7

Pitch and Periodicity Coding

Most sound sources involve a vibrating object of some kind. Regular vibration produces sound waves that *repeat over time*, such as the pure and complex tones described in Chapters 2 and 3. Within a certain range of repetition rates, we perceive the periodic sound wave as being associated with a *pitch*. It is important for us to be able to identify the repetition rate or fundamental frequency of sounds. The variation in the fundamental frequency of vowel sounds in speech can be used to convey prosodic information. Many musical instruments produce complex tones, and the fundamental frequencies of these tones can be varied to produce different melodies and chords. Less obviously, differences in fundamental frequency are very important in allowing us to *separate out* different sounds that occur at the same time, and *group together* those frequency components that originate from the same sound source. These aspects are discussed in Chapter 10.

This chapter will concentrate on how the auditory system represents—and then extracts—information about the periodicity of a sound waveform. The discussion will begin with our perceptions, and go on to describe the auditory mechanisms that may form the basis for these perceptions.

7.1 PITCH

7.1.1 What Is Pitch?

In auditory science, pitch is considered, as is loudness, to be an attribute of our *sensations*, and the word should not be used to refer to a *physical* attribute of a sound (although it often is). So you should not say that a tone has a "pitch" of 100 Hz, for example. When we hear a sound with a particular fundamental frequency, we may have a sensation of pitch that is related to that fundamental frequency. When the fundamental frequency increases, we experience an increase in pitch, just as an increase in sound intensity is experienced as an increase in the sensation of loudness.

The American National Standards definition of pitch reads as follows: "Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends mainly on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus." (ANSI, 1994, page 34).

I am not a big fan of this definition, because it seems to be too broad. It requires the words "low" and "high" to be associated with pitch or frequency, rather than with loudness or intensity, for example. I will adopt a narrower definition here, and define pitch as that aspect of auditory sensation whose variation is associated with *musical melodies*. In other words, if a sound produces a sensation of pitch, then it can be used to produce recognizable melodies by varying the repetition rate of the sound. Conversely, if a sound does not produce a sensation of pitch, then it cannot be used to produce melodies. This definition is consistent with what some researchers regard as an empirical test of the presence of pitch: If you can show that a sound can produce melodies, then you can be sure that it has a pitch (e.g., Burns & Viemeister, 1976).

7.1.2 Pure Tones and Complex Tones

In Chapter 2, I introduce the sinusoidal pure tone as the fundamental building block of sounds and, from the point of view of acoustics, the simplest periodic waveform. Broadly speaking, a complex tone is "anything else" that is associated with a pitch or has a periodic waveform, and usually consists of a set of pure-tone components with frequencies that are integer multiples of the fundamental frequency of the waveform (see Section 2.4.1). Within certain parameters, both pure and complex tones can be used to produce clear melodies by varying frequency or fundamental frequency, and hence they qualify as "pitch evokers" according to my definition. Furthermore, if the fundamental frequency of a complex tone is equal to the frequency of a pure tone, then the two stimuli usually have the same pitch, even though the *spectra* of these sounds might be very different. What seems to be important for producing the same pitch is that the *repetition rate* of the waveform is the same (see Fig. 7.1). In a sense, pitch is the *perceptual correlate* of waveform repetition rate.

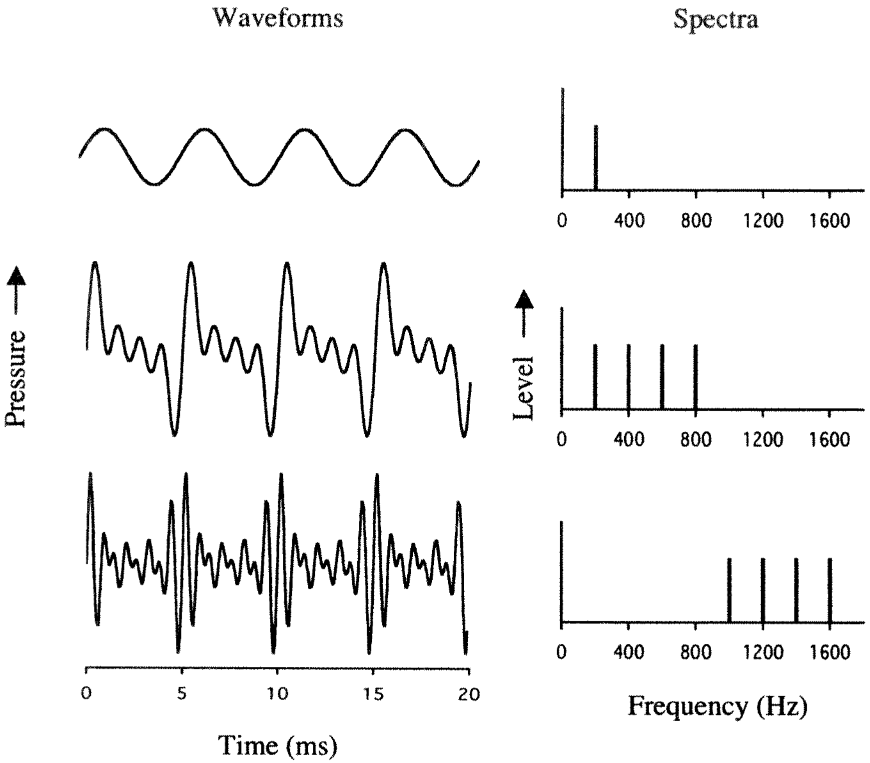

FIG. 7.1. The waveforms (left) and spectra (right) of three sounds, one pure tone and two complex tones, that have the same pitch. Notice that the similarity between the sounds is in the waveform repetition rate, and not in the overall spectrum.

Research on pitch perception has tended to be divided into research on pure tones and research on complex tones, and I am going to keep this distinction. I am not saying that the pitch of pure tones is in some way a dramatically different percept to the pitch of complex tones. After all, a pure tone is a complex tone containing just the first harmonic. Because of their simplicity, however, pure tones have a special place in the heart of an auditory scientist, and have been given much attention. Understanding how simple sounds are perceived can help our understanding of the perception of more complex stimuli, and in many cases it is possible to extrapolate findings from pure tones to complex tones.

7.1.3 The Existence Region for Pitch

What are the lowest and highest frequencies, or fundamental frequencies, that can be used to produce a pitch? Put another way, what is the frequency range, or *existence region*, for pitch? In regard to pure tones, there is a reasonable consensus on the upper limit. Studies have indicated that frequencies above about 4000–5000 Hz

cannot be used to produce recognizable melodies (Attneave & Olson, 1971). This is similar to the highest *fundamental* frequency of a complex tone that can be used to produce a pitch, as long as it has a strong first harmonic. Effectively the pitch percept is dominated by the first harmonic at these high rates, and because the frequency of the first harmonic is equal to the fundamental frequency, the upper limit is similar. It can surely be no coincidence that the highest note on an orchestral instrument (the piccolo) is around 4500 Hz. Melodies played using frequencies above 5000 Hz sound rather peculiar. You can tell that something is changing, but it doesn't sound musical in any way.

Complex tones, however, can produce a pitch even if the first harmonic is not present (see Section 7.3.1). In these situations, the range of fundamental frequencies that produce a pitch depends on the harmonics that are present. Using a complex tone consisting of only three consecutive harmonics, Ritsma (1962) reported that, for a fundamental frequency of 100 Hz, a pitch could not be heard when harmonics higher than about the 25th were used. For a 500-Hz fundamental, the upper limit was only about the 10th harmonic. It appears that the higher the fundamental, the lower the harmonic numbers need to be (not to be confused with a low number of harmonics!).

At the other end of the scale, for broadband complex tones containing harmonics from the first upward, melodies can be played with fundamentals as low as 30 Hz (Pressnitzer, Patterson, & Krumbholz, 2001). This value is close to the frequency of the lowest note on the grand piano (27.5 Hz). In summary, therefore, the range of repetition rates that evoke a pitch extends from about 30 Hz to about 5000 Hz.

7.2 HOW IS PERIODICITY REPRESENTED?

The first stage of neural representation occurs in the auditory nerve. Any information about a sound that is not present in the auditory nerve is lost forever as far as the auditory system is concerned. This section will focus on the aspects of auditory nerve activity that convey information about the periodicity of sounds.

7.2.1 Pure Tones: Place and Time

In Chapters 4 and 5 it is described how different neurons in the auditory nerve respond to activity at different places in the cochlea (*tonotopic* organization). In response to a pure tone, the firing rate of a neuron depends on the level of the pure tone (the higher the level, the higher the firing rate) and the frequency of the pure tone (the closer the frequency to the characteristic frequency of the neuron, the higher the firing rate). It follows that the firing rates of neurons in the auditory nerve provide information about the frequency of the pure tone. The characteristic

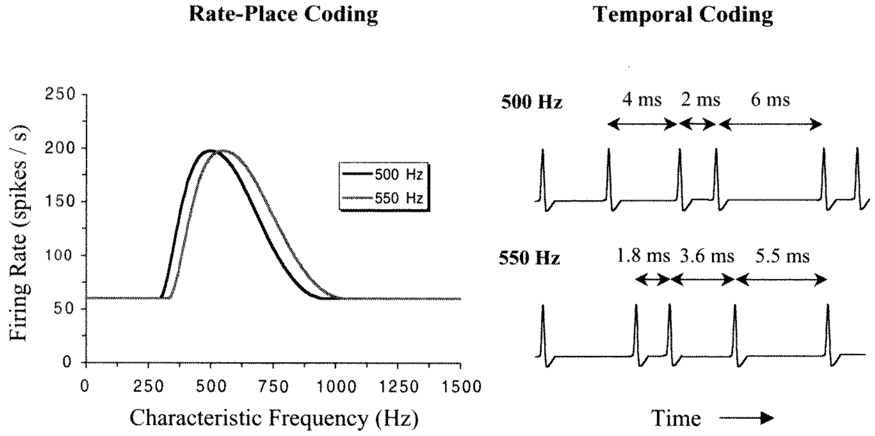

FIG. 7.2. Rate-place and temporal representations of pure-tone frequency. The figure shows simulated neural excitation patterns (left) and temporal spike patterns (right) for 500-Hz and 550-Hz pure tones. Both tones were presented at 60 dB SPL. The excitation patterns show the firing rates of high spontaneous rate fibers as a function of characteristic frequency. The temporal spike patterns show the response of a single auditory nerve fiber.

frequency of the neuron that produces the most spikes should be *equal* to the frequency of the tone (for low levels at least). More generally, however, frequency may be represented in terms of the *pattern* of activity across neurons with different characteristic frequencies.

The *rate-place* coding of frequency is illustrated by the excitation pattern representation, introduced in Sections 5.4.4 and 6.3.3. Figure 7.2 shows neural excitation patterns (expressed in terms of the simulated firing rates of high spontaneous rate nerve fibers) for two pure tones whose frequencies differ by 10%. The difference between these sounds is represented not only by the locations of the peaks of the excitation patterns, but also by the difference in firing rate at any characteristic frequency for which a response is detectable. Zwicker (1970) suggested that we might detect pure tone frequency *modulation* by detecting changes in excitation level at the point on the pattern where the level changes most (usually somewhere on the steeply sloping low-frequency side).

We are neglecting the other type of information in the auditory nerve, however—the information provided by the propensity of neurons to *phase lock* to the vibration of the basilar membrane. When a low-frequency pure tone is played to the ear, neurons tend to fire at a particular phase of the waveform, so that the intervals between neural spikes are close to *integer multiples* of the period of the pure tone (see Section 4.4.4). Different frequencies produce *different patterns* of spikes across time. For example, a 500-Hz pure tone (2-ms period) will tend to produce spikes separated by 2, 4, 6, etcetera ms. A 550-Hz pure tone (1.8-ms period) will

tend to produce spikes separated by 1.8, 3.6, 5.5, etcetera ms (Fig. 7.2). Any neuron that can respond to the tone tends to phase lock to it, and neurons will continue to phase lock even when they are saturated (see Section 6.3.4). Hence, the same temporal regularity will be present across a wide array of neurons.

So, the information about the frequency of a pure tone is represented by the pattern of neural activity across characteristic frequency, and by the pattern of neural activity across time. But which type of information is used by the brain to produce the sensation of pitch? There is certain amount of disagreement about that, but a couple of facts help sway the balance for many of us. The first is that our ability to discriminate between the frequencies of two pure tones, below 4000 Hz or so, may be too fine to be explicable in terms of changes in the excitation pattern. Figure 7.3 shows pure tone frequency discrimination as a function of frequency. The data show that we can just about detect the difference between a 1000-Hz pure tone and a 1002-Hz pure tone. My own modeling results suggest that the *maximum* difference in excitation level across the excitation pattern is only about 0.1 dB for these stimuli. Given that the smallest detectable change in pure-tone level is about 1 dB, we may be asking too much of the auditory system to detect such small frequency differences by changes in excitation level/neural firing rate alone. We are much worse at frequency discrimination at higher frequencies (Fig. 7.3), above the limit of phase locking, where pure tone frequency *is* presumably represented by a rate-place code. The other salient fact is that the breakdown of phase locking at about 5000 Hz seems to coincide neatly with the loss of the sensation of pitch (in

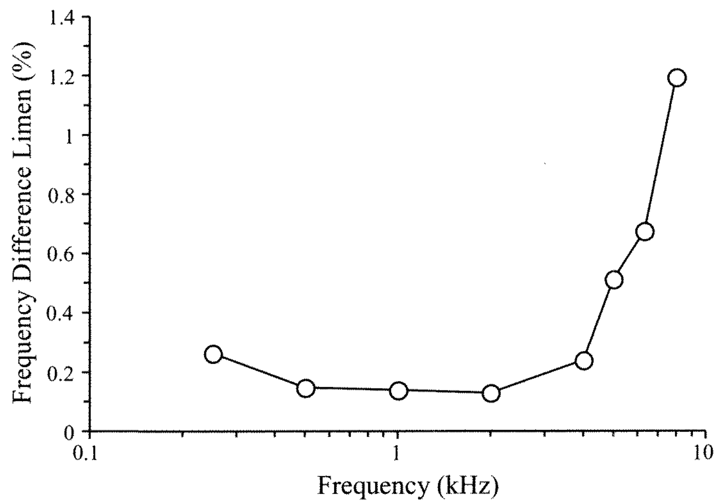

FIG. 7.3. Frequency discrimination as a function of frequency for a 200-ms pure tone. The smallest detectable increase in frequency, or frequency difference limen, is expressed as a percentage of the baseline frequency. Data are from Moore (1973).

terms of melody recognition) at around the same frequency. The implication is that phase locking, or temporal coding, may be *necessary* for the sensation of pitch.

My own view is that while it remains possible that the *frequency* of pure tones may be represented in part by firing rate across characteristic frequency (especially at high frequencies), the sensation of *pitch* is probably derived from the temporal pattern of firing.

7.2.2 Complex Tones

Figure 7.4 shows the excitation pattern of a complex tone with a number of equal-amplitude harmonics. In the region of the excitation pattern corresponding to the low harmonic numbers, there are a sequence of peaks and dips in excitation level. Each peak corresponds to a single harmonic, so the center frequency of the auditory filter giving a peak output is equal to the frequency of the harmonic. An auditory filter (or a place on the basilar membrane) tuned to a low harmonic responds mostly to that harmonic only: The other harmonics are attenuated by the filter. In a filter with a center frequency between two harmonics, both harmonics may receive substantial attenuation and the result is a dip in the excitation pattern. The first few harmonics are effectively *separated out* or *resolved* by the frequency selectivity of the basilar membrane. Each resolved harmonic has a separate representation in the cochlea. Furthermore, the representation in the cochlea is reflected in our perceptions. With practice, or an appropriate cue, such as a pure tone with the frequency of one of the harmonics presented separately, it is possible for listeners to "hear out" the first five or so harmonics of a complex tone as separate pure tones (Plomp & Mimpen, 1968).

As we go to the right of the excitation pattern, toward higher harmonic numbers and higher center frequencies, the difference between the peaks and dips decreases. Above about the tenth harmonic the pattern is almost smooth. Why is this? The spacing between harmonics is constant, 100 Hz in this example. However, the auditory filters get broader (in terms of the bandwidth in Hz) as center frequency increases (Section 5.4.3). A filter centered on a high harmonic may pass *several* harmonics with very little attenuation, and variations in center frequency will have little effect on filter output. These higher harmonics are not separated out on the basilar membrane and are said to be *unresolved.*

The resolution of harmonics depends more on harmonic number than on spectral frequency or fundamental frequency. If the fundamental frequency is doubled, the spacing between the harmonics doubles. However, the frequency of each harmonic also doubles. Because the bandwidth of the auditory filter is approximately proportional to center frequency, the doubling in spacing is accompanied by a doubling in the bandwidth of the corresponding auditory filter. The result is that the resolvability of the harmonic does not change significantly. Almost irrespective of fundamental frequency, about the first eight harmonics are resolved by the cochlea (see Plack & Oxenham, 2005, for a discussion of this issue). The number of the highest resolvable harmonic does decrease somewhat at

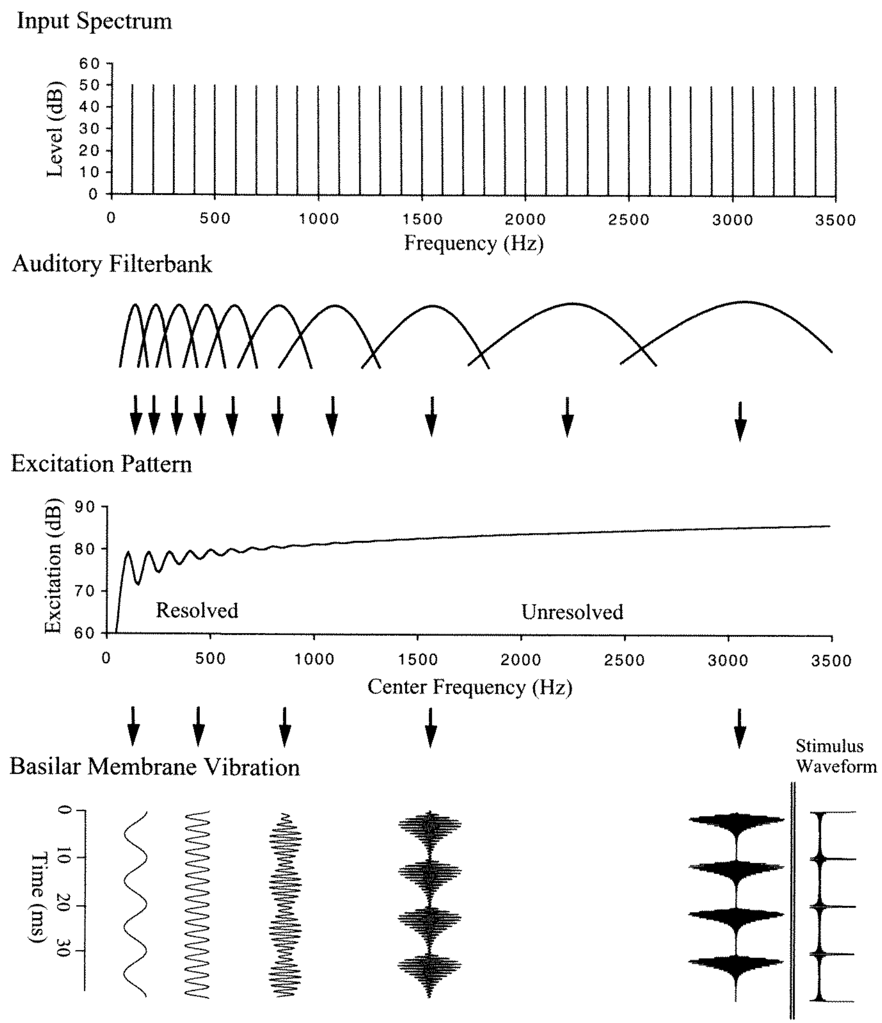

FIG. 7.4. The spectrum (top) excitation pattern (middle) and simulated basilar membrane vibration (bottom) for a complex tone consisting of a number of equal-amplitude harmonics with a fundamental frequency of 100 Hz. The auditory filters become broader as center frequency increases, hence, the high-frequency harmonics are not resolved in the excitation pattern. The basilar membrane vibration is simulated for five different characteristic frequencies (indicated by the origins of the downward-pointing arrows). The original waveform of the complex tone is also shown for reference (bottom right).

low fundamental frequencies (below 100 Hz or so), because the Q of the auditory filters is less at low center frequencies (i.e., the bandwidth as a proportion of center frequency is greater).

Information about the individual resolved harmonics is preserved in the auditory nerve both by the pattern of firing rates across frequency (rate-place, as illustrated by the excitation pattern), and by the phase locked response of the nerve fibers. At the bottom of Fig. 7.4 is a simulation of the vibration of the basilar membrane at different characteristic frequencies in response to the complex tone. A fiber connected to the place in the cochlea responding to a low harmonic will produce a pattern of firing synchronized to the vibration on the basilar membrane at that place, which is similar to the sinusoidal waveform of the harmonic. The time intervals between spikes (*inter-spike intervals*) will tend to be integer multiples of the period of the harmonic. The *individual frequencies* of the first few harmonics are represented by different patterns of firing in the auditory nerve.

The pattern of basilar-membrane vibration for the unresolved harmonics is very different. Because several harmonics are stimulating the same region of the basilar membrane, the pattern of vibration is the complex waveform produced by the addition of the harmonics. Because the spacing between the harmonics is the same as the fundamental frequency, the resulting vibration has a periodicity equal to that of the original complex tone. (The harmonics are *beating together* to produce amplitude modulation with a rate equal to the frequency difference between them; see Section 2.5.1.) The pattern of vibration at a place tuned to high harmonic numbers is effectively the waveform of a complex tone that has been *band-pass filtered* by the auditory filter for that place.

There is little information about fundamental frequency in terms of the distribution of firing rates across nerve fibers for complex tones consisting entirely of unresolved harmonics. Firing *rate* does not substantially change as a function of characteristic frequency. There is still information, however, in the temporal pattern of firing. Neurons will tend to phase lock to the *envelope* of the basilar membrane vibration (Joris & Yin, 1992), so that the time intervals between spikes will tend to be integer multiples of the period of the complex tone. That the auditory nerve can phase lock to the envelope means that the modulation rate of stimuli, such as amplitude modulated noise, is also represented in terms of the pattern of firing. It has been shown that sinusoidally amplitude modulated noise (a noise whose envelope varies sinusoidally over time, produced by multiplying a noise with a pure tone modulator) can elicit a weak, but demonstrably musical, pitch sensation, which corresponds to the frequency of the modulation (Burns & Viemeister, 1976).

Figure 7.5 illustrates the temporal pattern of firing that might be expected from nerve fibers tuned to a resolved harmonic (the second) and those responding to several higher unresolved harmonics. Time intervals between spikes reflect the periodicity of the harmonic for low harmonic numbers. Time intervals between spikes reflect the periodicity of the original waveform (equal to the envelope repetition rate) for high harmonic numbers (and for amplitude modulated noise).

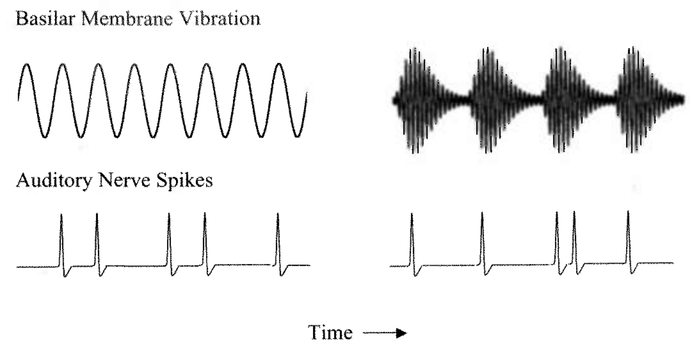

FIG. 7.5. Basilar membrane vibration and spike patterns in response to a resolved harmonic (the second harmonic, left) and in response to several unresolved harmonics (right). The nerve fiber tuned to the resolved harmonic phase locks to the fine structure. The nerve fiber tuned to the unresolved harmonics phase locks to the envelope.

It is clear from the bumps in the excitation pattern in Fig. 7.4 that there will be information about the resolved harmonics in the pattern of firing rates across characteristic frequency. As is the case for pure tones, however, it is thought that the *temporal* coding illustrated in Fig. 7.5 is more important for carrying information about the fundamental frequencies of complex tones to the brain. Our sensitivity to differences in fundamental frequency (less than 1% for resolved harmonics) suggests that we do not rely on rate-place information for resolved harmonics. Furthermore, unresolved harmonics and modulated noise simply do not produce any rate-place information in the auditory nerve concerning repetition rate.

7.3 HOW IS PERIODICITY EXTRACTED?

The repetition rate of the waveform of a pure tone is equal to the spectral frequency. The spectral analysis of the ear for all other periodic waveforms (complex tones), provides a number of frequency components, each of which contains information about the fundamental frequency of the waveform. In fact, the fundamental frequency is defined unambiguously by the frequencies of any two successive harmonics. How is the information in the auditory nerve used by the auditory system to derive periodicity and to give us the sensation of pitch?

7.3.1 The Missing Fundamental and the Dominant Region

Early luminaries of acoustics such as Ohm (1843) and Helmholtz (1863) thought that the pitch of a complex tone was determined by the frequency of the first harmonic, or fundamental component. The idea is that, if you hear a series of frequency components, you just extract the frequency of the lowest component and that gives you the fundamental frequency and the periodicity. This method works very well for most complex tones we encounter outside psychoacoustic laboratories. The fundamental frequency is equal to the frequency of the first harmonic, which is usually present in the spectra of sounds such as vowels and musical tones. However, Licklider (1956) showed that the pitch of a complex tone is unaffected by the addition of low-pass noise designed to mask the frequency region of the fundamental. Even if the fundamental component is inaudible or missing, we still hear a pitch corresponding to the basic periodicity of the complex. (Recall that the repetition rate of the waveform of a complex tone does not change when the fundamental component is removed; see Section 2.4.1 and Fig. 7.1.) The auditory system must be able to extract information about the fundamental frequency from the higher harmonics.

In fact, research has shown that the first harmonic may not even be the most important for determining the pitch of some complex tones. There is a region of low-numbered harmonics that tends to "dominate" the percept, so that frequency variations in these harmonics have a substantial effect on the pitch. For example, Moore, Glasberg, and Peters (1985) varied the frequency of one component in a complex tone, so that the component was "mistuned" from the harmonic relation. For small mistunings, changes in the frequency of the harmonic produced small changes in the pitch of the complex tone as a whole. The results for one listener are shown in Fig. 7.6. Note that the greater the mistuning of the harmonic, the greater the shift in pitch. This occurs for mistunings up to about 3%, after which the magnitude of the pitch shift *decreases* as the frequency of the harmonic is increased (see Section 10.2.2 for an explanation). Although there was some variability between individuals in the experiment of Moore et al., variations in the second, third, and fourth harmonics produced the largest effects for fundamental frequencies from 100 to 400 Hz.

Dai (2000) used a slightly different technique, in which the frequencies of all the harmonics were jittered randomly. The contribution of each harmonic to the pitch of the complex as a whole was determined by the extent to which a large jitter in that harmonic was associated with a large change in pitch. For a range of fundamental frequencies, Dai found that harmonics with frequencies around 600 Hz were the most dominant. The dominant harmonic *numbers* were therefore dependent on the fundamental frequency. For a fundamental frequency of 100 Hz, the sixth harmonic was the most dominant, and for a fundamental frequency of 200 Hz the third harmonic was the most dominant. For fundamental frequencies above 600 Hz, the first harmonic (the fundamental component) was the most dominant.

Whatever the precise harmonic numbers involved, it is clear from all the research on this topic that the *resolved* harmonics are the most important in pitch perception. For the complex tones that we hear in our everyday environment (e.g., vowel sounds), which usually have strong low harmonics, pitch is determined mainly by a combination of the frequency information from the resolved harmonics. Although they can be used for melody production, complex tones consisting entirely of unresolved harmonics have a weak pitch (i.e., the pitch is not very clear or salient). In addition, we are much better at detecting a difference between the fundamental frequencies of groups of resolved harmonics than between the fundamental frequencies of groups of unresolved harmonics (Fig. 7.7).

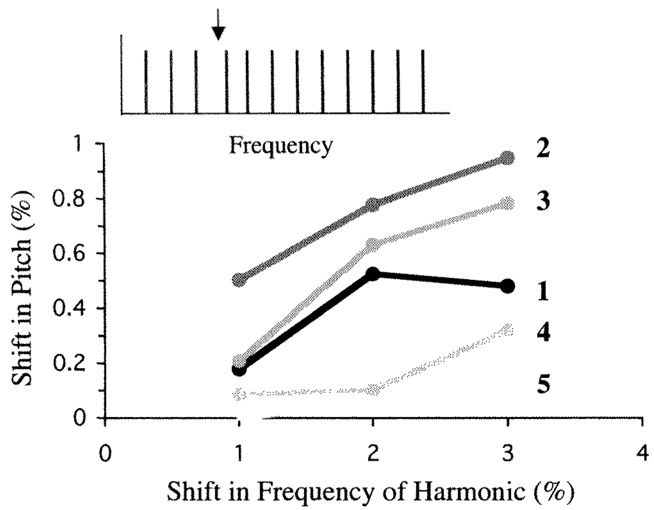

FIG. 7.6. The effects of a shift in the frequency of a single harmonic on the pitch of the whole complex tone. The harmonic number of the shifted harmonic is shown to the right. A schematic spectrum for the stimulus with a shifted fourth harmonic is shown above the graph (the arrow indicates the usual location of the fourth harmonic). The pitch shift was determined by changing the fundamental frequency of a complex tone with regular harmonic spacing until its pitch matched that of the complex tone with the shifted harmonic. The results are those of listener BG, for a fundamental frequency of 200 Hz, from the study by Moore et al. (1985). For this listener, the second harmonic was the most dominant, followed by the third, first, fourth, and fifth.

Just a final note, in case you ever end up doing research on pitch perception. It is *very* important to include low-pass masking noise in experiments than involve removing low harmonics from a complex tone. If the fundamental component, or any low harmonics, are simply removed from the waveform entering the ear, they can be *reintroduced* as combination tone distortion products on the basilar membrane (see Section 5.2.4, and Pressnitzer & Patterson, 2001). Although you may think that you are presenting a stimulus that only contains unresolved harmonics, thanks to non-linearities on the basilar membrane, the input to the auditory nervous system may well contain resolved harmonics. Unless you mask these components

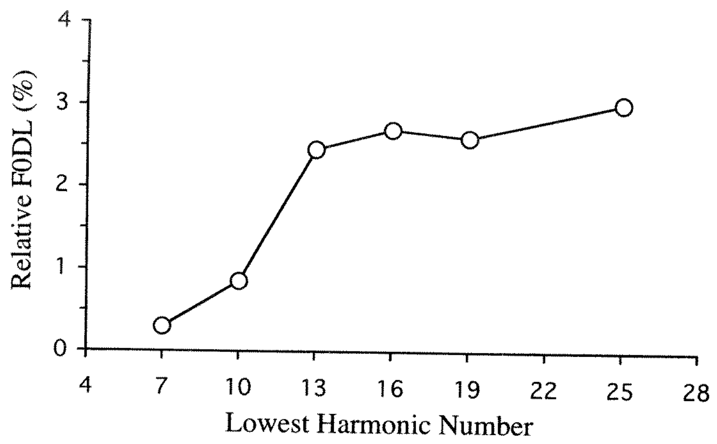

FIG. 7.7. The smallest detectable difference in fundamental frequency (the fundamental frequency difference limen, or F0DL, expressed in %) as a function of the lowest harmonic number in a group of 11 successive harmonics with a fundamental frequency of 200 Hz. When all the harmonics are unresolved (lowest harmonic numbers of 10 and above), performance is worse than when there are some resolved harmonics (lowest harmonic number of 7). Data are from Houtsma and Smurzynski (1990).

with noise, as Licklider did in the missing fundamental experiment, you cannot be sure that you are just measuring the response to the unresolved harmonics. Many experiments (including my own) have left themselves open to criticism by not taking this precaution.

7.3.2 Pattern Recognition

The low-numbered harmonics in a complex tone are resolved, in terms of the pattern of excitation on the basilar membrane, and in terms of the firing rates of auditory nerve fibers as a function of characteristic frequency. Because of this resolution, a neuron tuned to an individual low-numbered harmonic will tend to phase lock only to that harmonic. Information about the individual frequencies of resolved harmonics is present in the auditory nerve, and, indeed, we can be cued to "hear out" and make frequency matches to individual resolved harmonics. Because the individual frequencies of the low-numbered harmonics are available, it has been suggested that the auditory system could use the *pattern* of harmonic frequencies to estimate the fundamental frequency (Goldstein, 1973; Terhardt, 1974). For instance, if harmonics with frequencies of 400, 600, and 800 Hz are present, then the fundamental is 200 Hz. If harmonics with frequencies of 750, 1250, and 1500 Hz are present, then the fundamental is 250 Hz. Since the spacing between successive harmonics is equal to the fundamental frequency, any two

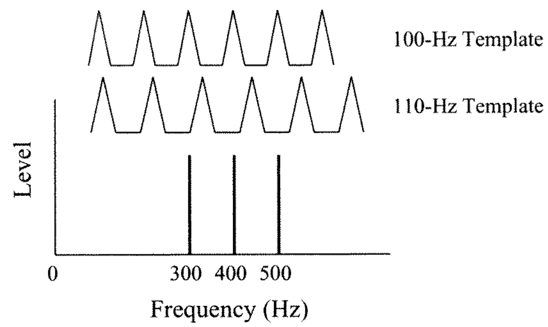

FIG. 7.8. How template matching may be used to extract the fundamental frequency of resolved harmonics. Harmonics of 300, 400, and 500 Hz produce a strong match to the 100-Hz template, but a weak match to the 110-Hz template.

successive resolved harmonics should be enough, and we can, indeed, get a sense of musical pitch corresponding to the fundamental of just two successive resolved harmonics (Houtsma & Goldstein, 1972).

Pattern recognition may be implemented as a form of *harmonic template*. For example, the auditory system may contain a template for a 100-Hz fundamental that has slots at frequencies of 100, 200, 300, 400, 500, 600, etcetera Hz (Fig. 7.8). When a group of harmonics is presented to the ear, the auditory system may simply find the best-matching template from its store. A complex that has frequency components close to the slots on a template (e.g., 99.5, 201, 299.5 Hz) is heard as having the best-matching fundamental (e.g., 100 Hz), even though the sound may not be strictly harmonic. That is how pattern recognition models can predict the pitch shift produced by mistuning a single harmonic—the best matching fundamental is also shifted slightly in this case.

Note that, in general terms, the pattern recognition approach does not rely on any specific mechanism for deriving the frequencies of the individual harmonics. The frequencies may be derived from either the rate-place representation or from the temporal representation, but the temporal representation seems more likely. The distinctive feature of pattern-recognition models is that the individual frequencies of the resolved harmonics must be extracted before the fundamental frequency can be derived. Pattern recognition models cannot account for the pitches of stimuli that do not contain resolved harmonics. However, it is clear that a weak, but clearly musical, pitch can be evoked when only unresolved harmonics are present. Furthermore, musical melodies can be played by varying the modulation rate of amplitude modulated noise (Burns & Viemeister, 1976). Neither of these stimuli contains resolved harmonics. The only information about periodicity available to the auditory system in these cases is in the temporal pattern of basilar membrane vibration, as represented by the phase-locked response of auditory nerve fibers. Although pattern recognition can account for the pitch of stimuli containing resolved harmonics, it cannot be an explanation for all pitch phenomena.

7.3.3 Temporal Models

Schouten (1940; 1970) proposed a purely temporal mechanism for the extraction of fundamental frequency. The unresolved harmonics interact on the basilar membrane to produce a waveform that repeats at the frequency of the fundamental (see Fig. 7.4). Schouten suggested that we derive the fundamental frequency by measuring the periodicity of the interaction of the unresolved harmonics. However, we know that the resolved harmonics are much more dominant, and important to the pitch of naturalistic stimuli, than the unresolved harmonics. In the absence of resolved harmonics, unresolved harmonics evoke only a weak pitch. The pattern recognition approach fails because we don't need resolved harmonics for pitch, and Schouten's approach fails because we don't need unresolved harmonics for pitch. Is there a way of using the information from *both* types of harmonics?

Many researchers are looking for ways in which the phase-locked activity in auditory nerve fibers may be decoded by the auditory system. Some modern theories of pitch perception suggest that there is a single mechanism that *combines* the information from the resolved and unresolved harmonics. As described in Section 7.2.2, neurons that have low characteristic frequencies will tend to phase lock to the individual resolved harmonics, whereas neurons with higher characteristic frequencies will tend to phase lock to the envelope produced by the interacting unresolved harmonics. In both cases, there will be inter-spike intervals that correspond to the period of the original waveform. For example, a neuron tuned to the second harmonic of a 100-Hz fundamental frequency may produce spikes separated by 5 ms (1/200 seconds), but the 10-ms interval corresponding to the period of the complex (1/100 seconds) will also be present. A neuron tuned to the unresolved harmonics will also produce a proportion of intervals corresponding to the period of the fundamental, because the neuron will phase lock to the envelope that repeats at the fundamental frequency. By picking the most prominent inter-spike interval (or perhaps the shortest common inter-spike interval) across frequency channels, an estimate of fundamental frequency (or indeed, the frequency of a pure tone) may be obtained (Moore, 2003, page 223). Figure 7.9 illustrates the idea in a schematic way: Neurons at each characteristic frequency contain a representation of the period of the fundamental, even though many other inter-spike intervals may be present.

An effective way of extracting periodicity in this manner is by a computation of the autocorrelation function (Licklider, 1951). An autocorrelation function is computed by correlating a signal with a delayed representation of itself (see Fig. 7.10).

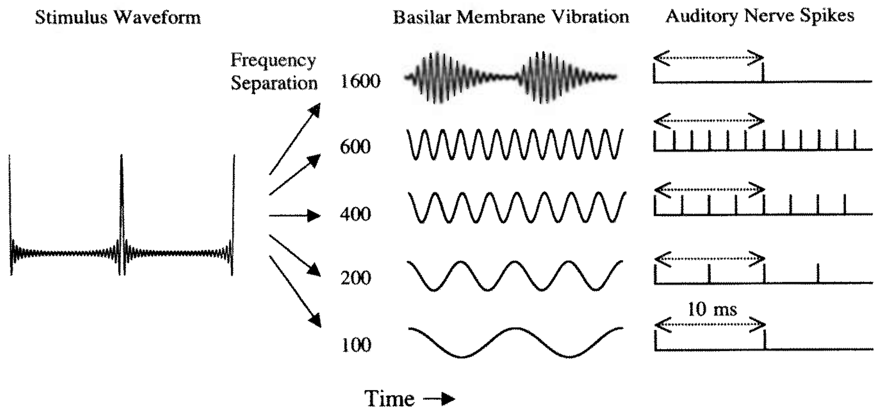

FIG. 7.9. Basilar membrane vibration, and phase locking in the auditory nerve, in response to a 100-Hz complex tone. The responses to the first, second, fourth, and sixth harmonics are shown, along with the response to a group of interacting unresolved harmonics. The characteristic frequencies of each place on the basilar membrane (in Hz) are shown to the left of the vibration plots. The temporal patterns of spikes are unrealistic for individual fibers (e.g., individual neurons would not phase lock to every cycle of the sixth harmonic), but can be taken as representative of the combined responses of several fibers. The important point is that the inter-spike interval corresponding to the period of the fundamental (illustrated by the arrows) is present at each characteristic frequency.

At time delays equal to integer multiples of the repetition rate of a waveform, the correlation will be strong. Similarly, if there are common time intervals between waveform features, then this delay will show up strongly in the autocorrelation function. For instance, a pulse train is a complex tone with a regular sequence of pressure pulses (see Fig. 2.12): If there is an interval of 10 ms between successive pulses, a delay of 10 ms will match each pulse to its neighbor. A simple addition of the autocorrelation functions of auditory nerve spikes across characteristic frequencies produces a summary function that takes into account the information from both the resolved and unresolved harmonics. Any common periodicity (see Fig. 7.9) will show up strongly in the summary function. The delay of the first main peak in the summary function provides a good account of the pitch heard for many complex stimuli (Meddis & Hewitt, 1991).

Although modern temporal models do a good job, there are still some niggling doubts. First, autocorrelation models do not provide a satisfactory explanation of why we are so much better at fundamental frequency discrimination, and why pitch is so much stronger, for resolved harmonics than for unresolved harmonics. This is true even when the fundamentals are chosen so that the harmonics are in the same

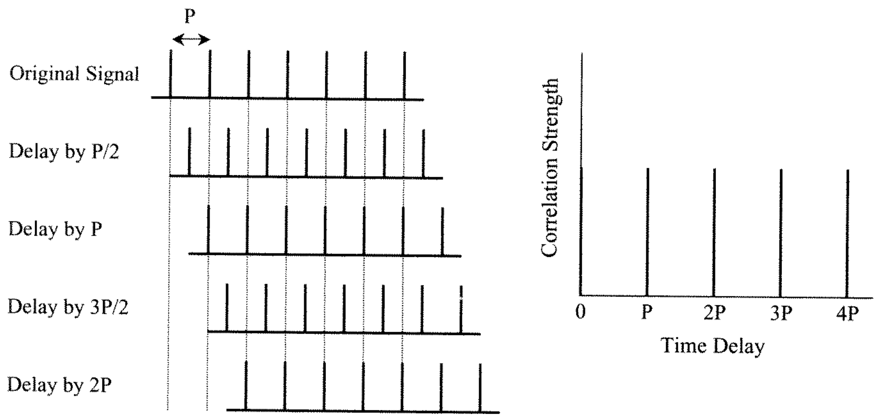

FIG. 7.10. How autocorrelation extracts the periodicity of a waveform. A pulse train, or set of neural spikes (top left), is subject to different delays. When the delay is an integer multiple of the period of the pulse train (P) the delayed version of the waveform is strongly correlated with the original (the timings of the original pulses are indicated by the dotted lines). Correlation strength is measured by *multiplying* the delayed version with the original version, and summing the result across time.

spectral region (Shackleton & Carlyon, 1994). Second, recent experiments with groups of unresolved harmonics suggest that the regularity of temporal information may be less important for these stimuli than the gross rate of temporal fluctuations. Removing pulses at random from a regular high-pass filtered pulse train does not affect the autocorrelation function significantly, because there are still sufficient intervals between pulses corresponding to the fundamental. Yet the pitch is heard to *decrease* in these situations (Carlyon, 1997). Some have suggested that there may be separate pitch mechanisms for resolved and unresolved harmonics (Carlyon & Shackleton, 1994). If so, that would open the door for the pattern recognition models. The debate rages on.

7.3.4 Neural Mechanisms

We are unclear, therefore, about the precise nature of the algorithm that is used to extract the repetition rate of a periodic stimulus. We are also unclear about *where* in the auditory system such an algorithm might be implemented. It must come after the inputs from the two ears are combined, because a pitch can be derived from just two harmonics presented to *opposite ears* (Houtsma & Goldstein, 1972). That could be almost anywhere among the brainstem auditory nuclei, but probably after the cochlear nucleus.

The maximum repetition rate to which a neuron will phase lock decreases as the signal is transmitted up the ascending pathways. In the medial geniculate body

the upper limit may be about 800 Hz (Møller, 2000), compared to about 5000 Hz in the auditory nerve. Some investigators think that, at some stage before this in the ascending auditory pathways, a representation of pitch by the temporal pattern of neural spikes is converted into a representation by firing *rate*, with different neurons tuned to different periodicities or different fundamental frequencies. Just as the spectrum of a sound is encoded by the distribution of firing rates across characteristic frequency (tonotopic organization), the fundamental frequencies of sounds may be encoded by the pattern of activity across neurons with different characteristic periodicities (*periodotopic* organization). It would be particularly nice to find a neuron that responds with a high firing rate to complex tones with a particular fundamental frequency, irrespective of the spectrum of the complex tone (for example, regardless of whether it contains only low harmonics or only high harmonics). Unfortunately, the evidence for such neurons is not conclusive. There is some excitement about neurons in the inferior colliculus that show tuning to periodicity, different neurons being sensitive to different modulation rates (Langner & Schreiner, 1988). However, the tuning may be too broad to provide the accurate representation of periodicity required for pitch.

A pitch template that forms the basis of a pattern recognition mechanism could be produced by combining the activity of neurons with different characteristic frequencies, spaced at integer multiples of the fundamental frequency, or possibly by combining the outputs of neurons that are sensitive to the different patterns of phase locking in response to the resolved harmonics. There is little evidence for either of these types of processing, although there are neurons in the auditory cortex that show double peaks in their tuning curves (reflecting input from neurons with two characteristic frequencies), some of which are positioned at harmonically related frequencies (Sutter & Schreiner, 1991).

There also is little evidence for a neural network that can perform autocorrelation as described in Section 7.3.3. These networks require a set of delay lines so that the neural firing pattern can be delayed, then combined (theoretically multiplied) with the original version. A large set of delays is needed, from 0.2 ms to extract the periodicity of a 5000-Hz stimulus, to 33 ms to extract the periodicity of a 30-Hz stimulus. Delays may be implemented by using neurons with long axons (spikes take longer to travel down long axons) or by increasing the number of synapses (i.e., number of successive neurons) in a pathway, since transmission across a synapse imposes a delay. A "pitch neuron" could then combine the activity from a fast path and from a delayed path in order to perform the autocorrelation. If the delays were different for different pitch neurons, each neuron would respond to a different periodicity. Langner and Schreiner (1988) have found some evidence for neural delay lines in the inferior colliculus of the cat.

Overall, then, the picture from the neurophysiology is a little unclear. We see several tantalizing glimpses of neural pitch mechanisms, but nothing we can be sure of at present. Researchers have some idea what they are looking for in terms of the neural response, but finding it is a different matter altogether.

7.4 SUMMARY

We are at the stage in the book where facts begin to be replaced by conjecture. We know quite a lot about the representation of periodicity in the auditory nerve, but how that information is used to produce a neural signal that corresponds to our sensation of pitch is still a matter for investigation. The perceptual experiments on human listeners help to guide the search for pitch neurons. We need to know what the perceptions are before we can identify the neural mechanisms that may underlie these perceptions.

1. Sounds that repeat over time (*periodic* sounds) are often associated with a distinct pitch that corresponds to the repetition rate. The range of repetition rates that evoke a musical pitch extends from around 30 Hz to around 5000 Hz.
2. The frequency of pure tones is represented in the auditory nerve by the pattern of firing rates across characteristic frequency (*rate-place* code) and by the pattern of phase-locked activity across time (*temporal* code). It is probable that the temporal code is used to produce the sensation of pitch.
3. The low harmonics of complex tones are *resolved* on the basilar membrane, and produce separate patterns of near-sinusoidal vibration at different places along the membrane. The higher harmonics are *unresolved* and interact on the basilar membrane to produce complex waveforms that repeat at the fundamental frequency. Neurons phase lock to the individual frequencies of the resolved harmonics, and to the envelope that results from the interaction of the unresolved harmonics.
4. Complex tones produce a clear pitch even if the first harmonic (the fundamental component) is absent. However, the resolved harmonics are dominant and produce a clearer pitch than the unresolved harmonics. In the case of most complex tones that we encounter in the environment, pitch is mainly determined by a *combination* of the information from the individual resolved harmonics.
5. Pattern recognition models suggest that the auditory system extracts the frequencies of the resolved harmonics, and uses the patterning of these harmonics to estimate the fundamental frequency. However, these models cannot account for the pitch of unresolved harmonics.
6. The period of a complex tone is reflected in the time intervals between spikes of nerve fibers responding to both resolved *and* unresolved harmonics. Modern temporal models assume activity is combined across fibers with different characteristic frequencies to estimate this period (and, hence, to provide the sensation of pitch).
7. The neural basis of pitch extraction is unclear. Somewhere in the brainstem, temporal patterns of firing may be converted into a code based on firing rate, with different neurons tuned to different periodicities. However, the evidence is inconclusive at present.

7.5 READING

For a comprehensive account of all aspects of pitch perception, I recommend:

Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (Eds.). (2005). *Pitch: Neural coding and perception.* New York: Springer-Verlag.

Moore and Houtsma provide excellent readable introductions:

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). London: Academic Press. Chapter 6.

Houtsma, A. J. M. (1995). Pitch perception. In B. C. J. Moore (Ed.), *Hearing* (pp. 267–295). New York: Academic Press.