5 Introduction to Pitch Perception

John Pierce

5.1 Pitch

Most musical instruments, except percussion instruments and some bells, have a clear pitch that is associated with the periodicity of the sound they produce. Nearly periodic sounds have nearly harmonic partials. A vibrating string or column of air produces nearly periodic sounds with nearly harmonic partials. One orchestral instrument can be tuned to the same pitch as another by ear, by judging when the two pitches are nearly the same. With greater precision, an instrument can be tuned to the same pitch as another by slowing the beats between the two simultaneous instrumental sounds.

But is pitch just periodicity, harmonicity, or the lack of beats? It is important to note that pitch is a perceptual measure, like loudness and some other terms we will use in this book. A machine that detects pitch would have to match human judgments of pitch. Just making a periodicity detector or a harmonicity detector would get us close, but it would not be enough because there are ambiguous cases where humans would assign a pitch and the machine would not. In this chapter we will look at various quantitative aspects of pitch.

5.2 Pitch and Brightness

In thinking about and experimenting with the pitch of musical tones, we must distinguish the sense of pitch from a sense of brightness or dullness. A sound is bright when its spectrum has many high-frequency partials. Musical tones ordinarily have no partials below their pitch frequency. Hence, tones of high pitch tend to be brighter than tones of low pitch.

Pitch depends on periodicity. Brightness or dullness depends on the distribution of total power between high and low frequencies. At the same pitch, the vowel /i/ (as in "beet") is brighter than the vowel /u/ (as in "boot"). At the same pitch the fraction of high-frequency power is greater in the spectrum of /i/ than in the spectrum of /u/. As another example, the sound of a trombone is brighter

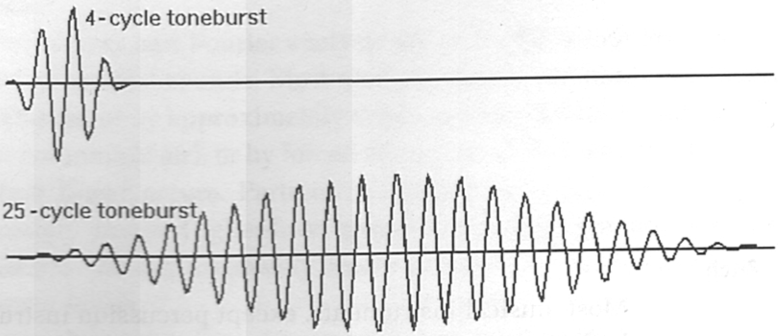

Figure 5.1 The upper tone burst would likely not produce a sensation of pitch, whereas the lower one would.

than the sound of a French horn, because the French horn is played facing away from the listener and the player places his hand in the bell.

A short click made up of low frequencies only sounds dull; a short click made up of high frequencies only sounds bright. Such a difference in brightness may or may not be associated with a difference in pitch. The waveforms in figure 5.1 are sine waves with a raised cosine envelope. The wave at the top is composed of four cycles of a sine wave. It sounds like a click, bright or dull depending on the frequency of the sine wave. It does not give any impression of a musical tone with a pitch. The waveform shown at the bottom has 25 cycles rather than 4. It gives a clear sensation of pitch, though there may also be a little sensation of click. If the number of cycles is made large enough, the "clicky" sensation disappears completely.

As the number of cycles in such a tone burst is increased from four, we begin to experience a sensation of pitch as well as a clicky sensation. Above some tens of cycles the clickiness fades away, and we have a sensation of pitch without click. The number of cycles necessary to give a pure pitch without a click increases somewhat with the frequency of the sine wave, but it lies in the range of tens of cycles. We will return to this sort of tone-burst waveform later.

Periodic waves are essential to the sensation of pitch. But we should keep in mind that a periodic sound wave can have a clear pitch only if it has a sufficient duration measured in cycles or periods.

5.3 Pitch and Partials

Musical tones have a number of harmonic partials whose frequencies are integer multiples of a fundamental frequency or pitch frequency, the frequency of a sine wave that matches the tone in pitch. The pitch frequency, the fundamental frequency (or, more commonly, fundamental), need not be present in the musical tone. In 1924, Harvey Fletcher of Bell Telephone Laboratories observed that for a sound with harmonic partials to be heard as a musical tone, its spectrum must include three successive harmonics of a common frequency. The pitch is given by that common frequency, whether or not it is present in the musical tone.

Orchestra chimes provide a striking example of a musical tone whose pitch frequency is entirely absent from the spectrum. The fourth, fifth, and sixth partials of the flexural mode have frequencies that are approximately the second, third, and fourth harmonics of a tone that just isn't there but is heard as the pitch frequency of the chime.

In the nineteenth century Helmholtz and his followers believed, erroneously, that the musical pitch of a sound is dependent on the actual presence in the sound wave of a sinusoidal component or partial of the pitch frequency. This misled Fletcher in the first of two papers he published in 1924 in *The Physical Review.* These papers recount experiments in listening to a number of instrumental sounds from which the fundamental or pitch frequency had been filtered out by a high-pass filter. Fletcher found that filtering out the fundamental, or even the fundamental and several successive low harmonics, did not change the pitch of the sound. Indeed, the identities of some sounds weren't much changed by the omission of lower harmonics. Instruments could still be easily recognized.

This is in accord with everyday experience. We hear the "right" pitches when we listen to a little radio that has a tiny speaker, which for low pitches puts out negligible power at the pitch frequency even when the pitch frequency is present in the original waveform. The telephone system rests on a frequency response that is not sufficient to carry the fundamental frequency of the lowest male voices, but we still hear voices as basically normal over the telephone. The most plausible and the true conclusion is that eliminating energy at the pitch frequency, and even at several low harmonics of the pitch frequency, doesn't necessarily alter our sensation of pitch.

In his first paper Fletcher proposed that the missing fundamental was re-created by nonlinearities in the mechanism of the ear. He soon abandoned this false conclusion. In the second paper he described experiments in synthesizing musical tones. It was these experiments that led to the assertion stated above, that a tone must include three successive harmonic partials in order to be heard as a musical tone, a tone that has the pitch of the fundamental, whether or not the fundamental is present.

In 1972, Terhardt published a theory of pitch perception, and later worked with others to define a computer algorithm that was capable of matching human pitch judgments on a number of sounds, including speech, bells, and spectra containing only octave partials. He defined a *virtual pitch* characterized by the presence of harmonics or near-harmonics, and a *spectral pitch* corresponding to individual audible pure-tone components. He stated that most pitches heard in normal sounds are virtual pitches, and this is true whether the fundamental is present in the spectrum or not. His algorithm calls for spectrum analysis using the Fourier transform, followed by the identification of important audible sinusoidal components, weighting the components, and allowing for masking effects (some components may cover up others nearby). The weighted sinusoids are then used to form many candidate virtual pitches that are the common divisors, or subharmonics, of the important component frequencies. These subharmonics are then combined and inspected to determine the most likely perceived pitch, and a measure of the saliency (how pitched it will sound, and how likely it is to be selected as the pitch) of that pitch is determined.

The mechanism of human pitch perception is different at low pitches and at high pitches. Fletcher's observations on pitch do not apply at very low frequencies, nor at frequencies well above 1000 Hz. Terhardt defines the crossover from virtual to spectral pitch to be at about 800 Hz, but his method depends on the selection of clear sinusoidal components in the spectrum. At very low frequencies we may hear successive features of a waveform, so that it is not heard as having just one pitch. For frequencies well above 1000 Hz the pitch frequency is heard only when the fundamental is actually present. Fletcher proposed that at high frequencies the pitch is given by the position of excitation along the cochlea, and at low frequencies by a time mechanism.

We should note that we do not sense any discontinuity or abrupt difference in sensation as we play notes from the lowest to the highest musical pitches. For periodic musical sounds, the two mechanisms overlap in frequency range. To investigate them we must tease them out with ingenious experiments.

One interesting experiment is to add successive equal-amplitude harmonics (integer multiples of a fundamental frequency), one by one, to a sinusoidal tone of some fundamental pitch frequency. It is

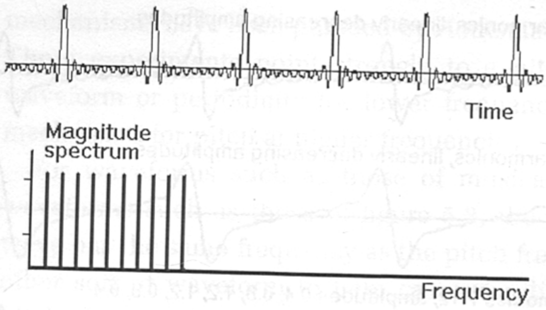

Figure 5.2 If we add successive harmonics to the fundamental frequency of a low-pitched sound, the sense of pitch grows stronger. However, a high-frequency component is also heard, associated with the wiggles between the peaks.

interesting to do this at frequencies of 55 and 440 Hz. One thing that we hear is a gradual *reinforcement* of the pitch frequency. But, in addition to the constant, low-pitch frequency, we get a sense of a higher frequency. This is most apparent in the 55 Hz tone.

Figure 5.2 shows the overall waveform of 10 successive harmonic cosine waves of equal amplitude. The big − to + excursion of every cycle accounts for a strong sensation of pitch at the fundamental frequency. But the strong wiggles between the peaks tend to create a sensation of a high pitch. It is as if we heard the shape of the waveform—a strong, low-pitched sound at the fundamental, and a weaker, high-pitched sound corresponding to the wiggles.

That seems to be how we hear sounds when the pitch is low enough. Somehow, the ear follows excursions of sound pressure. In some sense, it counts out the pitch at low frequencies.

We can reduce or get rid of the sensations of frequencies higher than the pitch frequency by choosing the relative amplitudes of the harmonics in a different way. Figure 5.3b shows the waveform for 12 sinusoids with relative amplitudes 1.2, 1.1, 1.0, 0.9, and so on. Figure 5.3a shows the waveform for six successive harmonics with relative amplitudes 1.2, 1.0, 0.8, 0.6, 0.4, and 0.2. The wiggles between the peaks are quite small in figure 5.3b, and moderate in figure 5.3a. The waveform of figure 5.3c contains only harmonics 7–12, with amplitudes 0.4, 0.8, 1.2, 1.2, 0.8, 0.4. We see that this last waveform has an envelope that goes periodically from great to small with the same periodicity as the other waveforms.

Suppose that at frequencies around 55 Hz and around 440 Hz we listen to sequences of three waveforms, the 12-partial waveform of figure 5.3a, the 6-partial waveform of figure 5.3b, and the waveform of figure 5.3c, which consists of only 5 high partials.

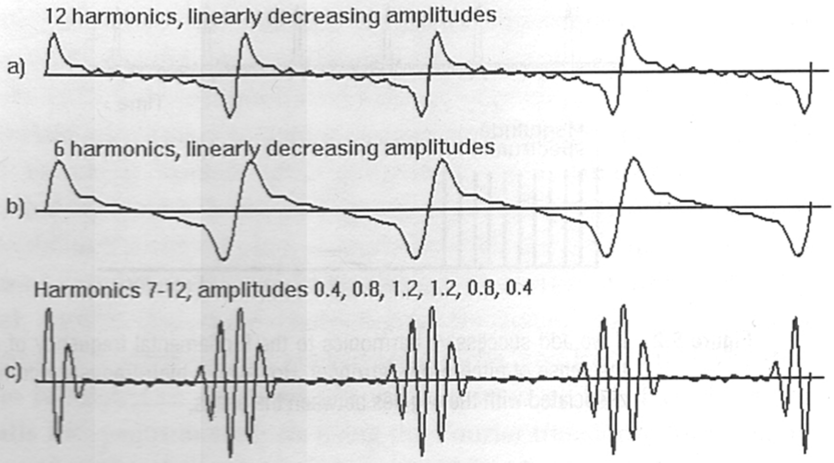

Figure 5.3 Waveforms with small wiggles. At around 55 Hz all three give the same pitches. Around 440 Hz the first two give a pitch frequency near 440 Hz, but the third gives higher frequencies.

The 12-partial tone includes frequencies from 1 through 12 times the pitch frequency. The 6-partial waveform includes frequencies from 1 to 6 times the pitch frequency. The third waveform is centered around a frequency 9.5 times the pitch frequency, and does not include the pitch frequency itself or any of its nearby harmonics.

What about the perceived pitches of these waveforms near 55 Hz and near 440 Hz? We could explore this question by playing a little tune with the three waveforms. Around 55 Hz we will hear the same tune for all three waveforms, with slightly different timbres. Waveform (c) has a somewhat "sharper" or more raucous timbre than the other two. Around 440 Hz we hear the same tune with waveforms (a) and (b). But with the third waveform we hear bright, faint sounds rather than the "correct" pitches that we get with (a) and (b).

Around 55 Hz the ear can "look at" the first 12 harmonics of the pitch frequency, or at the first 6 harmonics of the pitch frequency, or at harmonics 7 through 12 of the pitch frequency, and see a repetition rate equal to the pitch frequency. Around 440 Hz the ear can get the "right" pitch from the first 12 or 6 harmonics, but it can't get the "right" pitch from harmonics 7 through 12 of the pitch frequency. The time resolution of the ear isn't good enough to follow the envelope of waveform (c). Instead, we get a sound based on the frequencies of all the wiggles, small or large.

Experiments of this sort point strongly to what Licklider (1959) called a "duplex theory of pitch." Experiments supporting two pitch mechanisms have been pursued by Houtsma and Smurzinski (1990). These experiments point strongly to a pitch sensation based on waveform or periodicity for lower frequencies, and to some other mechanism for pitch at higher frequencies.

For waveforms such as those of musical instruments, and for waveforms such as those of figure 5.3, the envelope of the sound wave has the same frequency as the pitch frequency. We need some other sort of waveform to help cast more light on the question of pitch perception.

5.4 Experiments with Tone Bursts

Figure 5.4 shows patterns of tone bursts that are reminiscent of the waveforms of figure 5.1. These tone bursts are sine waves with a raised cosine envelope. In (a) all tone bursts have the same phase, and the fundamental frequency is given by the frequency of occurrence of tone bursts. In (b) every fourth tone burst is the negative of the other three. The four tone bursts constitute one period of the waveform, and the fundamental frequency is one-fourth of the tone-burst rate, that is, one-fourth of the fundamental frequency of waveform (a). This is exactly the same fundamental frequency as waveform (c).

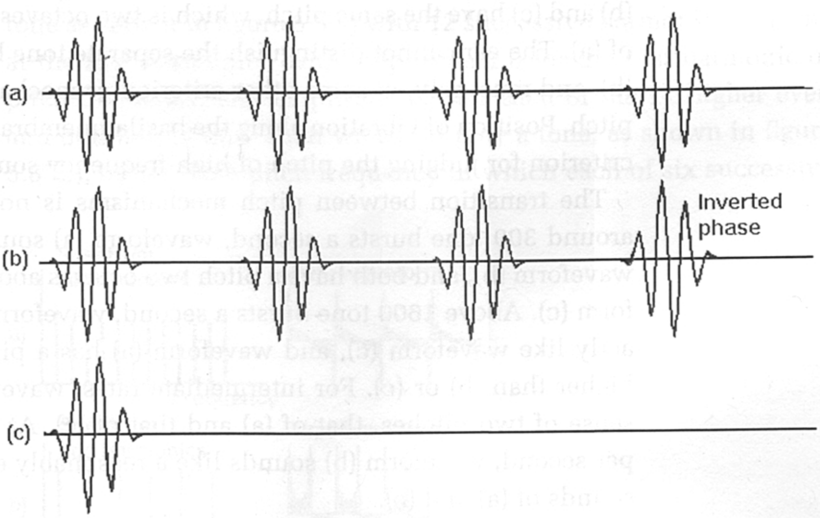

Figure 5.4 Three tone-burst sequences. Top: all tone bursts are the same. Center: every fourth tone burst is inverted. Bottom: one toneburst for each four of (a). At low rates the pitches of (a) and (b) are the same, and c) gives a pitch two octaves lower. At high rates (a) has a pitch two octaves above those of (b) and (c).

Indeed, the amplitude spectrum of (b) is the same as that of (c), but the relative phases of the sinusoidal components are different.

In using such waveforms to investigate pitch, it seems desirable to choose a number of cycles in the tone burst such that an individual tone burst gives no sensation of pitch. Suppose we choose to make the frequency of the sine wave of a tone burst 4000 Hz. By choosing a tone-burst rate for (a) of five tone bursts per second, we can easily hear successive tone bursts. We find that if the tone bursts are 64 cycles long, they give a strong sensation of pitch with no clickiness. However, tone bursts four cycles long give no sensation of pitch; they are strictly bright clicks.

In further experiments we can use 4000 Hz tone bursts that are four cycles long. At sufficiently high tone-burst rates we hear tones whose pitch increases with tone-burst rate. What are the relative pitches of the tone-burst patterns (a), (b), and (c)?

Up to around 300 tone bursts per second, (a) and (b) have the same pitch, which is two octaves above the pitch of (c). At low tone-burst rates the auditory system seems to follow the envelope of the waveform, and we can hear successive tone bursts rise and fall in amplitude. At low rates the auditory system seems to forget all about the preceding tone burst by the time the next one comes along.

At high tone-burst rates, above 1600 tone bursts a second or more, (b) and (c) have the same pitch, which is two octaves below the pitch of (a). The ear cannot distinguish the separate tone bursts in (a) and (b), and must rely on some other criterion or mechanism in judging pitch. Position of vibration along the basilar membrane is a plausible criterion for judging the pitch of high-frequency sounds.

The transition between pitch mechanisms is not abrupt. Up to around 300 tone bursts a second, waveform (a) sounds exactly like waveform (b), and both have a pitch two octaves above that of waveform (c). Above 1600 tone bursts a second, waveform (b) sounds exactly like waveform (c), and waveform (a) has a pitch two octaves higher than (b) or (c). For intermediate rates, waveform (b) gives a sense of two pitches, that of (a) and that of (c). At 640 tone bursts per second, waveform (b) sounds like a reasonably equal mix of the sounds of (a) and (c).

This seems to be conclusive evidence that there are two different mechanisms of pitch perception, a sort of counting mechanism valid at lower frequencies and another mechanism that takes over when the ear cannot "count" peaks in the time envelope of the exciting signal. Plausibly, this high-frequency pitch mechanism relies on the amplitude of excitation along the basilar membrane. The mechanisms appear to be about equally effective at a frequency around 640 Hz. This, which basically agrees with the work of Fletcher, Terhardt, and Licklider, is our overall conclusion about the pitch of periodic waveforms.

5.5 Odd Harmonics Only

Most musical instruments have spectra in which there are both even and odd harmonics. For sounds of low or moderate frequencies, the pitch of such tones is the fundamental frequency, even when there is no component of this frequency in the spectrum.

In the tone of a clarinet, even harmonics are much weaker than odd harmonics. Odd harmonics predominate in the spectra of closed organ pipes.

Digital sound generation enables us to create and explore various types of spectra. For example, we can synthesize sounds with no even harmonics. Will we hear the pitch frequency in the absence of the fundamental frequency? No! Even when the fundamental frequency is present, for sufficiently low pitch frequencies the pitch heard will be higher than the repetition frequency.

Figure 5.5 shows how to demonstrate this. We could first play a tone as shown in figure 5.5 a) with 12 successive harmonics (starting at the first harmonic), with amplitude 1.0 for the first harmonic or pitch frequency, and amplitude 0.7 for each of the 11 higher even and odd harmonics. Then we could play a tone, as shown in figure 5.5 (b), of the same pitch frequency in which each of six successive

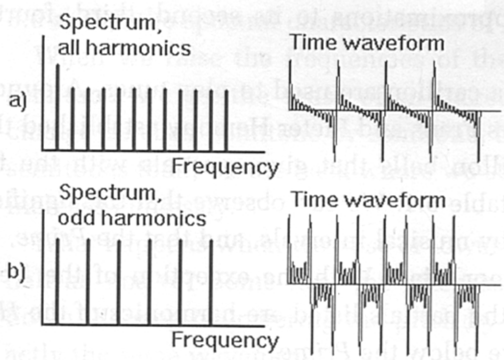

Figure 5.5 Two tones with a similar spectral range, power, and brightness. For a pitch frequency of 880 or 440 Hz the pitches heard are the same. For a pitch frequency of 55 or 110 Hz, the odd harmonic tone has a pitch greater than that of the upper tone.

odd harmonics has amplitude 1.0. The two tones have the same first harmonic, the pitch frequency or fundamental. The widths of the spectra of the tones are similar, so the brightnesses of the tones are similar. The total powers are nearly the same (each harmonic contributes a power proportional to the square of its amplitude).

Suppose the tones are played with pitch frequencies of 880, 440, 220, 110, and 55 Hz. At 880 and 440 Hz the two tones are heard as having the same pitch. At 110 and 55 Hz the odd harmonic tone is heard as having a higher pitch than the successive harmonic tone. When the pitch frequency is low, the loudness of the fundamental and other low harmonics is very much less than that of the higher harmonics. At low enough frequencies the fundamental and some other low harmonics cannot be heard. In the case of odd harmonics only, the frequency separation between successive odd harmonics is twice the pitch frequency. At low pitch frequencies the ear makes a pitch judgment that reflects the distance between the higher harmonics.

5.6 Chimes and Bells

The orchestra chime consists of a number of metal tubes that vibrate transversely after they are struck. The partials of a transversely vibrating tube are not harmonic. But, as we have noted earlier, the fourth, fifth, and sixth partials have frequencies very close to a harmonic series. That series has a pitch frequency that is 4.5 times the frequency of the lowest frequency of vibration, and the seventh partial is close to being the seventh harmonic of this absent frequency. We hear a pitch of this absent pitch frequency based on the presence of close approximations to its second, third, fourth, and seventh harmonics.

Bells in a carillon are used to play tunes. Around the year 1644, the brothers Frans and Pieter Hemony established the classical tuning of carillon bells that gives partials with the frequency ratios shown in table 5.1. We can observe that the significant partials are separated by musical intervals, and that the *Prime, Third,* and *Fifth* form a minor triad. With the exception of the frequencies of the *Third,* all the partials listed are harmonics of the *Hum tone,* which is an octave below the *Prime.*

If the Hemony third were a major third rather than a minor third above the prime, then the prime, third, and fifth would form a major triad, and all listed partials would be harmonics of a tone one octave

Table 5.1 The frequencies of a Hemony bell sound

PARTIAL	RELATION TO f_P, THE PERCEIVED PITCH
Hum tone	0.5 f_p, one octave down
Prime	f_p
Third	1.2 f_p (a minor third up)
Fifth	1.5 f_p (a fifth up)
Octave	2 f_p
Upper third	2.5 f_p
Upper fifth	3 f_p

below the hum tone. Might this not result in more pleasing carillon sounds? In 1987 a collaborative effort between the Dutch Institute for Perception in Eindhoven and the Royal Eijbouts Bell Foundry in Asten resulted in the casting of a four-octave transportable carillon of major-third bells. The major-third bell indeed sounds less dissonant than the traditional Hemony bell, and has been received with a good deal of favor.

5.7 Pitch and Unusual Tones

There has been a good deal of work relevant to pitch that makes use of computer-generated sounds. In 1964, Roger Shepard published material on what have come to be called Shepard tones. These tones consist of a number of sinusoidal components an octave apart, with a fixed envelope that goes to zero at low and high frequencies. Figure 5.6 shows the spectral characteristics of Shepard tones.

When we raise the frequencies of the sinusoidal components a semitone, we get the sense of an increase in pitch. This sense of change persists, semitone by semitone, but when we've reached 12 semitones total, we are back where we started. The pitch appears to increase endlessly.

What happens when we raise or lower the pitch by six semitones, half an octave? Some hear the pitch as going up; some, as going down. Raising or lowering the pitch by six semitones gives us exactly the same waveform.

We may note that Shepard tones are somewhat different from musical tones. Though they have many partials, they omit partials other than those an octave apart. Also, there is no lowest partial. As we

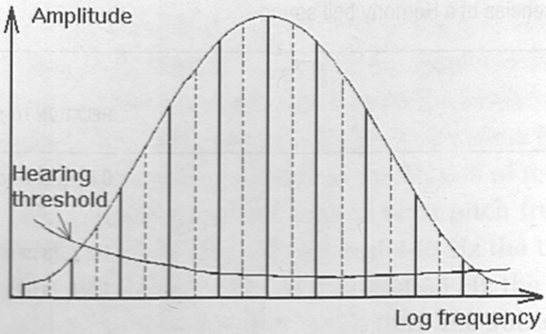

Figure 5.6 Frequency components of a Shepard tone. The envelope goes from left to right, from an inaudible level to an audible level and back down to an inaudible level. A tone has partials separated by an octave. (Modified from Shepard, 1964. Reprinted with permission, © 1964 Acoustical Society of America.)

continually increase the frequency of the octave partials in order to give a higher pitch, a partial will eventually attain audible amplitude, rise in amplitude, fall in amplitude, and sink into inaudibility. There are always more low and high partials available, but they're inaudible due to the imposed spectral envelope and the threshold of hearing.

Successive Shepard tones need not be separated by a semitone. Shepard tones separated by a tenth of an octave can be synthesized as nine equal-amplitude octave partials of a fundamental frequency A0 (27.5 Hz) and of successive fundamentals 0.1 octave, 0.2 octave, 0.3 octave, and so on, to 0.9 octave, then back to A0, and on up again. The sensitivity of human hearing provides the required "attenuation" at low and high frequencies.

Let us consider the pitch of another sort of computer-generated tone. What happens to the pitch of a tone when we double the frequencies of all its frequency components? Jean-Claude Risset has produced tones for which the pitch decreases a little when we double the frequencies of all partials. The tones are made up of successive partials a little more than an octave apart. That is, the ratio of the frequencies of two successive partials is a constant slightly greater than 2. When we abruptly double the frequencies of all partials, the ear hears each frequency component as "replaced" by a partial of slightly lower frequency. The lowest frequency vanishes, and a partial appears with a frequency slightly more than an octave above the previously highest partial. The ear can disregard this disappearance and appearance of partials.

To illustrate this, we could synthesize a tone with nine partials, each successive partial having a frequency 2.1 times that of the preceding partial. When the frequency is shifted from a "fundamental" of 27.5 Hz to a "fundamental" of 55 Hz, the perceived pitch obligingly falls a little.

Risset's tones started with a partial at some plausible and highly audible pitch frequency. The effect isn't as strong or foolproof in this more musical form. Risset also produced tones of ambiguous pitch, in which the pitch heard is dependent on context.

5.8 Concluding Remarks

Our sense of pitch is essential in almost all music. *Pitch frequency* can be related to the fundamental frequency of the sound wave, the frequency of the lowest harmonic or fundamental, whether or not a component of that frequency is actually present in a sound wave. We sense changes in pitch as similar both in notes of low pitch frequency and in notes of high pitch frequency.

There are two mechanisms of pitch perception. One is a sort of counting out of the cycles of a periodic musical tone. This is the dominant mechanism for pitch frequencies below a few hundred Hz. The other mechanism must be based on the place of excitation along the basilar membrane. This mechanism is dominant for pitch frequencies above perhaps 1600 Hz. In between these ranges of complete dominance, both mechanisms operate. For musical sounds, for which the fluctuation of the envelope is equal to the fundamental frequency (which may be present or absent), the mechanisms agree in their estimate of pitch. For tones of low pitch that have odd harmonics only, the pitch may be different from the frequency of the fundamental.

By using computers, tones with peculiar pitch properties have been devised. These peculiar properties do not occur in the tones produced by traditional musical instruments. They are useful experimentally, and may be useful musically.

References

In taking the work of Fletcher into account, it is important to note that he does not discuss his work on pitch in his book.

Fletcher, H. (1953). *Speech and Hearing in Communication*. New York: D. Van Nostrand. Republished in 1995 by the Acoustical Society of America, New York. The republished

version, edited by Jont B. Allen, includes a complete bibliography of Fletcher's publications, through which the 1924 *Physical Review* papers and other papers that discuss pitch can be found.

It appears that Schouten was not aware of Fletcher's earlier work, nor were many of his readers. Schouten is responsible for the term *virtual pitch,* used to describe pitch in the absence of the fundamental, as contrasted to the pitch of tones with a fundamental. Of course, no one can separate pitch and virtual pitch while listening to music.

Schouten, J. F., Ritsma, R. J., and B. L. Cardozo. (1962). "Pitch of the residue." *Journal of the Acoustical Society of America,* 134, 1418–1424.

Terhardt's pitch model seems to work well for most signals encountered in music, but does not exactly match human responses on particular types of specially designed tones, such as Risset's tones and the pulse trains described in this chapter. Here are the fundamental Terhardt references:

Terhardt, E. (1972). "Perception of the Pitch of Complex Tones." *Acustica,* 26, 173–199.

Terhardt, E., G. Stoll, and M. Seewann. (1982). "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals." *Journal of the Acoustical Society of America,* 71, 679–688.

Risset's paradoxical sounds:

Risset, J. C. (1969). "Pitch Control and Pitch Paradoxes Demonstrated with Computer-Synthesized Sounds," *Journal of the Acoustical Society of America,* 46, 88. (Abstract).

Mathews, M. V., and J. R. Pierce, eds. (1989). *Current Directions in Computer Music Research.* Cambridge, Mass.: MIT Press. Compact disc sound examples for the book are available from the publisher. Chapter 11, "Paradoxical Sounds," by Jean-Claude Risset, is particularly appropriate. Unfortunately, chapter 14, "Residues and Summation Tones—What Do We Hear?," by John Pierce, is erroneous, as are the sound examples.

Other references on pitch perception:

Greenberg, S., J. T. Marsh, W. S. Brown and J. C. Smith. (1987). "Neural Temporal Coding of Low Pitch. I. Human Frequency-Following Responses to Complex Tones." *Hearing Research,* 25, 91–114.

Houtsma, A. J. M., and Smurzynski, J. (1990). "Pitch Identification and Discrimination for Complex Tones with Many Harmonics," *Journal of the Acoustical Society of America,* 87, 304–310.

Licklider, J. C. R. (1959). "Three Auditory Theories." In S. Koch, ed., *Psychology: A Study of the Science,* vol. 1. New York: McGraw Hill.

Nordmark, J. O. (1968). "Mechanisms of Frequency Discrimination." *Journal of the Acoustical Society of America,* 44, 1533–1540.

Pierce, J. R. (1991). "Periodicity and Pitch Perception." *Journal of the Acoustical Society of America,* 90, 1889–1893.

6 What Is Loudness?

Max Mathews

What does "loud" mean? Can you measure it with a meter? How loud is "twice as loud"? In chapters 1 and 2 we considered how the ear and the brain work. In chapter 5 we discussed some fundamentals of pitch perception. Here we will concentrate not on the mechanics of how our sensors work but on how we eventually perceive auditory phenomena. As with the chapter on pitch, we will not consider physically measured things, such as intensity, but rather how we perceive intensity as loudness.

6.1 Equal Loudness Contours

Figure 6.1 shows an important, classical diagram known as the Fletcher–Munson curves. We were first introduced to these curves, which show equal loudness contours for sine wave sounds, in chapter 4. Frequency in Hz is shown left to right, from 20 Hz at the left to over 10,000 Hz at the right. Sounds of different frequencies whose intensities lie along any one of these curves sound equally loud. Along the ordinate (vertical axis) are the actual sound pressures (or intensity) in decibels.

As you can see, some of the equal loudness curves go up and down quite a lot. For example, in order for a soft sound at 50 Hz to sound as loud as one at 2000 Hz, the 50 Hz sound has to be about 50 dB more intense than the 2000 Hz sound. That is an enormous difference of 100,000 times in power.

What this tells us is that our auditory systems are much more sensitive at 2000 Hz than at 50 Hz. You may have noticed that you can barely hear the 60 Hz hum that is present in cheap audio equipment. If the equipment produced the same intensity of "hum" at 2000 Hz, it would drive you out of the room.

If you examine the curves of figure 6.1 a little more closely, you can see that the auditory system is most sensitive at a frequency of about 2000 Hz. Some of the curves show the maximum sensitivity nearer to 3000 Hz. Musically, 2000 Hz is a fairly high pitch; most singers cannot produce a pitch that high. But they might get up to

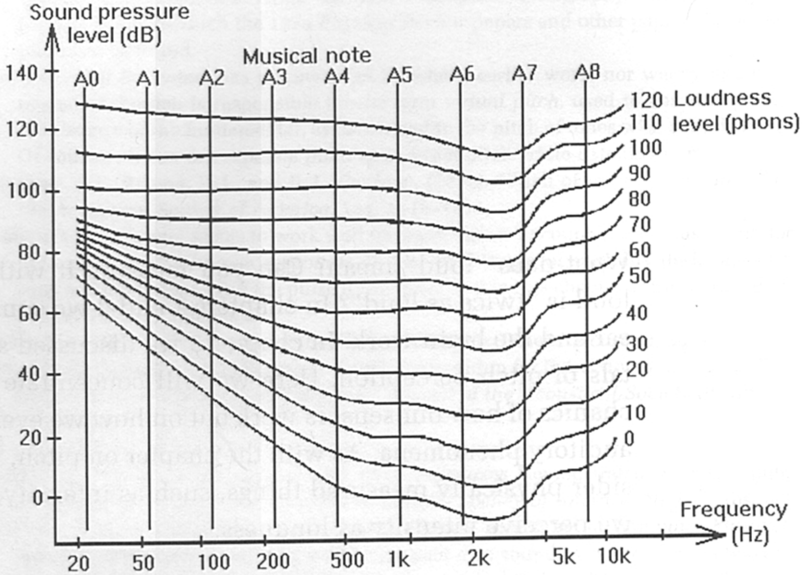

Figure 6.1 Fletcher–Munson equal loudness contours show the intensity at which a sine wave must be presented, as a function of frequency, in order to sound equally loud as sine waves at other frequencies on the same curve.

1000 Hz, and really blast you. Moreover, musical tones of lower pitch have harmonics in the range of 2000 Hz and above, and a part of the loudness of musical tones is contributed by their harmonics.

At low frequencies below 100 Hz, the auditory system is not very sensitive. And at high frequencies of 10,000 Hz and above, it also is insensitive. How fast the sensitivity of the ear falls off at high frequencies depends on how old you are. My hearing (at age 60) falls off very rapidly around 10,000 Hz, whereas most young people can hear sounds up to 15,000 to 20,000 Hz.

Musically, the most important frequencies lie between 100 and 5000 Hz. As discussed in chapter 5, musical tones with pitch frequencies below 100 Hz are heard chiefly through their harmonics. It's amazing how little we lose by eliminating all frequencies below 100 Hz and above 5000 Hz. AM radio stations don't broadcast above 5000 cycles. Small radio sets don't produce much intensity at 100 Hz and below. Lower and higher frequencies do make a difference, however, that we can appreciate in good hi-fi systems.

There are a lot of curves in figure 6.1. What are the differences between them? It's the intensity of the sound. Consider the lowest curve. For this curve, the intensity of sound at the auditory system's most sensitive frequency is 8 dB greater than the threshold of perception at that frequency. This bottom curve is close to the softest sound that the average person can hear at most frequencies.

The number attached to a particular curve is its loudness level in *phons,* defined by the shape of the curve itself. The intensities along the curve, including the minimum intensity, have a common loudness level, specified in the number of phons that labels the curve. These curves are 10 dB apart. What do the decibels refer to? A sound-level meter will read such intensities in SPL (sound pressure level), which is a measure of sound intensity. Sound-level meters are calibrated in SPLs.

If you read a number on a sound-level meter, for sine waves you will get SPLs in decibels above the threshold of hearing. To get the loudness level for a complicated sound such as speech, you must find the loudness level in phons of a sine wave that sounds equal in loudness to the sound you're trying to quantify.

The lower curves in figure 6.1 represent soft sounds. Normal speech sounds are 70 to 80 dB above the levels of the lowest curve. The ambient sound level in a good empty auditorium will be around 30 dB. It takes good sound insulation to get the ambient noise much below 30 dB.

We must remember that the curves of figure 6.1 are for sine waves. Speech and noise have many frequencies. For weak sounds, some frequency components may lie below threshold, so that their power, which is included in the SPL reading, may actually not contribute to the audibility of the sound. In an extreme case, a 20 Hz sound with a 30 dB SPL would be completely inaudible.

At the top of the set of curves, the sounds at 120 dB approach the threshold of pain. That hurts, and is dangerous. If you are exposed to levels like that for very long, you are apt to have damage to your hearing—you'll have what's called a "threshold shift." A little exposure will result in temporary shift, and you can suffer permanent damage through prolonged or repeated exposure. As a matter of fact, sound levels above 100 dB should be experienced with caution, and not for very long.

I once went to a Grateful Dead concert and sat behind the mixing console. The Grateful Dead concerts (and many others) are very carefully controlled by the engineer doing the sound mixing. The musicians don't have anything to do with it during the concert. All the lines from the pickups and microphones lead to the mixing console, which is always at a given location, 80 or so feet from the two towers of loudspeakers on the stage. The mixing console can adjust the levels of all the sounds of the instruments. I watched the sound-level meter during part of the concert. It never went below 100 dB, and it never went above 110 dB. Later, I asked why this was so. I was told that if the level went below 100 dB, the audience got restless. If it went above 110 dB, there was so much feedback to the stage that the performers couldn't play. I didn't ask about hearing damage.

Returning to the curves of figure 6.1, we should note that for soft sounds there is a great difference in the auditory system's sensitivity to different frequencies. The lower curves have a big dip in them. For loud sounds, even for 90 phons, there is much less variation with frequency. There is a moderate amount of variation for sounds of intermediate intensities.

Musically, this variation is important. For one thing, this tells us that if we want a musical instrument to be heard easily, its frequencies should lie between, say, 500 Hz and a few thousand Hz. If you have an instrument that produces low frequencies only, it can't be perceived as loud. If you have a high-frequency instrument, it won't be quite as loud as a midfrequency instrument. To get a very loud sound you need a lot of energy in the frequency range from 500 to a few thousand Hz.

Another conclusion is that when your are playing fortissimo, low frequencies contribute relatively more than in a soft passage. In connection with the things I have covered in this chapter, and in connection with other matters concerning the hearing of sound, I advocate the "I'm from Missouri" approach. You shouldn't believe things unless you can hear them. You should listen to sounds and decide whether or not to believe what you are told. Data are average data, and you weren't included in the averaging. The course I advocate is best carried out by listening to synthesized sounds.

6.2 Synthesizing and Listening to Examples

There are various ways of producing computer-generated sounds, but most are related by the ability to manipulate sine waves in frequency, phase, and amplitude. Some systems allow for control of noise and other characteristics. With computer-generated sounds we can try to confirm what has been said about various features of the Fletcher–Munson (equal loudness) curves. One can adjust the relative intensities of sine tones at 50 Hz and at 2000 Hz until they seem to sound equally loud, and compare the result with the published curves. In such a demonstration it is desirable to be able to see the waveform and a spectrum display, as well as to hear the sounds, so that you're sure about the sound to which you're listening.

In a lecture demonstration I use when teaching the psychoacoustics course, sounds are played at various relative levels in decibels SPL. The level of the sinusoidal sound in the room is adjusted by using a sound-level meter.

The class votes as to which frequency sounds louder. As the level of the 50 Hz sound is increased in steps, from a level equal to that of the 2000 Hz sound to a higher and higher level, the vote finally changes from the 2000 Hz tone sounding louder to the 50 Hz tone sounding louder. The procedure is tried with softer sounds, and can be used with other frequencies as well. The results come out strikingly close to those shown in figure 6.1. But they are not as precise, for the levels are adjusted in 10 dB steps.

The large change in relative intensities is sensible for a demonstration in a room in which the relative intensities at various locations may be different for different frequencies. The data for the curves of figure 6.1 were of course taken with earphones.

Another important demonstration is "How much is a decibel?" I do this at 440 Hz, the A above middle C. This frequency is chosen to avoid changes in intensity that can be caused by varying head position in a real room with resonances (remember that the wavelength of 440 Hz is about 2.6 feet). Successive sounds are played 1 dB apart; the difference in intensity is difficult to hear. Musically, a 1 dB difference in intensity doesn't produce any impact. A 6 dB difference is easy to hear, but not very impressive as a contrast in music. Six dB is a factor of 4 in the intensity or power of a sound, or a factor of 2 in pressure. Twenty dB is a factor of 100 in intensity or 10 in pressure—a pretty big change. Forty dB is a really big change, from barely audible in a "noisy" room to a good loud sound. It is about the largest range of musically significant sound in ordinary circumstances. In a quiet concert hall you can't do much better than 60 dB.

Musically, if you want to make a contrast, you have to do more than just change the intensity of the sound. You have to change the complexity of the sound, and perhaps the harmony. A composer must do more than just change intensity in going from soothing to startling.

Again we must remember that figure 6.1 is for sine waves. Most demonstrations of sounds are for sine waves. The voice, music, and all natural sounds are not sine waves; they are made up of many partials or frequency components, and often of nearly harmonic partials. We must take into account the data we get with sine waves, but we must not let ourselves be misled by them. We have considered the perception of equal loudness, and the perception of more or less loudness. But does it make sense to call one sound twice as loud as another, or half as loud as another, or ten or a hundred times as loud as another?

6.3 What's Twice as Loud?

In order to answer this question, you have to be willing to make a different kind of judgment. This new kind of judgment is harder to make, but if enough people agree, it will be one we can rely on. So far we have talked about listening to two sounds and saying which one is louder, or we have adjusted the level of one of two sounds until the sounds are equally loud.

We can present a sound, and then follow it with another sound that is more intense than the first by some number of decibels. Then we can ask, "Is the second sound more or less than twice as loud as the first?" This has been done both in classroom demonstrations and in controlled research labs. The judgments by vote showed that most people heard a 5 dB more intense sound as less than twice as loud, and most people said a 10 dB more intense sound was at least twice as loud. Thus, somewhere between 5 and 10 dB difference in intensity is twice as loud.

What do researchers find when they carefully run such an experiment with earphones, and repeat it many times with many people? What they find for 1000 Hz sine tones is shown in figure 6.2. At moderate-to-high levels a difference in intensity of around 10 dB doubles the loudness measured in *sones* (the sone is defined as a perceptual comparative loudness relative to a sine wave at 1000 Hz, 40 dB SPL). But at lower intensities, and especially near the threshold of hearing, the loudness in sones increases more rapidly with intensity in decibels. That makes sense, because if you compare a sound you can barely hear with a sound just a few decibels more intense, the latter seems bound to sound much louder.

Figure 6.2 holds only for the frequency 1000 Hz. Curves have been measured and plotted for several frequencies across the audio range but all curves look about the same in shape except for the one at 100 Hz. At high enough levels the curves have a slope of about 10 dB for doubling the loudness. Fletcher, Stevens, and others have

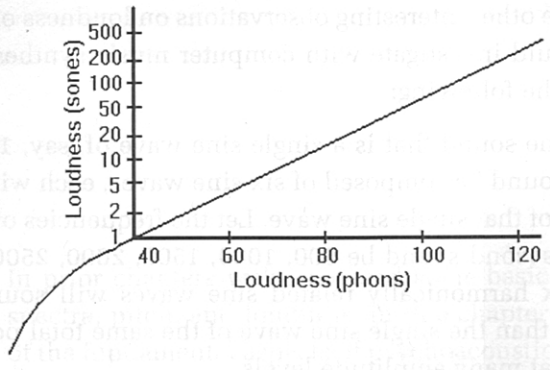

Figure 6.2 Loudness in sones versus loudness level in phons at 1000 Hz.

proposed that in the normal-to-high range, the loudness varies as the third root of the intensity; that would give about 9.4 dB for doubling the loudness, and 10 dB is close enough to this level to be a good rule of thumb. The region of loudness for which this linear variation of loudness versus intensity holds is the region that is of interest in music. In this region we can get, for sine waves, the loudness versus intensity level at any frequency from figure 6.1, the equal-loudness contour.

We have now done two things. We have investigated contours of equal loudness, with each contour labeled in phons. We also have made a scale of relative loudnesses in sones. The equal loudness contours involve only quantities that you can read on a meter, once you have made an adjustment for equal loudness. The relative loudness in sones cannot be read on a meter. It depends on consistent human judgment of what intensity ratio doubles the perceived loudness.

6.4 Complex Sounds

What about more complex sounds? We can adjust the level of a sine wave of some frequency so that the sine wave sounds as loud as some other sound. The loudnesses will not necessarily seem the same if we raise or lower the levels of the two sounds by the same number of decibels. Fletcher's experimental conclusion was that putting the same sound in both ears makes it sound twice as loud as if it were heard monaurally. Experiments also indicate that the loudnesses of two sounds in the same ear add, but only if the sounds are separated enough in frequency so that one doesn't interact with or mask the other.

Some other interesting observations on loudness of spectra, which you could investigate with computer music synthesis software, include the following:

Take one sound that is a single sine wave of, say, 1000 Hz. Let another sound be composed of six sine waves, each with one-sixth the power of that single sine wave. Let the frequencies of the sine waves in the second sound be 500, 1000, 1500, 2000, 2500, and 3000 Hz. The six harmonically related sine waves will sound a good deal louder than the single sine wave of the same total power. See if this is true at many amplitude levels.

Compare the loudness of the six harmonic waves above with a sound composed of six inharmonic frequencies (e.g., 500, 1100, 1773, 2173, 2717, 3141 Hz). Which do you think will be louder?

Reference

Moore, B. C. J. (1989). *An Introduction to the Psychology of Hearing.* Third edition. London: Academic Press.