. Its timbre and radiation patterns will therefore be very similar to that of a violin for notes transposed down two octaves.

However, the cello, and particularly the contrabass, do not usually present significant recording problems since they are not used for the principal parts of musical works but a bass accompaniment.

12.3 Wind Instruments

12.3.1 General Features of the Class

Wind instruments are made up a sound production device incorporated into a resonating air column, with provision for radiation of sound from the resonating column into the surrounding air. They are all of the sustained tone type. They are all therefore based on acoustic feedback from the resonating column to the sound production device. Their tone growth will therefore generally follow the pattern of Fig. 12.6 shown in the context of the growth of sound from bowed string instruments. From a physics point of view, the main difference between a wind instrument and a bowed string instrument is that the travelling wave used for the feedback to the sound producing element is in air rather than along a string.

However, the modes of vibration of the resonating air column will generally not be as close to being harmonics of the lowest mode as in the case of a stretched string. Furthermore, the basic sound producing element will not usually produce oscillations with a sawtooth waveform as in a bowed stretched string. The tones produced by wind sinstruments will therefore have very different timbres than that of the bowed string family of musical instruments.

Winds instruments are often classified into two types based on the material from which they are made; brass and wood. The musical justification for this is the distinctive difference in timbre of instruments made from these two vastly difference materials. However, from a physics point of view, an alternate classification is perhaps more appropriate; that in terms of the basic sound producing element of the instrument. This is the classification used in Olsen and which will be used here.

The basic sound producing element of all wind instruments may be called a "reed". In this meaning, a reed is any small stiff surface over which air flow will tend to set up oscillations, either in the air flowing over the reed, in the reed itself, or in both. Air flow over any reed produces sound due to the Bournoulli effect, considered in Section *11.7.4*.

Within such a classification scheme there are several distinct types of musical instruments; the air reed, where only the air flowing over a reed vibrates, the single reed in which only one reed vibrates against a solid surface, and the double reed in which two reeds vibrate against each other. These three distinct types of reeds will be considered in turn, starting with the type that is perhaps easiest to understand; the double reeded instruments.

12.3.1 Double Reed Wind Instruments

As in all wind instruments, the basic sound production mechanism in a double reeded wind instrument is the Bournoulli effect. The so-called "Bournoulli Principal" is that air under flow creates a vacuum proportional to its velocity. This can be thought of as a tendency for moving air to suck other things into its path, including solid surfaces. Among other things, it is the basis of modern airplane wing design (Fig. 12.7).

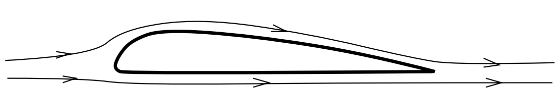


Figure 12.7 The profile of a modern airplane wing. It is deliberately designed to have a greater curvature on the top surface than the bottom so that the air over that surface has to move at greater velocity than that over the bottom surface. This produce a lift, even when the bottom surface is horizontal as shown. This minimizes the drag force compared to the lift force.

By having the wing curved so that there is a longer path length over the top surface than the bottom, the air flowing over the top surface has a longer way to go than the air flowing over the bottom surface. It therefore must travel at greater speed over the top than over the bottom. The air flowing over the top therefore creates a greater vacuum than that flowing over the bottom and so produces a lift on the airplane wing.

In the case of a double reeded wind instrument, air flow between two identical flat reeds causes a vacuum in the air between the two reeds (Fig. 12.8).

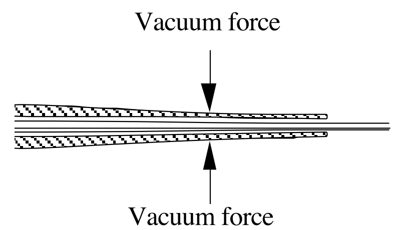


Figure 12.8 The profile of a typical double reed in a wind instrument. The reed in their relaxed state are typically separated by a small gap. Air flow through this gap results in a negative pressure causing the reeds to be pulled together to close the gap.

If the reeds are flexible, the vacuum caused by the air flow will pull them together. This, of course, closes the gap and reduces the air flow and therefore the vacuum force. The reeds therefore move apart again.

Clearly, an oscillation will be set up. Furthermore, if the air flow is strong enough and the reeds flexible enough, the reeds can be made to completely shut off the air flow for a short period of time. Thus the flow of the air from the end of the reed combination can be made into extremely sharp pulses, leading to an oscillation which is very rich in harmonics.

Thus both the frequency of the oscillation and its timbre depends on a variety of factors including the elasticity of the reeds, their mass and the strength of the air flow. Part of the set of skills of a double reed wind instrument player, in addition to blowing and tonguing techniques, is the use of a sharp knife to shave the reeds and proper moisture treatment to achieve a desired timbre from the instrument.

The attack of a double reeded instrument is distinguished by a particularly well-behaved exponential development to a saturation level as shown in Fig. 12.6. This is because any initial starting noises due to tonguing to get a particular note started are filtered out by the body of the instrument which completely encloses the reeds. The tones also develops fairly quickly compared to an instrument of the string family in the same register of frequencies. This is because the stored energy in the air column of a wind instrument is considerable less than that of a vibrating string and so the Q of its normal modes is less. A lower Q results in not only a faster decay after the note is finished but also a faster rise in the energy of the resonator (see Chapter 11).

The principal modern orchestral double reed wind instruments are the oboe, the oboe d'amore, the bassoon (& contrabassoon), the sarrusophone and the English horn.

(a) The Oboe

The oboe has a range of almost three octaves, from $B\flat_3$ to G_6 (233 to 1568 Hz). It's resonator is a long slightly conical tube, closed around the reed at the mouthpiece end and slightly flared at the mouth. The overall length of the instrument is slightly over 60 cm. Asd in all wind instruments, the effective length of the air column for resonance purposes is adjusted by opening and closing holes in the side of the column.

Because of the sharpness of the air pulses entering the resonating column from a double reed, the spectrum of sounds from an oboe is very rich in harmonics. However, the actual harmonics that are radiated from the instrument are determined by the normal modes of the resonating pipe. These produce several formants in the oboe sound, a sound which has been characterized as that of the vowel "a", with strong broad formants centered on about 1100, 3000 and 5000 Hz and a smaller one as high as 10 kHz.

Despite the presence of many harmonics, the starting transient of an oboe, as pointed out earlier, is well behaved; all the harmonics developing together with the basic oscillation to which they are associated. The starting transient for a tongued note can be as short as 40 ms, even for the lowest not and as short as 20 ms for the highest note. However, by gentle blowing the the notes can have a starting transient as long as 100 ms.

(b) The Bassoon & Contra Bassoon & Sarrusophone.

The bassoon is essentially a very large oboe with the resonating column doubled back on itself to make the instrument manageable. In this way a resonating column of overall length 2.4 m is achieved in an instrument which is only about 125 cm long.

The range of a bassoon is $B\flat_1$ to $E\flat_5$ (58.3 to 622 Hz), or about 2 octave down from the oboe. Its radiation properties for its tones are therefore very similar to those from an oboe transposed up two octaves. However, the formants, corresponding the the "a" vowel sound, are not transposed all the way down to two octaves but occur at about 500, 1200, 2000 and 3500 Hz.

The contrabassoon is an even larger version of the oboe. Its resonating tube is folded several times to get an overall length of almost 5 meters into a length of about 125 cm. Its fundamental range is B_0 to F_3 (30.8 to 174.8 Hz) or about one octave below the bassoon. Its sounds and radiating properties are therefore very similar to that of the same tones transposed up one octave to a bassoon (or transposed up three octaves to an oboe). The formants of the contra bassoon are about one octave down from the bassoon being at about 250, 450, 700 and 2000 Hz.

The sarrusophone is very similar to a bassoon in configuration but is made of brass and has a slightly greater flare at the mouth. This gives it a more "brassy" character than the bassoon but otherwise its acoustic properties are very similar. It comes in various sizes but the most common form is somewhat equivalent to the contrabass type having a range of $D\flat_1$ to B_3 (34.6 to 247 Hz).

c) The English Horn & Oboe d'Amore.

The English horn is very similar to the oboe (it's fingering is even the same) except that it has a hollow spherical bulb with a small opening at the mouth of the instrument. This considerably changes the timbre of the sound and introduces another lower formant, making the "a" vowel sound slightly darker than that of the oboe. Also, the resonances are at a lower slightly lower frequency than for the oboe, the sounds produced being somewhat like an oboe transposed down one-fifth (to 2/3). The Oboe d'Amore (or heckelphone) is similar to the English horn except that it is transposed down from the oboe by only a third (to 4/5).

12.3.2 Human Voice (vocal-cord reed)

The basic sound production mechanism in the human voice has similarities to that of a double reed wind instrument. A rough sketch of the human voice production system is shown in Fig. 12. 9.

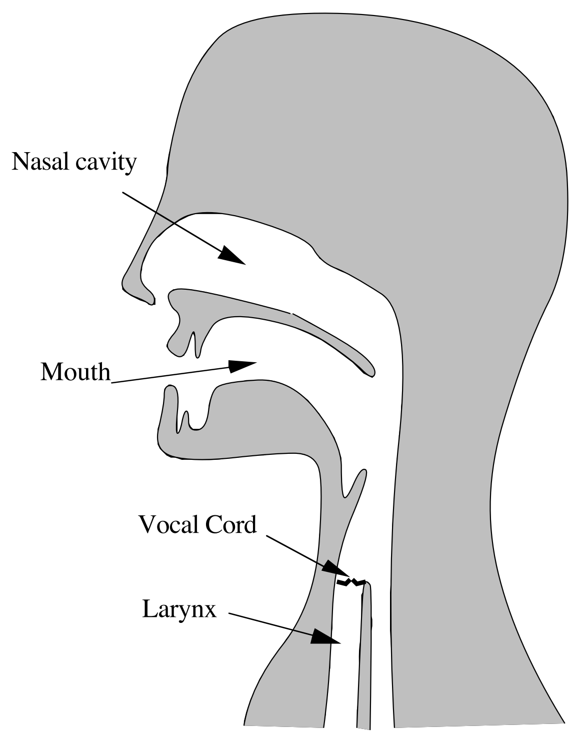


Figure 12.9 A sketch of the important components of the human voice system.

The vocal chords of a mammal resemble a double reed in that air is forced trough a slit in a membrane which is stretched by muscles. The vocal tract into which the air pulses from the vocal cords enter forms a resonating cavity which, like in the double reed instruments, is closed around the double reed and open at the far end.

However, in the case of the human vocal tract the resonating cavity is very complex and its normal modes can be altered considerably by the placement of the tongue, the opening of the mouth and the amount of air directed through the nose. It can therefore be "tuned" to give specific formants leading to different vowel sounds.

12.3.3 Lip Reed Instruments

Lip reed instruments are also related to the double reed instruments in that the lips which are pressed together form a double reed. Also, the lips are pressed tightly against a mouthpiece leading to a resonating cavity so that the resonating cavity is closed around the sound source. The other end of the cavity is open to let sound radiate into the surrounding air.

The principle modern lip-reed instruments are the bugle, the trumpet, the cornet, the French horn, the trombone and the tuba

12.3.4 Single Mechanical Reed Instruments

Single reed instruments differ from double reed instruments in that only one reed is allowed to vibrate. They may be thought of as devices in which one of the reed of a double reed instruments doesn't move. Their tonal characteristics can therefore be very similar to double reed instruments.

However, the single reeds of musical instruments are often of metal for durability and reliability in mechanically excited instruments. The tonal characteristics can therefore be very different from that from reeds made of other materials.

The principle single reed wind instruments are the clarinet, saxophone, bagpipe, harmonica, accordion and free-reed organ.

12.3.5 Air Reed Instruments

Air reed instrument differ from single and double reed instruments in that none of the surfaces that a stream of air is directed over moves. However, the air stream is set up so that it moves itself instead.

Another significant difference for air-reed instruments is that the basic sound source is not enclosed by the resonating system but forms part of the radiator.

The principle air-reed wind instruments are the whistle, fife, recorder, flageolet, ocarina, flute, piccolo and flue organ pipe.

12.3.6 Pipe Organ

The pipe organ is a combination of mechanical and air reeds devices.

Exercises and Discussion Topics.

1. Select a class of instruments of your choice and discuss the general mechanisms by which sound is produced in the instruments and radiated into a room. Relate the general properties to a specific instrument of your choice within that class.

CHAPTER 13

ACOUSTIC IMPEDANCE

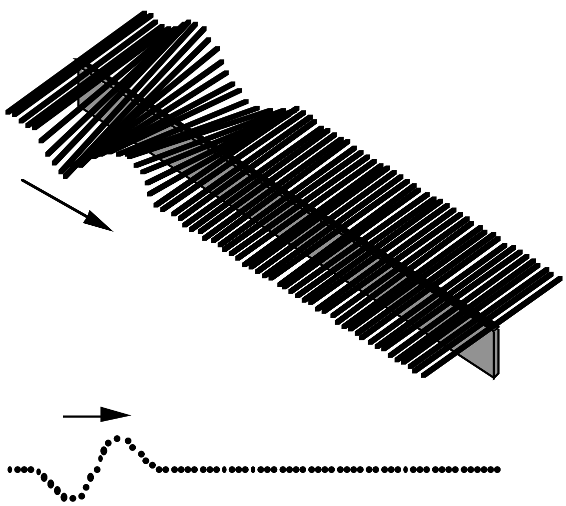
The subject of this chapter is how pressure oscillations in a device couple to the air in a room. This has two aspects; the propagation of sound away from a source and the reverse, the production of oscillations in a device by sound waves falling on it.

Both aspects are important in recording engineering. The first is that of a source, such as a musical instrument, propagating sound into the surrounding air in a room. The second is that of sound waves in the room interacting with a device such as the human ear or a microphone so as to transmit energy into that device and thereby produce a discernible signal. The subject therefore has relevance to the overall problem of how a musical instrument gets to be heard, or picked up by a recording microphone.

13.1 The Absorption of a Wave

13.1.1 The Absorption of a Transverse Wave

As an introduction to the principles involved, first consider the general problem of absorption of wave energy. This can be demonstrated by a torsional wave apparatus made up of steel rods suspended and clamped at their centers onto a steel spine (see Fig. 13.1).



Appearance of
Ends of Rods

Figure 13.1 Apparatus for demonstrating wave motion. It is made up of steel rods suspended by being clamped to a steel spine running through their centers, the steel spine itself loosely suspended on a support structure. The wave is actually a torsional wave along the spine but the resulting motion of the ends of the rods when viewed along the rod lengths, gives a vivid picture of transverse wave motion.

Since the whole assembly is suspended by loose supports for this spine, torsional waves can easily be introduced by giving a twisting motion about the spine to one of the end rods. If such a transverse wave is allowed to reach the end of the rod system it will be seen to reflect back along the system;

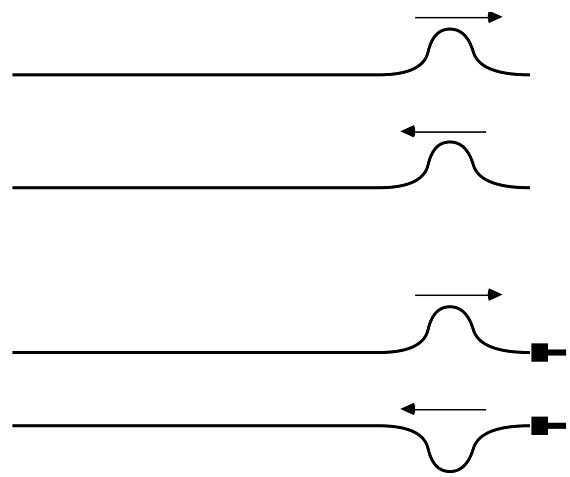


Figure 13.2 The reflection of a transverse wave in the rod system. The upper two pictures show the reflection that would occur with the right-end rod free for a pulse created by a sharp up and down movement of the rod to the far left. The lower two pictures show the reflection that would occur for the same pulse with the far right rod clamped.

When the far right rod is free to move, the reflected pulse will be appear like the incident pulse. When the far right rod is clamped, the reflected pulse will be inverted. However, in both cases the pulse is completely reflected and any energy it represents is retained and sent back along the system.

Because the incident and reflected wave are the same, no energy is taken out of the system when the pulse reaches the end. This should not be surprising. If the last rod is completely free there is no force acting on it from the right. Therefore the motion of the rod takes no work. If it is clamped so that it cannot move, it also cannot take out energy. This is because work is the product of both force and movement and if either is zero, the work is zero. How then, can the wave energy to be taken out of the system?

This is an important point for signals being propagated by waves. For the wave to deliver a signal to a receiving system, at least some of the energy of the wave must be deposited in that receiving system. For maximum signal, all the energy must be so delivered, with no reflected wave.

There is a simple device that will take the energy of a wave out of this torsional vibration system. It is a small nylon piston loosely fitting in a plastic tube and connected by a stiff wire to the end rod of the system (Fig. 13.3).

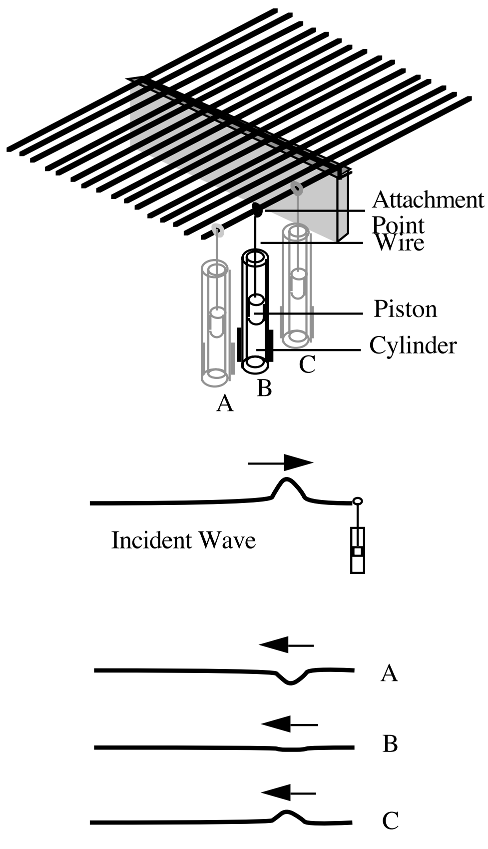


Figure 13.3 A device for taking the energy out of the torsional vibration system. The effects on a wave arriving at the end rod to which this device is attached is shown for various positions of the device.

The particular property of this system that makes it a wave damper is that the resistive force of the device is proportional to the velocity of the piston. This is because the resistance is due to the viscosity of the air forced to flow around the piston and for the modest velocities involved in this apparatus, resistive forces due to viscosity are proportional to velocity.

By moving the point of attachment of the upper end of the damper wire to the end rod, a point can be reached where a wave reaching the end rod gives practically no reflection. What determines the position of this point?

The position of the damper at which there is no reflection of the wave must be the position at which it reacts on the last rod just as would the next rod in the system if the system did not end. The question then becomes; how then does a rod in the system interact with the preceding one?

The interaction of one rod on the next is through the twist in the spine between the two rods (Fig. 13.4). If both rods have turned the same amount, then there is no net twisting of the spine and hence no torque. However, if one rod has turned more than the other (the rod to the right in the Fig 13.4), then the effect of the preceding rod is to twist that rod backwards from its forward motion.

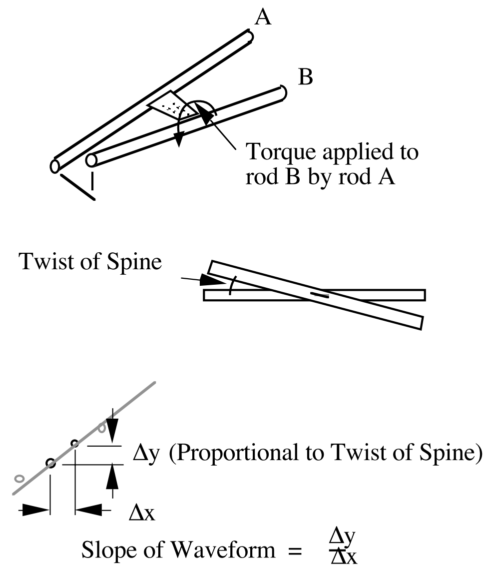


Figure 13.4 The connection between the torque applied by the spine and the slope of the waveform in the torsional wave apparatus. The torque is proportional to the twist of the spine which is proportional to negative Δy which is, in turn, proportional to the negative slope of the waveform.

The twist of the spine (difference in the angle of turn from one rod to the next) shows as a displacement upward of the end of one rod relative to the next. The twisting torque of the spine on the rod to the right is proportional to the twist of the spine. Looking at the ends of the rods, the twisting torque on the rod to the right will be therefore be proportional to the negative of the slope of the waveform.

It can also be seen that the velocity of the end of a rod is proportional to the negative of this slope (Fig 13.5).

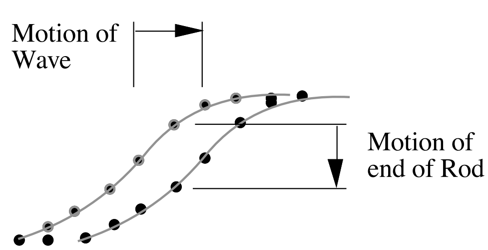


Figure 13.5 The connection between the slope of the waveform and the velocity of the end of the rod (and hence of the angular velocity of the rod). Note that the velocity of the rod downward is the greatest when the slope is the greatest positive.

It is seen in this figure that where there is no slope from one rod to the next there is no velocity of the rod end due to the wave motion. It is also seen that or a forward side of the wave where the slope is negative, the velocity of the rod ends will be upward.

Since the velocity of the end of a rod is in turn proportional to the angular velocity of the rod there is a final overall relation between the torque applied to the next rod in the system and the angular velocity of the next rod. It is

$$Torque \;= Constant \times Angular\ Velocity \qquad (13.1)$$

This can be rearranged to give

$$\frac{Torque}{Angular\ Velocity} \;=\; Constant$$

$$\equiv Wave\ Impedance \;\equiv Z_{wave} \qquad (13.2)$$

The constant which represents the ratio of torque to angular velocity is called the "impedance" of the wave system where impedance is usually written as Z. If then the damper is attached so that it also applies this impedance, then it will have the same effect on the last rod as would a continuation of the system and no wave would be reflected.

By increasing the distance of the damper from the spine, the velocity of the damper due to the angular velocity of the rod is increased and the torque resulting from the resistive force of the damper is increased. The ratio of torque to angular velocity due to the damper is therefore increased. By varying this distance, the impedance presented by the damper can be tuned to be equal to that for the rods in the wave action and there is then no reflection.

It should now be understandable why the reflections when the damper is not tuned are as shown in Fig. 13.3. When the damper does not apply sufficient torque, the last rod does not meet sufficient resistance to its motion. It then overshoots and produces a reflected positive pulse. If the damper applies too much torque, then the motion of the last rod is insufficient for the wave and a negative pulse is reflected.

The ratio of the amplitudes of the reflected wave and the incident wave can be determined by noting the ratio of torque to velocity which must exist at the termination. Considering first the simpler case of the reflection from an open end, there can be no torque at the end and so the reflected wave and the incident wave must add to give zero torque (Fig. 13.6a). The opposite extreme of a clamped end requires that the incident and the reflected wave add to give zero velocity. The result is that their torques then add. The required displacement waveforms are no inversion for the open end and inversion for the clamped end.

This requirement, that the ratio of torque and velocity that results from adding the incident and the reflected waves must be equal to the actual impedance of the termination, can also be used to calculate the ratio of the amplitudes of the reflected and incident waves for the general case. The result is

$$\frac{A_{ref.}}{A_{inc.}} \;=\; \frac{Z_{wave} - Z_{term.}}{Z_{wave} + Z_{term.}} \qquad (13.3)$$

When $Z_{term.} = Z_{wave}$, the ratio is, of course, zero. This means no amplitude of reflected wave.

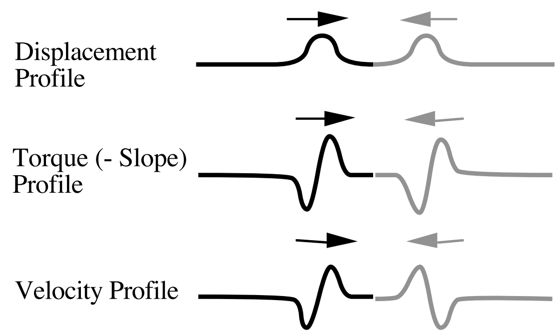


Figure 13.6a The displacement, torque (negative slope) and velocity profiles for the incident wave and the reflected wave when the end of the torsional vibration apparatus is left loose. Note that as they run together, the torques of the two waves subtract but the velocities add.

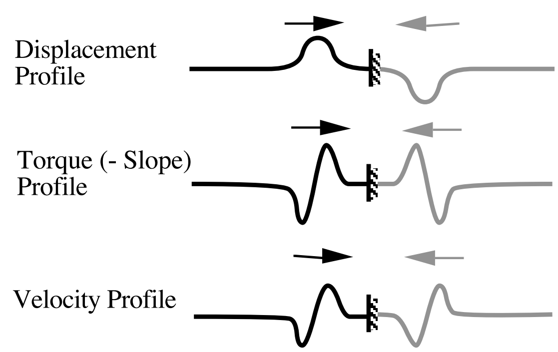


Figure 13.6b The displacement, torque (negative slope) and velocity profiles for the incident wave and the reflected wave when the end of the torsional vibration apparatus is clamped. Note that as they run together, the torques of the two waves add but the velocities subtract.

Taking the extreme case of $Z_{term.} = 0$ (open ended system) gives the ratio as 1, while taking the other extreme of $Z_{term.} = \infty$ gives a ratio of -1. Thus it is seen that the equation gives us the correct values for the three simple cases that could be analyzed without mathematics.

The energies in the incident and the reflected waves will be proportional to the squares of their amplitudes. The formula for the fraction of wave energy reflected is therefore

$$\frac{E_{ref.}}{E_{inc.}} = \left|\frac{\mathrm{Z}_{\mathrm{wave}} - Z_{term.}}{Z_{wave} + Z_{term.}}\right|^2 \qquad (13.4)$$

The formula for the ratio of reflected wave amplitude to incident wave amplitude for an impedance mismatch at the end of a wave system can be derived from considering that the ratio of the combined torques and velocities due to the two waves must be the termination impedance;

$$\frac{Torque_{inc.} + Torque_{ref.}}{Ang.\ Vel._{inc.}. + Ang.\ Vel._{ref.}.} = Z_{term.} \qquad (13.5)$$

The angular velocities in the incident and the reflected waves are related to the torques by

$$\frac{Torque_{inc.}}{Ang.\ Vel._{inc.}} = Z_{wave}\ ; \qquad (13.6)$$

$$\frac{Torque_{ref.}}{Ang.\ Vel._{ref.}} = -Z_{wave} \qquad (13.7)$$

(The negative sign for the case of the reflected wave is because the reflected wave is travelling backwards.)

Combining these equations to eliminate the angular velocities and rearranging gives

$$\frac{Torque_{ref}}{Torque_{inc.}} = \frac{1 - Z_{inc}.Z_{wave}}{1 + Z_{inc}.Z_{wave}}$$

$$= \frac{\mathrm{Z}_{\mathrm{wave}} - Z_{term.}}{Z_{wave} + Z_{term.}} \qquad (13.8)$$

Finally, since the ratio of the torques is equal to the ratio of the wave amplitudes,

$$\frac{A_{inc.}}{A_{ref.}} = \frac{Z_{wave} - Z_{term.}}{Z_{wave} + Z_{term.}} \qquad (13.9)$$

13.1.2 The Absorption of a Sound Wave

The same criteria that apply to the absorption of a wave in the torsional vibration apparatus also apply to the absorption of a sound wave in air. If, for example, the sound wave is to be absorbed by a wall then that wall must offer the same impedance (ratio of force to velocity) as that which exists in the sound wave. Taking the simple case of a square meter wall, then the force exerted by the sound wave is just the sound pressure in the wave. The ratio of this pressure to the air velocity in the wave was given in Chapter 3 and, for air at normal atmospheric pressure and 20°C, is 413 Pa-s/m. For the wall to completely absorb the sound it must therefore have an impedance to movement such that 413 N moves it at 1 m/s. If it has a higher or lower impedance, some of the sound energy will be reflected.

The ratio of pressure to velocity in a sound wave in an open medium is called the "Characteristic Acoustic Impedance" of the medium (sometimes it is called the "specific acoustic impedance") and is usually designated as r_A. It is called "characteristic" because it is based on fundamental properties of the medium. Different media have different characteristic acoustic impedances. For example, the characteristic acoustic impedance of helium at normal atmospheric pressure and 20°C is 172 Pa-s/m. From the physics of wave motion in any medium it can be shown that the characteristic acoustic impedance for a medium is given by

$$r_A = \rho c \qquad (13.5)$$

where ρ is the density of the medium and c is the wave velocity in that medium.

One of the most important cases of an impedance mismatch in sound transmission is that which occurs when sound in air is propagated into water. This is a situation which occurs in animal hearing where sound waves in air are used to excite sound waves in the cochlea so that the hair cells on the basilar membrane can produce signals to send to the brain. The cochlea is filled with a fluid which has an acoustic impedance very similar to water.

The acoustic impedance of water is about 1,500,000 Pa-s/m, about 3500 times that of air. Taking $Z_{term.}$ as the impedance of the water, the fraction of sound energy reflected at an interface between air and water becomes

$$\frac{E_{ref.}}{E_{inc.}} = \left|\frac{Z_{air} - Z_{water}}{Z_{air} + Z_{water}}\right|^2$$

$$= \frac{(1 - 3500)^2}{(1 + 3500)^2} = 0.99886 \qquad (13.10)$$

The fraction of energy which actually gets into the water is therefore only 0.00114 or about 0.1%. This corresponds to a 30 dB loss in sound level.

The intricate mechanism making up the middle ear is designed to overcome this problem. It is essentially an impedance transformation device. It is so well designed that in the range of frequencies over which the human ear is most sensitive (from about 2500 to 4500 Hz) the efficiency of transmission of sound energy to close to 100%. People with a functioning cochlea but no functioning middle ear will hear room sounds by bone conduction of sound waves through the head. Here the impedance mismatch is even greater than that for air to water resulting in about a 40 dB loss relative to that for a person with normal hearing.

As a finish to this section. consider again the problem of sound absorption in a room. The concept of impedance matching tells us that a perfect acoustic absorber is one which has the same characteristic acoustic impedance as air. One such absorber of course is more open air. An open window therefore is a perfect acoustic absorber.

However, open windows do not make practical acoustic absorbers for modern buildings. In fact, they will probably let in more sound then they let out. A more practical acoustic absorber would be one made up of wall material which has the same characteristic impedance as air. What this means is that this material must have a porosity to air flow which allows air to flow into it at a rate which is in proportion to the pressure forcing this flow. Furthermore, the proportionality constant must be such that the velocity of air flow is close to 1/413 meter per second for each Pascal of pressure causing this flow.

The amount of air flowing into the sound absorber will actually be very small. This can be seen by calculating the air motion for a 100 dB sound at 1000 Hz. A sound level of 100 dB corresponds to a sound pressure of about 2 Pa giving a air velocity p/ρA of only about 5 mm/s. The motion that would give this flow velocity at 1000 Hz is microscopic;

$$v = 2\pi f A$$

$$A = \frac{v}{2\pi f} = \frac{0.005}{2\pi \infty 1000} = 0.7 \text{ micron} \qquad (13.7)$$

A micron is about the smallest size that can be seen in the most powerful optical microscope. If a sound absorbing material has the right surface porosity to absorb these small motions, its surface can appear quite solid to the eye.

13.2 Acoustic Impedance in Standing Waves

13.2.1 The Ratio of Pressure to Velocity in a Standing Wave

In a travelling sound wave, the ratio of pressure to velocity is a constant 413 Pa-s/m for air in a normal room. However, for sound in a container such as a typical musical instrument, the relationship is not so simple. This was seen for the normal modes of vibration of air in a pipe in Chapter 10. Is the concept of acoustic impedance of any use in dealing with the phenomenon of normal modes in systems?

As an introduction to this subject, consider again the pressure and velocity patterns of the lowest mode of vibration of air in a closed pipe shown in Fig. 10.9 and repeated in Fig. 13.7 without the lines indicating the positions of the air in the pipe during the motions.

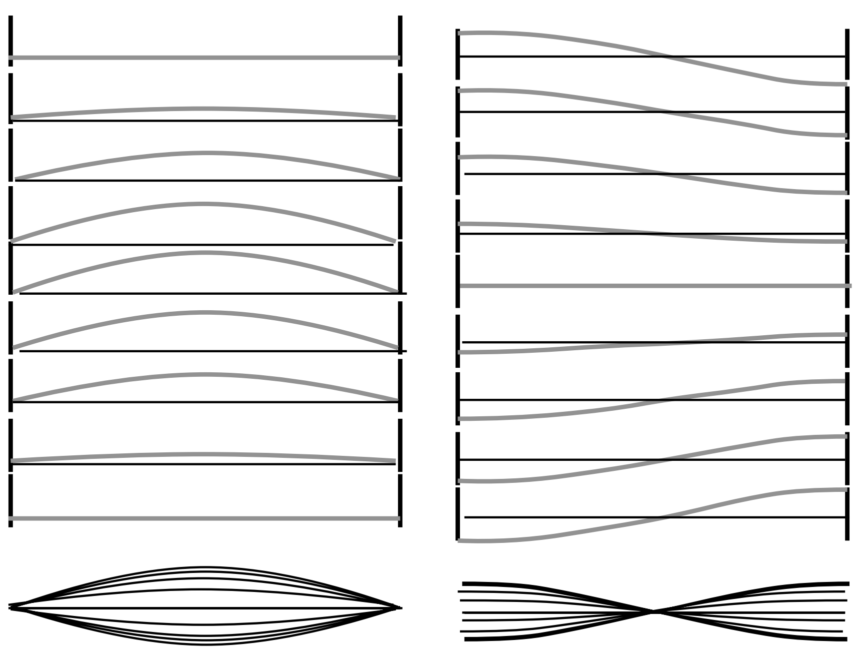


Figure 13.7 The motions involved for one half-cycle of the first standing wave mode of vibration of the air in a closed pipe. Successive diagrams downward represent successive instants in time. The diagrams are repeated side by side so that one set can be used to indicate the velocity patterns and the other to indicate the pressure patterns (both shown as shaded lines on the diagrams). The diagrams on the bottom give an overview of the velocity and pressure patterns for a complete cycle.

As in the travelling wave in open air, the pressure patterns and the velocity patterns will be related. If all the pressures are doubled, all the velocities will be doubled. Furthermore, since the standing wave can be thought of as two travelling waves moving in opposite directions, the ratio of the maximum pressure that can be achieved by the overlap of these two waves and the maximum velocity that can be achieved must still be the ratio of the pressure and the velocities in the two individual travelling waves or 413 Pa-s/m.

However, the ratio of pressure to velocity at particular points in the system is variable. At the center of the pipe it is zero, at the ends of the pipe it is infinity, and at all other points it is somewhere in between. At points 1/4 of the length of the pipe in from the ends it is actually the same as the ratio of the maximum pressure (at the ends) to the maximum velocity (at the center) or the same as that for open air. Thus for the center half of the pipe the impedance of the air is less than that of open air and for the other half it is greater.

Furthermore, the pressure and the velocity are not in phase, as they are for travelling waves. For the left half of the pipe the pressure leads the velocity by 90° while for the right half of the pipe it lags the velocity by 90°. What are the consequences of the pressure and the velocity being 90° out of phase?

Consider first a more common example of doing work by pushing on an object. If, for instance, you push on something it must move with the push for you to be doing work on it. If it moves toward you as you push on it, that object is doing work on you and the work that you are doing becomes negative.

Suppose that you are pushing on an object in an oscillatory fashion and that the object responds by oscillating with its velocity 90° out of phase to your applied force. The graphs for your force and the velocity will be as shown in the right hand side of Fig. 13.8.

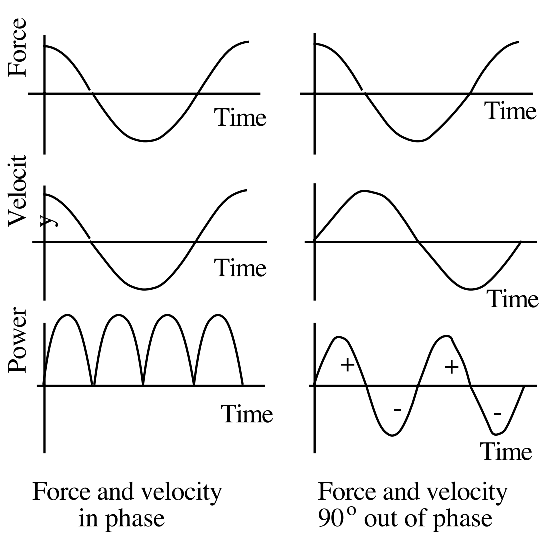


Figure 13.8 Force, velocity and power curves for force and velocity in phase and 90° out of phase. The power for in phase motion is always positive. For 90° phase the power alternates equally between being positive and negative.

If you follow the progress of the work you are doing it turns out to be alternatively positive and negative. Starting out with your force at a maximum positive manner and the velocity zero but growing, for one quarter of a cycle you are doing positive work. The object is moving in the direction you are pushing it. However, after this quarter cycle your force reverses (you pull on the object) and your work becomes negative. This is because the object is doing work on you by moving away from your pull, i.e. giving back the work you put into it).

After this quarter cycle of negative work, the velocity now reverses to be back in direction again with your pull. Your work therefore becomes positive again (you are now pulling and the object is moving towards you). After this quarter cycle, your force turns positive again while the object is still moving towards you. Your work is therefore again negative, cancelling out the work you did in the third quarter cycle. Thus in a complete cycle you do no net work.

This is, of course, the basic nature of systems with a steady oscillation. Energy is being continuously transferred from one element to another. In the case of a spring pulling on a mass, the energy goes from the spring into the mass and back, twice in each oscillation cycle.

The question then becomes, how do you get net energy into an oscillation? The answer is to have your force not at 90° phase relative to the the velocity of the motion. The most effective way for the force to transfer net energy to the oscillation in each cycle is to have the force and the velocity exactly in phase. In the left half of Fig 13.8, the force is always in phase with the velocity. Now when the force switches over from being positive to negative, the velocity switches over so that the energy input always stays positive.

Returning now to a sound wave, the velocity being in phase with the pressure is the necessary condition for the wave to be doing work on the air. We can now visualize the sound wave propagation as the overpressure of one particular region of air pushing away the adjacent air with a velocity in phase with the overpressure and thereby doing work on that adjacent air. This adjacent air in turn does work on the air adjacent to it and so the wave propagates.

This gives a new perspective on the direction of propagation of a wave. If we take a given direction as being a forward direction for velocities (say to the right) then for waves travelling in that direction the pressure and the velocities will be in phase. This corresponds to energy being propagated to the right. If however, the pressures and the velocities are 180^0 out of phase, then the work done in the "forward" direction will be negative. The wave will be delivering work from the right, not transferring it to the right. Such a wave would actually be moving from the right to the left.

Thus the phase relationship between pressure and air velocity in a travelling wave determines the direction of propagation of its energy. In phase means propagation in the direction for which air velocity is defined as being positive. Out of phase means propagation in the opposite direction.

Phases relationships halfway in between these two extremes, i.e. at ± 90°, must mean no propagation in either direction. This is, of course, the condition of a standing wave.

Thus there is a definite relationship between the pressures and velocities in a standing wave, the principle one being that they are always 90° out of phase with each other. To express the ratio of pressure to velocity as an impedance in such a situation the special term of "Reactance" is used.

13.2.2 Acoustic Reactance

From the above it is concluded that in a standing wave the pressure and velocity are at 90° phase relative to each other and so no net energy is propagated in an oscillation cycle. Yet there is a definite ratio between pressure and air velocity at any point in a standing wave; the amplitude of the pressure oscillation is still proportional to the amplitude of the velocity oscillation. To distinguish this ratio from that of the ratio when the pressure and velocity are in phase, the term "reactance" is used; the term "resistance" being used when the pressure and velocity are in phase. The reason for the term reactance is that a system in which the force and velocity are at 90° phase relative to each other absorbs no net energy in a cycle, giving back any energy it receives. it "reacts".

Unlike resistance, reactance can be positive or negative depending on the relative phase of the pressure and velocity oscillations. The convention used for the sign of a reactance in general is that if the force on a system leads the velocity by 90°, the reactance is said to be positive; if the force lags the velocity by 90°, the reactance is said to be negative. Simple common examples are an oscillating mass and an oscillating spring. When one oscillates a mass, an applied force generates an acceleration which takes some time to generate a velocity. The force therefore leads the velocity. The reactance of a mass is therefore positive. On the other hand, a spring will move initially with no force. The force only develops after there has been velocity for some time and the spring becomes compressed. The force therefore lags the velocity and the reactance of a spring is negative.

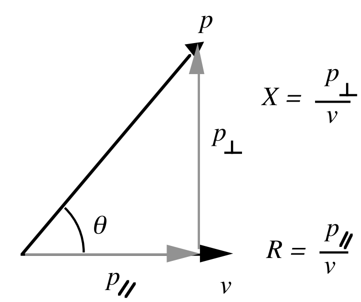
The acoustic impedance of the air in the lowest standing wave mode in a closed pipe is therefore positive reactance for the left-hand half of the pipe and negative reactance for the right-hand half (for positive directions to the right). What this implies is that, looking to the right from the left half of the pipe, the reactance of the air in the system is due to the mass of the air that is being moved in the standing wave. When looking to the right from the right hand half, the reactance is like that of a spring; the pressure is due to the air piling up against the walls of the pipe and springing back. Looking to the left, of course, the reactances are reversed.

13.2.3 The General Concept of Acoustic Impedance

Two categories of sound impedances have been introduced; resistive and reactive. Resistive impedances are when pressure and velocity are in phase and represent propagation or dissipation of energy. Reactive impedances are when pressure and velocity are 90° out of phase with each other and represent storage and return of energy to a driving element.

In general, there can be any phase relationship between acoustic pressure in air and the air velocity associated with it; nor just zero and plus and minus 90°. The relationship between pressure and velocity will in general be expressed by a phasor diagram with an arbitrary angle θ between the two phasors.

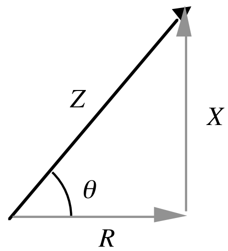
For such a situation, the pressure phasor can be thought of as being made up of two orthogonal components, one along the velocity phasor and one perpendicular to it. This is shown in fig. 13.9. When the component of the pressure phasor parallel to the velocity phasor is divided by the velocity phasor, one gets the resistive part of the impedance of the air. When the pressure phasor component perpendicular to the velocity phasor is divided by the velocity phasor, one gets the reactance of the air.



$$R = \frac{p}{v}\cos\theta \quad X = \frac{p}{v}\sin\theta$$

Figure 13.9 The relationship between the pressure and velocity phasors in a general acoustic impedance. The resistance *R* and the reactance *X* are defined as shown.

Thus the ratio of pressure to air velocity at any point in air can be thought of as being made up of a combination of resistance and reactance. Such a combination is called a "complex impedance" and can be represented as in fig. 13.10 where *R* represents the resistive impedance and *X* the reactive impedance. Here *Z* represents the total impedance and, in acoustics, is the ratio of the amplitude of the pressure oscillation to the amplitude of the velocity oscillation. The angle θ between *Z* and *R* is the phase angle between the pressure oscillations and the velocity oscillation in the air.



$$R = Z \cos\theta$$
$$X = Z \sin\theta$$
$$Z = \sqrt{R^2 + Z^2}$$

Figure 13.10 The relationship of resistance and reactance to complex impedance.

Thus, if we know the resistance and the reactance of a point in air, we can calculate the ratio of the pressure oscillation amplitude to the velocity oscillation amplitude by calculating Z from R and X and we can calculate the phase angle between the pressure and the velocity by evaluating the inverse tangent of X/R. (Note for example, that if X is negative then the phase angle will be negative, indicating that the velocity oscillation leads the pressure oscillation in phase.)

13.3 Acoustical Impedance and Acoustical Power.

13.3.1 Acoustic Impedance vs Characteristic Acoustic Impedance

Up to now in these notes, the concept of acoustic impedance has been used to discuss the ratio of pressure to air velocity in air at a point in a sound system. However, this is not what is normally meant by the term "acoustic impedance", a concept that was invented to deal with acoustic power in systems. The ratio of pressure to air velocity is, strictly speaking the "characteristic" or "specific" acoustic impedance and is of not much use in problems dealing with acoustic power.

The usefulness of the concept of impedance in power problems lies in the connections between force, velocity and power in mechanics or voltage, current and power in electricity. These are simple stated.

In mechanics

$$Power = Force \times velocity \qquad (13.8)$$

In electricity

$$Power = Voltage \times current \qquad (13.9)$$

If there is a proportional relationship between force and velocity (such as actually occurs in many mechanical systems) or between voltage and current (as occurs in many electrical conductors), then the power equations can be rewritten as below.

$$R_{mech.} = \frac{Force}{velocity} \;; \qquad Power = R_{mech.} \times velocity^2$$

giving the usual form of the equation for mechanical power;

$$P = R_m \, v^2 \qquad (13.10)$$

$$\mathrm{R}_{\mathrm{elec.}} = \frac{Voltage}{current} \;; \qquad Power = R_{elec.} \times current^2$$

giving the usual form of the equation for electricity;

$$P = R\, I^2 \qquad (13.11)$$

The reason that the characteristic acoustic impedance is not much use in power problems is that the product of pressure time velocity is not power but intensity;

$$Pressure \times velocity = \frac{Force}{area} \times velocity$$
$$= \frac{Force \infty velocity}{area} = \frac{Power}{area}$$
$$= Intensity \qquad (13.12)$$

To get an equation connecting acoustic power to acoustic pressure we need an acoustic "current";

$$Pressure \times acoustic\ current = Power \qquad (13.13)$$

It is easily seen that we can get an acoustic current from the air velocity by multiplying it by the area over which the air has this velocity;

$$Power = Intensity \times area =$$
$$= Pressure \times velocity \times area$$
$$= p\, U \qquad (13.14)$$

where p is the pressure and U is the acoustic current defined as velocity times area.

The acoustic current U is simply the volume flow rate (in cubic meters per second) of the air in the sound wave. If now there is a ratio of p/U which we can call the acoustic resistance R_A, then the acoustic power becomes

$$Acoustic\ Power = R_A\, U^2 \qquad (13.15)$$

The unit for acoustic resistance is called the "acoustic ohm" (in analogy with the electrical ohm which is the ratio of voltage to current).

As an example, consider a sound wave travelling along a tube of 1 cm^2 cross-section (about 11 mm inside diameter). The pressure to velocity ratio will still be as for the wave in open air but the volume flow rate will be $v \times 10^{-4}$ m^3. The acoustic resistance will therefore be

$$R_A = \frac{p}{v \infty 10^{-4}} = \frac{r_A}{10^{-4}}$$

$$= 4{,}130{,}000 \text{ ohm} \qquad (13.16)$$

The acoustic resistance for a pipe of 10 cm^2 cross-section (about 36 mm inside diameter) would be only 413,000 ohm, leading to the reasonable result that the larger diameter pipe has less acoustic resistance than the smaller diameter tube.

In should be noted that the unit used here for acoustic impedance is the modern SI acoustic ohm which has the units Pa-s/m^3. Many textbooks, and some publications, particularly American, still use the older "centimeter-gram-second" or cgs acoustic ohm which has the units dyne-s/m^5. The connection between this unit and the newer one is that a device with an acoustic impedance of one cgs acoustic ohm will have a resistance of 10^5 Si acoustic ohms (i.e. the cgs acoustic ohm is much bigger than the SI acoustic ohm). One reason that many workers prefer the older cgs unit is that the acoustic impedance of many devices of importance (such as the human ear) gives very large numbers in SI units. For example, the human ear at 1000 Hz has an acoustic impedance of about 400 cgs ohms but an impedance of 40,000,000 SI ohms. However, when dealing with engineered devices such as loudspeakers, the SI unit is much more practical since it gives answers for power which are in the familiar unit of watts whereas the cgs would give answers in units of erg per second.

The concept of complex impedances involving both resistances and reactances introduced for the specific acoustic impedance can now be transferred to true "acoustic impedance" in acoustic ohms. As with specific acoustic impedance, there can be, in general, any phase relationship between acoustic pressure and the volume flow associated with it in a sound system. The relationship between pressure and flow will again be expressed by a phasor diagram with an arbitrary angle θ between the two phasors (fig. 13.11). The only difference from the case for characteristic acoustic impedance is that the two orthogonal components of pressure are now along and perpendicular to the flow phasor U instead of the velocity phasor v.

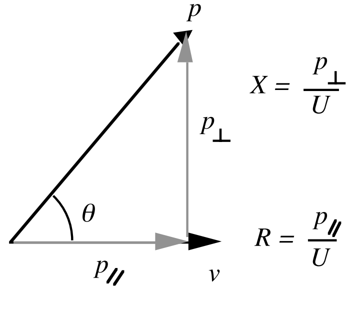


Figure 13.11 The relationship between the pressure and air flow phasors in a general acoustic impedance. The resistance R and the reactance X are defined as shown.

Again, the ratio of pressure to air flow in any acoustic system can now be thought of as being made up of a combination of resistance and reactance. As with the characteristic acoustic impedance, this complex impedance can be represented as in fig. 13.10 where, again, R represents the resistive impedance, X the reactive impedance and Z the total impedance, which is now the ratio of the amplitude of the pressure oscillation to the amplitude of the flow oscillation. Now the angle θ between Z and R in fig. 13.10 is the phase angle between the pressure oscillation and the flow oscillation.

Repeating, the difference between this acoustic impedance and the previous "characteristic" or "specific" acoustic impedance is that the true acoustic impedance deals with the ratio of the sound pressure on a system to the overall flow of air through that system and hence with overall power while the specific acoustic impedance deals with the ratio of pressure to air velocity at a specific point in a system and is related to sound intensity at that point.

Acoustic impedances are used in acoustic power problems of systems as follows. The power is, as shown above, the product of the in-phase component of the pressure and the air flow rate;

$$\textit{Acoustic Power} = p\, U \cos\theta \qquad (13.17)$$

This can be simply related to the resistive part of an impedance and the air flow rate;

$$P = U^2 \frac{p}{U} \cos\theta = U^2 Z \cos\theta$$

$$= U^2 R_A \qquad (13.18)$$

One way of looking at this is by considering the resistance and the reactance of a system to be in series (fig. 13.12).

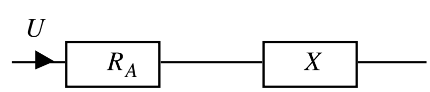


Figure 13.12 The electrical equivalent of a complex acoustic impedance. The resistance R and the reactance X are regarded as being "in series", meaning that the air flow U is regarded as flowing through one component of the complex impedance and then the other.

In this picture the same U (the acoustic "current") flows through the resistance and the reactance. The flow through the reactance involves no net work in a cycle and all the acoustic power is therefore dissipated in the resistance.

A simple example of the use of the complex acoustic impedance of a system is the Helmholtz resonator of Chapter 10. Here the acoustic reactance has two parts; that due to the inertance M and that due to the acoustic capacitance C_A. By analogy with an inductor where the electrical reactance is ωL, the acoustic reactance of an inertance is

$$X = \omega M. \tag{13.19}$$

Also in analogy with the electrical reactance of a capacitance, the acoustic reactance of an acoustic capacitance is

$$X = -\frac{1}{\omega C_A}. \tag{13.20}$$

At resonance the reactance of the inertance and the reactance of the capacitance balance. Therefore at resonance;

$$\omega M = \frac{1}{\omega C_A} \quad ; \quad \omega = \frac{1}{\sqrt{MC_A}} \tag{13.21}$$

13.3.2 The Acoustic Power of a Vibrating Disk; An Example of the use of Acoustic Impedance

As an illustration of the general technique of applying the acoustic impedance concept to acoustic power problems, consider the problem of the sound power radiated by a circular disk moving in and out of an infinite flat baffle (fig. 13.13). This is one of the most important general problems in acoustics. It is easy to see the relevance of this in loudspeaker design but the problem also has relevance to the leakage of sound through orifices and the radiation patterns of musical instruments such as trumpets and horns.

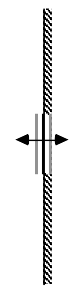


Figure 13.13 Schematic of the vibrating disk in an infinite baffle.

The problem is of such general interest that graphs are available which allow one to estimate the acoustic impedance to any diameter disk oscillating at any frequency in any acoustic medium. This is done on one graph by scaling the frequencies and the disk dimensions together and expressing the answer in terms of the characteristic impedance of the medium involved (fig 13.14).

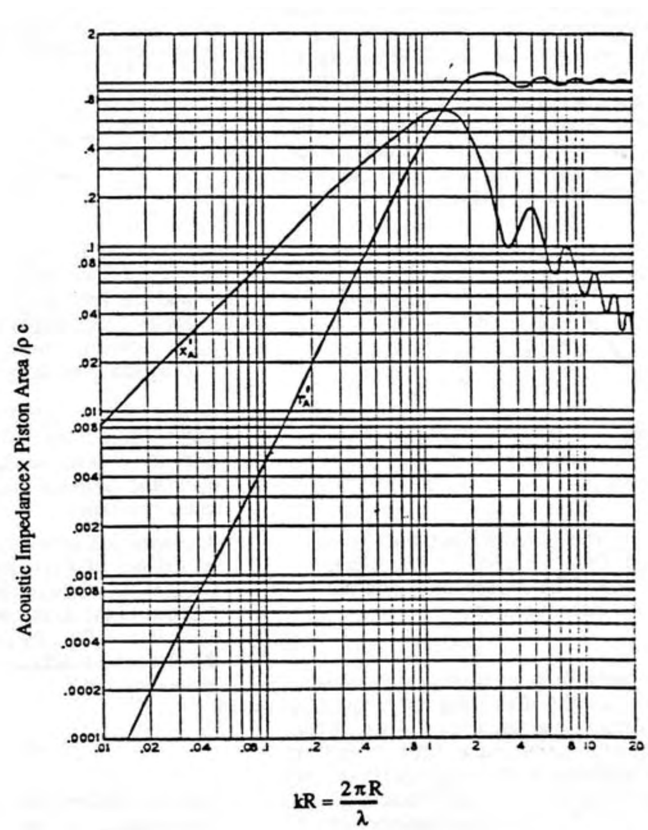


Figure 13.14 The acoustic resistance r_A' and the acoustic reactance x_A' in normalized units of a piston of radius R set in an infinite plane baffle. (The normalized units are explained in the text.)

The normalized units are obtained by dividing the actual acoustic impedance by the specific acoustic impedance of the medium and multiplying by the area of the piston. The scales on the graph are normalized so that the graph is good for all frequencies of sound and all media. The horizontal coordinate is the wave number of the sound multiplied by the piston radius which is equivalent to the piston circumference divided by the wavelength of the sound. The vertical scale is acoustic ohms divided by the density of the medium and the velocity of sound in the medium for a unit area of the piston.

Thus to get the actual acoustic impedance from this graph one must multiply the coordinate by the velocity of sound and the density of the medium and divide by the area of the piston. The horizontal scale is ratio of the circumference of the piston to the wavelength of the sound. (At normal room temperature and pressure (20^{o}C and 100 kPa), the density of air is about 1.2 kg/m^3 and the velocity of sound in air is 340 m/sec.)

As a specific example showing the use of the graph, consider the problem of the acoustic impedance that air presents to a disk of 20 cm diameter (such as would approximate a loudspeaker cone) moving back and forth by a total extension of 1 cm (an amplitude of vibration of .5 cm) at a frequency of 100 Hz (fig. 13.15).

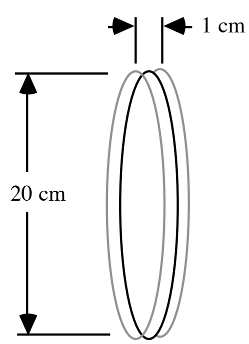


Figure 13.15 A schematic of a loudspeaker emitting a low frequency tone.

The first number that is needed is the coordinate on the horizontal axis of the graph. This is in units of kR where k is the "wave number" or $2\pi/\lambda$ for the sound to be produced. For 100 Hz, λ is 3.4 m and the horizontal coordinate on the graph therefore becomes $2\pi \times 0.1/3.4$ for $R = 0.1$ m. This is 0.18 or 0.2 to an accuracy good enough for an acoustic calculation.

The vertical coordinate of the graph for r_A' at this point is about 0.02. This means that the acoustic impedance the air presents to the motion of this disk is only 0.02 to that which air would normally present to a plane wave passing through it.

To evaluate the actual acoustic impedance that the disk sees we therefore have to calculate the acoustic impedance presented by the same area of air to a plane wave. This would be $413/\pi R^2$ or $413/\pi \times 0.01$ giving a result of 13,200 ohms. The actual acoustic impedance to the vibration of the disk is therefore 13200×0.02 or 264 ohms.

To use this to calculate the acoustic power radiated into the air by the disk, we need to know one other thing; either the pressure seen by the disk or the air flow created by the disk. We do not know the pressure seen by the disk since the acoustic resistance calculation tells us that the relationship between velocity and pressure is not the same as for a plane wave in air. However, we can say that the moving disk must move the air directly in contact with it at a velocity equal to that of the disk itself. This gives a flow rate equal to the disk velocity times the disk area.

The disk velocity can be calculated by noting the amplitude and frequency of the motion. The peak value of the velocity will be given by

$$v = \omega A = 2\pi f A = \pi \text{ m}^3/\text{s} \qquad (13.22)$$

The peak value of the volume flow rate U will be

$$U = v \times area = \pi \times \pi R^2$$
$$= 0.01\,\pi^2 \approx 0.1 \text{ m}^3/\text{s} \qquad (13.23)$$

This allows the use of the equation;

$$Acoustic\ Power = U^2 R_A \qquad (13.24)$$

Putting in the values for U and R gives the acoustic power:

$$Peak\ Acoustic\ power = 0.1^2 \times 260$$
$$= 2.6 \text{ watts.} \qquad (13.25)$$

This is the peak acoustic power, which occurs when the acoustic current is at a peak value. The average acoustic power, which is the quantity that we are normally concerned with, can be shown to be just one half the peak power for a pure single frequency of oscillation such as we have assumed here (see appendix). The average sound power radiated from the disk will therefore be 1.3 watts.

This would perhaps not seem to be a very great power. However, calculate what such a sound source would do in an average room. The equation for the intensity of a sound source in a room is, from the considerations of room reverberation;

$$I = \frac{TNc}{13.8V} \qquad (13.26)$$

where T is the room reverberation time, N is the sound power of a source in watts, c is the velocity of sound and V is the room volume. This gives, for a room which would have a reverberation time of about 1 second for this frequency and a volume of 10 × 8 × 3 or 240 m^3, a sound intensity level of ;

$$I = 0.13 \text{ Watt/m}^2 = 111 \text{ dB.} \qquad (13.27)$$

This would be a thunderous sound! It would make you thankful that the acoustic impedance seen by the disk is lowered by a factor of 50 for otherwise the sound level would be 50 times higher or about 130 dB.

However, a careful look at the graph tells you that the acoustic impedance seen by the disk can approach that of a plane wave in open air at higher frequencies. Specifically, consider the point for $kR = 2$, which corresponds to a frequency 10 times greater or 1000 Hz. Here the acoustic impedance seen by the disk would be 13,200 ohms. The same flow rate through this impedance would of course now take less amplitude at this frequency (the velocity of a vibration is proportional to the frequency of the vibration). The amplitude of the motion of the disk would now have to be only 0.5 mm for a total excursion of only 1 mm.

What this is telling us is that a 20 cm disk vibrating back and forth a total extent of only 1 mm at 1000 Hz in an infinite baffle will radiate 62.5 watts of sound into a room, causing a sound level in a room such as a normal lecture room of about 130 dB.

That sound level is well beyond the threshold of real pain and would cause almost instant physical damage to the ears. Why then do people buy 100 watt per channel stereo sound systems?

First of all, they are not buying 100 watt per channel of sound power into the room. For reasons of avoiding mechanical resonances in the speaker itself, the typical modern loudspeaker system is deliberately designed to have an efficiency of seldom greater than 1 %. Thus 100 watt per channel sound systems deliver only about 1 watt per channel of sound power into a room.

Still, that leaves about 110 dB of sound intensity with only one channel operating (or about 113 dB with two channels operating) for a moderately sized lecture room. For a room such as a typical living room of volume 75 m^3, even an acoustically "dead" one with a reverberation time of only 0.5 seconds, the sound level with both channels operating would be 118 dB.

100 Watt per channel stereo systems would therefore seem a wasteful luxury. Yet many people buy them with good justification. The reason for this cannot be found in the sustained sound levels from a record but rather in the peak sound powers during transients. In fact, systems rated at 100 watt per channel do not deliver 100 watts of average power per channel. Rather the rating applies to the peak power the system is capable of delivering.

Because of the nature of power averaging, the peak power in music with sharp transients can reach as high as 100 times the average power. (See appendix on average power vs peak power.) To properly reproduce the transient sound, this peak power must be delivered.

This means that in musical sounds which have an average intensity of 95 dB, which is about the maximum that would ever be desired by a normal listener, there can be transients corresponding to 115 dB of intensity. These transients need not be perceived as a loud noise. If they are very short in duration, they are perceived as only a sharp click and may not even show up on a VU meter or a sound level meter used to measure sound levels in a room. 100 watt per channel stereo systems are perfectly justified when one wants faithful reproduction of such transients.

Returning to the acoustic impedance presented by air to a vibrating disk, for low values of frequency (kR much less than 1 or the wavelength much greater than the circumference of the disk) the reactance is much greater than the resistance. What this means is that the pressure exerted by the disk is much greater than just that to move the air through the resistive part of the impedance and thereby create sound in the room. However, due to the large reactive part most of this pressure is out of phase with the disk velocity. It is, in fact leading the velocity. At kR = 0.1 the pressure will be leading the velocity by about 86^{o}.

A pressure (or any force) leading a velocity indicates that a mass in being accelerated. At low frequencies, a loudspeaker cone therefore is putting most of its force into accelerating and decelerating the air around it rather than radiating power out into the room. A loudspeaker at low frequencies is therefore very inefficient in coupling its motion to the air in the room to produce sound.

On the other hand a vibrating disk at high frequencies (kR >1 or the wavelength less than the circumference) becomes very effectively coupled to the air; the impedance the air presents is that of a plane wave.

What this means is that , in fact, a plane wave will radiate away from the disk. The sound from the disk is therefore beaming forward; a highly undesirable feature in a loudspeaker. For this reason loudspeaker cones are never made of rigid flat disks that vibrate uniformly over their whole surface at high frequencies but are made of flexible material so that at high frequencies, only the center region takes part in the vibration. Loudspeaker cones are therefore deliberately made to be relatively ineffective sound power radiators at all frequencies.

13.3.3 Acoustic Impedance in a Trumpet

Another use of the acoustic impedance presented by air to a vibrating disk relates to instruments with bells such as a trumpet. As previously mentioned in Chapter 11, the reflection of an input sound of the mouthpiece from the bell back to the mouthpiece is a very important feedback mechanism for allowing the player to hold a note and to build up the sound intensity of the note.

The wave of sound produced inside a trumpet at the mouthpiece propagates along the tube of the trumpet very much as a simple plane wave in a small tube. The impedance of the air in this tube is very high because the tube is very small. For example, for a 1 cm diameter tube the acoustic impedance would be $414/\pi R^2$ where $R =$ 0.005 M. This gives about 5.3 million acoustic ohms.

At the mouth of the trumpet, the tube expands. If this expansion is gradual enough, the sound wave remains almost planar and the acoustical impedance drops slowly. If this is done at a slow enough rate, there will be no reflection from the impedance change. (This is the principle of the horn which makes it a very effective radiator of sound.)

If the trumpet has a bell of say 15 cm diameter, the impedance of the sound wave when it reached the bell would be down by the ratio of the area of the beginning of the tube to the area presented by the bell. This would be a factor of 225 which is equal to the square of the ratio of the radii. The acoustic impedance would then be about 25,000 ohms.

However,at the bell there is a sudden change. Now the sound wave is presented with the impedance seen by a vibrating disk as discussed in the previous lecture. The impedance of this disk will depend on the frequency of the sound that has arrived at the bell. If this frequency is high enough, then we are on the part of the impedance graph where the impedance is the same as that for a plane wave in open air. There is no impedance change as the sound leaves the bell and so no reflection. The horn becomes a very efficient radiator of a plane wave going straight forward. This is again why horns are very good radiators of high frequency sounds in a forward direction.

However, to develop a note this is no help at all. The player needs the reflection for feedback to build up a resonance. Thus, notes with a high frequency fundamental are practically impossible. Once a resonance has been built up with the players lips, the sound produced may have many high frequency harmonics which are very efficiently radiated by the bell but the fundamental upon which the sound is based must not have a good impedance match at the bell. (A good impedance match means no reflection back from the bell.)

The graph of fig. 13.14 can be used to valuate the frequencies involved. For a R of 7.5 cm, and a kR of 2 where, from the graph, it can be seen that the impedance match with open air is perfect, k = 2/0.075 or 26.67. This gives a wavelength of $2\pi/k$ or 23.6 cm. The frequency for this wavelength is 340/0.23 or about 1500 Hz.

Thus, it would be practically impossible to play a note with a fundamental at 1500 Hz on this trumpet.

This is of course faster than anyone could vibrate their lips anyway. However, let us see how the situation changes as we go down in frequency. At 750 Hz, the value of kR would be about 1 for an acoustic impedance, according to the graph, of about 0.4 that for open air. From the equation for the fraction of sound energy reflected at an impedance mismatch;

$$\frac{E_{ref.}}{E_{inc.}} = \left|\frac{Z_{wave} - Z_{term.}}{Z_{wave} + Z_{term.}}\right|^2 \qquad (13.29)$$

the fraction of sound reflected is

$$\frac{E_{ref.}}{E_{inc.}} = \left(\frac{1 - 0.4}{1 + 0.4}\right)^2 = 0.18 \qquad (13.30)$$

Thus, at this frequency we would expect about 18% of the sound arriving at the bell to be reflected back to the players lips.

Going down in frequency to half again at 375 Hz, we get a value of kR of about 0.5 and an acoustic impedance of 0.12 of that for a plane wave. The fraction of sound energy now reflected back to the players lips from the bell is;

$$\frac{E_{ref.}}{E_{inc.}} = \left(\frac{1 - 0.12}{1 + 0.12}\right)^2$$
$$= 0.62 = 62\% \qquad (13.31)$$

Thus, playing the lower notes on a trumpet is much easier than playing the high notes.

There is a great deal of technique involved in getting high notes on a trumpet, in addition to having a "good lip". One of the techniques involved is directing the pulses of air from the lips towards the side of the mouthpiece. It seems that this helps to build up the resonance in the mouthpiece itself, thereby relieving some of the demands made on the reflection from the bell. In fact if one looks at the dimensions of the standard trumpet mouthpiece, and estimates the Helmholtz resonance frequency, one get values in the range of those near the top of the trumpet range.

13.3.4 The Q of a Helmholtz Resonator

As a final example of an application of the concepts of acoustic impedance, consider the question of the Q of a Helmholtz resonator. Here it is necessary to introduce yet another equivalent definition of the Q of an oscillator; that of the ratio of the reactance to resistance in the oscillator

$$Q = \frac{X}{R} \qquad (13.32)$$

Here the reactance is that of the inertive part of the oscillator (the mass in a mechanical system, the inductance in an electrical system or the inertance in an acoustic system). For the Helmholtz oscillator of Chapter 10 (frequency = 214 Hz) this becomes

$$M = 236 \text{ Pa-s/m}^3 \; ;$$
$$\omega M = 2\pi \times 214 \times 236$$
$$= 317 \text{ k Ohm} \qquad (13.33)$$

For the resistive part it can be noted that the air moving back and forth at the opening of the neck of the bottle is equivalent to the motion that would be created by a piston. For a diameter of 16 mm and a frequency of 220 Hz, kR on the impedance diagram for a disk becomes

$$kR = \frac{2\pi \text{ x } 220 \text{ x } 0.008}{340} = 0.033$$

r_A' for this kR is about 0.0005. R_A for 16 mm dia of open air, is

$$\frac{413}{\pi \text{ x } 0.008^2} = 2 \text{ x } 10^6 \text{ Ohm.}$$

R_A for the neck of the bottle is therefore about 10^3 Ohm. The theoretical Q for the bottle as a Helmholtz resonator is therefore

$$Q = \frac{317000}{1000} = 317$$

This is higher than what was obtained experimentally in Chapter 11. There must therefore be other energy dissipating factors then just the radiating sound wave which reduce the Q of the system. One of these would be the viscosity of the air in the neck of the bottle.

13.4 Analysis of Systems Using Acoustic Impedance

The concept of impedance is a very powerful tool for analyzing oscillating systems. This is the principle reason it is so important in electrical engineering and why it has been brought into the subject of acoustics.

In general systems can be broken down into three parts; a source, a reacting system and a receiver. The performance of this overall system is then analyzed in terms of the individual impedances of the source, the system and the receiver (fig. 13.16) and the nature of any signal put into the source.

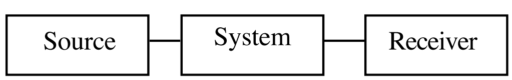


Figure 13.16 A schematic of a complete system.

From Chapter 9 the nature of any input from the source can be described by the amplitudes and phases of the Fourier components of its input. By knowing the impedances of the system for these components, the overall performance of the system can be understood.

What is needed, then, are the impedances of the source, system and receiver for the various frequencies. This is often expressed in a graph which gives the locus of the head of the impedance phasor for the various frequencies. Typical graphs for a system are shown in fig. 13.17.

The diagram at the top is fairly standard; the impedance always has positive resistance and positive reactance. In electrical systems this would mean that at all frequencies the system is resistive and inductive. In acoustic systems it means that the system is resistive and inertive (primarily reacting through its mass). The diagram beneath is perhaps puzzling. The reactance going negative at the low frequencies (below about 1400 Hz) is easy to understand; it means that at low frequencies, the capacitance of the system is providing the principal reactance. Above 1400 Hz, the inertance (or inductance) takes over. However, what about the resistive component going negative at about 4000 Hz?

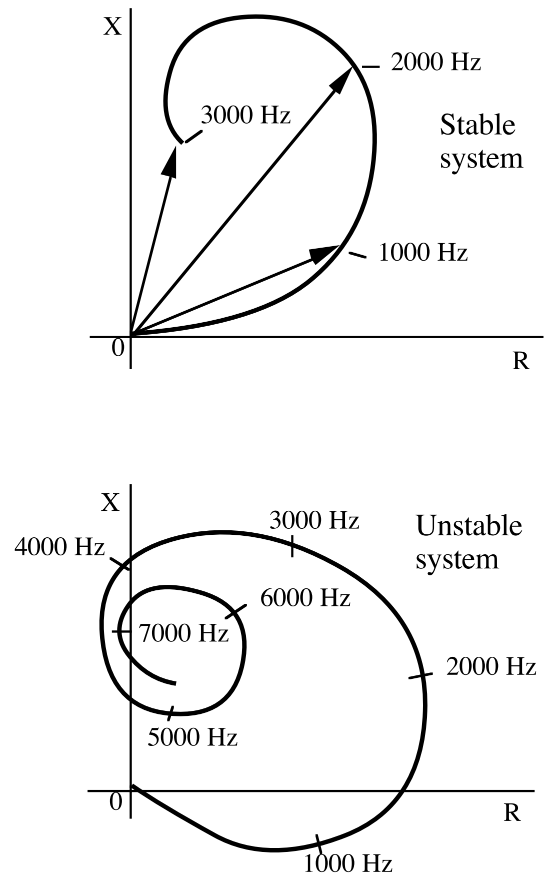


Fig 13.17 Typical graphs for the impedance of two types of systems at various frequencies. The heavy curve is the locus of the head of the phasors representing the impedance for the various frequencies. Frequency then becomes a parameter along the length of this locus.

A negative resistance corresponds to the pressure across a device being in phase with the velocity in the sense that when a velocity occurs, a pressure occurs that will drive that velocity. (The pressure being in phase with the velocity up to now has been considered to be an externally applied pressure; the pressure being generated by the system being opposite to this and thereby generating "resistance".

A negative resistance therefore means that the system aids flow through it. This means that energy is not dissipated but enhanced. This, of course, is the consequence of positive feedback in a system.

The diagram to the right is therefore that of a typical fed-back system. It can be seen that, given the right input source impedance to match the system impedance, the system will oscillate at a frequency somewhere in the negative resistance side of the diagram.

The graph for the impedance of an acoustic system will generally vary with the point in the system chosen for the input. Thus connecting sources to different points in the system can give different feed-back oscillation frequencies.

Exercises and Discussion Topics

1. Describe the motion of air as a sound wave with a pure tone is passing through it. What is the relative phase of the pressure oscillation and the velocity oscillation? What is the relationship between the pressure and velocity amplitudes? What changes between pressure and velocity when a wave goes in the opposite direction? Distinguish clearly between the actual air velocity and the wave velocity of the sound.

2. How does pressure and velocity in sound wave oscillations lead to the transmission (or absorption) of energy by a sound wave? What is the connection between intensity, pressure and air velocity in a sound wave?

3. What is meant by the term "characteristic acoustic impedance" of air? Is it resistive or reactive? Why is it a useful concept? What are the connections between intensity, pressure, velocity and characteristic acoustic impedance?

4. The density of hydrogen is 0.09 kg/m^3 and the velocity of sound in hydrogen is 1270 m/s. How does the acoustic impedance of hydrogen compare with that of air?

5. The density of water is 1 kg/liter and the velocity of sound in water is 1500 m/s. How does the acoustic impedance of water compare with that of air? Discuss the problem of getting sound waves in air to vibrate the fluid surrounding the basilar membrane (so that you can hear the sound) using these figures.

6. a)What does a hard reflecting surface do to the pressure and the air velocity in a sound pulse when it bounces off that surface? What is happening at the instant the sound pulse is being reflected?

   b)What does a opening in the confining walls of a tube do to the pressure and air velocities in a sound pulse arriving at that opening? What is happening at the instant the sound pulse is at the opening?

7. What is the relative phase between the pressure oscillations and the velocity oscillations in air near a hard sound reflecting surface? What is the relative phase near an opening in a cavity? Justify your answers by short statements. How else does the relationship between pressure and velocity oscillations change in going between these two extremes?

8. Using the figures of problem 5, what would be the percentage of sound energy reflected from perpendicular incidence of a sound wave in air onto water?

9. The density of steel is 7900 kg/m^3 and the velocity of sound in steel is 5000 m/s. What percentage of the sound energy is reflected from perpendicular incidence of a sound wave from air onto steel?

10. What is meant by the term "acoustic impedance"? What is the difference between resistive acoustic impedance and reactive acoustic impedance (i.e. acoustic reactance)? What are the two kinds of acoustic reactance? Why is sound intensity related to "characteristic acoustic impedance" while sound power is related to simply "acoustic impedance"? What part of the acoustic impedance determines the sound power?

11. a)Sketch a graph of the acoustic impedance as seen by a source at various points in a horizontal completely closed tube when the source is driving sound waves to the right at the fundamental resonant frequency of the tube.

    b)Repeat part a) for a horizontal tube which is open at both ends.

12. A child's eardrum is measured to have an acoustic impedance of 45 MegOhm with an impedance phase factor of 30 degrees. What is the acoustic resistance and the acoustic reactance? Is the eardrum behaving as a mass to be moved or as a pierced opening?

13. Where would you place a high impedance sound source in a tube with an open end so that it will excite all normal modes of oscillation of air in the tube? Explain in one sentence why. Where would you place the source to eliminate the second mode above the fundamental? What other modes would you eliminate by placing the source at this point?

14. Describe the general features of the acoustic impedance of the air surrounding a vibrating flat surface as the vibration frequency increases from very low to very high. Why would it be predominantly reactive at low frequency and predominantly resistive at high frequency?

15. a) Given the acoustical impedance chart of fig. 13.14, what will be the acoustic impedance a loudspeaker made up of a flat plate 25 cm in diameter vibrating at 60 Hz? What would it be at 600 Hz?

    b) What would be the acoustic power radiated if the amplitude of the speaker movement was 1 cm (1 cm to both sides of it's equilibrium position) at 60 Hz?

    c) What would be the amplitude of the speaker movement for the same air flow rate at 600 Hz? What would be the sound power radiated by the speaker at this frequency and this amplitude of movement?

    d) What would be the decibel level of reverberant sound in a room of 10 meters by 6 meters by 2.5 meters with a reverberation time of 1.2 seconds for these two frequencies?

16. Why are very high peak power levels sometimes necessary in sound systems in order to satisfactorily handle transients?

Answers

4) 0.277 times; 5) 3600 times; 8) 99.89% 9) 99.996% 12) Resistance = 39 Mohm, Reactance = 22.5 Mohm 13) 250 15) a) 840, phase angle 86°, 8060, phase angle 47° b) 2 Watt peak, 1 Watt average c) 1 mm, 200 Watt peak, 100 Watt average d) 113 dB, 133 dB

APPENDIX

PEAK POWER VERSUS AVERAGE POWER

The power involved in any electrical or acoustic device is always given by the equations;

$$Power = p\,U \quad \text{(Acoustical)}$$

or

$$Power = V\,I \quad \text{(Electrical)} \qquad \text{(A13.1)}$$

This power is constantly changing as the pressure and air flow change in a sound or as the voltage and the current changes in an electrical circuit. The perceived intensity of a sound is related to the average intensity of the sound or indirectly, to the average power output of the sound source. The average power involved in an oscillation is therefore a matter of some importance.

It will not be proved here but for a pure sound of one frequency, the relationship between the average power and the peak power is quite simple; the average power is just half the peak power. This can perhaps be believed from the graphs shown in fig. A13.1. In that figure the situation discussed in this chapter where there was 8 m^3 of sound air flow through a 4 m^2 area and a sound pressure associated with this flow of 826 Pa (very high values but suitable for the trial calculation to be given here). Suppose that these flows and pressures were actually peak values in the sound oscillations. The result of these pressure and flow values would be a power level in the sound of as also shown in the diagram. The average sound power would be half the peak value or 3304 watts.

When dealing with average powers in oscillating systems, it is convenient to define values of the oscillating parameters which will give average power when used into equations. For example, in sound it is convenient to define values of the pressure and air flow which will give the average power when used in the equations;

$$Average\ Power = R_A\, U^2_{eff} \qquad \text{(A13.2)}$$

and

$$Average\ Power = \frac{p^2_{eff}}{R_A} \qquad \text{(A13.3)}$$

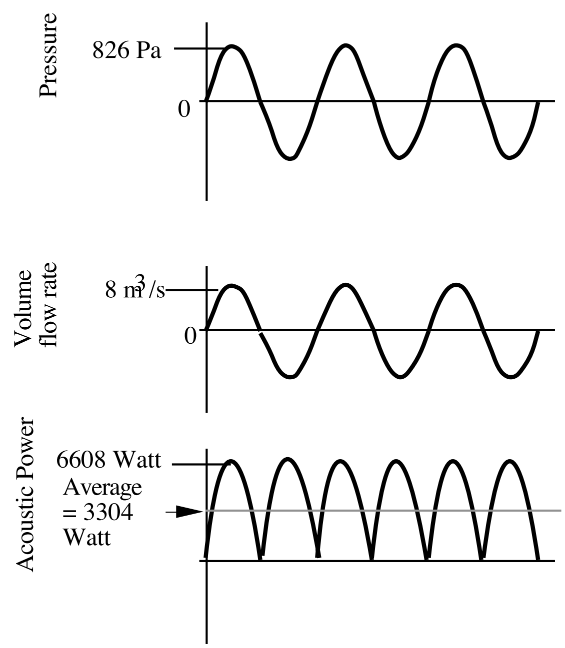


Figure A13.1 Pressure, air flow rate and power for a sinusoidal oscillation through a resistance.

Taking the case in fig. A13.1, a power of 3304 watts in an acoustic resistance of 103.5 ohms would require an effective flow rate given by

$$3304 = U^2_{eff} \times 103.5$$

$$U_{eff} = 5.65\ m^3/s \qquad \text{(A13.4)}$$

This flow can be thought of as the effective flow through the resistance. The relationship between this value and the peak value of 8 m^3/s can be seen to be;

$$\frac{U_{eff}}{U_{peak}} = \frac{5.66}{8} = \frac{1}{\sqrt{2}} \qquad \text{(A13.5)}$$

This "effective" flow rate is generally referred to as the RMS (for Root Mean Square) value of the flow rate. The term comes from the fact that what you are averaging when you average the power is the square of *U*. The square root of this average of the square is called the "root mean square." Thus for a simple pure sound of one frequency the RMS value of the acoustic current will be the peak value divided by the square root of 2.

The same considerations can be applied to the pressure. To get an acoustic power of 3304 watts into an acoustic resistance of 103.5 ohms will take

$$3304 = \frac{p_{eff}^2}{103.5}$$

$$p_{eff} = 585 \text{ Pa} \qquad \text{(A13.6)}$$

Again this may be thought of as an effective pressure. Its value is seen to be related to the peak pressure by the same sort of equation as for the acoustic air flow;

$$\frac{p_{eff}}{p_{peak}} = \frac{1}{\sqrt{2}} \qquad \text{(A13.7)}$$

This "effective" acoustic pressure is also called the RMS acoustic pressure for the same reason that the effective acoustic flow is called the RMS flow. Again, <u>for a pure sound of only one frequency</u>, the RMS value of pressure is the peak pressure divided by the square root of 2.

We can now freely use these RMS values of pressure and sound air flow to calculate <u>average</u> acoustic powers. In the case we have here;

$$\begin{aligned} \textit{Average power} \quad &= \frac{p_{eff}^2}{R_A} \\ &= \frac{585^2}{103.5} = 3304 \text{ Watt} \\ &= U_{eff}^2 \times p_A \\ &= 5.65^2 \times 103.5 = 3304 \text{ Watt} \\ &= p_{rms} \times U_{rms} \\ &= 585 \times 5.65 = 3304 \text{ Watt} \end{aligned} \qquad \text{(A13.8)}$$

In the case of acoustic pressure, the effective or RMS pressure is so important that when one refers to the acoustic pressure one is generally referring actually to the RMS pressure, without even saying the "RMS". Thus, if we have a sound intensity of say 1 watt per m^2, we can calculate the acoustic pressure from the relation;

$$I = \frac{p^2}{413}$$

$$p = \sqrt{413\,I} \qquad \text{(A13.9)}$$

for $I = 0.01$ Watt/m^2 (100 dB), $p = 2.03$ Pa

Since the intensity we are usually referring to is the <u>average</u> acoustic intensity, the pressure that we calculate is actually the RMS acoustic pressure. For a pure tone sound of just one frequency, the actually pressure would have peak values of $\pm\sqrt{2}$ times this or ± 2.88 Pa.

A similar situation occurs in electrical power. The so-called "110 Volt" system used for house wiring has a voltage swing from +155 Volt to -155 Volt. The "110 Volts" actually refers to the RMS voltage of this voltage swing. Similarly for the electrical current drawn. If a device such as an electric iron draws 10 Amperes of current it is actually drawing about 14 Amperes at peak. The "10 Amperes" refers to RMS or effective Amperes.

In the case of very complex oscillations, typical for sound, there are no such simple relationship between peak values and the RMS values. The RMS values of the pressure is now determined by measuring the average intensity by some metering device and calculating the RMS pressure that would give such intensity by the equation;

$$p_{rms} = \sqrt{413 I} \qquad \text{(A13.10)}$$

In such an RMS determination there can be very high spikes which do not contribute very much to average power. Consider as an example which can be easily calculated, a pressure which goes to 10 Pa for 1 ms and then stays at 1 Pa for 99 ms. A graph of this variation would be as shown in fig. A13.2

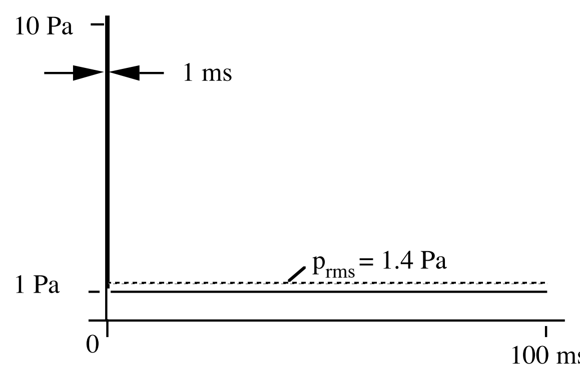


Figure A13.2 A graph of a sudden pressure pulse.

Calculating the average intensity for this period using the equation $I = p^2/413$, we get for 10 Pa an intensity of $10^2/413$ or 0.242 watt/m^2 for 1 ms. We then get an intensity of 1/413 or 0.00242 watt/m^2 for 99 ms.

The total energy delivered in the 100 ms is therefore $0.242 \times 0.001 + 0.00242 \times 0.099$ or 0.000481 Joules/m^2. The average intensity is 0.000481/0.1 or 0.00481 watts/m^2.

Converting this back to an RMS intensity, we get;

$$p_{rms} = \sqrt{413 \infty 0.0048}$$

$$= 1.4 \text{ Pa} \qquad \text{(A13.11)}$$

Thus the RMS pressure is only about 1/7 th of the peak pressure.

More important from the point of view of audio playback systems, the accurate reproduction of this transient peak would require that the system be able to deliver about 50 times as much peak power as average power.

This is the reason that good sound systems are capable of delivering much more power than is required for simply sustaining high levels of sound in a room. They need the power for the occasional transient which will make high demands on the system for accurate reproduction.

In the above the acoustic power was calculated from a knowledge of the acoustic current. A similar concept is used for acoustic power related when the pressure applied to a system is known. Here, however, the connection to the resistance is not so simple if the impedance is complex;

$$Power = pU\,cos\,q\ =$$

$$\frac{p^2}{Z}\ \cos\,\theta = p^2\frac{R}{Z^2} \qquad \text{(A13.12)}$$

To get a simple form for the equation connecting power to pressure, the concept of "admittance" is sometimes used;

$$Y\ =\ \frac{U}{p}\ \ = \frac{1}{Z} \qquad \text{(A13.13)}$$

The components of admittance are "conductance" and "susceptance" related to the admittance in the same way that resistance and reactance are related to impedance (fig. A13.3)

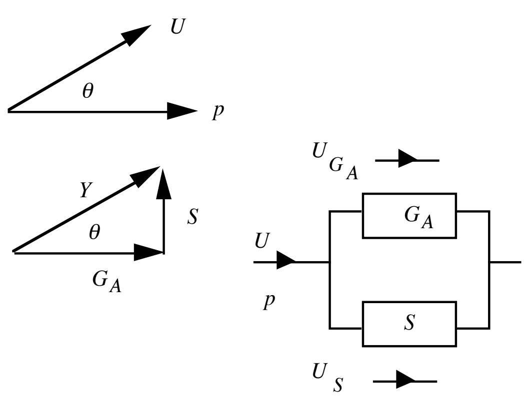


Figure A13.3 The graphical relationship between conductance, susceptance and admittance.

The conductance is the ratio of the current in phase with the pressure to the pressure and the susceptance is the ratio of the component of current 90^o out of phase with the pressure. In this concept, the current is divided into two parts (in the impedance concept the pressure is divided into two part; that across the resistance and that across the reactance.)

The acoustic power is then all due to the current through the conductance, the current through the susceptance taking no net energy in a complete cycle. The power dissipated by the system is then;

$$P = p\ U\cos\,\theta = \ p^{\,2}\frac{U}{p}\ \cos\,\theta$$

$$= p^2\ Y\cos\,\theta = p^2\ G \qquad \text{(A13.14)}$$