

Interference effects and phase sensitivity in hearing

BY BRIAN C. J. MOORE

*Department of Experimental Psychology, University of Cambridge,
Downing Street, Cambridge CB2 3EB, UK*

Published online 28 March 2002

This paper reviews interference effects in the auditory system, particularly effects occurring in the outer ear and the inner ear (cochlea). Sounds enter the ear canal both directly and after reflections from the pinna. This results in complex spectral patterns, which vary systematically with the direction of incidence of the sound source relative to the head. Evidence is described indicating that these spectral patterns are used in the localization of sounds in space. The cochlea behaves like a limited-resolution frequency analyser. When the components of a complex sound are closely spaced in frequency, they can interfere on the basilar membrane (BM) within the cochlea. Interference effects on the BM are complex, as they are influenced by a physiologically active mechanism which introduces strong nonlinearities, including level-dependent amplification. Interference effects on the BM play a role in many aspects of auditory perception, including the perception of consonance and dissonance, the perception of pitch, the perception of changes in phase, and the perception of timbre. Interference effects in the cochlea may also play a role in producing the spectral regularity observed in sounds reflected from the ear (otoacoustic emissions).

Keywords: hearing; interference; phase sensitivity;
pitch; timbre; otoacoustic emissions

1. Introduction

This paper describes interference effects occurring at two stages in the auditory system, the outer ear and the inner ear. To understand the nature of these effects, it is necessary to have a basic knowledge of the anatomy and physiology of the peripheral auditory system. Hence, I start with these topics. Figure 1 shows the structure of the human peripheral auditory system. It is traditionally considered to be composed of three parts: the outer, middle and inner ear. The outer ear is composed of the pinna and the auditory canal or meatus. Reflections from the pinna, head and upper torso interfere with the direct sound from the source and significantly modify the sound entering the ear canal at medium and high frequencies. The nature of this interference and its perceptual consequences are described later in this paper.

Sound travels down the meatus and causes the eardrum, or tympanic membrane, to vibrate. The eardrum forms the outer boundary of the middle ear. These vibrations

One contribution of 15 to a special Theme Issue 'Interference 200 years after Thomas Young's discoveries'.

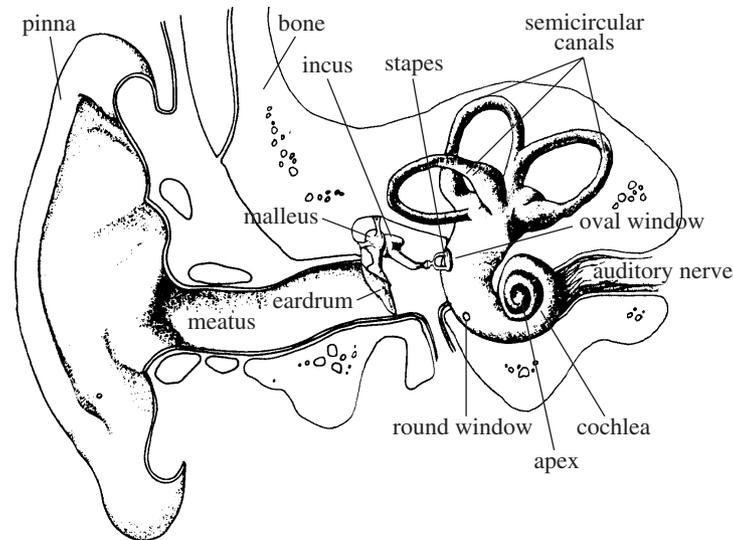


Figure 1. Illustration of the structure of the peripheral auditory system showing the outer, middle and inner ear. Redrawn from Lindsay & Norman (1972).

are transmitted through the middle ear by three small bones, the ossicles, to a membrane-covered opening in the bony wall of the spiral-shaped structure of the inner ear, the cochlea. This opening is called the oval window and it forms the inner boundary of the middle ear. The major function of the middle ear is to ensure the efficient transfer of sound energy from the air to the fluids in the cochlea. If the sound were to impinge directly onto the oval window, most of it would simply be reflected back, rather than entering the cochlea. This happens because the acoustic impedance of the oval window is very different to that of air. The middle ear acts as an impedance-matching device which *reduces* interference effects that would otherwise be caused by reflections from the oval window.

The cochlea is shaped like the spiral shell of a snail. However, the spiral shape does not appear to have any functional significance (except for saving space), and the cochlea is often described as if the spiral had been ‘unwound’. The cochlea is filled with almost incompressible fluids, and it has bony rigid walls. It is divided along its length by two membranes, Reissner’s membrane and the basilar membrane (BM). The start of the spiral, where the oval window is situated, is known as the base; the other end, the inner tip, is known as the apex. It is also common to talk about the basal end and the apical end. Because of the incompressible fluids, inward movement of the oval window results in a corresponding outward movement in a membrane covering a second opening in the cochlea, the round window. Such movements result in pressure differences between one side of the BM and the other (i.e. the pressure is applied in a direction perpendicular to the BM) and this results in movement of the BM.

A third membrane, called the tectorial membrane, lies close to and above the BM, and also runs along the length of the cochlea. Between the BM and the tectorial membrane are *hair cells*, which form part of a structure called the organ of Corti (see figure 2). They are called hair cells because they appear to have tufts of hairs,

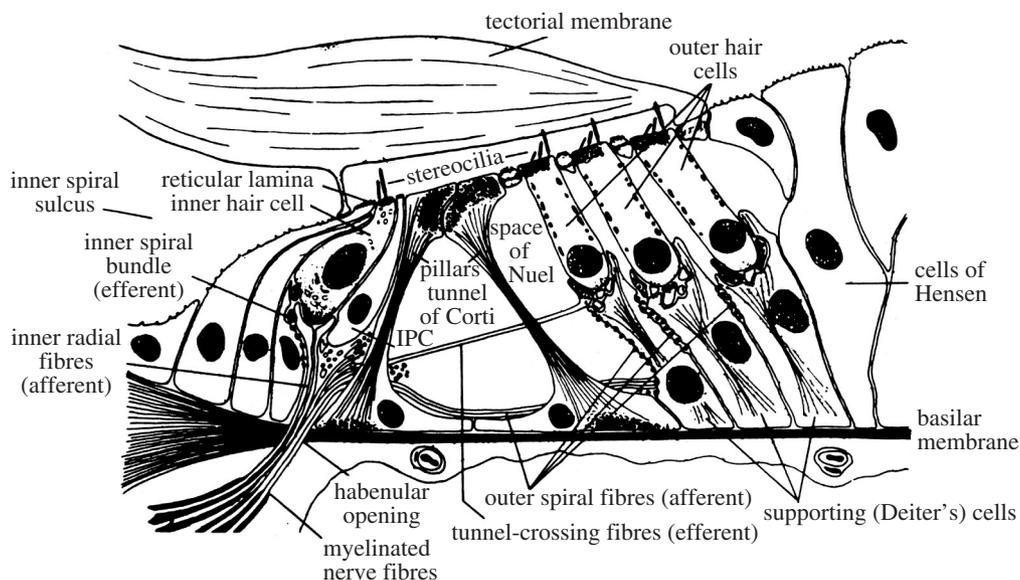


Figure 2. Cross-section of the organ of Corti as it appears in the basal turn of the cochlea. (IPC, inner pillar cell.) Adapted from Ryan & Dallos (1984).

called stereocilia, at their apices. The hair cells are divided into two groups by an arch known as the tunnel of Corti. Those on the side of the spiral shape are known as outer hair cells (OHCs), and they are arranged in three to five rows. The hair cells on the other side of the arch form a single row, and are known as inner hair cells (IHCs). It appears that the stereocilia of the OHCs actually make contact with the tectorial membrane, but this may not be true for the IHCs. The tectorial membrane appears to be effectively hinged at one side (the left in figure 2). When the BM moves up and down, a shearing motion is created; the tectorial membrane moves sideways (in the left–right direction in figure 2) relative to the tops of the hair cells. As a result the stereocilia at the tops of the hair cells are moved sideways. The movement of the stereocilia of the IHCs leads to a flow of electrical current through the IHCs, which in turn leads to the generation of action potentials (nerve spikes) in the neurons of the auditory nerve. Thus the IHCs act to transduce mechanical movements into neural activity.

The main role of the OHCs is probably actively to influence the mechanics of the cochlea. The OHCs have a motor function, changing their length, shape and stiffness in response to electrical stimulation (Ashmore 1987; Yates 1995), and they can therefore influence the response of the BM to sound. The OHCs are often described as being a key element in an active mechanism within the cochlea. The function of this active mechanism is described in more detail below. The exact way in which the active mechanism works is complex, and is still not fully understood, although the existence of such a mechanism was predicted by Gold (1948) over 50 years ago. For recent reviews, the reader is referred to Camalet *et al.* (2000), Eguluz *et al.* (2000) and Ashmore & Mammano (2001).

When the oval window is set in motion by a sound, a pressure difference occurs between the upper and lower surface of the BM. The pressure wave travels almost

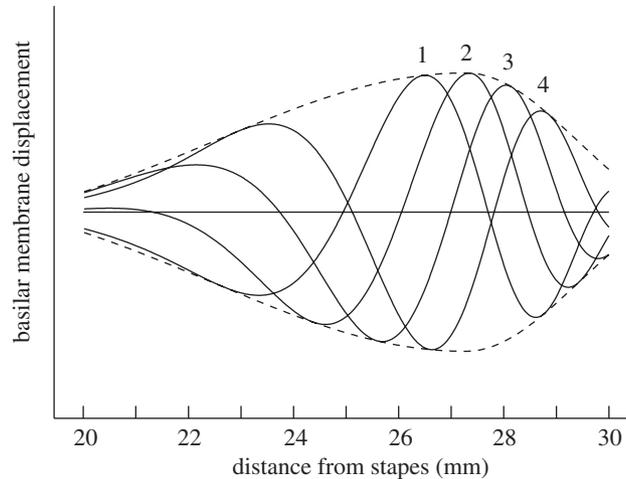


Figure 3. The solid lines show the instantaneous displacement of the BM at four successive instants in time (labelled 1–4), derived from a cochlear model. The pattern moves from left to right, building up gradually with distance, and decaying rapidly beyond the point of maximal displacement. The dashed line represents the envelope traced out by the amplitude peaks in the waveform. The envelope shown here is representative of what would be observed in a dead ear. In a healthy living ear, the envelope would have a much sharper peak.

instantaneously through the incompressible fluids of the cochlea. Consequently, the pressure difference is applied essentially simultaneously along the whole length of the BM. This causes a pattern of motion to develop on the BM. The pattern does not depend on which end of the cochlea is stimulated. The response of the BM to stimulation with a sinusoid takes the form of a travelling wave which moves along the BM from the base towards the apex (von Békésy 1960). The amplitude of the wave increases at first with increasing distance from the base and then decreases rather abruptly. The basic form of the wave is illustrated in figure 3, which shows schematically the instantaneous displacement of the BM for four successive instants in time, in response to a low-frequency sinusoid. The four successive peaks in the wave are labelled 1–4. This figure also shows the line joining the amplitude peaks, which is called the envelope. The envelope shows a peak at a particular position on the BM.

The response of the BM to sounds of different frequencies is strongly affected by its mechanical properties, which vary progressively from base to apex. At the base the BM is relatively narrow and stiff. This causes the base to respond best to high frequencies. At the apex the BM is wider and much less stiff, which causes the apex to respond best to low frequencies. Each point on the BM is tuned; it responds best (with greatest displacement) to a certain frequency, called the characteristic frequency (CF) or best frequency, and responds progressively less as the frequency is moved away from the CF. The CF decreases monotonically with distance from the base. It is now believed that the tuning of the BM arises from two mechanisms. One is referred to as the passive system or passive mechanism. This depends on the mechanical properties of the BM and surrounding structures, and it operates in a roughly linear way. The other mechanism is the active mechanism. This depends on the operation of the OHCs, and it operates in a nonlinear way. The active mecha-

nism depends on the cochlea being in good physiological condition, and it is easily damaged. When the OHCs operate normally, the BM shows sharp tuning, especially for low-input sound levels. The travelling wave illustrated in figure 3 is representative of what might be observed in a dead cochlea, when only the passive system is operating. In the living cochlea, the envelope of the travelling wave would have a much sharper peak.

A second function of the active mechanism is to provide level-dependent amplification or gain on the BM. The gain is greatest for low-level inputs (levels below *ca.* 30 dB sound pressure level (SPL)), and decreases progressively with increasing level for levels up to 90–100 dB SPL (Sellick *et al.* 1982; Robles *et al.* 1986; Ruggero *et al.* 1997). This level-dependent gain means that the response on the BM is compressive. For example, if the input level of a sinusoid is increased from 50 to 60 dB SPL, the response on the BM at the place tuned to the frequency of that sinusoid may increase by only *ca.* 2.5 dB. The compression appears to be very fast acting, *i.e.* the gain changes very rapidly when the input amplitude is changed (Recio *et al.* 1998). The compression is strongest for mid-range input levels (30–90 dB SPL), and is weak or absent for very low sound levels (below *ca.* 30 dB SPL). We shall see later that interference effects on the BM can be strongly influenced by the compression.

When a complex sound is presented to the ear, the BM acts as a kind of frequency analyser: the sinusoidal frequency components in the sound are separated or resolved, each producing its maximum response at a specific place on the BM. However, the resolution of the BM is limited. If two components are too close together in frequency, they will not be resolved, but will interfere on the BM. Perceptual consequences of this interference are the main topic of this paper.

2. Interference effects in the sound entering the ear

Some of the sound from a given source enters the meatus directly, but some enters the meatus after reflection from one or more of the folds in the pinna. There are also some reflections from the shoulders and the upper part of the torso. When the direct sound and the reflected sound are added together, this results in interference effects, giving a change in the spectrum of the sound reaching the eardrum.

Say, for example, that the sound reflected from the pinna has a time delay of 75 μ s (this is the time taken for a sound to travel a distance of *ca.* 2.5 cm), and that the sound source is a broadband white noise, which has a ‘flat’ spectrum (power spectrum magnitude independent of frequency). For a frequency of 6667 Hz (period 150 μ s), the reflected sound is delayed by one-half of a period relative to the leading sound. Therefore, the direct and reflected sounds cancel, giving a dip in the spectrum at 6667 Hz. For a frequency of 13 333 Hz (period 75 μ s), the reflected sound is delayed by one whole period, so the direct and reflected sounds add, giving a peak in the spectrum. Because there are usually several reflections from the pinna, the spectrum of the sound reaching the eardrum is complex, containing multiple peaks and dips. The spectral pattern varies systematically with the direction of the sound source (Wightman & Kistler 1989; Blauert 1996).

The filtering produced by this interference effect is often characterized by measuring the spectrum of the sound source and the spectrum of the sound reaching the eardrum. The ratio of these two (usually expressed in decibels) gives what is called the ‘head-related transfer function’ (HRTF). This is illustrated in figure 4, which

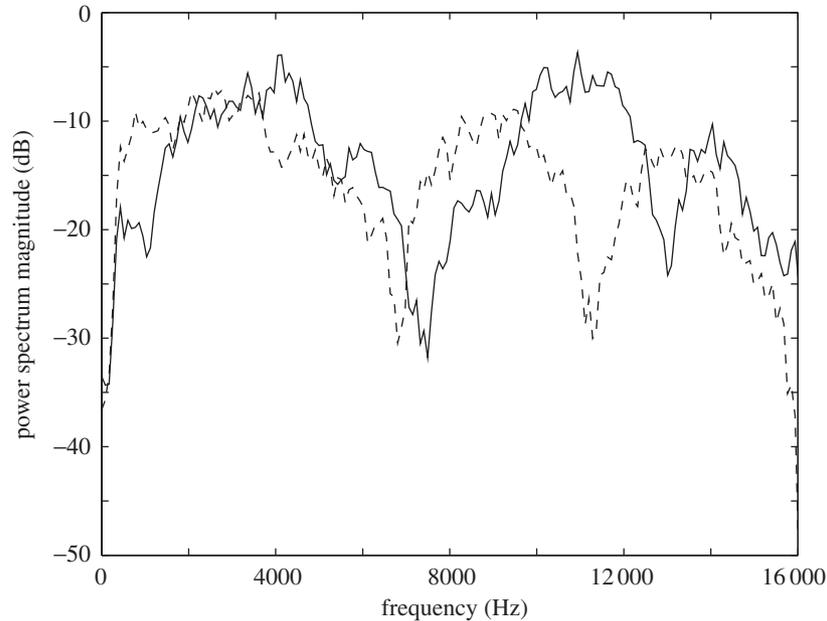


Figure 4. Power spectra recorded close to the eardrum for bursts of white noise presented in an anechoic room. In one case (solid line) the sound source was directly in front of the head (0° azimuth and 0° elevation). In the other case (dashed line) the sound source was directly behind the head (180° azimuth and 0° elevation).

shows the result of recordings made in an anechoic chamber with a small probe microphone close to the left eardrum of the author. The source signal was white noise. The figure shows the spectra at the eardrum for two locations of the source. In one case (solid line) the sound source was directly in front of the head (0° azimuth and 0° elevation). The spectrum shows peaks at *ca.* 11 000 Hz and 14 000 Hz, and dips at *ca.* 7500 Hz and 13 000 Hz. In the other case (dashed line) the sound source was directly behind the head (180° azimuth and 0° elevation). The spectrum shows peaks at *ca.* 9000 Hz and 13 000 Hz, and dips at *ca.* 6500 Hz and 11 000 Hz. The spectral magnitude at *ca.* 500–1000 Hz is lower for the sound at the front than for the sound at the rear; this is a result of reflections from the shoulders and torso. At some frequencies there is a difference in power spectrum magnitude between the two positions of 25 dB.

Differences in the spectral shape of sounds, which result in differences in the distribution of vibration along the BM, are often heard as changes in timbre, or tone quality, especially when the differences occur at low frequencies. For example, each vowel sound is characterized by a particular pattern of spectral peaks and dips, and this pattern determines the perceived identity of the vowel (this is described in more detail later). However, the spectral changes introduced by interference in the meatus are not usually heard as changes in timbre, especially when they occur at high frequencies. Rather, they are associated with the perceived location of the sound source (Blauert 1983). Human listeners probably do not use changes in spectral shape at low frequencies for sound localization, because these changes are produced by reflections from the shoulders and torso (and perhaps nearby surfaces), and the position of these

relative to the head is not fixed; therefore, the spectral changes at low frequencies do not provide consistent information about the location of a sound relative to the head. However, the pinna is fixed relative to the head, so spectral changes produced by reflections from the pinna are reliable indicators of location. These changes occur mainly at frequencies above 6000 Hz, since it is only at high frequencies that the wavelength of sound is sufficiently short for it to interact strongly with the pinna.

Since it is the spectral *patterning* of the sound which is important, the information provided by the pinna is most effective when the sound has spectral energy over a wide frequency range. When sounds at the eardrum include the type of spectral patterning that would normally be provided by the pinna, the sounds are usually heard as outside the head (Batteau 1967). However, when the sounds lack such patterning (as can happen for sounds presented by insert earphones, for example), the sounds may be heard as inside the head.

If the listener is to make efficient use of spectral cues associated with the direction of a sound source, then it is necessary to distinguish spectral peaks and dips related to direction from peaks and dips inherent in a sound source or produced by reflections from nearby surfaces. Thus one might expect that a knowledge of the sound source and room conditions would also be important. However, knowledge of the spectrum of the sound source might not be *essential*, for two reasons. Firstly, the spectral peaks and dips introduced by the pinna are quite sharp, whereas for many sound sources the spectrum at high frequencies is relatively smooth. Secondly, the two ears provide separate sets of spectral cues, so the difference between the sound at the two eardrums can be used to locate unfamiliar sound sources. Even for sound sources in the median plane (all points equidistant from the two ears), asymmetries in pinna shape may result in interaural disparities, which can be used for localization if the stimuli contain audible components with frequencies above 8–10 kHz (Searle *et al.* 1975; Middlebrooks *et al.* 1989).

Although knowledge of the spectrum of the sound source may not always be necessary to make use of pinna cues, Plenge (1972, 1974) has presented evidence that we do, in fact, make comparisons with stored stimulus patterns in judging the location of a sound source. He showed that if subjects were not allowed to become familiar with the characteristics of the sound source and the listening room, then localization was disturbed. In many cases the sound sources were heard inside the head. This was particularly true for sound in the median plane (the locus of all points equidistant from the two ears). However, such familiarity does not seem to require an extended learning process. We become familiar with sound source characteristics and room acoustics within a very few seconds of entering a new situation.

3. Interference effects on the basilar membrane

Each point on the BM can be considered as a filter with a certain centre frequency and a certain bandwidth; the latter is often specified as the equivalent rectangular bandwidth (ERB). The centre frequency changes progressively with position along the BM, having extreme values of *ca.* 50 Hz at the apex and 15 000 Hz at the base. The ERB increases with increasing centre frequency. In humans, the ERB can be estimated from masking experiments (Fletcher 1940; Moore 1995). Its value in hertz, for normally hearing people, is given approximately by the following equation:

$$\text{ERB} = 24.7(4.37F + 1), \quad (3.1)$$

where F is the centre frequency in kHz (Glasberg & Moore 1990). For centre frequencies above 2 kHz, the ERB is *ca.* 12% of the centre frequency.

Generally, interference between frequency components in a complex sound is strong when the components are separated by less than one ERB, while it is weaker when the components are separated by more than one ERB. However, for high input sound levels the patterns of vibration on the BM can spread over a considerable proportion of its length, and in this case interference can occur between components that are quite widely spaced in frequency.

When two sinusoidal components in a complex sound are separated sufficiently in frequency, the wave pattern produced on the BM may show two distinct peaks. Perceptually, this is associated with the ability to ‘hear out’ the individual components, i.e. to hear a pitch corresponding to each component. In such a case, the components are said to be resolved. For components with equal amplitudes, the frequency separation required for them to be resolved corresponds to *ca.* 1.25ERB (Plomp 1964; Plomp & Mimpen 1968; Moore & Ohgushi 1993).

It has been claimed that the ear is ‘phase deaf’, i.e. insensitive to the relative phases of the components in a complex sound (Ohm 1843). In a harmonic complex tone (a tone where the frequencies of all of the components are integer multiples of a common fundamental component with frequency F_0), the lower harmonics (up to about the seventh) are separated by more than 1.25ERB. It appears that human listeners are indeed insensitive to the relative phases of these lower harmonics (Helmholtz 1863). However, Helmholtz did not rule out the possibility that phase changes might be audible for the higher harmonics of a complex tone, which will interfere on the BM. We shall see later that Helmholtz was correct in this conjecture. Under some conditions, a change in phase of a single high harmonic of only 2° can be audible. Such phase sensitivity generally occurs only when harmonics interfere strongly on the BM.

(a) *Interference between two sinusoids*

When two sinusoids of the same level but with different frequencies are presented simultaneously, their perceived quality depends strongly on their frequency separation. For very small frequency separations, the sound fluctuates regularly in loudness. These loudness fluctuations occur because the two sinusoids interfere on the BM, moving alternately in and out of phase at a rate equal to the frequency separation of the two tones. The fluctuations are referred to as beats. When the frequency separation of the two sinusoids is increased above *ca.* 20 Hz, the individual fluctuations are no longer perceived, but the sound takes on a rough, dissonant and unpleasant quality (Terhardt 1974*a*). This roughness and dissonance tends to increase with increasing frequency separation of the two sinusoids (i.e. with increasing beat rate), but is only perceived when the two sinusoids interfere strongly on the BM. If the frequency separation of the sinusoids is increased sufficiently, the roughness fades away, and the tones are perceived as consonant. Presumably, at this point, the sinusoids no longer interfere strongly on the BM.

The role of interference on the BM in these perceptual effects is supported by the finding that the frequency separation of the two sinusoids at which roughness (or dissonance) is maximal depends on the centre frequency, and corresponds to *ca.* 0.44ERB for centre frequencies up to 2000 Hz (Plomp & Levelt 1965; Plomp

& Steeneken 1968; Kameoka & Kuriyagawa 1969; Plomp 1976). For example, for a centre frequency of 2000 Hz, maximum roughness is perceived when the tones are separated by *ca.* 100 Hz. For higher centre frequencies, the frequency separation giving maximum roughness is lower than 0.44ERB, probably because the central auditory system becomes less sensitive to rapid fluctuations in amplitude when the rate of those fluctuations exceeds *ca.* 120 Hz (Kohlrausch *et al.* 2000; Moore & Glasberg 2001). The frequency separation at which the tones first become consonant also depends on centre frequency and is *ca.* 1.2ERB (Plomp & Levelt 1965).

(b) *Interference between many sinusoids: effects on pitch*

A harmonic complex tone is usually perceived as having a single pitch corresponding to F_0 , even if the fundamental component is not present (Schouten 1940; de Boer 1976; Moore 1997). This pitch has been given various names, including ‘the pitch of the missing fundamental’, ‘residue’ pitch, ‘virtual’ pitch and ‘low’ pitch. The low pitch is most clearly heard when the tone contains many harmonics, but it can be heard when the tone contains only two (Smooenburg 1970) or three (Schouten *et al.* 1962) harmonics. The low pitch may be determined either from the pattern of frequencies of the lower, resolved harmonics (Goldstein 1973; Terhardt 1974*b*), or from the time pattern of the waveform evoked on the BM by the higher, unresolved harmonics (Schouten 1940; Schouten *et al.* 1962). Action potentials (spikes) in the auditory nerve tend to be evoked by the most prominent peaks in the waveform on the BM. Many theorists believe that the pitch of complex tones is at least partly determined by the pattern of inter-spike intervals in the auditory nerve (Patterson 1987*b*; Meddis & Hewitt 1991; Cariani & Delgutte 1996*a, b*; Moore 1997); the pitch is assumed to correspond to the interval that occurs most often, and this usually is equal to the period of F_0 .

The waveform of a harmonic complex tone can be strongly influenced by the starting phases of the harmonics, and this may be expected to influence the low pitch under some circumstances. This is illustrated in figure 5, which shows the effect of changing the relative phase of one harmonic in a complex tone containing the 9th, 10th and 11th harmonics. Panels (a) and (c) show waveforms of the individual harmonics, and panels (b) and (d) show the waveforms of the complex tones produced by adding together the harmonics. For the case illustrated in the top half of the figure, all three harmonics start in ‘cosine’ phase; this means that the harmonics have their maximum amplitudes at the start of the waveform. Correspondingly, a peak in the envelope of the complex tone occurs at the start of the waveform, and a new peak in the envelope occurs for every 10 oscillations in the ‘fine structure’. At the point in time marked by the vertical dashed line, a peak in the waveform of the centre component (the 10th harmonic) coincides with minima in the waveforms of the two other harmonics. This gives a minimum in the envelope of the complex tone. Thus, the waveform of the complex tone has an envelope with distinct peaks and dips and is described as ‘amplitude modulated’. It might also be described as a ‘peaky’ waveform.

For the case illustrated in parts (c) and (d) of figure 5, the phase of the highest component is shifted by 180°; that component starts with a minimum rather than a maximum in its waveform. As a result, the amplitude of the complex tone at the start of the waveform is not as high as for the case when all harmonics started in

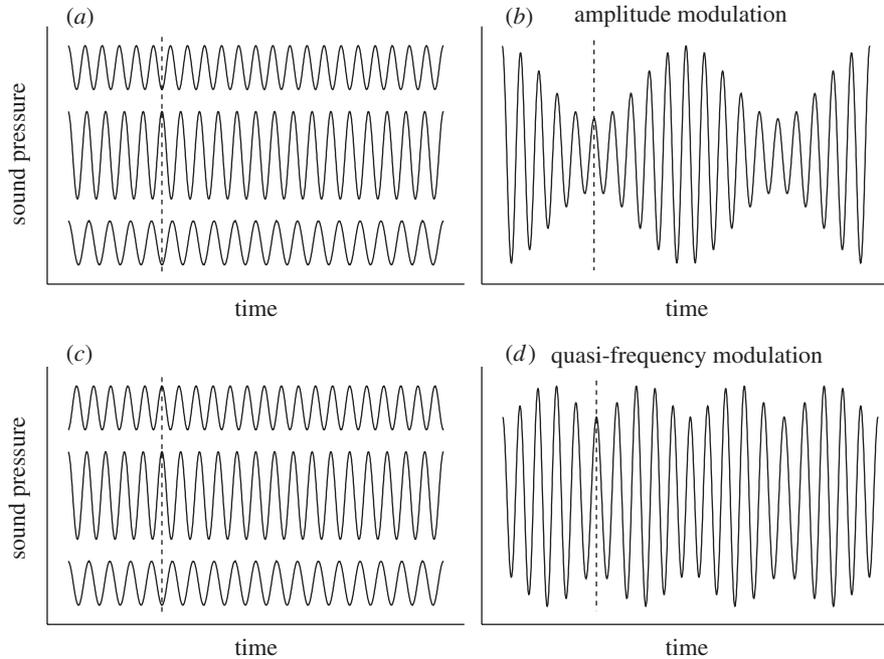


Figure 5. The effect of changing the relative phase of one component in a complex tone containing just three harmonics: the 9th, 10th and 11th. (a), (c) Waveforms of the individual sinusoidal components. (b), (d) Waveforms of the complex tones produced by adding together the sinusoidal components. For one phase ((a), (b)), the waveform has an envelope with distinct peaks and dips; this waveform is sometimes called ‘amplitude modulated’. For the other phase ((c), (d)), the envelope is much flatter; this waveform is sometimes called ‘quasi-frequency modulated’.

cosine phase. At the point in time marked by the vertical dashed line, a peak in the waveform of the centre component (the 10th harmonic) coincides with a minimum in the waveform of the ninth harmonic and a peak in the waveform of the eleventh harmonic. The minimum in the envelope of the complex tone is less deep than the minimum when all harmonics started in cosine phase. Thus, the envelope is much flatter, and the envelope actually shows two maxima for each period of the waveform. This waveform is sometimes called ‘quasi-frequency modulated’, as the time between peaks in the fine structure fluctuates slightly; however, this is not easily visible in the figure.

If the harmonics have low harmonic numbers (say the second, third and fourth), they will be resolved on the BM. In this case, the relative phase of the harmonics is of little importance as the envelope on the BM does not change when the relative phases of the harmonics are altered. Human listeners seem to be relatively insensitive to the relative phase across different points on the BM (Patterson 1987a; Uppenkamp *et al.* 2001). However, if the harmonics have high harmonic numbers (as in figure 5), they interfere strongly on the BM. Changes in the relative phase of the harmonics can then result in changes in the envelope of the waveform on the BM and changes in the pattern of inter-spike intervals in the auditory nerve. If the pitch theories referred to above are correct, this should affect the perceived low pitch.

For tones containing only high harmonics, phase can indeed affect the pitch that

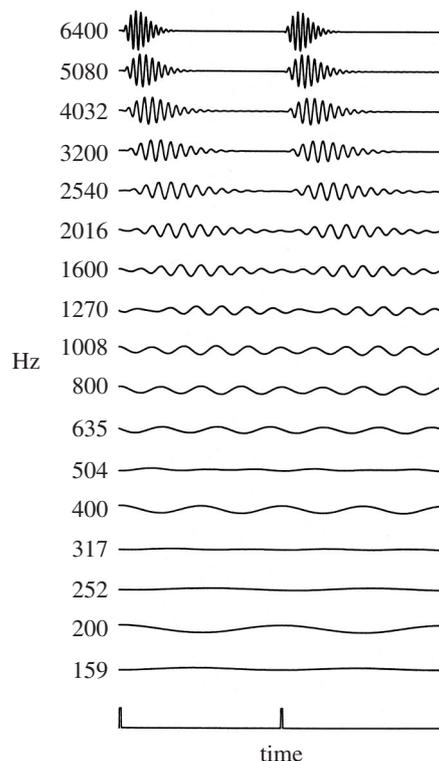


Figure 6. A simulation of the responses on the BM to periodic impulses of rate 200 pulses per second. Each number on the left represents the characteristic frequency (CF) at a specific point on the BM. The waveform which would be observed at that point, as a function of time, is plotted opposite that number.

is heard (Ritsma & Engel 1964; Moore 1977; Patterson 1987*a*; Shackleton & Carlyon 1994). For example, a change in phase that changes the waveform from ‘amplitude modulated’ (figure 5*a, b*) to ‘quasi-frequency modulated’ (figure 5*c, d*) can result in the perceived pitch going up by an octave, presumably as a consequence of the doubling of the envelope repetition rate. The pitch is also more clear for the amplitude-modulated tone than for the quasi-frequency-modulated tone (Moore 1977). Effects of relative phase on the clarity of pitch for complex tones with many harmonics have also been reported (Lundeen & Small 1984).

In summary, when the harmonics in complex tones are closely spaced in frequency (relative to the ERB), they interfere on the BM and the resulting waveform depends on the relative phase of the components. This strong interference occurs for the high harmonics. When the complex tone contains only high harmonics, changes in phase can affect the magnitude of envelope fluctuations and the number of envelope peaks per second, and these can affect both the clarity of the pitch and its value.

(c) *Effects of phase on the audibility of high harmonics*

For complex tones containing many harmonics, changing the relative phase of a single harmonic can create the sensation of a pitch corresponding to the frequency of

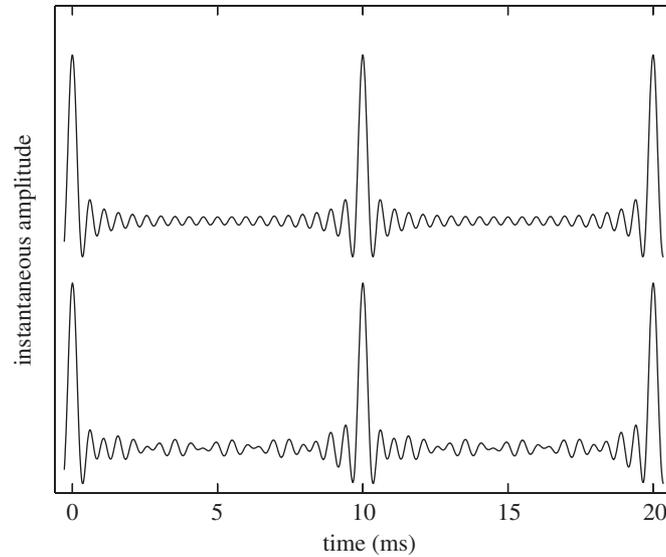


Figure 7. The upper trace shows the waveform resulting from summing the first 20 harmonics of a fundamental frequency of 100 Hz. All harmonics had equal amplitude and they were added with a phase of 90° at time zero (cosine phase). The lower trace shows the waveform when the phase of the 15th harmonic was shifted by 30° .

that harmonic. This effect depends strongly on interference of the harmonics on the BM. Consider as an example a train of brief impulses, with silent intervals between the pulses, as illustrated by the trace at the bottom of figure 6. This is equivalent to a complex tone with many equal-amplitude harmonics, each starting in cosine phase (90°). It has a strong low pitch and a ‘buzzy’ quality. In this example, the F_0 was 200 Hz, so the waveform repeats with a period of 5 ms. The other traces in figure 6 show simulations of the waveforms that would be evoked at different places along the BM; the CF of each place is indicated in hertz at the left of each trace. On the BM, the lower harmonics are resolved, so at places tuned to the frequencies of the lower harmonics the waveforms are nearly sinusoidal. However, at places tuned to higher frequencies individual harmonics are not resolved. The waveforms at places on the BM responding to higher harmonics are complex, but they all have a repetition rate equal to the F_0 of the sound. Furthermore, for this cosine-phase stimulus, there are distinct low-amplitude portions between the major waveform peaks.

Consider now the effect of changing the phase of a single high harmonic. The effect on the stimulus waveform is shown in figure 7. The upper trace shows the sum of the first 20 harmonics of a 100 Hz fundamental with cosine starting phase, all harmonics having equal amplitude. The lower trace shows the waveform resulting from changing the starting phase of the 15th harmonic (at 1500 Hz) by 30° . The phase change results in a change in the low-amplitude portions of the waveform between the peaks. Figure 8 illustrates the effects that might be observed on the BM at the place tuned to 1500 Hz. The upper trace shows the simulated waveform at that place when all harmonics have cosine starting phase. The lower trace shows the effect of shifting the phase of the 15th harmonic by 30° . The phase-shifted component ‘appears’ in the low-amplitude portion of the waveform between the major peaks.

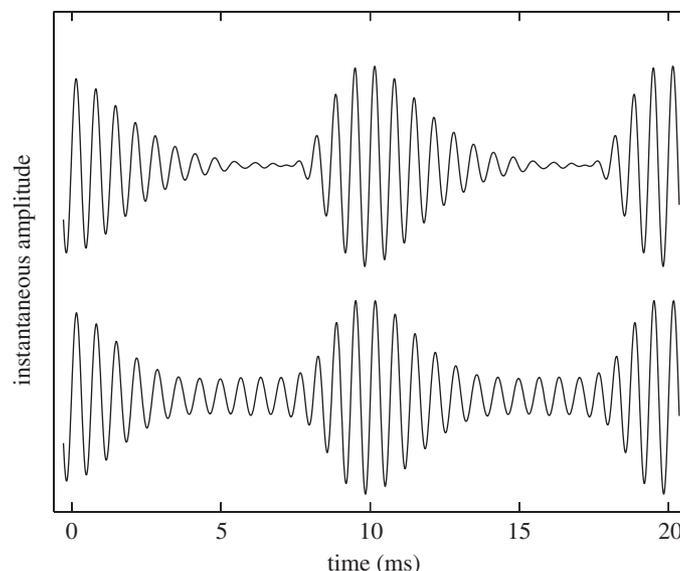


Figure 8. Output of auditory filter at 1500 Hz. A simulation of the waveform that would be observed on the BM at the place tuned to 1500 Hz in response to the input waveforms shown in figure 7. When all harmonics are added in cosine phase (top) there are distinct low-amplitude portions between major peaks. When the phase of the 15th harmonic is shifted by 30° (bottom), a 1500 Hz sinusoid appears in the low-amplitude portions.

Correspondingly, human listeners hear a 1500 Hz pure tone ‘pop out’ from the buzzy complex tone.

Duifhuis (1970, 1971) was one of the first researchers to report that shifting the phase of a high harmonic in a pulse train can result in people hearing a pitch corresponding to the phase-shifted harmonic. One special case occurs when the phase is changed by 180° . One can think of this phase change as equivalent to adding an out-of-phase harmonic to the original cosine-phase stimulus, the amplitude of the added harmonic being double that of the original harmonic of the same frequency. One-half of the amplitude of the added harmonic is ‘used up’ in cancelling the original harmonic, and the other half then results in the presence of a harmonic with opposite phase to the original harmonic. Another special case occurs when a high harmonic is removed from a periodic pulse train (or a cosine-phase complex tone with many equal-amplitude harmonics). Removing the harmonic results in the perception of a tone corresponding to the frequency of that harmonic: we hear a pitch corresponding to a sinusoidal component that is not present in the power spectrum! The explanation for this is that the removed harmonic ‘appears’ in the low-amplitude portions of the waveform on the BM at the place tuned to the frequency of the harmonic; the effect is similar to that illustrated in figure 8.

Moore & Glasberg (1989) measured the limits of the ability to detect a change in phase of a single harmonic in a complex tone with 20 equal-amplitude harmonics. Thresholds for detecting the phase change are shown in figure 9, for complex tones where each harmonic had a level of 70 dB SPL. Thresholds are plotted as a function of the harmonic number of the phase-shifted harmonic, for three different F_0 s: 50, 100 and 200 Hz. Individual results are shown for three subjects. Generally, performance

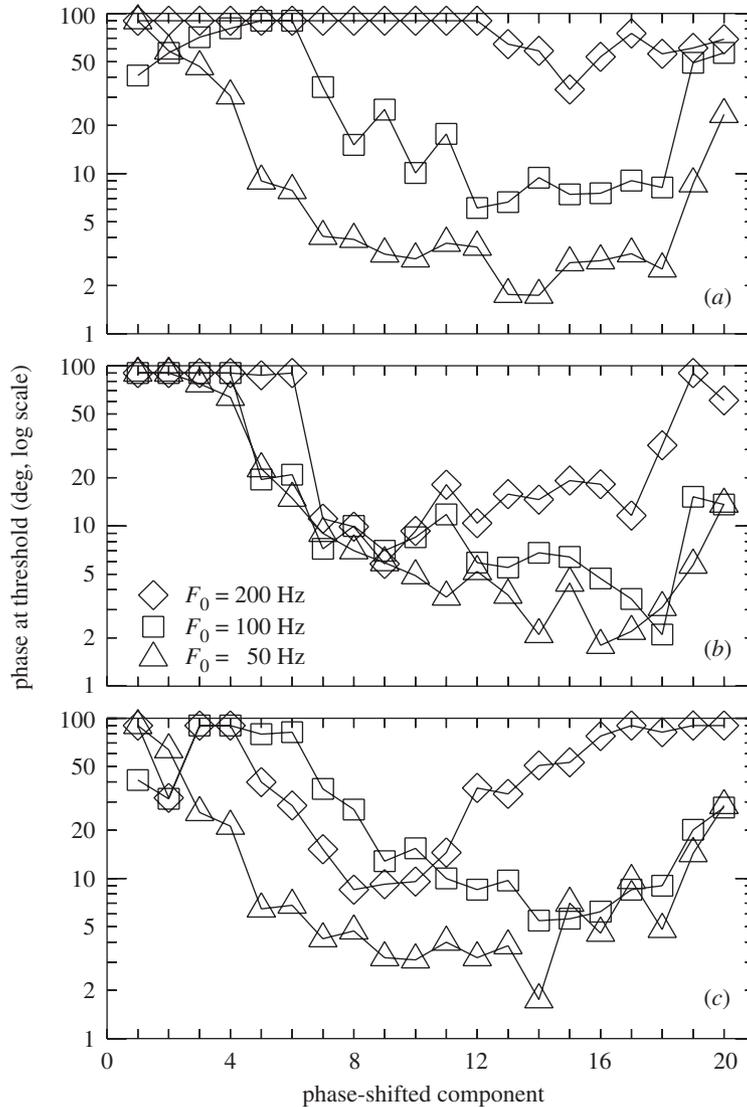


Figure 9. Thresholds for detecting a shift in phase of a single component in a harmonic complex tone, plotted as a function of the harmonic number of the shifted component. Each panel shows results for one subject. The parameter is the fundamental frequency of the complex tone: 50 Hz (triangles), 100 Hz (squares) or 200 Hz (diamonds). Points plotted at 90° indicate that the phase change was not detectable. Data from Moore & Glasberg (1989).

was better for the lower F_0 , and for harmonic numbers between about 8 and 16. For these conditions, subjects reported that the phase-shifted component appeared to 'pop out', and was heard as a pure tone. For the 50 Hz F_0 , some thresholds were as low as 2° .

The finding that performance was better for the lower F_0 s can be explained in two ways. Firstly, for a given place on the BM, the lower the F_0 the greater the number of harmonics that interact at that place. The greater number of harmonics

leads to a more highly modulated waveform on the BM, and hence more distinct low-amplitude portions. This improves the ability to detect the phase-shifted harmonic in those low-amplitude portions. Secondly, the auditory system may be able to extract information *selectively* from the low-amplitude portions of the waveforms on the BM. This is sometimes described as ‘listening in the dips’ or listening at selected intervals in time. This form of selective listening is probably limited by the temporal resolution of the auditory system. Information from the waveform dips can only be selectively extracted when the dips are sufficiently long. The temporal resolution of the auditory system can be modelled as a sliding temporal integrator or ‘window’ which smoothes the representation of sounds flowing from the cochlea (Penner & Cudahy 1973; Moore *et al.* 1988; Oxenham & Moore 1994; Moore 1997). The effective duration of this window appears to be *ca.* 8–11 ms (Oxenham & Moore 1994; Peters *et al.* 1995). Such an integrator allows dip-listening to be very effective when $F_0 = 50$ Hz (period 20 ms), somewhat effective when $F_0 = 100$ Hz (period 10 ms) and rather ineffective when $F_0 = 200$ Hz (period 5 ms).

The relatively poor phase discrimination for the low harmonic numbers can be explained by the fact that low harmonics are resolved on the BM. As a result, the waveforms on the BM at places tuned to the low harmonics are not altered when the phase of a single low harmonic is shifted. The relatively poor performance for harmonic numbers approaching 20 can be explained by the fact that the complex tones contained only 20 harmonics. At the place on the BM tuned to the frequency of the 19th or 20th harmonic, the waveform in response to the ‘reference’ cosine-phase stimulus would not have shown distinct low-amplitude portions, because this requires the interference of many harmonics with frequencies both above and below the CF of the relevant place.

Moore & Glasberg (1989) found that, for intermediate harmonic numbers (10–15), thresholds for detecting the phase change were strongly dependent on overall sound level, being large for low levels and small for high levels. This effect may depend partly on the operation of the active mechanism in the cochlea, which applies a large amount of gain to low-amplitude portions of the waveform on the BM, and less gain to high-amplitude portions. This happens because the gain can change very rapidly, *i.e.* the compression is fast acting (Recio *et al.* 1998). When the overall level was high, the low-amplitude portions of the waveforms, which contain information about the phase-shifted harmonic (see figure 8), would have received much more gain than the high-amplitude portions. This would be expected to enhance the audibility of the phase changes. In contrast, when the overall level was low, the low-amplitude and high-amplitude portions would have received similar gains, as the active mechanism does not apply compression at low levels. The enhancement of low-amplitude portions of the waveform by fast-acting compression may account for the finding that, at high levels, detection of phase changes was possible for some harmonics of the 200 Hz F_0 even though the period in this case was too short to allow selective listening in the dips. Probably, in this case, the detection cue was a change in the overall magnitude at the place on the BM tuned to the frequency of the phase-shifted harmonic.

(d) Phase effects in masking

By manipulating the phases of the components in a complex sound, it is possible to create waveforms that differ markedly in their envelopes after filtering on the BM.

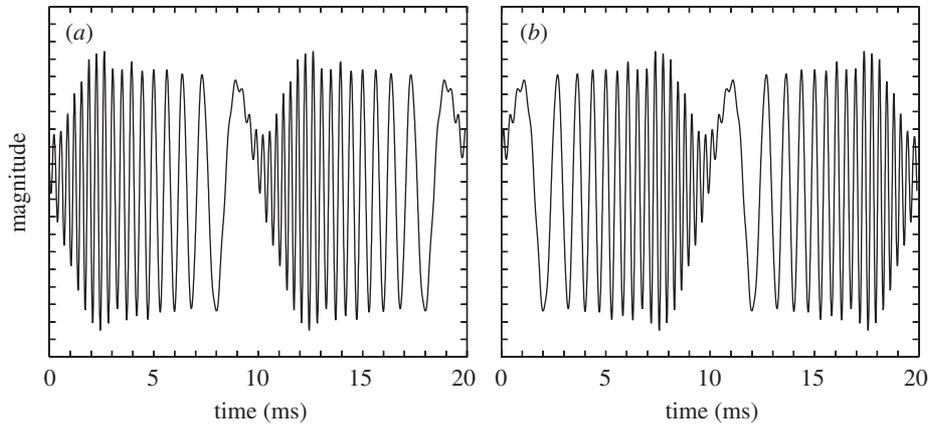


Figure 10. Waveforms of a Schroeder-positive stimulus (a) and a Schroeder-negative stimulus (b). The two waveforms were composed by adding together a large number of equal-amplitude harmonics of a 100 Hz fundamental frequency. The phases of the components were chosen according to one of the two equations given in the text.

This can happen even when the physical envelopes of the stimuli are very similar. The envelopes on the BM can be explored by using the complex sounds as maskers, as described below. The results can then be used to infer the phase response at specific places along the BM.

Kohlrausch & Sander (1995) used complex tones containing many equal-amplitude harmonics of a 100 Hz fundamental. The phase, θ_n , of the n th harmonic was determined according to either of two equations described by Schroeder (1970):

$$\theta_n = +\pi n(n-1)/N, \quad (3.2)$$

$$\theta_n = -\pi n(n-1)/N, \quad (3.3)$$

where N is the number of harmonics in the complex sound. In these equations, if a plus sign is used (Schroeder-positive), the phase increases progressively with increasing harmonic number, while if a minus sign is used (Schroeder-negative phase), the phase decreases progressively with increasing harmonic number. Both of these phases give waveforms with almost flat envelopes; one waveform is a time-reversed version of the other (see figure 10). The Schroeder-negative waveform is like a repeated upward frequency sweep (up chirp), while the Schroeder-positive waveform is like a repeated downward frequency sweep (down chirp).

Kohlrausch & Sander (1995) presented evidence that the two waveforms give very different responses on the BM at a place tuned to 1100 Hz; the Schroeder-positive phase leads to a very 'peaky' waveform, while the Schroeder-negative phase leads to a waveform whose envelope remains nearly flat. They showed this by using the complex tones as maskers, and measuring the threshold for a very brief (2.5 ms duration at the 6 dB down points on the Hanning envelope) 1100 Hz sinusoidal signal presented at various times during the 10 ms period of the masker. For the Schroeder-negative phase, the threshold of the signal varied little with the time of presentation. For the Schroeder-positive phase, the threshold varied over a considerable range, indicating that the internal representation of the masker was fluctuating within one period.

The difference in BM response to the Schroeder-positive and Schroeder-negative stimuli is assumed to occur because the phase response of the BM at a specific place

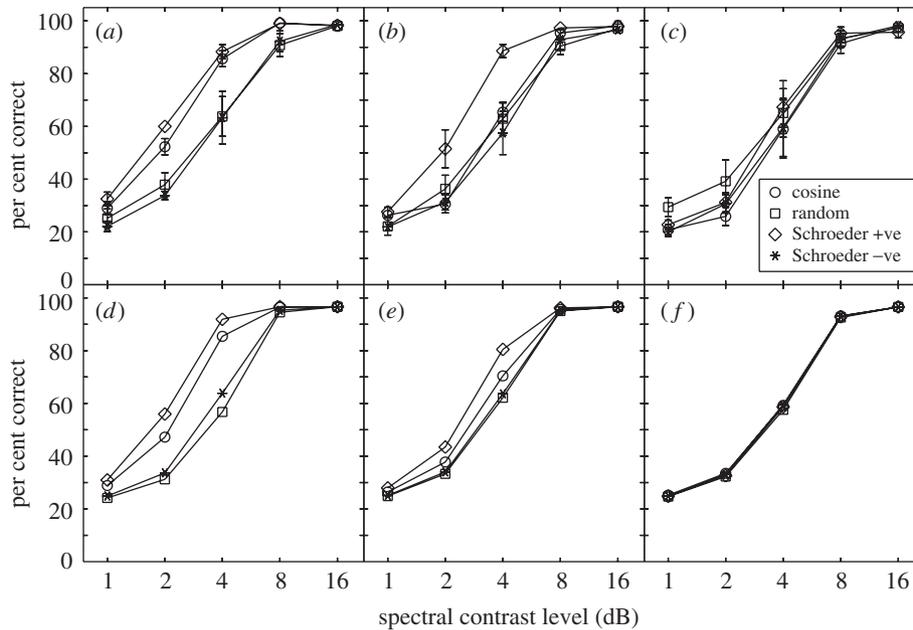


Figure 12. (a)–(c) Results of Alcántara *et al.* (1996). Percentage correct scores are plotted as a function of spectral contrast (1, 2, 4, 8 and 16 dB) for the four phase conditions (cosine, random, Schroeder-positive and Schroeder-negative). Each part shows results for one overall level. Error bars indicate plus and minus one standard error across subjects. Error bars are omitted when they would be smaller than the symbol used. (d)–(f) Predictions of a model that is described in the text. Adapted from Alcántara *et al.* (1996). (a), (d) 85 dB SPL; (b), (e) 65 dB SPL; and (c), (f) 45 dB SPL.

were complex tones containing the first 35 harmonics of a 100 Hz fundamental. All of the harmonics below 3000 Hz were equal in amplitude except for three pairs of successive harmonics, at frequencies corresponding to the first three formants of six vowels, which were incremented in level relative to the background harmonics by 1, 2, 4, 8 and 16 dB; the amount of increment was referred to as spectral contrast. The spectrum of one of the stimuli is illustrated schematically in figure 11.

The components in the harmonic complexes were added in four different starting phase relationships; cosine, random, Schroeder-positive and Schroeder-negative. These phases were chosen to give several distinct temporal patterns at the outputs of the BM filters (i.e. the temporal pattern at a given place on the BM would be different for each of the four phases). The stimuli were presented at three overall levels; 85, 65 and 45 dB SPL.

The mean results (averaged across subjects and vowels) are shown in figure 12*a–c*. Performance was similar for the random and Schroeder-negative phases and did not vary as a function of level. Performance for the cosine and Schroeder-positive phase conditions was better than for the other two phase conditions, but decreased as the level was reduced. Performance for all four phase conditions was equivalent for the lowest level.

Alcántara *et al.* (1996) argued that the good performance observed for the cosine- and Schroeder-positive-phase stimuli at high levels occurred because these stimuli

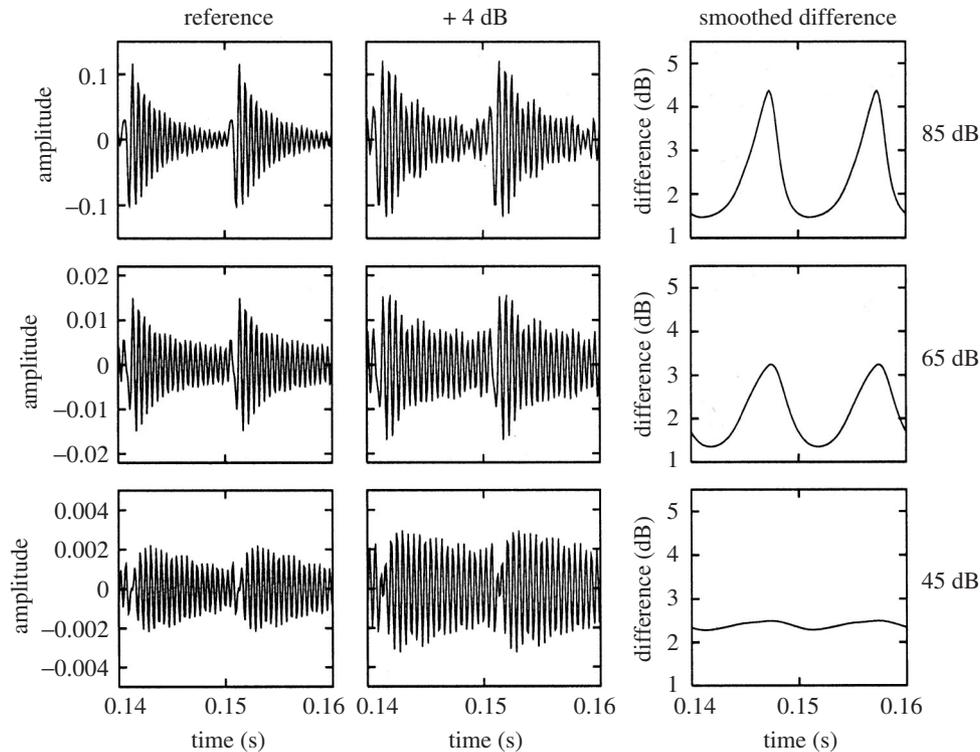


Figure 13. Temporal waveforms at the output of a nonlinear BM filter with a characteristic frequency of *ca.* 2150 Hz for a cosine-phase reference stimulus (left column) and a test stimulus with 4 dB spectral contrast (middle column). The right-hand column shows the variation in internal contrast as a function of time for two periods of the stimulating waveforms. The time axes are aligned with respect to one another. Each row shows results for one level. Note that the amplitude scale in each of the rows is linear, and varies across rows. Adapted from Alcántara *et al.* (1996).

produce waveforms on the BM with distinct low-amplitude portions. To account for their results, they used a model including the following stages.

- (1) A bank of nonlinear BM filters with level-dependent amplitude characteristics, a compressive input–output function, and a realistic phase response (Strube 1985, 1986).
- (2) A sliding temporal integrator to model the temporal resolution of the auditory system.
- (3) An across-channel comparator.
- (4) A decision device.

It was assumed that performance depended on the difference in response between channels (corresponding to different places on the BM) responding mainly to the incremented harmonics and channels responding mainly to the background harmonics.

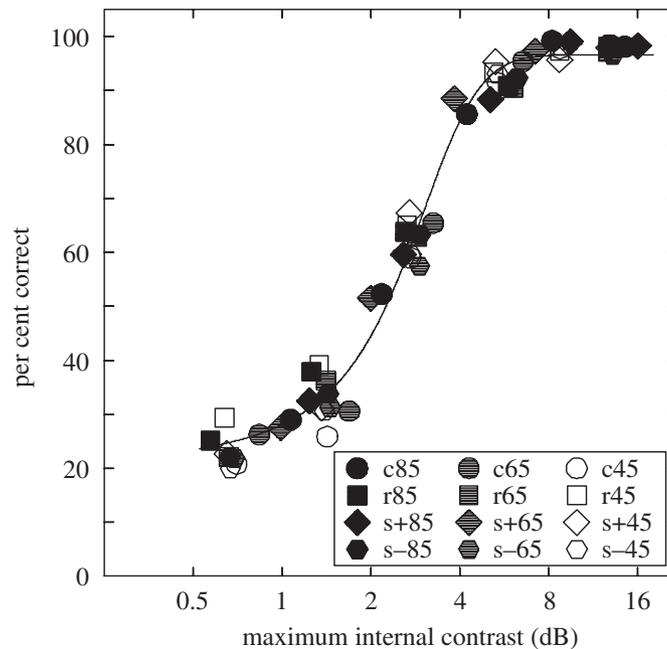


Figure 14. Results of Alcántara *et al.* (1996) showing the percentage correct for each condition (a specific combination of spectral contrast level, phase and overall level) plotted as a function of the maximum internal contrast derived from the model. The phases are indicated as follows: c, cosine; r, random; s+, Schroeder positive; s-, Schroeder negative. The numbers refer to the overall level of the stimuli in dB SPL. Adapted from Alcántara *et al.* (1996).

The operation of the model is illustrated in figure 13. The left column shows responses for ‘reference’ cosine-phase stimuli, which have zero spectral contrast. When formant harmonics are incremented in level, this produces changes in the low-amplitude portions of the waveforms at points on the BM tuned close to the formant frequencies, as illustrated in the middle column of figure 13. The right column of figure 13 shows the difference in response for these two cases, after smoothing each response by the sliding temporal integrator. The maximum value of the difference was called the ‘maximum internal contrast’ and it corresponds to the peak values in the right column of figure 13. If performance depends on the maximum internal contrast, then the results for all the different conditions should collapse onto a single function if they are expressed as per cent correct versus maximum internal contrast. Figure 14 shows that this was indeed the case. Figure 12*d–f* shows the predictions of this model for mean performance as a function of spectral contrast, overall level and phase. The predictions correspond very well with the data in parts (a)–(c).

Alcántara *et al.* (1996) found that the effects of level could not be accounted for using the BM model without compression as the initial stage; it was important to use a model with a realistic nonlinear input–output function. This suggests that the compressive nonlinearity on the BM plays an important role in producing the observed phase and level effects.

(f) The role of interference in otoacoustic emissions

If a low-level click is applied to the ear, then it is possible to detect sound reflected from the ear, using a sensitive microphone in the ear canal (Kemp 1978). The early part of this reflected sound appears to come from the middle ear, but some sound can be detected for delays from 5 to 60 ms following the instant of click presentation. These delays are far too long to be attributed to the middle ear, and they almost certainly result from activity in the cochlea itself. The reflected sounds are known as transient evoked otoacoustic emissions.

Although the input click in Kemp's experiment contained energy over a wide range of frequencies, only certain frequencies were present in the reflected sound. Kemp (1980) suggested that the reflections are generated at points on the BM, where there is a gradient or discontinuity in the mechanical or electrical properties of the BM or the organ of Corti. The response is nonlinear, in that the reflected sound does not have an intensity in direct proportion to the input intensity. In fact, the relative level of the reflection is greatest at low sound levels; the emission grows *ca.* 3 dB for each 10 dB increase in input level. This nonlinear behaviour can be used to distinguish the response arising from the cochlea from the linear middle ear response. Sometimes the amount of energy reflected from the cochlea at a given frequency may exceed that which was present in the input sound (Burns *et al.* 1998), which indicates involvement of an active mechanism. Many ears emit sounds in the absence of any input and these can be detected in the ear canal (Zurek 1981). Such sounds are called spontaneous otoacoustic emissions, and their existence indicates that there is a source of energy within the cochlea that is capable of generating sounds.

Otoacoustic emissions can be very stable in a given individual, both in waveform and frequency content, but each ear gives its own characteristic response. Responses tend to be strongest between 500 and 2500 Hz, probably because transmission from the cochlea back through the middle ear is most efficient in this range. Otoacoustic emissions can be measured for brief tone bursts as well as clicks, and it is even possible to detect a reflected component in response to continuous stimulation with a sinusoid. This component is called a stimulus-frequency otoacoustic emission.

A remarkable feature of evoked otoacoustic emissions is that their frequency spectra are roughly periodic; they contain a fairly regular sequence of peaks and valleys. This is illustrated in figure 15, which shows the magnitude of the sound level recorded in the ear canal in response to a sinusoid as a function of the frequency of the sinusoid (Shera & Guinan 1999). For each curve shown, the earphone was driven with a constant voltage. For the highest input voltage, which resulted in a sound that was *ca.* 50 dB above the threshold for detection, the sound level in the ear canal varies smoothly with frequency, as the otoacoustic emissions are small relative to the input. However, at low levels there are distinct oscillations with frequency. These result from interference between the otoacoustic emissions and the input stimulus.

In humans, the peaks in the otoacoustic emissions occur at intervals, Δf , given roughly by $\Delta f/f = 1/15$, when the centre frequency, f , is above 1 kHz. To explain this regularity, Zweig & Shera (1995) have proposed that evoked otoacoustic emissions originate through a novel reflection mechanism. Travelling waves moving in the normal 'forward' direction (from base to apex), reflect off random irregularities in the micromechanics of the organ of Corti, giving a backward-travelling wave. The reflections can be evoked by any part of the forward-travelling wave, but the dominant

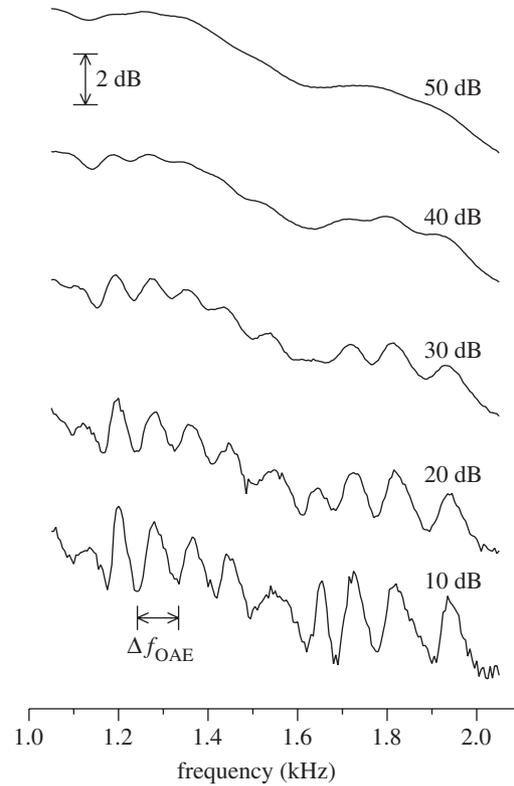


Figure 15. The sound level recorded in the ear canal in response to a sinusoid, as a function of the frequency of the sinusoid. For each curve shown, the earphone was driven with a constant voltage. The number by each curve indicates the approximate amount by which the stimulus level exceeded the detection threshold at 1300 Hz. Adapted from Shera & Guinan (1999), and reproduced with permission.

reflections occur when the irregularities are located within a broad region around the peak of the forward-travelling wave. This happens simply because the amplitude is much higher at the peak than elsewhere.

Within the peak region, there may be many reflectors, arrayed in an irregular manner. This irregular array of reflectors can be considered as a superposition of many spatial frequencies. However, it is also the case that, within this region, the wavelength of the travelling wave varies only slowly. When the spatial period of the reflectors matches half the wavelength of the travelling wave at its peak, wavelets reflected at different locations within the peak region combine in phase, and the reflectance is large. The condition that the spatial period matches half of the wavelength arises because then the total distance traversed by a wavelet travelling ‘round trip’ between reflectors (there and back) is one wavelength. When the spatial period of the reflectors does not match half the wavelength of the travelling wave at its peak, wavelets reflected at different locations combine out of phase, and tend to cancel one another. The peak region thus acts as a spatial frequency filter that selects irregularities arrayed within a narrow range of spatial frequencies. The phase of the net reflected wave rotates monotonically with the frequency of the input at a rate

determined by the wavelength of the travelling wave in the region of its peak. The reflected wave results in pressure changes at the stapes, and these propagate through the middle ear to create a sound in the ear canal. Interference between this sound and the stimulus tone creates the observed spectral periodicity in the otoacoustic emission. The spectral spacing of peaks in the emissions, Δf , in a given frequency region around frequency f , is proportional to the wavelength of the travelling wave at its peak, when the input stimulus has a frequency close to f . Therefore, measurements of otoacoustic emissions can be used to determine properties of the travelling wave in humans non-invasively.

4. Concluding remarks

This paper has reviewed a variety of interference effects occurring in the outer ear and cochlea. The effects in the outer ear are relatively straightforward, involving a linear superposition of sounds entering the ear canal directly and sounds entering the ear canal after reflections from the pinna, head and torso. The resulting spectral patterns play a strong role in the localization of sounds in space. Interference effects in the inner ear (cochlea) are more complex as they are influenced by the active mechanism, which introduces strong nonlinearities. Interference effects in the cochlea play a role in many aspects of auditory perception, including the perception of consonance and dissonance, the perception of pitch, the perception of changes in phase, and the perception of timbre. Interference effects in the cochlea may also play a role in producing the spectral regularity observed in otoacoustic emissions.

I thank José Alcántara, Brian Glasberg, Aleksander Sek and Chris Shera for assistance with figures. I also thank Hedwig Gockel, Michael Stone and George Zweig for helpful comments on an earlier version of this paper.

References

- Alcántara, J. I. & Moore, B. C. J. 1995 The identification of vowel-like harmonic complexes: effects of component phase, level, and fundamental frequency. *J. Acoust. Soc. Am.* **97**, 3813–3824.
- Alcántara, J. I., Holube, I. & Moore, B. C. J. 1996 Effects of phase and level on vowel identification: data and predictions based on a nonlinear basilar-membrane model. *J. Acoust. Soc. Am.* **100**, 2382–2392.
- Ashmore, J. F. 1987 A fast motile response in guinea pig outer hair cells: the cellular basis of the cochlear amplifier. *J. Physiol.* **388**, 323–347.
- Ashmore, J. F. & Mammano, F. 2001 Can you still see the cochlea for the molecules? *Curr. Opin. Neurobiol.* **11**, 449–454.
- Batteau, D. W. 1967 The role of the pinna in human localization. *Proc. R. Soc. Lond. B* **168**, 158–180.
- Blauert, J. 1983 *Spatial hearing*. Cambridge, MA: MIT Press.
- Blauert, J. 1996 *Spatial hearing: the psychophysics of human sound localization*. Cambridge, MA: MIT Press.
- Burns, E. M., Keefe, D. H. & Ling, R. 1998 Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies. *J. Acoust. Soc. Am.* **103**, 462–474.
- Camalet, S., Duke, T., Julicher, F. & Prost, J. 2000 Auditory sensitivity provided by self-tuned critical oscillations of hair cells. *Proc. Natl Acad. Sci. USA* **97**, 3183–3188.

- Cariani, P. A. & Delgutte, B. 1996a Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* **76**, 1698–1716.
- Cariani, P. A. & Delgutte, B. 1996b Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch and the dominance region for pitch. *J. Neurophysiol.* **76**, 1717–1734.
- de Boer, E. 1976 On the ‘residue’ and auditory pitch perception. In *Handbook of sensory physiology* (ed. W. D. Keidel & W. D. Neff), vol. 3, pp. 479–584. Springer.
- Duifhuis, H. 1970 Audibility of high harmonics in a periodic pulse. *J. Acoust. Soc. Am.* **48**, 888–893.
- Duifhuis, H. 1971 Audibility of high harmonics in a periodic pulse. II. Time effects. *J. Acoust. Soc. Am.* **49**, 1155–1162.
- Eguiluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J. & Magnasco, M. O. 2000 Essential nonlinearities in hearing. *Phys. Rev. Lett.* **84**, 5232–5235.
- Fletcher, H. 1940 Auditory patterns. *Rev. Mod. Phys.* **12**, 47–65.
- Glasberg, B. R. & Moore, B. C. J. 1990 Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **47**, 103–138.
- Gold, T. 1948 Hearing. II. The physical basis of the action of the cochlea. *Phil. Trans. R. Soc. Lond. B* **135**, 492–498.
- Goldstein, J. L. 1973 An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* **54**, 1496–1516.
- Helmholtz, H. L. F. 1863 *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: F. Vieweg.
- Kameoka, A. & Kuriyagawa, M. 1969 Consonance theory. I. Consonance of dyads. *J. Acoust. Soc. Am.* **45**, 1451–1459.
- Kemp, D. T. 1978 Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* **64**, 1386–1391.
- Kemp, D. T. 1980 Towards a model for the origin of cochlear echoes. *Hear. Res.* **2**, 533–548.
- Kohler, A. & Sander, A. 1995 Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J. Acoust. Soc. Am.* **97**, 1817–1829.
- Kohler, A., Fassel, R. & Dau, T. 2000 The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J. Acoust. Soc. Am.* **108**, 723–734.
- Lindsay, P. H. & Norman, D. A. 1972 *Human information processing*. Academic.
- Lundeen, C. & Small, A. M. 1984 The influence of temporal cues on the strength of periodicity pitches. *J. Acoust. Soc. Am.* **75**, 1578–1587.
- Meddis, R. & Hewitt, M. 1991 Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J. Acoust. Soc. Am.* **89**, 2866–2882.
- Middlebrooks, J. C., Makous, J. C. & Green, D. M. 1989 Directional sensitivity of sound pressure levels in the human ear canal. *J. Acoust. Soc. Am.* **61**, 89–108.
- Moore, B. C. J. 1977 Effects of relative phase of the components on the pitch of three-component complex tones. In *Psychophysics and physiology of hearing* (ed. E. F. Evans & J. P. Wilson), pp. 349–358. Academic.
- Moore, B. C. J. 1995 Frequency analysis and masking. In *Hearing* (ed. B. C. J. Moore), pp. 161–205. Academic.
- Moore, B. C. J. 1997 *An introduction to the psychology of hearing*, 4th edn. Academic.
- Moore, B. C. J. & Glasberg, B. R. 1989 Difference limens for phase in normal and hearing-impaired subjects. *J. Acoust. Soc. Am.* **86**, 1351–1365.
- Moore, B. C. J. & Glasberg, B. R. 2001 Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* **110**, 1067–1073.

- Moore, B. C. J. & Ohgushi, K. 1993 Audibility of partials in inharmonic complex tones. *J. Acoust. Soc. Am.* **93**, 452–461.
- Moore, B. C. J., Glasberg, B. R., Plack, C. J. & Biswas, A. K. 1988 The shape of the ear's temporal window. *J. Acoust. Soc. Am.* **83**, 1102–1116.
- Ohm, G. S. 1843 Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Annln Phys. Chem.* **59**, 513–565.
- Oxenham, A. J. & Moore, B. C. J. 1994 Modeling the additivity of nonsimultaneous masking. *Hear. Res.* **80**, 105–118.
- Patterson, R. D. 1987a A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* **82**, 1560–1586.
- Patterson, R. D. 1987b A pulse ribbon model of peripheral auditory processing. In *Auditory processing of complex sounds* (ed. W. A. Yost & C. S. Watson), pp. 167–179. Hillsdale, NJ: Erlbaum.
- Penner, M. J. & Cudahy, E. 1973 Critical masking interval: a temporal analog of the critical band. *J. Acoust. Soc. Am.* **54**, 1530–1534.
- Peters, R. W., Moore, B. C. J. & Glasberg, B. R. 1995 Effects of level and frequency on the detection of decrements and increments in sinusoids. *J. Acoust. Soc. Am.* **97**, 3791–3799.
- Plenge, G. 1972 Über das Problem der Im-Kopf-Lokalisation. *Acustica* **26**, 213–221.
- Plenge, G. 1974 On the differences between localization and lateralization. *J. Acoust. Soc. Am.* **56**, 944–951.
- Plomp, R. 1964 The ear as a frequency analyzer. *J. Acoust. Soc. Am.* **36**, 1628–1636.
- Plomp, R. 1976 *Aspects of tone sensation*. Academic.
- Plomp, R. & Levelt, W. J. M. 1965 Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**, 548–560.
- Plomp, R. & Mimpen, A. M. 1968 The ear as a frequency analyzer. II. *J. Acoust. Soc. Am.* **43**, 764–767.
- Plomp, R. & Steeneken, H. J. 1968 Interference between two simple tones. *J. Acoust. Soc. Am.* **43**, 883–884.
- Recio, A., Rich, N. C., Narayan, S. S. & Ruggero, M. A. 1998 Basilar-membrane responses to clicks at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* **103**, 1972–1989.
- Ritsma, R. J. & Engel, F. L. 1964 Pitch of frequency modulated signals. *J. Acoust. Soc. Am.* **36**, 1637–1655.
- Robles, L., Ruggero, M. A. & Rich, N. C. 1986 Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves and response phases. *J. Acoust. Soc. Am.* **80**, 1364–1374.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S. & Robles, L. 1997 Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* **101**, 2151–2163.
- Ryan, A. F. & Dallos, P. 1984 Physiology of the cochlea. In *Hearing disorders* (ed. J. L. Northern), pp. 15–31. Boston: Little Brown.
- Schouten, J. F. 1940 The residue and the mechanism of hearing. *Proc. Koninkl. Ned. Akad. Wetenschap* **43**, 991–999.
- Schouten, J. F., Ritsma, R. J. & Cardozo, B. L. 1962 Pitch of the residue. *J. Acoust. Soc. Am.* **34**, 1418–1424.
- Schroeder, M. R. 1970 Synthesis of low peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans. Inform. Theory* **IT-16**, 85–89.
- Searle, C. L., Braida, L. D., Cuddy, D. R. & Davis, M. F. 1975 Binaural pinna disparity: another auditory localization cue. *J. Acoust. Soc. Am.* **57**, 448–455.
- Sellick, P. M., Patuzzi, R. & Johnstone, B. M. 1982 Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. *J. Acoust. Soc. Am.* **72**, 131–141.

- Shackleton, T. M. & Carlyon, R. P. 1994 The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Shera, C. A. & Guinan Jr, J. J. 1999 Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J. Acoust. Soc. Am.* **105**, 782–798.
- Smooorenburg, G. F. 1970 Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* **48**, 924–941.
- Strube, H. W. 1985 A computationally efficient basilar-membrane model. *Acustica* **58**, 207–214.
- Strube, H. W. 1986 The shape of the nonlinearity generating the combination tone $2f_1 - f_2$. *J. Acoust. Soc. Am.* **79**, 1511–1518.
- Terhardt, E. 1974a On the perception of periodic sound fluctuations (roughness). *Acustica* **30**, 201–213.
- Terhardt, E. 1974b Pitch, consonance and harmony. *J. Acoust. Soc. Am.* **55**, 1061–1069.
- Uppenkamp, S., Fobel, S. & Patterson, R. D. 2001 The effects of temporal asymmetry on the detection and perception of short chirps. *Hear. Res.* **158**, 71–83.
- von Békésy, G. 1960 *Experiments in hearing*. McGraw-Hill.
- Wightman, F. L. & Kistler, D. J. 1989 Headphone simulation of free field listening. I. Stimulus synthesis. *J. Acoust. Soc. Am.* **85**, 858–867.
- Yates, G. K. 1995 Cochlear structure and function. In *Hearing* (ed. B. C. J. Moore), pp. 41–73. Academic.
- Zurek, P. M. 1981 Spontaneous narrowband acoustic signals emitted by human ears. *J. Acoust. Soc. Am.* **69**, 514–523.
- Zweig, G. & Shera, C. A. 1995 The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J. Acoust. Soc. Am.* **98**, 2018–2047.