

Pitch pattern recognition and grouping

A pitch is generally heard when sounds have a waveform that repeats. Vibrations of things such as stretched strings, reeds, columns of air in pipes, engines and so on produce such 'periodic' sounds. The rate of repetition of the sound's waveform corresponds to the rate of the vibration that is producing the sound, so that we hear lower pitches when vibrations are relatively slow, and higher pitches from faster vibrations.

Place and periodicity theories. Some vibrating objects, such as tuning forks and some whistles, produce waveforms that are a single sine-wave tone. There is a long-standing debate about how the pitch of such sounds is perceptually extracted in the auditory system. Periodicity theories hold that the time interval between repeats of the waveform is measured, while place theories argue that information in the sound's spectrum is used, so that pitch corresponds to the peak in the spectrum. It has been difficult to decide between these theories for the case of *sine-wave tones*.

Periodic complex tones. Most natural vibrations, including those of musical instruments and voices, give rise to complex sounds that contain more than one sine-wave component. The frequencies of these components are related to the frequency of the vibration, f , measured in Hz. There can be a component with a frequency of f , the 'fundamental frequency' while other components have 'harmonic frequencies', which are f multiplied by an integer ('harmonic number'), i.e., $2f$, $3f$, $4f$, $5f$ etc. The pitch of complex sounds is generally the same as the pitch of a sine-wave tone whose frequency is f Hz. Complex sounds usually have a waveform that repeats itself at the rate of the vibration, so *periodicity theory* can still account for their pitch. *Place theorists* have argued that the pitch is determined by finding the fundamental frequency, f , in the sound's spectrum. However, complex sounds can be produced in which there is no component at f Hz. Such sounds are common in everyday listening e.g., the notes of a 'cello, or a piano played softly. Nevertheless, these sounds still have the same pitch as that of a sine-wave tone whose frequency is f Hz. This observation is known as '*the problem of the missing fundamental*'.

Fine structure theory. One attempt to explain how the pitch of the missing fundamental comes about was proposed by Schouten et al. (1962). They thought that pitch must somehow be extracted at the *analysis stage* of auditory processing, from the vibration pattern at some point on the basilar membrane. The ideas of *auditory filters* and the filter-bank model of hearing were used to get an idea of the nature of the vibration patterns at different points on the membrane. These patterns show that the *lower frequency components* of complex sounds, such as $3f$, $4f$ & $5f$, are generally '*well resolved*' for

sounds with diverse fundamental frequencies. The effect of this is that the vibration patterns at corresponding points on the basilar membrane are simple sine waves, each with a frequency of one of the components. As these patterns *do not repeat at the fundamental frequency* Schouten et al. thought that the fundamental's pitch could not be obtained from these components. However, *components at higher frequencies* are '*badly resolved*', so that the vibration pattern at a particular point is like a mixture of more than one sine wave. The frequencies of the sine waves in the mixture are those of a few adjacent, high harmonics in the complex sound. These patterns *do repeat at the fundamental's frequency*, so Schouten et al. proposed that a sound's pitch is 'extracted' in a high-frequency channel, by finding the repetition rate of the vibration pattern at a corresponding point on the basilar membrane. This is called 'fine-structure' theory, as it was thought that the details of the pattern are used to obtain its repetition rate.

Dominant harmonics. Fine-structure theory predicts that harmonics at higher frequencies will be more important for determining a sound's pitch than will the lower-frequency harmonics. This prediction of the 'dominance' of high harmonics was tested by Ritsma (1967) by first playing a complex sound with a fundamental and a full range of harmonics extending from low to high frequencies. He then selected trios of adjacent harmonics and shifted each of their frequencies upward by a small amount. At the same time he shifted the frequencies of all the other components downward by a small amount. Ritsma reasoned that the pitch should rise if the selected trio contained any 'dominant' harmonics, but should fall if the selected trio contained no dominant harmonics. Listeners' judgements indicated that the pitch rose when some trios were selected, but fell when other trios were selected, indicating that some harmonics are dominant. However, these dominant harmonics are not the badly resolved, high frequency ones that fine-structure theory would lead us to expect. The dominant harmonics are those that are well resolved. It was found that the pitch would generally be heard to rise if trios containing $3f$, $4f$ or $5f$ were shifted upwards in frequency, otherwise the pitch was generally heard to fall.

Central pitch. Place theory, periodicity theory and fine-structure theory all have difficulty explaining the phenomenon of central pitch, which was discovered by Houtsma and Goldstein (1972). They used sounds with just two harmonics and no fundamental, e.g., $3f + 4f$, $4f + 5f$, or $5f + 6f$ etc. Although the pitch of these sounds is rather weak, listeners can identify melodies played by varying the (missing) fundamental of each note's harmonics, even when the harmonic numbers are randomly varied from note to note (e.g., maybe $3f + 4f$ for one note, $5f + 6f$ for the next note, etc.). The crucial experimental condition was one in which each note's two harmonics were presented to different ears. Under these conditions there is no frequency component or repetition

period on the basilar membrane that corresponds directly to the fundamental frequency. Nevertheless, Houtsma and Goldstein found that their melodies could still be identified.

Synthesis stage. Thus, experiments on dominance and central pitch indicate that pitch is not extracted from a single frequency channel in the way that fine-structure theory indicates. These findings show that pitch is extracted at a synthesis stage of perception where information from several frequency channels is combined. This is the idea behind modern 'pattern recognition' models of pitch perception.

Pattern recognition models are generally 'computational theories' which produce their predictions about pitch in quite diverse ways, but they all have much the same underlying principle. Essentially they compute the fundamental from the frequencies of harmonics that are well resolved. For example, in Terhardt's (1974) model the frequency of each harmonic is divided by a series of integers (2, then 3, then 4 etc.) to give the frequencies of 'subharmonics'. If there is one frequency that is a subharmonic of each harmonic, then this is the sound's pitch. Thus if a sound has sine-wave components at 800 Hz, 1000 Hz and 1200 Hz, then the pitch is 200 Hz because this is 800 divided by 4, 1000 divided by 5, and 1200 divided by 6. Moore (1989) has argued that this kind of 'frequency division' arises because the 'neural spikes' that indicate the firing of the auditory-nerve's fibres occur at a range of subharmonic frequencies across the range of nerve fibres that are stimulated by a harmonic.

Mistuned harmonics. When the frequency of one of the components of a harmonic series is shifted slightly it loses its harmonic relationship with the other components. Small *increases in the component's frequency* ('mistunings') cause the pitch of the whole sound to rise slightly, to a pitch slightly above the fundamental's pitch. However, when the increase becomes a little larger a *perceptual segregation* occurs, and two sounds are heard. The shifted harmonic is heard as a sine tone with a high pitch corresponding to its frequency, while the other harmonically related components are heard as a separate, complex sound, which has the fundamental's pitch (Moore, 1986). Components in different frequency channels are thus grouped together by the pattern recognition process as long as they are 'harmonically related' by sharing the same fundamental frequency, otherwise components are perceptually segregated. This '*harmonicity*' is therefore a useful '*grouping heuristic*' (Bregman, 1990) because a sound source with a particular rate of vibration will have harmonically related components, which are not harmonically related to components of other sources that vibrate at other rates.

Competition among grouping heuristics. Parts of sounds that start or stop at the same time might also come from the same source, so this '*isochrony*' might also serve as a

'grouping heuristic' at the *synthesis stage of perception*. This was shown to be the case by Darwin and Ciocca (1992). They used sounds similar to those in Moore's (1986) experiment, i.e., complex sounds with a frequency-shifted (mistuned) harmonic. They shifted the frequency of the harmonic by just enough to increase the pitch of the whole sound, but not by enough to segregate the harmonic perceptually. Next, the shifted harmonic was extended in time so that it started well before the remaining components in the complex. This had the effect of abolishing the original pitch shift, and of making the shifted harmonic segregate from the rest of the complex sound. Thus various '*grouping heuristics*', such as isochrony and harmonicity, seem to be used at the synthesis stage of perception, and they seem to *compete with each other* in determining whether parts of sound are *perceptually grouped or segregated*.

Further reading. There are several suggestions on the course reading list.

A. J. Watkins, April 2002