2:

ACOUSTICS AND PSYCHOACOUSTICS

Introduction

The raw material that we are working with is sound. Our purpose is to develop effective ways to use sound to convey useful information. It is the relationships that we can construct using sound that enable its articulation. The design space, therefore, is largely constrained by the *type* and *richness* of of the sonic relationships employed.

The objective of this chapter, therefore, is to investigate the properties of sound that afford the construction of such relationships. The working assumption is that the better we understand such properties, the more effectively we can use sound. It is sometimes said that design is choice. If that is so, then the quality of design is affected by the richness of the alternatives that one has to choose from and the criteria used in selection. Here we hope to provide the acoustic and perceptual foundation to support both.

There are three classes of relationships that can be used to encode meaning into sound:

- *intrasound relationships:* These are relationships established among the parameters of individual sounds themselves. For example, a message or datum may be encoded by establishing a particular relationship among the pitch, timbre and duration of a sound.

- *intersound relationships:* These are relationships that are established between or among sounds. It is through the pattern of sounds that meaning is conveyed. A simple example would be assigning specific meaning to a particular motif.

- *extrasound relationships:* These are relationships that are established between sound and entities that exist outside the domain of sound. For example, the sound of knocking on a door tells us about someone wanting to enter, rather than something about sound.

While these classes are not mutually exclusive, and can be used in combination, they provide a useful vocabulary for our study. First, they help us categorize the work of others for comparative purposes. Second, they help us in the process and rationalization of design. Third, they help guide our study of raw materials. They help us know what to look for, and to recognize properties of sound that can be exploited (or avoided).

"Utterances" in the vocabulary of nonspeech audio take the form of one or more audio *events.* Our working model is that of each event constituting an instantiation of a sonic *object.* One can best think of an object as having some more-or-less invariant parameters that identify it as emanating from a particular *source,* and one or more other "run-time," or *instantiation* variables - that is, parameters that may change on or during each instantiation.

The notion of the source of a sonic object is important. It is closely related (if not identical) to *timbre.* These are the parameters of the object that cause us to associate the sound as coming from a trumpet or flute, or a closing door or breaking glass[1]. Within the design space of nonspeech audio, one can choose from among the repertoire of object classes (i.e., sources or timbres), and then set the parameters that control their instantiations, such as pitch, duration and loudness.

Just as we have classes of sonic relationships, so do we distinguish between two classes of applying parameters to the instantiation variables:

- *fully formed objects:* These are sonic objects whose instantiation variables are fully specified when the sound is initiated. For example, if we want to create an audio display of an existing three dimensional data set, we might use one dimension to determine the object used, and the other two to determine the pitch and duration. When the object is invoked, all parameters are fully specified.

- *evolutionary objects:* These are sonic objects in which one or more instantiation variables are updated (continuously or discretely) during the life-span of the object. For example, we may want to monitor a sensor by mapping its temperature to the pitch of a sustained sound object. In this case, for example, as the temperature goes up, so does the pitch.

Like the categorizations discussed above, these distinctions don't specify what mappings to use. That is a much larger topic. Our purpose, in this chapter is to lay out some of the *options* that exist to be used in such mappings. We now proceed to look at the properties of sound and our perceptual system. Hopefully the above discussion will change what follows from a basic course in (psycho)acoustics, to a hunt for properties and opportunities that can be exploited in later designs.

Our discussion has to do with *applied* acoustics and psychoacoustics. Thus, although we cover a wide range of topics, this discussion is not comprehensive in either depth or breadth. For more information on acoustics, see Lindsay & Norman (1977), Kinsler, et al. (1982), Roederer (1975), or Askill (1979). For more information on psychoacoustics, see Lindsay & Norman (1977), Boff & Lincoln (1988), Scharf & Buus (1986), Scharf & Houtsma (1986), Evans (1982), Benade (1960). For more information on music acoustics, perception and cognition, see Deutsch (1982), Pierce (1983), Mathews & Pierce (1990) and Dowling & Harwood (1986). Finally, the Acoustical Society of America (500 Sunnyside Blvd., Woodbury, N.Y. 11797) distributes a compact disk of auditory demonstrations which is an excellent compliment to this chapter. The disk costs $20. US. Payment must be included with the order, and be drawn on a US bank.

---

[1] Note, however, that later in the chapter on Everyday Listening, we argue that there are special properties of sounds that emanate from everyday objects such as doors, compared to sounds such as would emanate from a musical instrument.

Acoustics

Sounds are pressure variations that propagate in an elastic medium (for our purposes, the air). Our ears are essentially very sensitive barometers that sense these pressure variations and transforms them into a form which can be accommodated by the brain. There are a number of useful ways to analyze these pressure variations. One is to represent them as graphs of waves.

Waveforms

Figure 1 shows the *waveform* of a simple sound, with the amplitude of pressure variation on the abscissa, and time on the ordinate.

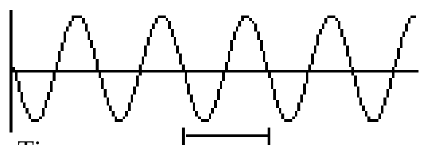


Figure 1. A simple waveform.

This kind of graph shows sound in the *time domain.* This is a periodic wave, that is, it repeats itself. The time it takes to complete a complete cycle is called the *period* of the wave, and is equal to 1/frequency. The actual *wavelength* of the signal is the distance that sound can travel in the interval of one period. This can be expressed as:

$$l = c\,p \qquad (1)$$

or

$$l = c\,/\,f \qquad (2)$$

where l is the wavelength, c is the speed of sound[1], p is the period of the wave and f is the frequency of the vibration.

The type of wave shown is a *sine wave.* It can be considered the simplest kind of waveform (for purposes which we will soon discover.)

> *Sound Example 2.1: A sine wave. A sine wave is presented at three different frequencies: 100 Hz., 1,000 Hz and 10,000 Hz. The sequence is presented twice.*

Fourier analysis and spectral plots

Sine waves are very rarely the result of natural events. In fact, there are few mechanical devices that create them (even the sound produced by a tuning fork is complex, particularly at the beginning). However, because of their mathematical properties they are extremely useful for acoustic theory. In particular, *Fourier analysis* allows the expression of any (well, almost any) complex wave as the sum of a number of sine waves of different frequencies, amplitudes and phases (see Figure 2).

[1] 334 metres/sec. (1130 feet/sec.) can be used for the speed of sound in air. But note that the speed of sound varies with the temperature. This value assumes an air temperature of 21^{o} C., or 71^{o} F.).

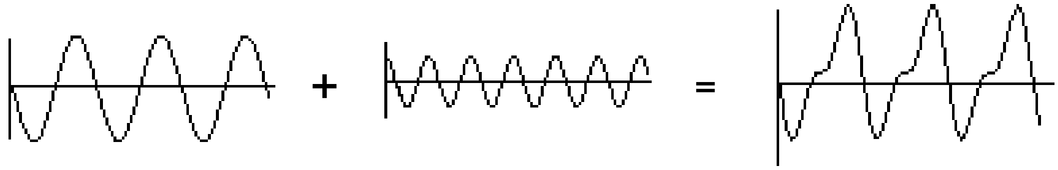


Figure 2   Two sine waves of different frequencies, amplitudes, and phases may be added to create a complex wave. Conversely, complex waves may be analyzed into their components via Fourier analysis.

When a wave is Fourier analyzed, the results may be shown in a *spectral plot* (Figures 3 to 6 show both spectral plots and waveforms). The spectrum of a wave is a two dimensional representation showing the amplitudes of sine waves of different frequencies that comprise the wave. Each spike on a spectral plot corresponds to a sine wave. So, in Figure 3, the spike in the spectral plot on the left corresponds to a single sine wave, shown on the right.

1. Sine wave

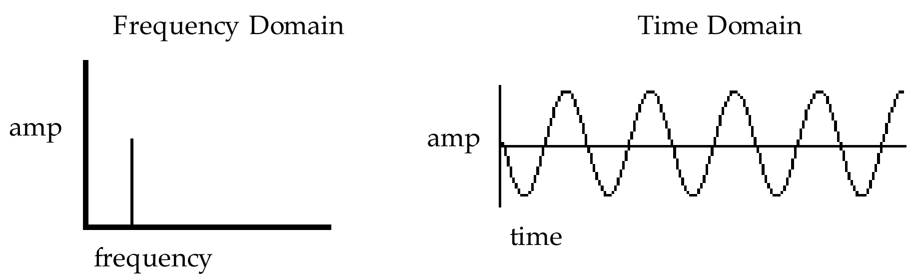


Figure 3.   A sine wave shown in the frequency domain (as a spectral plot) and the time domain (as a waveform).

More Complex waves

Sounds in nature are more complex that the examples that we have seen thus far. Figure 4 shows a more complex wave. The different spikes of energy seen in the spectral plot are called *partials*, with the lowest frequency being the first partial, the next higher the second partial, and so on.

This is an example of a special class of sound in which the frequencies of the partials are integer multiples of the lowest, *fundamental*, frequency. When this is the case, the partials above the fundamental are also called *harmonics* of the fundamental frequency, and the class of sound is called *harmonic.*

Harmonic sounds are periodic (their period is the same as its fundamental), and have a definite pitch. Harmonic sounds are relatively rare in the natural environment, but most musical instruments are designed to produce harmonic or near harmonic sounds. They have a clear pitch and can blend with other harmonic sounds in pleasing ways.

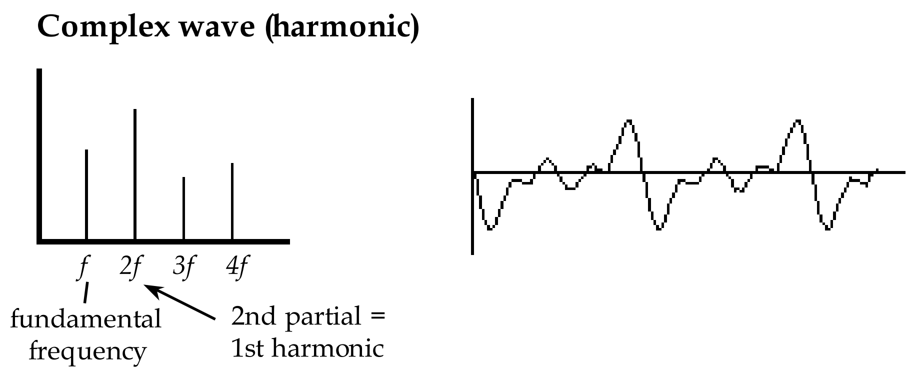


Figure 4. A complex wave that is harmonic.

Sound Example 2.2: Illustrating the sine wave components making up a complex harmonic wave. A 200 Hz complex tone consisting of 20 harmonics is played. The lowest 10 harmonics are cancelled and restored. Note that when nothing is changing, you hear a single tone, but on the transitions when harmonics switch on or off, the individual components can be heard.

Figure 5 shows another complex sound made up of five sine components. In this case, the partials are not integral multiples of the fundamental frequency. Consequently, the tone is *inharmonic,* and one does not call the partials harmonics Sounds with this property are by far the most typical of sounds in the everyday world. Note that in this simple example, the waveform does not repeat itself as obviously as the sine wave does, and in fact may be periodic only at very low frequencies (one over the product of the component frequencies).

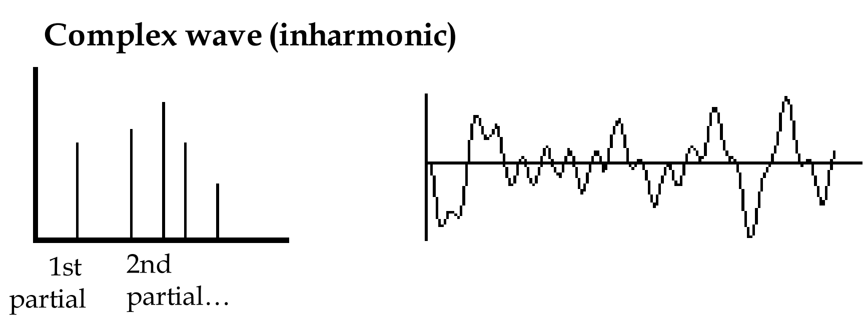
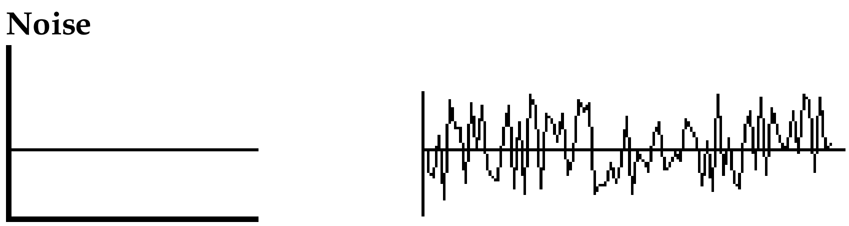


Figure 5 A complex wave.

Sound Example 2.3: A complex inharmonic wave. Note that the pitch is not well defined and one would not consider the sound musical (in traditional terms, anyhow).

Noise is made up of energy at many different frequencies. Figure 6 shows the flat spectrum and rough waveform that characterizes white noise, which contains equal amounts of all frequencies. Not all noise is white; for instance a noise might be *band limited*, containing energy in a certain band of frequencies. *Bandwidth* refers to the range of frequencies involved in a tone or noise.



.

Figure 6. The flat spectrum and rough waveform of white noise.

> *Sound Example 2.4 White Noise: this is sound that has equal amplitude at all frequencies in the audible spectrum..*

Sound, Obstacles, Bending and Shadows

Sound waves have the property that they will not bend around obstacles in their path whose size is greater than the sound's wavelength. For example, if there is a 1 metre wide column between you and a sound source, the column will cast a *sound shadow*, preventing you from hearing all frequencies whose wavelength is less than 1 metre. (Well, almost - nothing is perfect, and in practice you will hear some of these frequencies, but greatly attenuated. This is because some of the cut off frequencies will bounce off of other objects and get to your ears by an indirect route.)

To find the frequency above which sounds will be cut-off by a given obstacle, we need only find the frequency whose wavelength is the same as the minimum dimension of the obstacle. Remembering that a wave's period is the inverse of its frequency, from equation (2), this can be expressed as:

cut-off frequency = speed of sound / obstacle dimension *(3)*

Therefore, we can calculate the cut-off frequency for our example of standing behind a 1 metre wide column:

cut-off frequency = 334 / 1

= 33

As we shall see in later sections, sound shadows have an important role to play in sound localization (which partially relies on shadows cast by the head and outer ear, or *pinnae*.

> *Sound Example 2.4 Revisited: Sound Shadows. Replay the white noise of example 2.4. Hold up and then remove a book between your ear and the speaker. Try holding up objects of different size. Listen for the change in high frequencies. Notice that the larger the object, the lower the cut-off frequency.*

Phase: its Implication on Sound and Representations

As a rule, frequency domain representations of sound can be more easily correlated to perception than can time domain representations. For example, two sounds whose waveforms look similar may sound quite different. On the other hand, two waveforms that look quite different may sound similar, or the same (Plomp, 1976).

Figure 7 shows two waveforms which have the same spectra and sound identical. Nevertheless, their waveforms differ. In this example, the different waveforms result from a change in the phase relationship between partials. We tend not to perceive phase differences between the harmonics of periodic sounds (Cabot, Mino, Dorans, Tockel & Breed, 1976). The lesson to be learned here is one that (most) computer musicians learned a long time ago: specify sounds in the frequency domain, and resist the temptation to build a program that allows you to "draw" waveforms.

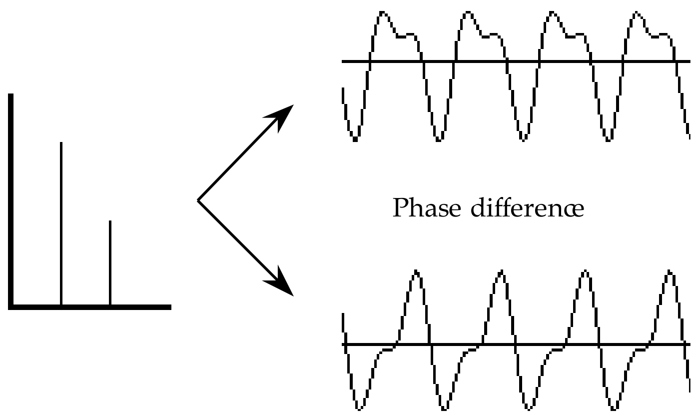


Figure 7. The same spectrum can give rise to very different waveforms, depending on the phase relations between partials.

Amplitude and Decibels

Humans are sensitive to an immense range of sound intensities -- for instance, the most intense 2000 Hz. sound tolerable is about 1,000,000,000,000 times as intense as the weakest that can be detected (Lindsay & Norman, 1977). For this reason, intensity is commonly expressed in the logarithmic *decibel* scale, in which:

$$dB = 10 \log (I / I_0)$$

where I_0 is (usually) taken as the international standard reference level 0.0002 dynes/cm^2. Note that this implies (among other things) that two sounds separated by 10*x* dB have an intensity ratio of 10^x. Figure XX is a table which correlates sound pressure level, as measured in dB, with a number of real-world sounds.

Sound Example 2.6: The Decibel Scale: Broadband noise is reduced in 10 steps of 6 dB, 15 steps of 3 dB and 20 steps of 1 dB. Demonstrations are repeated once.

The Inverse Square Law

Consider sound as radiating an equal amount in all directions from a from a point in space. Thought of in this way, it can be seen that at any given distance, the energy emanating from the source is distributed uniformly over the surface of an imaginary sphere whose radius is determined by the distance. For a constant intensity at the source, the intensity at a distance from the source will diminish proportionally to the inverse of the area of the sphere having a radius equal to that distance. Generally stated:

In a free field, sound intensity is inversely proportional to the square of the distance from the source.

Sound Pressure Level (dB SPL)	
220	- 12"" cannon 4 m in front and below muzzle
.	
.	
.	
140	
130	
	- Threshold of pain
120	- Rock band or loud discoteque
	- Hammering on steel, 1m
110	
	- subway station, express passing
100	
	- Electric power station
90	
	- Average factory
80	- very loud home stereo
80	
70	- Ordinary conversation, 1m
	- Department store, noisy office
60	
	- Quiet residentail street
50	
	- Average residence
40	
30	
20	- Quiet whisper, 1.5 m
10	- Out of door minimum
0	- Threshold of hearing

Figure XX: *Correlation between SPL as expressed in dB and some real world sounds (after XXX)*

This is known as the *Inverse Square Law.* It will be important later when we are interested in positioning sounds in space, and want to give the illusion of distance. One important point to keep in mind however, is that the sphere and uniform radiation on which the law is based is imaginary, and is accurate only in idealized situations, namely a *free field.* In reality, sound hits the floor, walls, furniture, etc. It may get absorbed by the surface, resulting in energy loss, or it may be reflected in a different direction, thereby creating another mini sound source that contributes to the overall intensity at another location. Like most laws, in practical terms this one is better considered a guideline.

Simple Variation: Envelopes

Virtually all natural sounds change over time. Perhaps the most obvious dimension of change is loudness, for how else could a sound have a start and a finish? Typically, this change is not a simple switch on and then off. The loudness contour of most sounds is more complex.

We can see this contour by looking at the outline of the graph of a sound's waveform. This can be seen in Figure 8A, for example. If we traced the "silhouette" of the waveform, the result would be a sketch of the sound's *envelope*[1]. Since the envelope is always symmetrical on either side of the x axis, we normally draw only the positive side. In specifying envelopes to sound synthesis equipment, for example, this is virtually always the case.

Using just the positive side, Figure 8B shows the same waveform outlined by a representation of its envelope. Notice that the envelope is drawn as a straight line approximation of what is actually a continuous curve. This is also the norm, and will usually be adequate for our purposes, as demonstrated in studies of timbre and data compression by Grey (1977). Notice that the segments of the envelope are labeled Attack, Decay, Sustain, and Release (ADSR). These are common terms used for specifying envelopes with many synthesizers, although many more sophisticated systems permit much finer control.

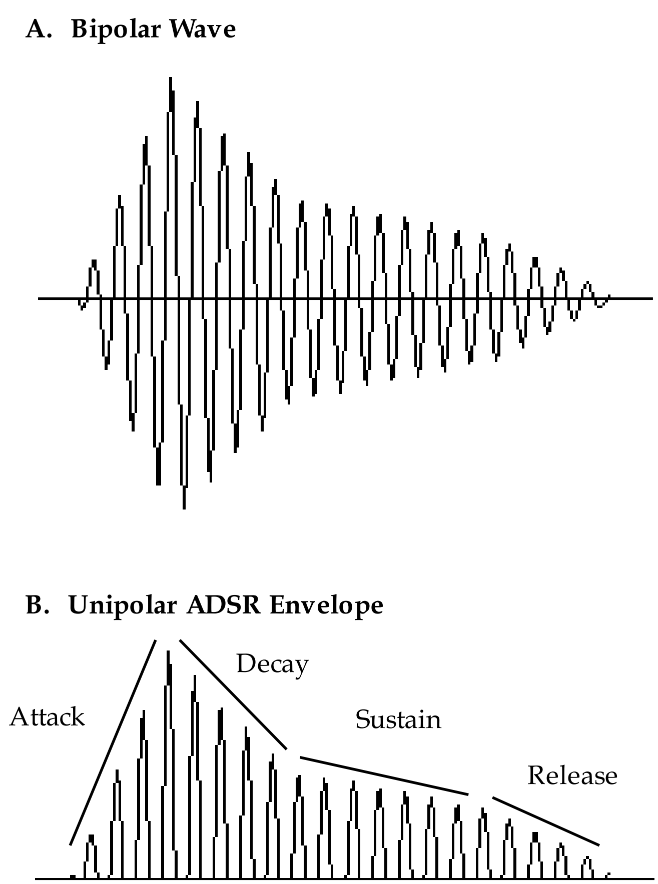


Figure 8. Attack, decay, sustain and release of a sound.

[1] Actually, what we get by tracing the waveform's outline is the sound's *amplitude* envelope, not that of its *loudness*. While amplitude and loudness are related, they are not the same, as our discussion of psychoacoustics will explain. However, it is the concept of an envelope that is important at this stage.

> *Sound Example 2.6: Amplitude Envelopes: A periodic complex tone of 440 Hz is played three times, each time with a different envelope. The example is presented twice.*

Spectral Variation over Time

We have already seen that most sounds are made up of spectral components, or partials, of different frequencies. Typically, the relative amplitude of these partials does not remain constant throughout the duration of a sound. That is, it is usually incorrect to think of the envelope simply as a function which controls the scaling of a waveform's amplitude over time.

A more accurate picture of what happens can be obtained by thinking of a waveform in the frequency domain, and then using a separate envelope to describe each partial. An example of this is shown in Figure 9. This type of representation is known as a *time varying spectral plot*.

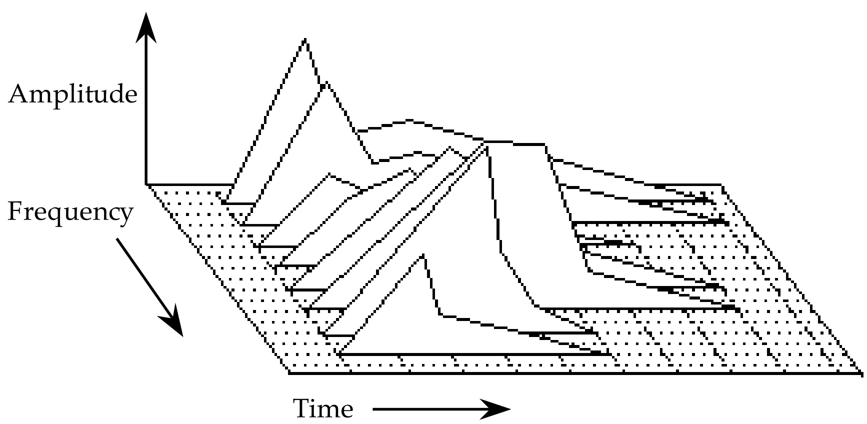


Figure 9. *A time varying spectral plot of a complex sound.*

Since the lower frequencies usually have the higher amplitudes, they apear at the back of the plot, so as to not obscure our view of the lower amplitude high frequencies plotted in the foreground. Notice that while each spectral component has its own envelope, we no longer have a graph of a single envelope describing the shape of the sound as a whole. As a first approximation, we could derive the overall envelope as the sum of the envelopes of each of the partials; however, in reality, the amplitudes do not sum linearly, especially in the higher frequencies where spectral components fall within about 1/3 of an octave of each other. This has to do with the perceptual system and something called the *critical band*, discussed later in this chapter.

> *Sound Example 2.7: The Effect of Time-Varying Spectra on Timbre:*
>
> *(a) The sound of a Hemony carilon bell, having a strike-note pitch of around 500 Hz (B4), is synthesized in eight steps by adding succesive partials with their original frequency, phase and temporal envelope.*
>
> *(b) The sound of a guitar tone with a fundamental frequency of 251 Hz. is analyzed and resynthesized in a similar manner.*

Helmholtz Revisited

The study of acoustics has been dominated for over a hundred years by the pioneering work of Helmholtz (1885/1954). One of his main theories, and one which is still widely taught, is that timbre can be explained solely in terms of spectra, independent of time. More recent work has demonstrated that his theories are inadequate, and that the relative temporal change among partials is an important determinant of the timbre of a sound. In fact, these relative time varying characteristics appear to be as important (and sometimes more so) than the frequencies of the partials actually present in the sound.

Risset trumpet sounds: real vs synthetic

Figure XX. *Characterizing Brass Sounds by Distribution of Spectral Enery over Time*

This is illustrated in the work of Risset (Risset & Wessel, 1982). He demonstrated that convincing brass sounds could be synthesized using spectral components that were foreign to natural brass sounds. What he found was that it was the relative rise and fall of energy in regions of the spectrum, rather than the specific components, that determined the brass-like character. Specifically, he showed that brass instruments are characterized by an increase in spectral content with frequency -- that is, low frequency partials build up faster than high ones, as is illustrated in Figure XX.. A convincing trumpet sound can be synthesized by following this rule, even if the partials used are for the most part foreign to the actual spectrum of a trumpet. This illustrates two important observations relevant to the understanding and design of sound. First, the traditional account of timbre is inadequate in that it ignores variations in time which are crucial for perception. Second, independent and dynamic control of spectral components is critical for the effective exploitation of timbre.

Sound Example 2.8: Time and Timbre: Synthetic and recorded brass sounds are presented. The spectrum of the synthetic instruments differs from the natural ones; nevertheless, they all sound "natural." This is because of the overall behaviour of the spectral components, rather than what specific components make up the sound.

Spectograms

In light of the importance of the distribution of spectral energy in the perception of timbre, it is worth briefly introducing a graphical representation of sound in which this is highlighted. The representation is known as a *spectrogram.* An example is illustrated in Figure XX.

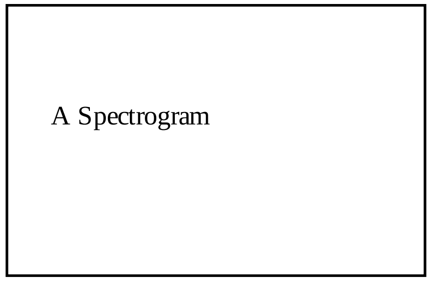


Figure XX. *A spectrogram. The sound energy in each 1/3 octave in tha audible range is plotted over time. The darker the plot, the more energy.*

Rather than show the amplitude envelopes of individual partials, a spectogram shows the energy in (usually 1/3 octave) bands of the audible range. The darker the plot in each band, the more the energy is there. This type of representation makes it very easy to get a sense of how the spectral energy is distributed over time.

Formants vs Partials

In Figure XX, notice that there are two regions in the spectrum where there is pronounced energy. Many sounds can be characterized by the presence and location of similar bands of spectral activity. These bands, known as *formants,* are most frequently encountered in studies of *phonemes* of speech. However, they are characteristic of most sounds that are produced by mechanical means.

Formants are important to our understanding of sound, and especially timbre. To understand them, we need to slightly modify the model of sound. We have been using a simplified model in which mechanical vibrations were transformed directly into atmospheric pressure variations (sound waves). Important to the understanding of formants is the introduction of the concept of *resonance*.

Think about an elastic band that is stretched between your fingers and then plucked, This is a good example of the simplified model that we have been using thus far. Now, if you hold one end of the same elastic against a match box, stretch it to the same tension and length, and pluck it exactly the same as before, the sound will be different. The reason is the same as why acoustic guitar strings are stretched over a hollow body rather than a flat plank: the body produces a resonator that modifies the spectrum of the vibrating string. These resonances occur in particular regions of the spectrum, and a large part of the craft of instrument design is determining the strength and frequency of these resonant frequencies.

Formants occur when partials of the primary vibration (for example, the guitar string) fall within these resonant regions, and are accentuated relative to those that fall outside the resonant frequencies. We can illustrate this further with another example using our voice.

Pick a pitch that you are comfortable with, then say - and sustain - the sound "ee" as in beet. Now, without changing the pitch or loudness of your voice, change to the sound "aw" as in hot. What you are doing is changing the properties of the resonators of your vocal tract. What you are *not* doing, is changing the spectrum of the sound wave leaving your vocal cords. But since the spectrum of the wave leaving your vocal cords must pass through the "resonator" of your mouth, certain spectral components get attenuated while others get accentuated - all depending on whether they fall within the resonant regions (determined by the tongue, lip and jaw position).

The voice is interesting, in that the resonant frequencies, and hence the formant frequencies, can be varied over time. Whereas this is also true for many electronic instruments, it is rare with mecahnical instruments. The resonant frequencies of a violin are a function of the materials used and its shape. The same is true for other instruments, such as the piano or bassoon. Therefore, because they are rooted in the physical structure of the instrument, *formant frequencies are fixed, and do not vary with the frequency of the vibration that is exciting the system.* (See, for example, Slawson, 1968 for more on this.)

One consequence of this is that for a given instrument, the spectral properties of a note at one pitch will typically be different than those at another (since being at different pitches, different parts of the spectrum will fall within the resonant regions of the instrument). Therefore, we cannot just analyse one note of an instrument and then transpose it up and down and expect the effect to be one of all notes coming from the same instrument. This is illustrated dramatically in the next sound example.

> *Sound Example 2.9: Change in timbre with transposition. A three octave scale on a basoon is presented, followed by a three octave scale of notes that are simple transpositions of the instrument's highest tone. This is how the basoon would sound if all of its tones had the same relative spectrum.*

Summarizing thus far...

- Sounds are pressure variations that propagate as waves in the atmosphere.
- The waveform of a sound can be displayed in the time domain as amplitude by time.
- The time for one cycle of a repetitive wave is called the period, and is the inverse of the wave's frequency (the number of repetitions per second).
- Complex waves may be analyzed according to the sine waves that could be combined to produce them, the results may displayed in the frequency domain on a spectral plot.
- When the frequencies of the partials of a complex sound are integer multiples of the sound's fundamental frequency, the sound is harmonic.
- Most complex waves are inharmonic. The individual frequency components of the wave are called partials.
- Noise contains energy at many different frequencies, and is not periodic. Band-limited sounds contain a certain range of frequencies; the size of this range is its bandwidth.
- Intensity is commonly expressed using the logarithmic decibel scale, in which dB = 10 log (I/I_0).
- Sounds cannot bend around obsticles whose size is longer than their wavelength. Consequently, te have the existance of frquency sensitive *sound shadows.*
- Inverse Square Law: In a free field, sound intensity is inversely proportional to the square of the distance from the source.
- The overall amplitude variations of a tone are commonly referred to as the tone's attack, decay, sustain and release.
- A more precise description of a sound's evolution over time is given by a time varying spectral plot with the dimensions time, frequency, and amplitude.
- The physical properties of sound generators typically create resonant frequencies that accentuate partials that fall within them. The resulting regions of high spectral energy are known as *formants.* Formant frequencies are independent of the fundamental frequency and spectral make-up of the exciting vibrating body.

Psychoacoustics

Up to now, we have been primailly discussing the physics or acoustics of sound. However, our body does not take these stimuli from the physical world and just linearly map them into the perceptual. The mapping between stimulus and percept is complicated by nonlinearities, interactions, and bounded by the limitations of our sensory and cognitive capability.

In this section we look at these mappings and interactions, building upon the basic understanding of acoustics developed in the previous section. We outline fundamental topics of psychoacoustics that are important for designing audio interfaces, discuss each topic briefly in turn, and explore some of the design implications of each. We begin with two concepts are prerequisites for what follows: *just noticeable difference (JND)* and *critical band.*

Just Noticeable Difference (JND)
• to come

Critical Bands
• Critical bands are frequency regions within which sound energies interact. They seem to be explicable in terms of the vibratory response of the basilar membrane. Simplified, frequencies which stimulate the same hair within the basilar membrane typically fall within the same critical band.

- Critical bands are roughly a third of an octave wide. For a centre frequency above about 500 Hz they are 20-30% wider. Below 500 Hz they are narrower. Critical band (*CB*) around a given centre frequency f (both in Hz) can be approximated by the following equation (Zwicker & Terhardt, 1980):

$$CB = 25 + 75 (1 + 1.4 f^2)^{0.69}$$

- Critical bands interact with loudness (see below): Energy is summed within a critical band, loudness without.

- Some aspects of masking (see below) can be explained by critical bands. Detecting a pure sound that is masked by a bandpass noise centered around its frequency becomes more difficult as the bandwidth is increased, up to a critical value (corresponding to a critical band) when further increases do not affect detection.

- Critical bands may help explain consonance and dissonance. Sounds within a critical bandwidth sound rough, those without do not.

Implications for design:
Multiple sounds interact differently depending on their frequency separation. Knowing the nature of these interactions is crucial for designing systems in which more than one sound may be heard at once.

Pitch
Pitch is the sensation most directly associated with the physical property of frequency. Roughly speaking:

- The pitch of a sine tone is a function of its frequency.

- The pitch of a harmonic tone is a function of its fundamental frequency.

• The pitch of an inharmonic tone or noise is a function of its amplitude-weighted average frequency (brightness[1]).

• The JND for pitch discrimination is about 1/30 of a critical band.

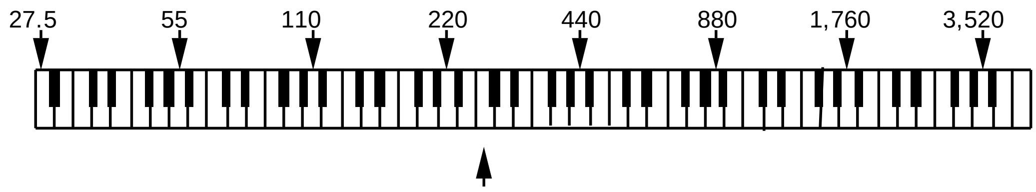


Figure XX: The relationship of frequency to musical pitch. The frequency corresponding to the nominal fundamental frequency of each "A" on the piano keyboard is shown. Note that each octave between successive A's corresponds to a doubling of frequency, not a fixed number of Hz.

But:

• The functional relation between pitch and frequency is *not* a linear one, rather *logarithmic.* Doubling the frequency nominally raises the pitch of a sound by an octave. Thus, in terms of pitch, the difference between 110 Hz and 220 Hz is the same as that between 440 Hz and 880 Hz. This is illustrated in Fig. XX. We say that the function is nominal, however, since other factors besides frequency affect pitch, such as those discussed below.

• Pitch is affected by loudness. The pitch of sine tones under about 1000 Hz. goes down as the sounds get louder; the pitch of sine tones over about 1000 Hz. goes up as the sounds get louder.

• Pitch is affected by timbre. Bright sounds (those with a relatively great amount of high-frequency energy) sound higher than dull ones of the same fundamental frequency.

• Pitch is not a single perceptual dimension. That is, unlike a ruler, as you go further and further along, you don't necessarily get further and further from where you started. For example, perceptually, two notes separated by an octave (12 semitones) will often be judged closer in pitch than two notes separated by just one semitone. There are two properties of pitch at play here. The first is *pitch height*, which is how high or low a sound is. The second is *pitch chroma*, or *pitch class*, which is the note type (such as A, B, or C#). We can discuss these two concepts of pitch using three notes as an example: A440, the A# a semitone above it, and A880, the octave above it. In terms of pitch height, the A# is closer to A440 by 11 semitones. However, in terms of chroma, the A440 is closer.

Pitch height and chroma can be visually represented together as a helix, as shown in Fig. XX. Here, the relationship between C and D is shown. Pitch height is along the spiral. Pitch chroma is represented by the vertical lines linking notes of the same pitch class.

[1] We will discuss brightness later in this section when discussing *timbre.*

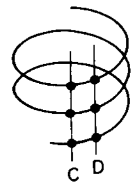


Figure XX: The pitch helix, or spiral: Pitch height increases as one goes along the spiral. The closer two points on the spiral, the closer their pitch height. The vertical lines indicate notes of the same chroma, in this case, C's and D's. Because of the perceptual similarity of notes having the same chroma, despite large differences in pitch height, pitch height alone should not be used in mapping data to pitch.

- Most people are bad at making absolute pitch judgments. The ability to do so, known as *perfect pitch*, is rare and cannot be taught reliably .
- People have varying degrees of ability at making relative pitch judgements, especially with tones which fit on the musical scale. This skill varies with musical ability, and can be taught reasonably reliably.
- People's ability to hear high pitched sounds, especially soft ones, drops off as they get older.

Pitches, Intervals, Scales and Ratios

From what we have read, we now know that the relationship between frequency and pitch is not liner, and that musical intervals are expressed as frequency *ratios*, rather than fixed values. For example, an octave above frequency *f* is 2/1*f*, while an octave below is 1/2*f.*

But what about musical intervals besides the octave? Here we fall into a bit of a black hole. For a detailed guide, the reader is referred to Chapter 8 of Backus (1969), for example. In what follows, we will introduce a few terms that the reader may encounter, and try and give a quick and dirty introduction to the issues.

Western music divides the octave into twelve steps. In the perfect world, the interval between each step would be the same. In this case, the ratio between each step would be $\sqrt[12]{2}$. Rather than expressing the intervals as a ratio, a linear scale is often encountered. In this case, the octave is said to consist of 1200 *cents*, and each semitone representing and interval of 100 cents. Since the time of J.S. Bach, most pianos and other musical instruments have been tuned such that their notes, or intervals, fall as closely as possible to this uniform sized semitone. This is known as the *tempered scale.*

Nature, however, is not quite so perfect. We can try to generate the notes of the western scale from from the pitces of natural harmonics. We start with a note of frequency *f.* The first harmonic, *2f* gives us the octave above. The second harmonic, *3f*, gives us a new note: the fifth, or *dominant* of the pitch associated with *f.* If *f* corresponded to the note C4 (C in the 4th octave), then *2f* corresponds to C5 and*3f* corresponds to G5, and G5:C5 = 3:2. In general, the fifth above any note is 3/2 times the frequency of that note. Going around what is known as the circle of 5ths, we should be able to generate all 12 pitches in the scale using this interval (since as we go clockwise around the circle, each succesive pitch is separated by a 5th).

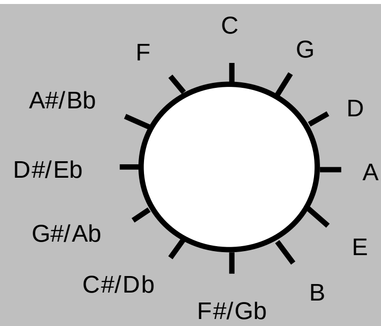


Figure XX: The Circle of Fifths

All 12 notes of the western scale can be generated from a start note, and the interval of a 5th (7 semitones). Each new note is a fifth above the previous one. If we use the ratio of 3:2 for a fifth, however, the note that we get after 12 steps does not have an exact octave relationship with the start note.

Without working through all of the arithmatic, suffice it to say that by the time we get all the way around the circle, the 5th above F does not have a power of 2 relationship with C. That is, using the 3:2 ratio, we do't really get a circle.

So, the tempered scale and uniform semitone are "artificial" and result from adjusting the ratios found in nature in order to "close" the circle of 5ths. The result makes diatonic music possible, but removes us from the world of simple ratios.

Implications for design:
The logarithmic relation between the musical scale and frequency should be used if an equal-interval scale of pitch is desired. Pitch encodings, especially musically based, which encode data based on relative pitch relationships, as in a melody or chord, can be used effectively. The absolute encoding of data as pitch is to be avoided if absolute judgements are to be made.

It is important to be aware of the myriad of interactions between pitch and other attributes of sound when using pitch, particularly if the aim is to convey orthogonal data dimensions using attributes of sound. Similarly, if pitch is used to represent magnitude, it is important to be aware of the cyclic nature of musical tones.

Finally, most pitch cues should be kept in the middle of the audible range. That way they will be reproducible on even poor quality equipment and be audible by most people.

Loudness
Loudness is the sensation most directly associated with the physical property of sound intensity. Roughly speaking, loudness (*L*) relates to intensity (*k*) according to the power law:

$$L = k\, I^{0.3}$$

where loudness is measured in *sones* (an international standard defines the loudness of a 1000 Hz. tone at 40 dB as 1 sone), and intensity as decibels. This relationship means that a 10dB increase in intensity results in a doubling of loudness.

But:

- Loudness depends on frequency. In general, tones at a given amplitude with frequencies between about 1000 and 5000 Hz. are perceived as louder than those that are higher or lower. More exact functions of loudness with frequency can be found in the form of *equal-loudness contours*. The equal loudness contours for sine waves, also known as the *Fletcher-Munson Curves*, are illustrated in Figure 11 (Fletcher & Munson, 1933). (Remember that the curves in the figure apply to sine waves only.)

• Loudness is affected by bandwidth. Within a critical band, energy is summed; outside of a critical band, loudness is summed. Sounds with a large bandwidth are louder than those with a narrow bandwidth, even if they have the same energy.

• Loudness depends on duration. For sounds shorter than about a second, loudness increases with duration. For example, for each halving in duration, the intensity of noise bursts below 200 m.s. must be increased by about 3 dB in order to maintain constant loudness (Scharf & Houtsma, 1986, p. 15.11). However, for sounds longer than a second, loudness remains constant.

• People are very bad at making absolute judgements about loudness.

• Our ability to make relative judgements of loudness are limited to a scale of about three to five different levels.

• Remember that perceived loudness is affected by the location of the sound source with respect to the listener due to the inverse square law and the effect of any obstacles that may be creating sound shadows.

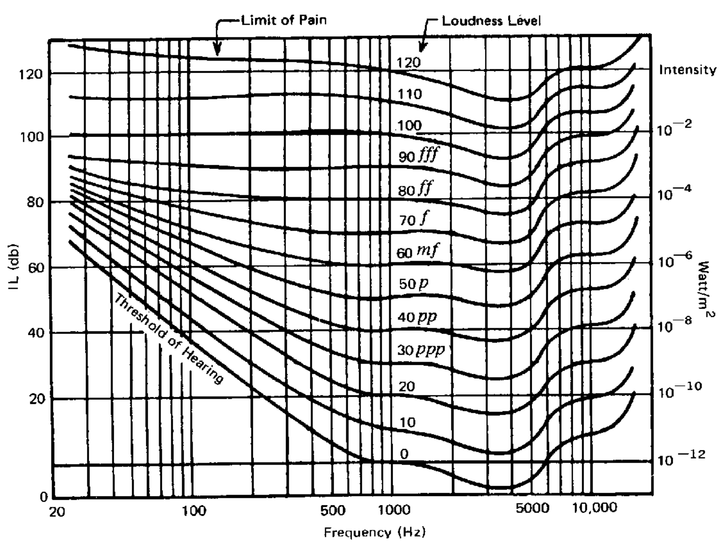


Figure 11: The equal loudness contours for sine waves, after Fletcher and Munsen (1933). From Roederer (1975,) p. 83.

Implications for design:
As with pitch, if loudness is to be used to encode data the proper function must be used to translate amplitude to loudness. Also, the interactions between loudness and frequency, bandwidth, and timbre should also be recognized, particularly if several of these variables are to be manipulated at once.

Finally, remember that users are certain to adjust – and misadjust – playback loudness. Any data encoded by loudness must take this into account and be tolerant of a large margin of error. Alternatively, calibration procedures to help the end user set levels appropriate for their environment must be provided.

Duration, Attack Time and Rhythm.
For most sounds, perceived duration corresponds pretty well to physical duration.

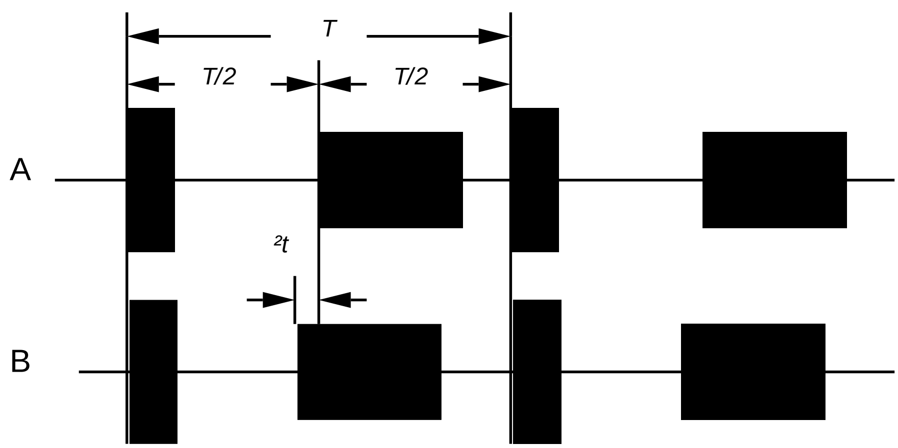
• Perception of changes in duration. The just noticeable difference (JND) for perceiving changes in the duration between two sounds is a change of about 10% for sounds longer than 50 msec., and proportional to the square root of the duration for sounds of less than 50 msec. (Scharf & Buus, 1986, p. 14.62)

• This JND is independent of the level and the spectral bandwidth of the signal.

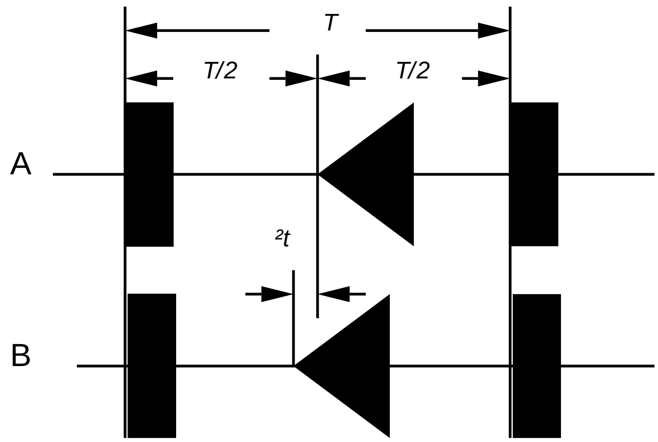
But:

• Duration affects rhythm. For instance, if two notes of equal duration alternate so that their onsets occur every 4 time units, we hear a regular rhythm. If the duration of one is changed to be 3/4 the duration of the other, however, our perception changes: Even though the onsets are spaced equally, we don't hear it that way. Instead the duration gives the long one extra "weight" that perceptually "pulls the beat off center" (Terhardt, 1978).

• The perceived onset of a sound depends on the slope of its attack. People don't hear sounds as starting until they have built up a certain amount of energy. Hence, in an orchestra, the tam tam player must anticipate an entry and start playing before, say, the trumpets (Terhardt, 1978, Wessel, 1979, Vos & Rasch, 1981).



Two streams of sounds are shown, A and B. Each consists of a short sound followed by a longer sound. All sounds have rectangular envelopes (i.e., sharp onsets and offsets). In A, the onsets of the sounds are equally spaced. Consequently, one would expect a regular rhythm to be perceived if the pattern repeats. However, this is not true due to the difference in length of the two sounds. In order for the two sounds to repeat such that the long sound is perceived to fall midway between the onsets of the short sounds, the longer sound must start early by an experimentally determined value Δt, as shown in B (after Terhardt, 1978).



Two streams of sounds are shown, A and B. Each consists of a short sound followed by a second. The first has a rectangular envelope (i.e., sharp onset and offset). The second has a slower gradual onset. In A, the onsets of the sounds are equally spaced. Consequently, one would expect a regular rhythm to be perceived if the pattern repeats. However, this is not true. Despite when the physical onset of the second sound begins, the perceptual onset does not occur until a certain "mass" of acoustic energy has been generated. In order for the two sounds to repeat such that the triangular sound is perceived to fall midway between the onsets of the short sounds, it must start early by an experimentally determined value Δt *, as shown in B (after Terhardt, 1978).*

Implications for design:
Once again, though physical dimensions of sounds can be specified independently, their perceived qualities may interact. In particular, care should be taken if duration and attack are to be manipulated independently. Likewise, rhythm involves more than the spacing of onset times, and manipulating rhythm separately from durations of component sounds may be difficult.

Microvariation and Spectral Fusion
We are seldom in an environment where we are exposed to only one sound source. In a concert, for example, several instruments may sound at once; in driving a car, there is the sound of the motor, the radio, the turn signals, and the passenger speaking. Often the partials of the different sounds overlap. How then do we know to which sound source a particular partial belongs? For example, does a particular time-varying spectral component belong to the clarinet or the oboe?

The acoustic cues which enable us to make these judgements are largely a result of the time varying nature of the sounds. In addition to the comparatively large scale amplitude changes described by envelopes, the spectral components of virtually all natural sounds also have microvariations in both pitch and amplitude. Vibrato and tremolo are two examples of such microvariation.

Perhaps the most important cue in deciding whether two partials belong to the same sound is the degree of *coherence* in their respective variation. When the variation of spectral components is coherent, they are normally judged to be derived from the same sound source. This process is known as *spectral fusion.*

This phenomenon is illustrated in in Figure 9. Part A of the figure represents several partials sounding simultaneously. There is no variation in the frequency of the partials. They are, therefore, coherent. Consequently, they are fused together perceptually and heard as a single complex sound.

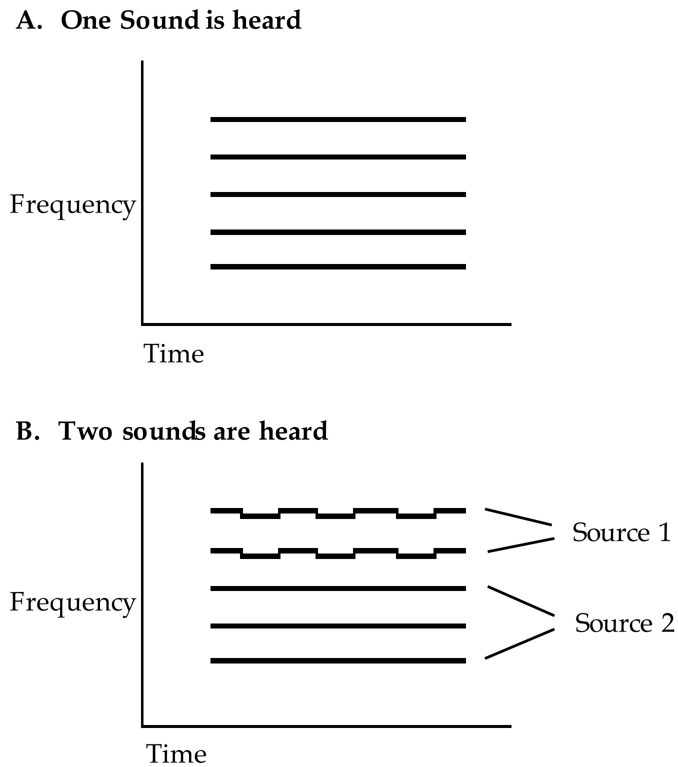


Figure 9: Coherence of variations results in spectral fusion and grouping of partials by source.

In part B of the figure, coherent microvariations (in this case vibrato) has been introduced on two of the partials. The partials now cluster into two groups, according to coherence: those with vibrato and those without. What will be heard are two distinct sounds from two different sources.

An understanding of spectral fusion is particularly important to those concerned with synthesizing non-speech sounds. In contrast to natural sounds, those generated using digital synthesizers are frequently, if not normally, in perfect tune with no microvariations in pitch or amplitude. The result is that two sounds meant to be distinguished may fuse into one complex sound. At best, the distinction of sources is made more difficult. In order to avoid this, special care must be taken to introduce the appropriate cues when sounds are to be distinguished. Conversely, manipulating these same cues gives us the means to merge two sounds that would otherwise be separate.

Timbre,

Timbre, or *tone color*, is defined as "... that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" (American National Standards Institute, 1973). Given the negative nature of this definition, it is perhaps not surprising that timbre has been called "the psychoacoustician's multidimensional wastebasket category" (McAdams & Bregman, 1979).

- Timbre is influenced by the spectrum of a sound, and changing the partials of a sound will generally have the effect of changing its timbre.

- To the extent that a sound's waveform is determined by its spectral components, there is a relationship between a sound's waveform and its timbre.

- However, this is not a very useful notion for at least two reasons. First, the spectrum of virtually all sounds in nature is not static, so it makes little sense to talk about the waveform of a violin, for example. Second, it has been shown that properties of harmonics' dynamic behaviour can be far more important to timbre than the actual harmonics themselves. (Add discussion of Risset.)

- In short, the theories of Hemholtz (and followers) about spectrum and timbre are incomplete, to say the least.

- Timbre also depends on temporal characteristics, such as the attack and decay of the sound (see our discussion of time varying spectra and timbre in the acoustics section above).

- Timbre is clearly multidimensional, but the dimensions are only partially known.

- Grey (1977) found 3 dimensions of timbre using multidimensional scaling. There physical correlates were 1) spectral energy distribution (brightness), 2) the relative amount of high-frequency energy in the attack, and 3) the degree to which amplitude functions of partials are synchronous.

- Although timbre research has been motivated recently by advances in computer music, many computer musicians are turning to physical modelling to understand and control sound. This move leads to research on everyday listening.

Implications for design: Given the range of audible differences that are referred to as timbre, there is obviously great potential to vary timbre to encode information. But given how little is known about what the perceptual dimensions of timbre might be, using timbre is also a tricky undertaking. Certainly systems which use more than one manipulation of timbre at a time should be founded on good theory or, better, should be tested experimentally.

Masking

- The loudness of a sound is context dependent. Sounds may *mask* each other, as when conversations in a noisy car become impossible.

- Masking depends on loudness; in general a loud sound masks a soft sound.

- Masking also depends on frequency. In general, sounds mask higher frequencies more than lower ones. Sounds within a critical band mask each other more than those more than a critical bandwidth apart.

- Sounds can be masked by other sounds that come before or after them; the masking sound does not have to be played at the same time as the masked sound.

- Masking can be predicted with great accuracy if the characteristics of ambient environmental noise is known (see Patterson, 1982 for a good example of sound design to avoid masking in a noisy environment).

- Complex tones are more difficult to mask than pure tones.

Implications for design: This again refers to context effects in hearing. Obviously an audio cue will be useless if it is masked. One can reduce the probability of a sound being masked by increasing the complexity of its spectrum. Of course, this needs to be done with care, since this could have the knock-on effect of increasing the chance that the tone may become a masker rather than a maskee - a situation that is equally undesirable. A way to reduce the probability of one signal masking another is to avoid sustained signals. In cases where the cue must be sustained over time, use repeating short sounds rather than a long one. The intent in this approach is to leave "space" where otherwise masked sounds can be heard.

Remember, however, that there must be enough space to ensure that forward and backward masking can't hide intended signals. There is a great deal of literature about masking; see in particular Boff & Lincoln (1988).

Auditory Streaming

Auditory Streaming refers to the tendency for separate sounds to be assigned to perceptual "streams, which correspond to different sources of sound. Streaming takes place both in grouping sequential sounds and in "fusing" partials in the perception of complex timbres.

• Figure 10 shows an example of streaming. In A, notes that are relatively close in frequency are alternated slowly. In this case, one perceptual stream is heard. In B, notes that are relatively far apart in frequency are alternated rapidly. In this case, two streams are heard; one of repeating low notes, and one of repeating high notes.

• Comparisons of timing are much more difficult between streams than within.

• Streaming is determined by:

• Frequency separation

• Tempo - the faster the tempo, the stronger the effect

• Timbre

• "Common Fate," the tendency of components to share similar temporal or frequency patterns.

Implications for design: Understanding streaming is crucial in developing good auditory cues, be they sequential or simultaneous. In particular, if sequences of sounds break into undesirable streams, or fail to be heard as separate streams, the intended message may be lost. For more information on streaming, see Bregman (1981) and McAdams & Bregman (1979).

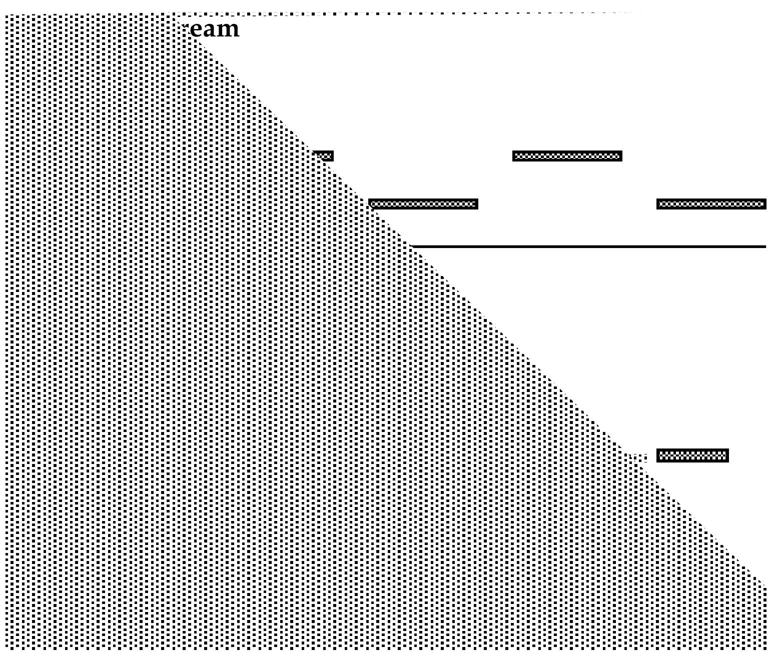


Figure 10: A typical example of streaming.

Changing Sounds

Changing sounds are the most noticeable and the most informative. Most simply, all sounds involve changes of air pressure; if such changes are not large enough or fast enough, we won't hear anything. But even if we can hear a sound, what we are most sensitive to are changes, transitions, or *transients*. Sounds that don't change quickly fade into the perceptual background.

- Change conveys information, and we tend to seek and attend to informative stimuli. Static sounds are moved to the background.

- Sensitivity to change holds at all levels. Steady state tones are less noticeable than changing ones (think of the rapid changes of sirens); steady rhythms -- which involve transients -- are less noticeable than changing ones (think of the steady "tick tick" of a clock). We seem to judge changes by the next higher derivative.

- Consider sitting in a restaurant when the air conditioner goes off. Two things become apparent: First, it may be that only when it goes off that you will become aware that it was ever on. This is even more surprising given the second observation, that it was incredibly loud when it was on, and yet you "didn't hear it." It was constant, carried no information, so you pushed it into the background of your consciousness.

- Consider walking down a street and hearing somebody walking behind you. Soon you will ignore the sounds, even though they contain many transients. But they will rapidly return to your attention should the person start running.

Implications for design: Audio monitors can be designed so that in normal circumstances they produce relatively steady sounds which fade into the background. Changes associated with abnormalities can be signaled by sudden changes in the sounds which will attract attention. Why use the steady state part at all? Think again about the air conditioning example. If somebody asked you "is the air conditioning on?" you could shift your attention and give the answer. The point is that auditory displays can be designed so that information is constantly available, but not obtrusive except in times of rapid and important change.

Psychoacoustic Illusions: Infinite Rising in Pitch

This effect is the audio equivalent to the rotating "candy-cane" like barber's pole: the pitch of the sound seems to rise (or fall) continuously. Like the barber's pole, the effect works by sneaking spectral components in at one end, and out at the other.

The effect was first demonstrated by Shepard (1964), and is properly known as *Shepard Tones.* His demonstration involved a chromatic scale (i.e., discrete notes rising step-wise at constant intervals of a semi-tone. Risset (1969) later demonstrated the same effect that more directly followed the barber-shop pole. Rather going up or down in discrete steps, the pitch changes was continuous, constituting the perceived effect of an infinite glissando – either up or down.

The mechanism underlying the effect is illustrated in Figure 12. It consists of a set of coherent sine waves, each separated by an octave. The entire set of waves is raised or lowered in pitch over time. The loudness of the sine waves varies, depending on their pitch. The amplitude/frequency function is a bell curve. Tones are loudest in the middle of the audible range, and softest at the extremes. Consequently, tones fade in and out gradually at the extremes. Since they are octaves (therefore harmonic) and coherent (therefore fuse spectrally), the overall effect is of a single rising or falling complex tone.

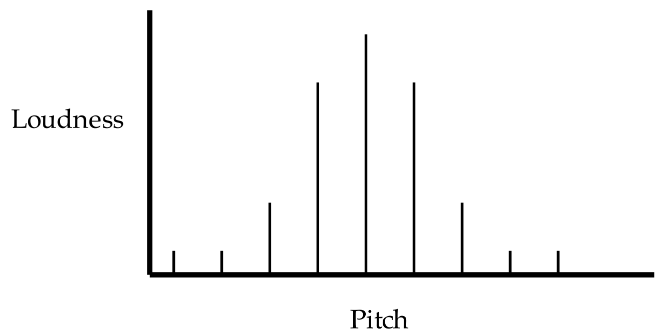


Figure 12: Shepard Tones - The illusion of an infinite rising or descending pitch.

Sound Example 2.?: Infinite Glissando - a Psychoacoustic Illusion. Three examples are provided. An upward gliss, a downward gliss, and one that glisses in both directions.

Psychoacoustic Illusions: Infinite Accelerando/Deccelerando
Furthermore, in 1974, a researcher at Bell Labs, Kenneth Knowlton (Risset 1986) demonstrated that the technique underlying Shepard tones could also be applied to tempo/rhythm. The result was the illusion of an infinite accelerando of a beat, such as a drum.

Sound Example 2.?: Infinite Accelerando and Deccelerando- a Psychoacoustic Illusion. First we hear the accelerando, then a deccelerand

Implications for design: The use of the techniques pioneered by Shepard (as applied to pitch, tempo, or both) have to potential to provide effective feedback as to the direction and rate of some background process – especially when one does not know how long the process will take. That is, since the effect is infinite, the feedback will be effective no matter how long the process takes, always continuing to rise or fall without ever falling outside of the limits of perception.

Spatial Hearing
The location of a sound is an important and potent cue that can be used when working with nonspeech audio cues. An attractive notion is to be able to distribute sounds in space (in one, two or three dimensions) in order to get many of the same types of benefits that one achieved from the spatial distribution of graphical objects in the desktop metaphor of graphical user interfaces and window managers. Ludwig, Pincever and Cohen (1990) and Cohen and Ludwig (1991), for example, have introduced and discussed the notion of *audio windows* to achieve these purposes.

The use of spatial cues is also useful in collaborative work. If events are accompanied by sounds, the location of those sounds can, for example, provide the cue as to who initiated specific events. Through spatialization, distributed workers can occupy a shared acoustic ecology that provides many of the awareness generating prompts and cues that are found and exploited in the natural everyday world.

There are three aspects of sound localization that we want to understand in order to make full use of available resources:

- the direction of the source, which is generally expressed in trms of its angular displacement, or *azmith,* on the horizontal and median (vertical) planes;
- the distance to the source
- if the source is not stationary, its movement.

Unfortunately, achieving all of these benefits is easier said than done. Sound localization is a complex issue. What follows below is a brief introduction of some of the issues, and some pointers to where help may be found. A relatively current detailed discussion of spatial hearing is provided by Blauert (1983).

Note: Need to include info on HRTFs, and the differences between 2-channel, stereo and full binaural 3D audio, along with implications for need to track head orientation & position and how modern IMUs and signal process open this up – when used with headphones. Design Implications for AR and VR, and tied very much to Cherry's Cocktail Party Effect.

*** ref ****

Begault, D.R. (1994). *3-D Sound for Virtual Reality and Multimedia.* Boston: Academic Press.

"The Cocktail Party Effect" (Cherry, 1953; Egan, Carterette & Thwing, 1954)

- We rely on binaural hearing to focus in on single sound source from among many.
- Spatially separating sounds helps clarity and lessens interference.
- It is easier to listen separately to two complex sounds if they occupy different positions in space, rather than a common one (Koenig, 1950; Schultz, 1962)

Azmith

- horizontal plane: resolution
- median plane: resolution
- perceptual mechanisms
  - timing cues
  - amplitude cues
- ambiguities
- head turning

Distance Cues

• amplitude

• reverb

• equalization

The Doppler Effect
The Doppler effect is an effect whereby the perceived pitch of a sound source changes due to its relative motion towards or away from the listener. This is an effect that we hear every day, such as when a car or train passes us: the pitch goes up as the sound source approaches, and goes down as it moves away. It is an important effect to understand if one wants to be able to simulate moving sound sources.

Why does the pitch change? As we already know, the pitch that you perceive a sound is primarily determined by the frequency at which succesive waves reach your ear. The shorter the interval between waves, the higher the pitch, and *vice versa.* Consider the case illustrated in Figure XX.

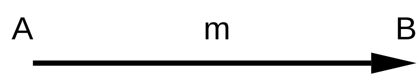


Figure XX: The Doppler Effect. A sound source of fixed frequency moves from A to B at 1/2 the speed of sound. A sound wave emitted at m will arrive in 1/2 the time of one that was emitted at A, thereby raising the perceived pitch by an octave.

Assume that a sound source is moving from point *A* to point *B* at half of the speed of sound. If a sound wave was initiated at point *A,* then by the time it reached point *B,* the source would have travelled half way from *A* to *B* (to point *m*). Since it only has half as far to travel, a second wave initiated at *m* would reach *B* in half the time that it would have had the source remained stationary at its original position, *A*. The effect of the early arrival would be a rise in the perceived pitch of the sound by an octave (since the period between successive waves has been halved).

Conversely, if the sound is moving away from the listener, the pitch goes down, since the interval between succesive waves is lengthened due to the added distance that the wave must travel compared to a stationary source. The effective period of the sound wave is 3/2 that of the stationary case (remember, we are assuming that the source is travelling at 1/2 the speed of sound), so the sound will be perceived a fifth lower than it would if it were stationary.

More formally, the frequency, f', at which a moving source of frequency f will be perceived can be stated as:

$$f' = f\,(c / (c + r \bullet v_S))$$

where: c is the speed of sound,
r is the unit vector in radial direction, and
v_S is the velocity of the sound source.

Note that while we have decribed the effect in terms of the source moving and a stationary listener, a comparable effect occurs if the source is stationary and the listener is moving. The next formula takes into account motion of both the source and the listener. Beware, however. Unlike the previous equation, we have simplified the math by making the assumption that the vector of motion is on the line intersecting the source and listener.

Note also that, as seen in the example above, for an object moving at constant speed, the magnitude of the pitch shift in approaching the listener is greater than that in moving away. In moving away, a negative value must be specified for velocity. As an exercise, verify the above example by calculating the frequency of a source of A 440 Hz moving towards and away from a stationary listener at half the speed of sound. Use either the formula above or the following one (Truax, 1978):

$$f' = f_s \times (v - v_o)/(v - v_s)$$

where: v is the velocity of the sound in the medium
v_o is the velocity of the observer relative to the medium
v_s is the velocity of the source relative to the medium
f_s is the frequency of the source while statinary

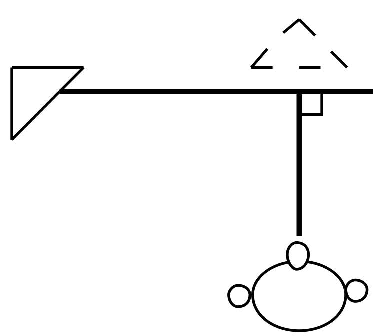
The net implication of all of this is that the frequency of simulated moving sounds must be shifted in accordance with the Doppler effect in order for motion to or away from the listener to be realistic. Examples of how this can be accomplished can be found in Chowning (1971) and Bosi (1990), for example.

Multi-channel vs stereo sound

- Multi-channel sound is where there are distinct channels (speaker, amp, etc.).
- Sounds are heard to be coming from a particular location because there is a real sound source at that location
- Stereo is the use of two real sound sources to create the illusion of a virtual (phantom) source *which is perceived to be as real as the real sources.*

Precedence / Hass effect (Snow, 1954; Hass, 1972; Wallach, Newmann & Rosenzweig, 1973)

- importance of listener position when creating a phantom source between two real sources
- technique of stereo is to present the same signal to two speakers
- if speaker on line at right angles to, and intersecting midpoint of, the line between speakers, the phantom will be heard.
- position of virtual source will depend on relative loudness of sound presented by each speaker.
- figure below illustrates case of placing phantom midway between real sources.



Stereophony: creating a phantom speaker using two real speakers.

- effect is lost if listener is closer to one speaker than the other
- problem is, with speakers, both speakers are heard by both ears

- if not equidistant, the sound image will be localized to the closest speaker only (although both speakers will contribute to loudness)

- when listening to a stereo, if you hear sounds from the far speaker when sitting close to the other, you are hearing *two channel* effects, not stereo (i.e., the sound you hear from the far speaker is not being presented from the near speaker.

Acoustic "Holograms" and other such phenomena

- we regularly hear of systems that can localize sounds "anywhere in a room," generating phantom sources using a low number of loudspeakers (often just two)

- any such system would break the rules of psychoacoustics as currently understood

- any such system that we are familiar with exploits some effect, or trick, has several limitations, and is not at all general (see Gardner, 1973, for an excellent example, which may well be of use in computer applications).

Multispeaker Approaches

- to avoid the problems of mobility of the listener when using speakers, some practitioners have designed multispeaker distribution systems, that is, used real sources

- see Fedorkow, Buxton and Smith (1978) for an example

Binaural Presentation

- if the problems of the precedence effect are due to both ears hearing the sounds from both speakers, then an alternative approach is to get around this by using headphones.

- such binaural presentation works well in terms of being able to place sounds at specified angles relative to the ears

- however, while the angle may be correct, there are often front/back ambiguities

- note that rotating our head is a prime strategy in localization

- since the headphones are attached to our head, they rotate too, so cue doesn't work

- solution is to track head position and orientation, and have real-time adaptive presentation of sounds

- sounds can then appear at specified locations in space, independent of head position

- first example was the *convolvatron* (Wenzel, Wightman & Foster, 1988a&b)

- a similar system has been developed by Bo Ghering

- work well, and are becoming less expensive

- but, have all of the problems of headphones, etc.

- see Wenzel, E.M., Wightman, F.L. & Kistler, D.J. (1991)

- see Kendall & Martins (1984) & Kendall et. al (1986) for additional background.

Perception of Emotion

see Dougherty (1993)

Summarizing Psychoacoustics

- Spectral components having coherent micro variations are heard as emanating from a single source through the process of *spectral fusion.*