PART II

Pitch perception of complex sounds in humans

Chapter 5

Introduction to pitch perception

Pitch perception

Common sounds with a well defined pitch are e.g. sounds produced by our vocal chords or by musical instruments. Such complex periodic sounds contain a fundamental frequency and a large number of harmonic overtones. The pitch of these sounds is related to the repetition rate of the waveform of the sound and covaries with its fundamental frequency. Research on pitch perception has concentrated on two questions: i) which physical parameters determine pitch and ii) how these physical parameters are processed in the auditory nervous system. Psychophysical experiments have elucidated many aspects of the first question. About the second question much remains unknown. It is, for example, still unknown whether, at the peripheral level of the auditory nerve, temporal patterning of neural firings (temporal coding) or the rate-place patterning of neural firings (place-rate coding) is used to transmit information concerning frequency to the central nervous system. To understand what is meant by place-rate and temporal coding, some basic knowledge about the working mechanism of the inner ear is required.

In the inner ear, the different frequency components of a complex sound excite different parts along the basilar membrane (von Békésy, 1960). High frequencies excite the basal part of the basilar membrane, while low frequencies travel through the cochlea and have their maximal response at the apical part of the basilar membrane. Due to the approximately logarithmic cochlear frequency map and the limited frequency resolving power of the cochlea, the relatively widely spaced lower harmonics of a complex sound are resolved, while the higher harmonics, which are spaced more closely, remain unresolved.

A resolved harmonic causes the basilar membrane to vibrate nearly sinusoidally at the place of resonance. The sensory hair cells, which are stimulated by the basilar membrane vibration, excite the auditory nerve fibres during positive deflection of the basilar membrane (resulting in a deflection of the hair cell bundle towards the taller stereocilia) and inhibit the auditory nerve fibres during negative deflection. Because of this phase locking behaviour, the temporal pattern of nerve fibre activity contains information about the frequency of a resolved component (temporal coding), as long as stimulus frequency is not too high (< 5 kHz, Rose *et al*., 1967). The mean overall activity of the nerve fibres increases during stimulation, because the excitation during positive deflection predominates the inhibition during negative deflection (e.g. Rose *et al*., 1967). Since most auditory nerve fibres are innervated by only one (inner) hair cell (Spoendlin, 1972) located at a certain position of the basilar membrane with a defined resonance frequency, also the increased activity gives information about the frequency of the component (place-rate coding).

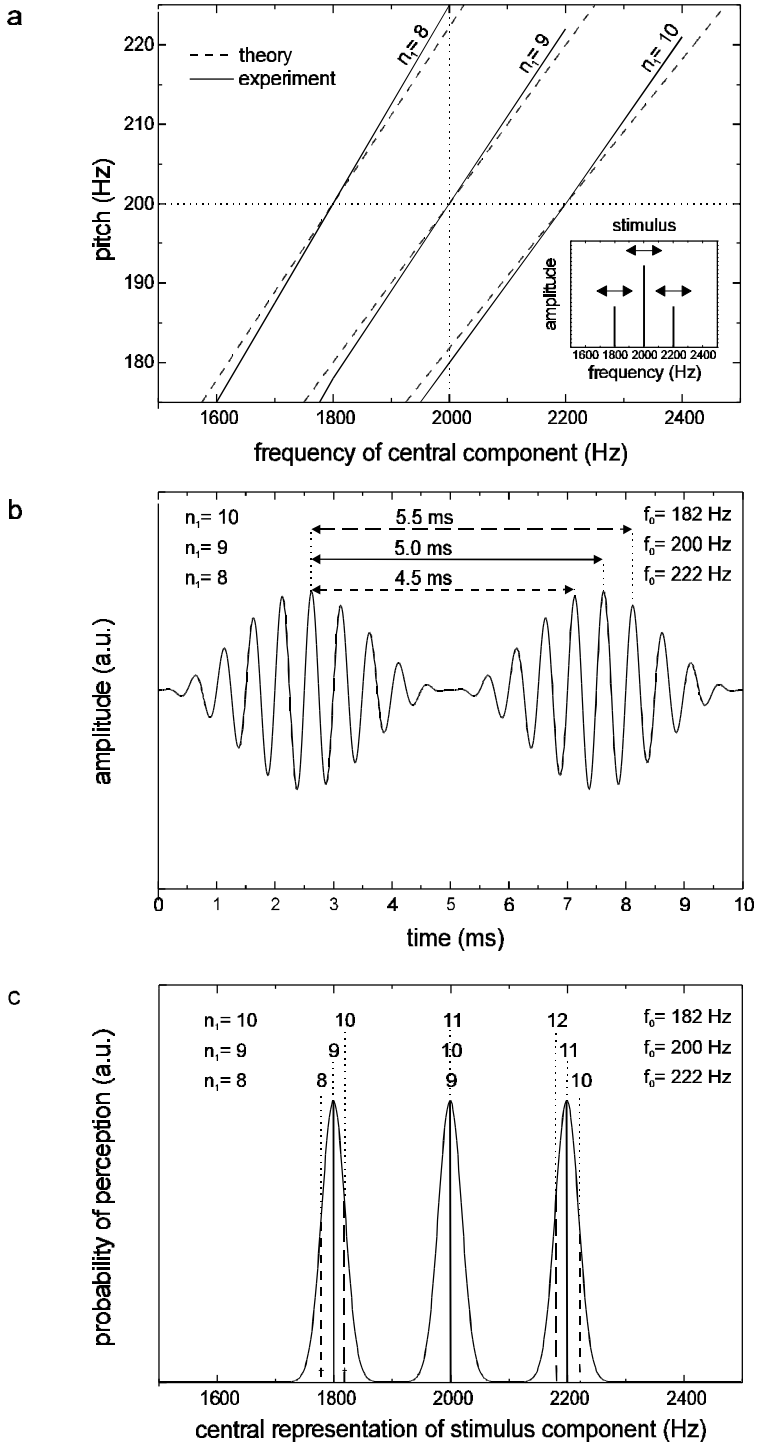
Unresolved harmonics produce an amplitude modulated basilar membrane vibration of two or more superimposed frequency components. The amplitude modulation of this vibration depends on the phase relation of the harmonics and its frequency equals the frequency difference between the harmonics. For a harmonic complex sound, this is the frequency of the fundamental. The auditory nerve fibres can, under certain loudness and spike

rate restrictions, code the amplitude modulation and temporal fine structure of the vibration of the basilar membrane in their temporal activity pattern (Horst *et al.*, 1986).

Theoretically there are several ways in which the central nervous system can extract the pitch of a complex sound, which is related to the frequency of the fundamental: Firstly, the activity of nerve fibres representing the fundamental component gives information about the fundamental, both via the place of activation along the basilar membrane and through their temporal firing pattern. Secondly, the amplitude modulated temporal pattern of the activity of nerve fibres innervated by the interference pattern of unresolved harmonics contains information about the fundamental frequency. Thirdly, the nerve fibres innervated by resolved harmonics contain, via rate-place or temporal coding, information about the frequencies of these harmonics. Since these are all multiples of the fundamental frequency, the combined activity of several auditory channels can be used to extract the fundamental pitch.

In the history of auditory research on pitch perception, all three sources of information were proposed to be essential for the perception of pitch (for a review see e.g. de Boer, 1976). It is now clear that periodic sounds containing no energy at the fundamental frequency evoke a well-defined pitch corresponding to this frequency, also if no energy at this frequency is reintroduced by aural nonlinearities (Seebeck, 1841; Schouten, 1938; Licklider, 1954). This means that the first mentioned source of information is not essential for hearing the fundamental pitch. The first theory dealing with this so-called 'pitch of the missing fundamental' was the residue theory of Schouten (1940). In this theory the fundamental pitch is extracted using the temporal activity pattern of nerve fibres stimulated by unresolved frequency components. Physiological experiments showed that, indeed, the temporal firing pattern of the auditory nerves, which are stimulated by unresolved frequency components, contains information about the temporal fine structure of the basilar membrane vibration (e.g. Horst *et al.*, 1986). Further, psychophysical experiments showed that the pitch of the missing fundamental can be perceived if only unresolved frequency components are present in the stimulus (Hoekstra, 1979; Moore and Rosen, 1979; Houtsma and Smurzynski, 1990).

Figure 1 (right). Ambiguity of three tone complexes. **(a)** Pitch of three tone complexes as a function of the frequency of the central component of the complex sound (after Schouten et al., 1962). The inset shows the amplitudes of the frequency components of the (AM) stimulus. The frequency difference of the components is, for all stimuli, 200 Hz. n_1 = estimated harmonic number of the first frequency component. [The difference between the theoretical predictions (dashed lines) and the experimental data (solid lines) is caused by the influence of combination tones (Smoorenburg, 1970)] **(b)** Temporal waveform of the stimulus with a central component of 2000 Hz. The ambiguity of the stimulus can be explained if the central nervous system extracts the time interval between the most prominent peaks in the temporal waveform of unresolved frequency components. **(c)** Central representation of resolved frequency components of the stimulus with a central component of 2000 Hz. The Gaussians centred around the frequencies of the components represent the limited accuracy with which the frequencies of the components are assumed to be represented in the central nervous system. The ambiguity of the sound can be explained by mismatches of the harmonic numbers.


a
theory
experiment
n1= 8
n1= 9
n1= 10
pitch (Hz)
220
210
200
190
180
stimulus
amplitude
1600 1800 2000 2200 2400
frequency (Hz)
1600
1800
2000
2200
2400
frequency of central component (Hz)
b
n1= 10
n1= 9
n1= 8
5.5 ms
5.0 ms
4.5 ms
f0= 182 Hz
f0= 200 Hz
f0= 222 Hz
amplitude (a.u.)
0 1 2 3 4 5 6 7 8 9 10
time (ms)
c
n1= 10
n1= 9
n1= 8
10
9
8
11
10
9
12
11
10
f0= 182 Hz
f0= 200 Hz
f0= 222 Hz
probability of perception (a.u.)
1600
1800
2000
2200
2400
central representation of stimulus component (Hz)

However, this pitch is very weak and it is now clear that low, resolved frequency components are dominant in determining the pitch of the missing fundamental (de Boer, 1956; Plomp, 1967; Ritsma, 1962; Moore *et al*., 1985). Strong evidence that information derived from totally resolved frequency components is used for the perception of pitch comes from experiments showing that the missing fundamental of a two component complex sound can be heard if each component is presented to a different ear (Houtsma and Goldstein, 1972). The discovery of the importance of the low order harmonics in pitch perception resulted in the development of harmonic pattern recognition models (Goldstein, 1973; Wightman, 1973; Terhardt, 1979). The basic idea of these models is that, to extract the fundamental pitch, the central nervous system combines the activity of several groups of auditory nerve fibres, which are stimulated by different resolved harmonics.

The mechanism by which the central nervous system combines information from different auditory channels to extract pitch is not known. To be able to answer this question, it is important to know which code, place-rate or temporal coding, is used to represent the frequencies of the harmonics. For resolved harmonics, the neural code delivering information about their frequencies can be based on place or timing information, or both. Several observations, however, such as the covariation of the precision with which frequencies can be perceived with the precision to be expected from neural time processing (e.g. Moore, 1973; Goldstein and Srulovicz, 1977), suggest that timing information is used to represent frequencies up to 4-5 kHz. It has been suggested that the nervous system compares inter spike intervals (ISI's) in the auditory channels to extract the pitch of complex sounds (Licklider, 1951; Goldstein and Srulovicz, 1977; Srulovicz and Goldstein, 1983; Delgutte and Cariani, 1992). The most prominent ISI present in several auditory channels could, e.g. determine pitch (Moore, 1982). An estimation of the fundamental frequency on the basis of most prominent ISI's can also explain that a weak pitch sensation is perceived if a complex sound contains only unresolved frequency components (Hoekstra, 1979; Moore and Rosen, 1979; Houtsma and Smurzynski, 1990).

Chapter 6 of this thesis: Ambiguity of the pitch of complex sounds.

The pitch of a complex is ambiguous when only a few harmonics are present. This is clearly shown in experiments of, e.g., Schouten *et al.* (1962). In these experiments, the frequency components of three tone complexes with a fundamental frequency of 200 Hz were equidistantly shifted in frequency. The pitch of these inharmonic complex sounds rises if the frequency components are shifted upwards and falls if they are shifted downwards (Fig. 1a). The results show that even for a purely harmonic complex (see e.g. Fig. 1a at a centre frequency of 2000 Hz) a number of discrete pitches (of around 180, 200 and 225 Hz) can be perceived.

The ambiguity of these stimuli can be explained by the residue theory of Schouten (1962) as well as harmonic pattern recognition models (e.g. Goldstein, 1973). In the former case the temporal fine structure of the interference pattern contains information about the different possible pitches (Fig. 1b). In the harmonic pattern recognition models, the ambiguity of the pitch results from the spread in the central representations of the frequencies of the components. In the optimum processor theory of Goldstein (1973), e.g., the fundamental pitch

is extracted by estimating the harmonic numbers of the frequency components (Fig. 1c). The limited accuracy of the central representation of the stimulus frequencies (in Fig. 1c represented by Gaussians centred around the stimulus frequencies) can cause mismatch of the harmonic numbers, resulting in the ambiguity of the pitch.

In the experiments described in the next chapter, the relative pitch strength of the fundamental pitch and alternative pitches is measured and compared with theoretical predictions derived from a harmonic pattern recognition model. In this model the technique of sub-harmonic summation (see e.g. Terhardt *et al.*, 1982; Hermes, 1988; Cohen *et al.*, 1995) was applied to extract the fundamental pitch and alternative pitches.

This sub-harmonic summation procedure can readily be interpreted as a central mechanism that looks for the most prominent ISI's present in different auditory channels. Fig. 2 shows the ISI-histograms for nerve fibres stimulated with 1800, 2000 and 2200 Hz (see e.g. Rose *et al.*, 1967). The most prominent ISI has a length of $1/f$ seconds (with f = stimulus frequency). Because the nerve fibres do not respond at each cycle of stimulation, but can miss one or more cycles, ISI's with a length of multiples of $1/f$ seconds are also found. In the insets of Fig. 2 it is clearly visible that all three fibres respond with ISI's of around 5 ms, corresponding to intervals where the auditory nerve spikes at every 9th, 10th or 11th cycle, respectively. Also, intervals of around 4.5 ms and 5.5 ms are present in all three fibres. A central mechanism, determining pitch on the basis of the most prominent ISI's in different auditory channels would perceive the pitch of the missing fundamental (200 Hz) and the alternative pitches of around 222 and 182 Hz. The sub-harmonic summation procedure can account for such a central mechanism if an n-th order sub-harmonic (with n = integer) is interpreted as an ISI in which (n-1) cycles are missed. The results presented in chapter 6 indicate that high harmonics contribute less to the pitch perception than low harmonics. Since the number of ISI's decreases with increasing values of n, this is expected on the basis of this mechanism.

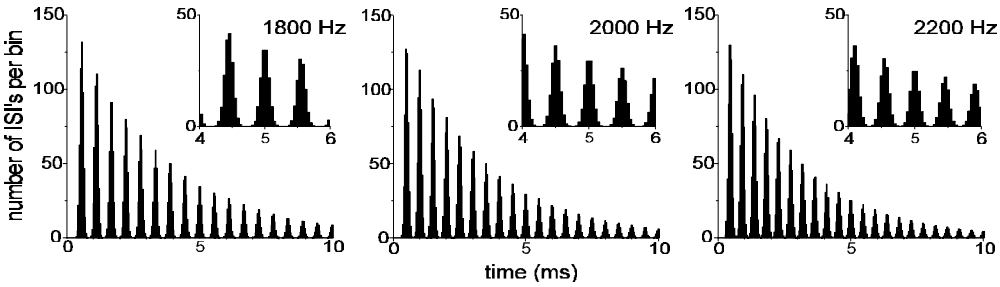


Figure 2. Simulation of ISI-histograms of nerve fibres stimulated with 1800, 2000 and 2200 Hz. The insets clearly show that all three fibres respond with ISI's of around 5 ms, but that ISI's of around 4.5 ms and 5.5 ms are present as well in all three fibres.