Section of Otology

President J C Ballantyne FRCS

Meeting 5 November 1976

President's Address

The Hearing Ear: Variations on a Theme of Helmholtz

by John Ballantyne FRCS
*(Royal Free Hospital,
Pond Street, London NW3 2QG)*

For over a century, much of the work in physiological acoustics and many of our clinical tests of hearing have been directed to the ear's responses to pure tones, yet in reality we hardly ever hear them. Indeed, are they not, in Frank McGuckin's words, 'the figments of man's electronic ingenuity'?

This is not to deny the essential part played by them in our investigations of hearing, but pure tones do not exist in nature or in art; and it is the purpose of this address to trace the development of our knowledge about the things we hear and how we hear them and to convince you (I hope) of the veracity of the statement which I have just made. And the story begins with Helmholtz.

Helmholtz

Physicist, physiologist, physician and philosopher, this intellectual giant of a man was 'one of the last [of the] great universalists of science' (Margenau 1954; see Helmholtz 1954), who belonged to 'a dying age in which a full synthetic view of nature was still possible, in which one man could not only unify the practice and teaching of medicine, physiology, anatomy and physics, but could also relate these sciences significantly and lastingly to the fine arts'.

At the age of 21 he acquired his doctorate with a thesis in which he made the fundamental observation that nerve fibres originate in ganglion cells; he measured the speed of nerve impulses; he discovered the law of conservation of energy; and before he was 30, he had invented the ophthalmoscope. He imitated 'the vowels of the human voice' (Helmholtz 1954, p 123) with tuning forks and organ pipes; and in 1878 he published an article entitled 'The Telephone and the Quality of Sound', this only two years after the telephone was invented by Alexander Graham Bell. But it is for his great work 'On the Sensations of Tone', which appeared in 1863, that we know Helmholtz best. Subtitled 'A Physiological Basis for the Theory of Music', the larger part of it is devoted to an explanation of music in terms of acoustics and hearing; but more than one-third deals with 'the nature of sound, the analysis of voice sounds into component frequencies, the synthesis of voice sounds out of component frequencies, and the mechanism of speech and hearing' (Bergeijk *et al.* 1961, p 222).

It may surprise you to know, therefore, that Helmholtz himself made light of his great contribution to cochlear physiology, referring to it as 'an hypothesis that may be entirely dispensed with'; and that less than one page in a hundred (Helmholtz 1954, pp 129, 146–8) is devoted to his famous 'place' theory of pitch perception. I make no excuse for re-stating it here, and I will do this in his own words (Helmholtz 1954, pp 129, 146):

> In reality, if we suppose the dampers of a pianoforte to be raised, and allow any musical tone to impinge powerfully on its sounding board, we bring a set of strings into sympathetic vibration, namely all those strings, and only those, which correspond with the simple tones contained in the given musical tone . . .
>
> The radial fibres of the basilar membrane may be approximately regarded as forming a system of stretched strings . . . Consequently any exciting tone would set that part of the membrane into sympathetic vibration, for which the proper tone of one of its radial fibres . . . corresponds most nearly with the exciting tone; . . . the parts of the membrane in unison with higher tones must be looked for near the round window, and those with the deeper, near the vertex of the cochlea.

In other words Helmholtz postulated, with some anatomical support from the observations of Waldeyer and Preyer (Helmholtz 1954, p 147), that the radial fibres of the basilar membrane formed a series of tuned resonators, and that each and every pitch would cause a resonant vibration of its own particular 'place' on the membrane.

Many of you will know the story about the famous Italian tenor, Caruso, whose voice was so powerful that it could shatter a drinking glass from the length of a room. The natural period of vibration of the glass was determined by stroking a dampened finger around its rim, and the note emitted was then copied exactly by the singer. The glass 'resounded' – violently – and it broke into pieces. The idea of a series of sharply-tuned resonators is, of course, far too simple; but it remained more or less unchallenged until Békésy propounded an alternative concept of the 'travelling wave'.

Békésy

Békésy was another man of great versatility. Born in Budapest, he was awarded a baccalaureate in chemistry by the University of Berne when he was only 21; and three years later he obtained a PhD degree in physics, at the university of his native city – for a thesis on interference microscopy. Many honours followed, including no less than four honorary doctorates in medicine; and he became a Nobel laureate in 1961.

Békésy's interest in hearing was aroused, almost fortuitously, by his early work in the research laboratories of the *Hungarian Post and Telegraph*. Constant complaints from other countries about the Hungarian telephone service focused his attention on telephone receivers, and this led him to study the relative sensitivities of the receiver and the human ear. 'He soon satisfied himself', wrote Hallowell Davis (1973), 'that the ear was far superior; so good in fact that he was fascinated to try to discover just how nature did such an amazing job of acoustical engineering'.

These first stirrings of wonder were followed by dissections of the human cochlea, and subsequently by his cochlear models; and it was in these models that he observed that a 'travelling wave' (Fig 1) appears in response to acoustic stimulation. This produces a displacement of the cochlear partition, and invariably the wave progresses from base to apex. Its amplitude increases as it moves, to a position where the maximum

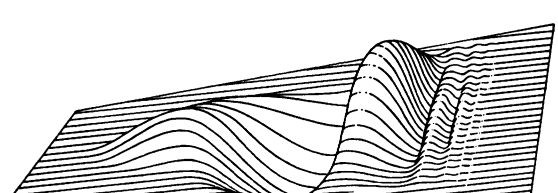

Fig 1 *The travelling wave of Békésy*

displacement is reached; and the location of this amplitude maximum is dependent upon frequency, passing from base to apex as the frequency of the stimulating tone is lowered.

In short, Helmholtz's hypothesis of a place principle was upheld, but the mechanism on which it was based – namely, a resonance mechanism – was replaced by a travelling-wave mechanism.

Rutherford

Totally at odds with any place theory was the 'telephone theory' of Rutherford, who was professor of physiology at Edinburgh when he first presented his ideas to the public. He suggested that the basilar membrane vibrates as a whole in response to acoustic stimulation and that the frequency of the stimulating tone is represented by the rate of firing in the fibres of the auditory nerve, pitch being analysed only when these impulses reach the brain. But even in Rutherford's day, it was known that the human ear could appreciate tones as high as 20 000 cycles per second (Hz), and it was also known (and, indeed had been known to Helmholtz a quarter of a century earlier) that no nerve fibre could conduct electrical impulses at anything approaching this rate. In fact, the refractory period of nervous action is such as to limit the uppermost frequency of such a system to something less than 1000 Hz.

It is evident, then, that neither of the classical theories of pitch perception can fit all the known facts; and it was not until 1949 that Wever put forward a sort of compromise solution, in the form of his 'volley theory'.

Wever

Wever suggested that the place principle probably accounts for the perception of high frequencies, which stimulate the hair-cells only in the basal turn of the cochlea; and that the telephone theory probably holds good for the low frequencies, which stimulate the whole of the organ of Corti and are represented in the auditory nerve by action potentials that are directly synchronous with the wave-forms of the applied signals.

However, in Wever's opinion, neither of the older theories could account satisfactorily for the perception of tones in the highly critical frequency range between 400 Hz and 5000 Hz; and he suggested that, within this range, groups of fibres fire asynchronously, in such a way that the frequency of the signal is presented to the central nervous system by the sequential firing of impulses in groups of fibres.

An example is given in Fig 2. The upper tracing represents a sine wave produced by a tone of an intermediate frequency, say 2000 cps. The individual lines A to D represent four nerve fibres, each one responding only to every fourth cycle. None of these fibres singly can conduct impulses at the rate of the stimulating tone shown in the sine wave, i.e. of 2000 cps; but if they are each conducting at different times (that is to say, asynchronously) in response to the same stimulating tone, the brain will receive from the combined activity of these four fibres a synchronous relay of electrical potentials which reproduce the original frequency of the sine wave. This is shown in the lowest tracing. Wever's theory fits many of the known facts and it has been supported by direct recordings of action potentials in the auditory nerve and its individual fibres.

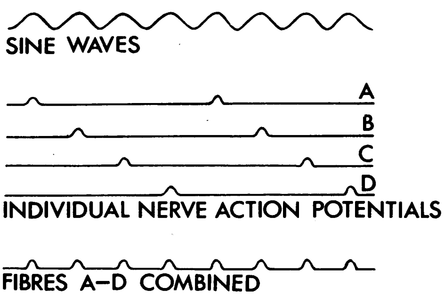

Fig 2 *The volley theory of Wever*

There may therefore be several mechanisms involved in our perception of pitch, and it must be presumed that there is a gradual transition from one mode of action to the next, from Rutherford to Wever to Helmholtz – in that order, from the lowest to the highest tones – and it is of interest that it is in cases of high-tone deafness that the ideas of Helmholtz have found greatest support from histopathological studies.

The Sounds of Music

So far I have discussed only those mechanisms whereby the cochlea is able to analyse simple, single pure tones. In fact, we spend only an insignificant part of our lives listening to them. As John Mills reminds us in his book 'A Fugue in Cycles and Bels' (1935), if the ear were not able to respond to more than one sound wave at a time 'there would be neither music nor speech as we know them'. An understanding of the ways in which the ear analyses the sounds of music can take us quite a long way towards an understanding of how it copes with the much more complicated sounds of speech. Let me begin, therefore, by discussing how the ear can distinguish a number of different musical instruments, each one playing the same note.

There are three 'families' of musical instruments (Fig 3), percussion instruments, stringed instruments and wind instruments, and almost anyone can tell the difference between them. What physical characteristics enable their sounds to be distinguished by the human ear? Joseph Fourier, French mathematician and friend of Napoleon, showed that all sounds, however complicated, could be broken down into a number of constituent sine waves, each with its own particular frequency and amplitude. Musical sounds are complex sounds; and all musical instruments are, essentially, complex resonators which, in accordance with their shape, emphasize certain overtones or harmonics representing simple multiples or ratios of the fundamental frequency. The differences in their tonal qualities which enable the listener to distinguish them by ear are due to differences in the harmonics accompanying the fundamental tone; and it is the ear's ability to recognize these overtones – and their relative intensities, in comparison with the fundamental and with one another – which permits the listener to distinguish one instrument from another.

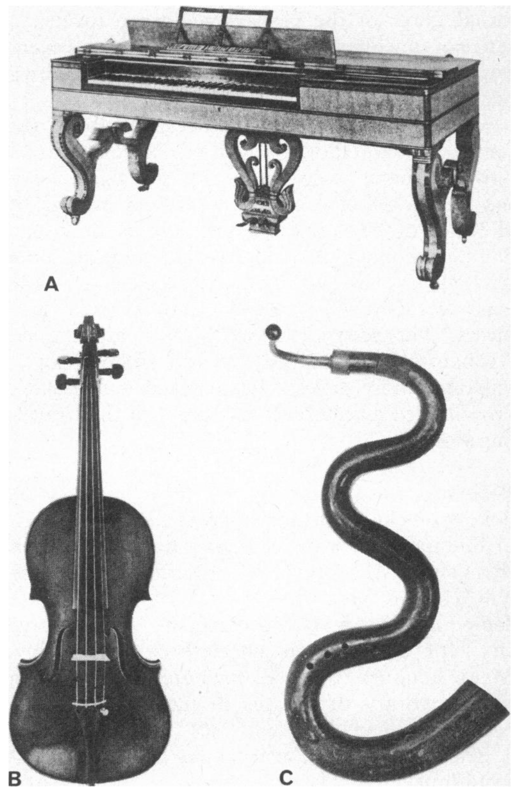

Fig 3 *Three families of musical instruments.* A, *percussion.* B, *string.* C, *wind*

Fig 4 shows the acoustic spectra of a piano and a violin. The acoustic spectrum of any instrument shows the distribution of sound energy at all frequencies throughout a given range, and Fig 4A shows the spectrum of the note A (440 cps) played on a piano. Note that the fundamental is very

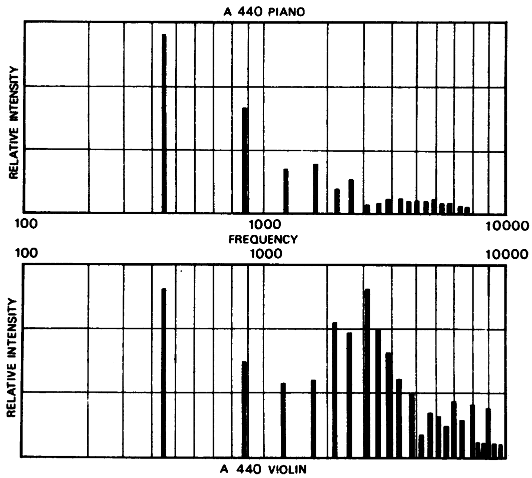

Fig 4 *Acoustic spectra.* A, *A 440 piano.* B, *A 440 violin*

strong: almost twice the intensity of the second harmonic, which is an octave higher; that the next four harmonics, at 1320, 1760, 2200, and 2640 cps, are considerably weaker, but more or less the same as one another; and that the remaining twelve harmonics, up to the eighteenth, are insignificantly small. Although these are recordable, masking makes them quite inaudible, and such components could be eliminated entirely without altering at all what we hear.

Compare this with Fig 4B, which shows the acoustic spectrum of a violin, and see how different it is. Note particularly that when the note A (440 cps) is played on a violin, the seventh harmonic is as strong as the fundamental; that the fifth, sixth and eighth are not much weaker; and that at least five other harmonics contribute significantly to the sound structure. Hence, at a peripheral level, the cochlea is capable of analysing the relative intensities of the constituent parts of complex tones. This is a Fourier analysis, and Helmholtz himself recognized it (1954, p 148).

But what happens when we listen to a multitude of notes, of widely differing pitches and on a variety of instruments, all played together? Fig 5 shows eight bars from one of Mozart's horn concertos (scored for two clarinets, two bassoons and strings, in addition to the solo instrument), with the oscillographic tracing; is it conceivable that the cochlea alone could analyse such complex sounds? In 1947 Hallowell Davis suggested that 'the various frequencies . . . in a complex sound are sorted out and cause the basilar membrane to vibrate in an equally complex pattern'. It is now well known, however, that the basilar membrane cannot possibly meet all the requirements for a series of highly-tuned resonators.

Békésy (1957) suggested that there may be 'an inhibitory mechanism which suppresses the weaker stimuli and thus sharpens considerably the sensation around the maximum'; and Capps (1967) produced some evidence that this inhibitory mechanism may lie in the efferent fibres of the cochlear nerve.

But the most convincing evidence that the ear does not always act as a simple frequency analyser of the Fourier type (that is, by breaking down complex sounds into a series of sine waves, each stimulating its own particular 'place' on the basilar membrane) has come from some fascinating work by Shcouten and his colleagues (1962) in Eindhoven. They have shown, for example, that when the acoustic output of certain organ pipes is analysed, the fundamental tone perceived by the listener may be entirely missing from the acoustic spectrum. That is to say, only overtones are

Fig 5 *Mozart's horn concerto No 3 in E flat major (K 447); last movement, bars 149–156. The oscillograph trace is shown between wind instruments (above) and stringed instruments (below). (Courtesy Professor Charles Taylor and BBC Publications)*

present in the spectrum, yet the note which the listener hears subjectively has the pitch of the fundamental tone. In other words, he is hearing a pitch whose 'place' on the basilar membrane cannot be vibrating at all; and this simply could not happen if frequency were analysed exclusively by place.

Schouten himself refers to this as the 'residue phenomenon', and many interesting examples of it are to be found in other studies. For instance, Bergeijk *et al.* (1961) have shown that when a sine wave of, say, 3000 cps is turned on and off one hundred times a second, prominent volleys can be observed in the auditory nerve, not at 3000 cps but at a rate of 100 cps; and the sound that we actually perceive when we listen to such a wave has a pitch of 100 Hz – not 3000 Hz. 'This is the same as Schouten's residue phenomenon, in which we hear the fundamental pitch while listening to only its higher overtones' (Bergeijk *et al.* 1961, p 161), and it points strongly towards a 'pitch-extracting mechanism different from and subsequent to the basilar membrane and operating in the time domain' (Schouten *et al.* 1962).

Tonndorf (1962) also believes that 'under certain conditions [the ear] might also recognize temporal features [in] an auditory signal' – high frequencies having a short travel time, low frequencies a long one. He puts forward what he calls a 'modified place concept' in which the ear's

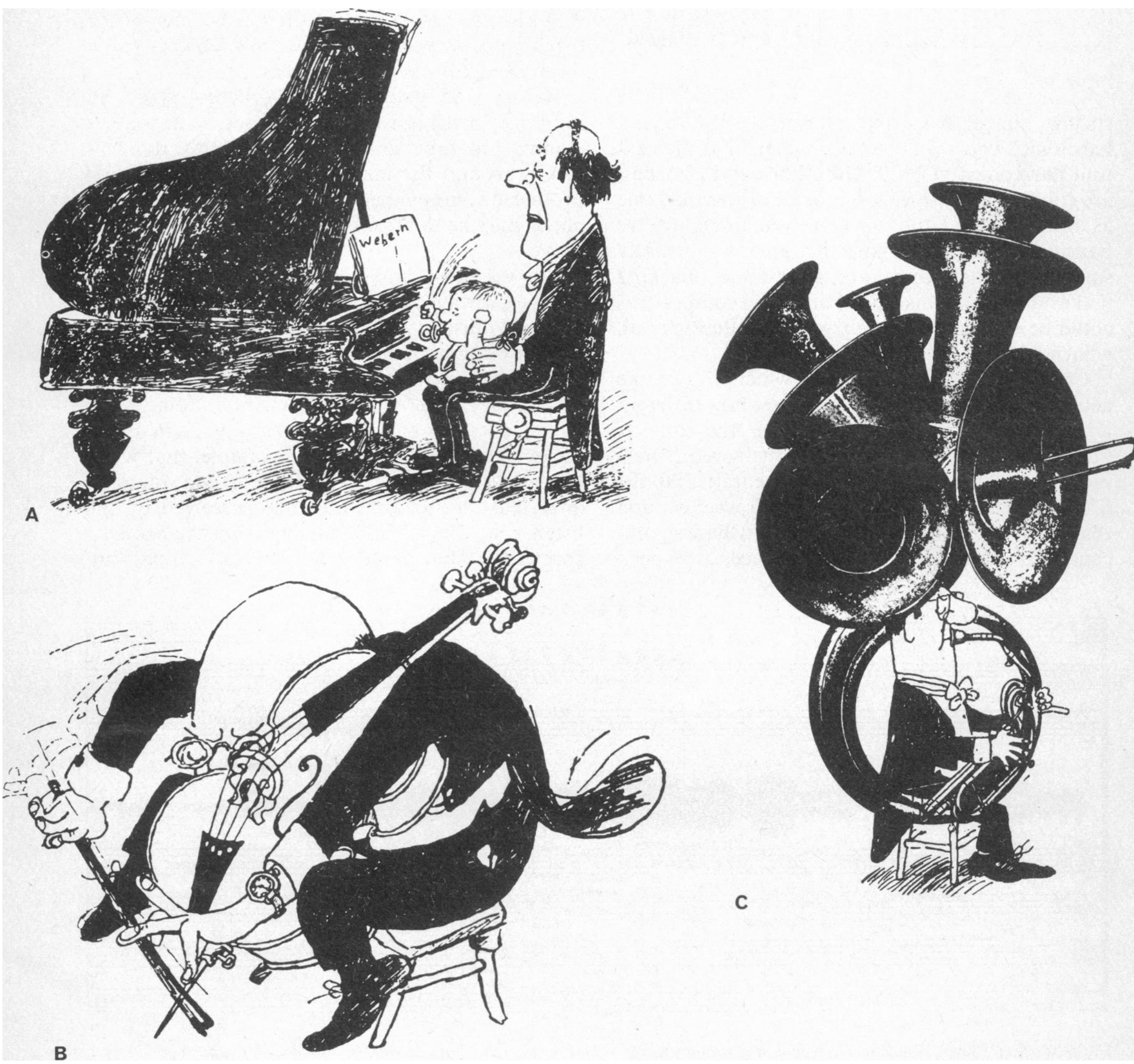

Fig 6 *Different ways of playing musical instruments.* A, *hitting.* B, *scraping.* C, *blowing.* (*Courtesy Gerard Hoffnung and Dobson Books Ltd*)

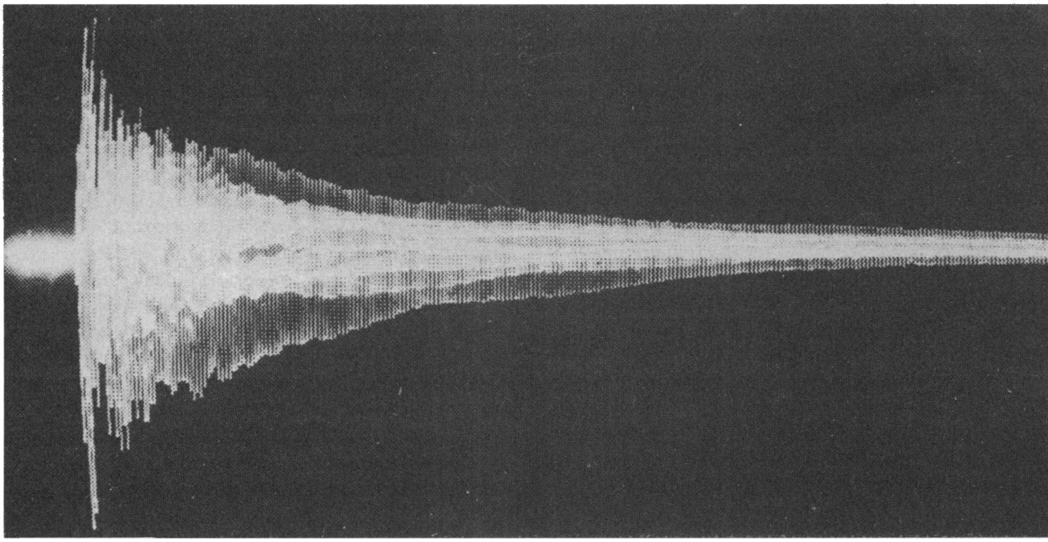

Fig 7 *Oscillograph tracing of note produced by plucking the top string of a Spanish guitar.* (*Courtesy Professor Charles Taylor and BBC Publications*)

response to simple sine waves probably represents one extreme, approaching a pure Fourier (or frequency) type of analysis; in which most other signals, including complex harmonic sounds, occupy a position intermediate between frequency analysis and time analysis; and in which the opposite extreme, approaching a pure time analysis, is occupied by the so-called 'transients'.

The acoustic spectra that I showed for the piano and the violin showed the 'steady-state' sounds emitted by these instruments: that is to say, the sounds produced after the piano string had been struck by the hammer, and after the violin string had been set into vibration by the bow. But before this steady state was reached, there was a moment of impact as the instrument was struck or bowed. For there are several ways in which a musical instrument can be played: it can be hit, it can be scraped, or it can be blown (Fig 6A, B, C).

As each instrument is brought into action, it will emit a brief, sudden, complex sound – as the hammer hits the string, the horsehair scrapes the catgut, or the blast of air is forced into the tube; and it is this first split second of a note which is known as the 'starting transient' – 'starting' because it comes at the beginning, 'transient' because it soon disappears. This is shown by the oscillographic trace (Fig 7) of the note produced by plucking the top string of a Spanish guitar; this note dies away very quickly.

Richardson (1954) wrote that 'In spite of their evanescent nature, the view is now held that it is these transients which enable the listener to distinguish the sounds of different musical instruments.' This is certainly an overstatement, yet there can be no doubt about the extreme importance of musical transients; and they play a vital part, too, in our discrimination of the highly complex sounds of speech.

Pitch Discrimination

The human ear responds to a total frequency range of approximately ten octaves (Fig 8), from 20 Hz to 20 000 Hz. The minimum perceptible difference in terms of pitch varies in different parts of this range. For example, above 500 Hz, the minimum perceptible difference remains approximately constant as a ratio; but below 500 Hz, the ear can discriminate between two frequencies to about 3 Hz, regardless of the pitch. At the very lowest end of our hearing range, it may be practically impossible to distinguish between two adjacent notes; but does this really matter? As Dr Scholes has put it, in his 'Oxford Companion to Music' (1950), 'Even 16-foot C, with 32 vibrations per second, may be out of tune on the pedal bourdon of a church organ without any member of the congregation finding his devotions greatly disturbed thereby'.

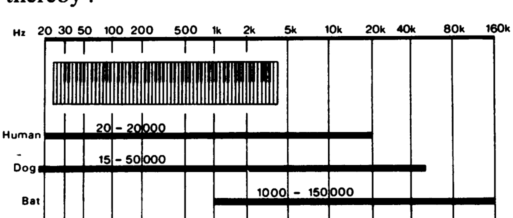

Fig 8 *Frequency range of human hearing compared with that of dog and bat*

There is enormous variation in the ear's ability to recognize the exact pitch of a sound, either in absolute terms, or in relation to other sounds; and there is a continuum of skills which ranges from 'tone deafness', through 'relative pitch', to 'absolute pitch'. Tone deafness is a defect of pitch discrimination in which the relationship of one musical tone to others cannot be accurately assessed or imitated. Most people have relative pitch, in so far as they are able to say, when given a certain reference tone, that a second sound is higher or lower in pitch; and with a modicum of training, most people can learn to recognize the commoner musical intervals.

Anyone who is accustomed to listening to tuning forks and audiometers should recognize the interval of an octave. With a little more musical experience, most of us can quickly learn such common intervals as the fifth (C to G) and the third (C to E). Even Pythagoras, in the sixth century BC, knew these intervals. If we put these notes together (C–E–G–C), we have the major chord of C; and if we recognize this chord, we do so because we are able to relate each note to the key note, C: we have relative pitch.

The professional musician will usually learn to recognize any interval within the accepted musical scale of twelve semitones to each octave; and when he can do this, he is said to have perfect pitch. But it is still relative, and not absolute, for the notes can be recognized only by relating them to some other reference tone. There is however, one way in which a musician might have a built-in reference tone constantly at his service: I refer, of course, to tinnitus.

Everyone knows that Beethoven was deaf, and was driven to distraction by his tinnitus; some of you will know that Schumann had the note A for ever sounding in his ears; but I doubt whether many of you will know that the Bohemian composer, Smetana, not only became totally deaf but also suffered from tinnitus – to such an extent that he wrote it into his music, in his autobiographical first string quartet. This was written exactly one hundred years ago, and it was subtitled 'From My Life' – 'about things that tortured me'.

The term 'absolute pitch' refers to the faculty of recognizing and defining, in absolute terms, the pitch of any tone (or tones), without any reference tone for comparison. My old piano teacher in Liverpool was blessed with absolute pitch, and he could name every note of any discord, instantaneously and with absolute certainty. To the practising musician, the advantage of absolute pitch is that it makes memorizing very much easier. There was one occasion when this pianist came to stay with us. He had been commissioned, for the first time, to play Rachmaninov's second piano concerto. He had heard a recording of it, played by the composer himself, but he had never seen the score. Just before he left Liverpool he bought the score, and he studied it in the train. When he arrived in Bath, he sat down and played it – all 46 pages of it – without the score. This would have been quite impossible without absolute pitch.

Von Bülow once memorized Stanford's Irish Symphony in the train between Hamburg and Berlin, and then conducted it without a score at a concert of the Berlin Philharmonic Society.

The Reverend Sir Frederick Ouseley, Professor of Music at Oxford for 34 years, was also remarkable for his sense of absolute pitch. He would say that the wind was whistling in D, or that a clock with a two-note chime was striking in B minor; and when he was only five years old, he was recorded as saying: 'Only think, papa blows his nose in G'.

There are several theories about absolute pitch and there are some fascinating papers about it, mainly in the *Journal of the Acoustical Society of America*, and notably by Bachem (1955) and Corliss (1973), both of whom have it; and the one fact which emerges, beyond any reasonable doubt, is that absolute pitch is inborn – or not. According to Bachem, less than one in ten thousand of the population has genuine absolute pitch (Ballantyne 1956).

Pitch discrimination can therefore vary enormously, from that of the general who failed to rise for the National Anthem at Queen Victoria's funeral, simply because he did not recognize the tune; to the reverend musical baronet, who, forgetting the number of a friend's house, recognized it by the pitch of the door-scraper!

In considering the vocal instrument, the human voice, and speech, I want to discuss the various components of the voice and to compare the vocal tract with other musical instruments; and I will take them in the same order.

The Speaking Voice

Like any other wind instrument, the vocal instrument emits a fundamental tone, which is generated in the larynx; and a number of harmonics, which are generated in the vocal tract above the cords. The whole tract acts as an acoustical resonating system but, to quote Denis Fry (1971):

> 'It has peculiar properties in that not only the effective length but also the cross-sectional area of the (air) column can be varied in a great variety of ways. The length can be changed . . . and the shape of the tube can undergo a very large number of modifications through movements of the tongue, soft palate and lips. All such adjustments alter the resonances of the whole vocal tract'.

The frequencies at which these resonances (or harmonics) occur depend upon the configuration of the whole tract, and each resonance is referred to as a 'formant'.

The laryngeal tones are referred to as phonation. This is of course a most important element in speech, because it is the sound generated by vibrations of the vocal cords that makes speech audible to the listener; and the frequency of the laryngeal tone is determined by the frequency of these vibrations, which average about 120 Hz in male speakers and about an octave higher in female speakers.

Adrian Fourcin, at University College in London, has made invaluable contributions to the study of laryngeal function by his development of the laryngograph (Fourcin & Abberton 1971), a device which permits the investigation of cord movements without any interference with normal speech. Two surface electrodes are placed on the neck, one on either side, at the level of the larynx; and one of the features of the human voice that can be demonstrated by the laryngograph is the fact that there are fluctuations in the laryngeal frequencies. In other words, the voice has intonation. This is important, because intonation patterns play a vital part in the expressive elements of speech. They enable us, for example, to distinguish a question from a statement; and they can be demonstrated visually. Fig 9, which shows the fundamental frequency (Fx) display of the sound [o], or 'oh', with rising, with flat and with falling characteristics, illustrates this.

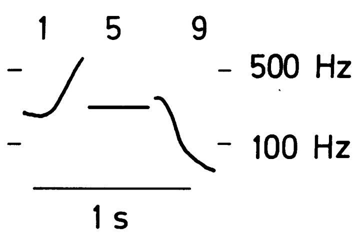

Fig 9 *Intonation patterns of sound 'oh' with rising, flat and falling characteristics.* (*Courtesy Dr Adrian Fourcin*)

There are also formants, or harmonics, superimposed upon the fundamental laryngeal tones. Their acoustical features are determined by the length, calibre and shape of the mobile structures of the vocal tract; and Fig 10 shows the various patterns assumed by the oral and pharyngeal cavities during the production of three English vowel sounds – [i], or 'EE'; [a:], or 'AH'; and [u], or 'OO'. When these are represented graphically (Fig 11), you will see that their formant structures (represented by the thickness of the small black horizontal figures) vary considerably. Note, too, that the amplitude of these formant harmonics diminishes as their frequency (shown on the vertical axis) rises.

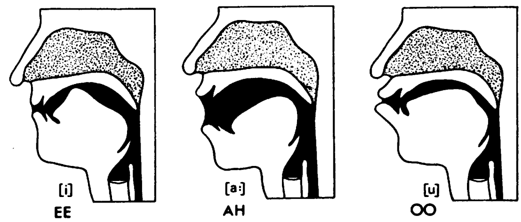

Fig 10 *Shape of oral and pharyngeal cavities for three vowel sounds*

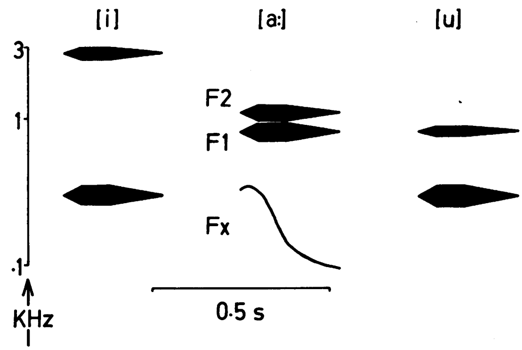

Fig 11 *Formants of three vowel sounds.* (*Courtesy Dr Adrian Fourcin*)

There is one other aspect of speech analysis which I have not yet discussed, and not yet compared with musical instrumental analysis. This is the question of transients. Just as musical instruments can be distinguished from one another largely by differences in their starting transients, so the transients of the voice are vitally important to our understanding of speech; so much so, in fact, that speech has been referred to as 'a process involving a concatenation of constantly changing transients' (Winckel 1974). I will not enumerate all the complicated classes of consonant sounds, but I would draw attention to the importance of the transients in distinguishing similar-sounding words, such as 'goat' and 'coat' with the hard 'g' and 'c'. The first is voiced, the second unvoiced; and the differences between these two sounds, as actually perceived by our ears, are due essentially to the fact that in the first the transients are strong, and in the second they are much weaker. The stylized spectrograms in Fig 12 show that this striking difference is in the transients. In the human voice, as in other musical instruments, they contain a great deal of information.

I have compared the fundamental (laryngeal) frequencies with the fundamental tones of other musical instruments. I have compared the formant harmonics of the human voice with the overtones of piano and violin; and I have compared the transients of the voice with the starting transients

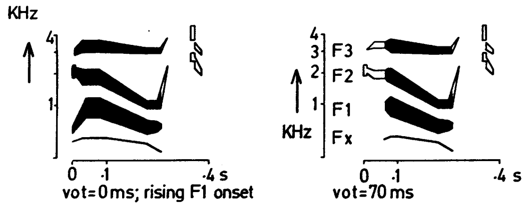

Fig 12 *Stylized spectrograms of 'goat-coat' synthetic stimuli.* (*Courtesy Dr Adrian Fourcin*)

of instrumental music. We can now put all these elements together, and ask another question: Do individual voices have sufficiently strong 'personalities' to enable us to distinguish them with certainty – the one from all others – by ear, or by recordings? Is each voice so unique as to allow precise, positive identification? Are there, for example, any visual means of identifying a voice with reasonable assurance? The question is more than hypothetical because, not so long ago, police departments all over the world were investigating the potential value of what they called 'voice prints', as a supplement to fingerprints or even as a substitute for them. Solzhenitsyn (1971), in 'The First Circle', made great play on this. Roitman is talking:

'Well, you see, we have a device for reproducing visible speech . . . which turns out what are called "voice prints" . . . In the voice prints speech is measured simultaneously in three dimensions: in frequency, across the tape; in time, along the tape; and in volume, by the density of the trace . . . By this means, each sound is recorded as a unique shape that can be easily identified'.

In fact, of course, this is not strictly true. But what he is talking about is the spectrographic analysis of speech (Fig 13), which gives a visual presentation of the three parameters of its acoustical structure – in the present example, for the words 'wrapped in a warm coat'. The voice shown in the upper trace is my identical twin brother's; that in the lower, my own.

Yet another question is this: Are there any other ways in which the laryngograph can help us to study different voice patterns? It is in fact also possible to use this device to make frequency distribution curves for different voices, and Fig 14

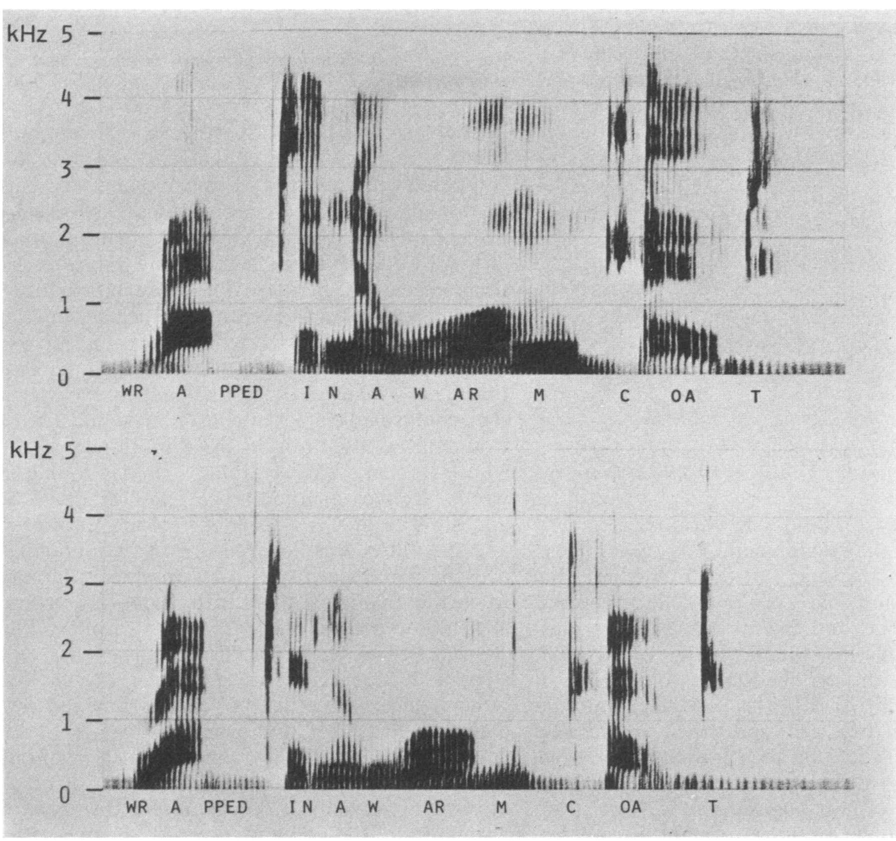

Fig 13 *Spectrograms for RB (above) and JB (below), speaking the words 'wrapped in a warm coat'. RB and JB are identical twins.* (*Courtesy Dr Adrian Fourcin*)

shows these curves for the two voices in the spectrograms, my brother's and my own. They are indeed very similar, but there are also differences so great that Dr Fourcin telephoned me to ask which of us was the heavier smoker!

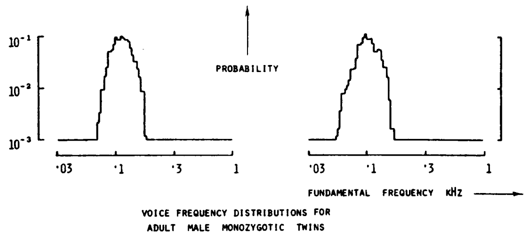

Fig 14 *Frequency distribution curves for RB* (*left*) *and JB* (*right*)

The Singing Voice

As Dr Scholes has said, there are 'many different qualities of vocal production, from the almost pure tones of an English Cathedral choir-boy to the rich tone of a fine contralto'. The difference lies in the harmonics, the former having few and the latter, many. There is also a wide frequency range in singers, from the lowest bass to the highest coloratura soprano. But even within a single vocal range, there are many differences which enable the listener to distinguish one singer from another even in the same song: these include phrasing, accent, tempo and sometimes even key; and every singing voice has its own characteristic quality, depending, at least in part, on individual physical attributes. But there are additional factors which may help the listener to distinguish one from another, and one of these is 'vibrato' due to modulations in frequency and amplitude about the so-called 'pitch frequency' of the voice.

In a course of Christmas Lectures given at the Royal Institution about five years ago Charles Taylor, Professor of Physics at University College in Cardiff, has this to say about it (1976):

'It seems that the "feed-back" mechanism, which enables us to maintain constant the pitch of the sound we are producing at a given moment, involves listening to ourselves. There is, however, a definite small interval of time which must elapse between our hearing system detecting that a drift of pitch is taking place and the correcting action being applied by the appropriate muscles. This turns out to be about 0.14 second. There is thus likely to be a wobble at about this rate which is inherent in the control mechanism and it is a curious fact that the corresponding frequency – about 7 Hz – is somewhere about the optimum for pleasant vibrato.'

This is illustrated in Fig 15.

He was referring to averages, but there are considerable individual differences in these frequency and amplitude modulations, and in the rate of 'swing' of different singers. These may help the listener to distinguish one singer from another; and Winckel (1974) has measured some of these differences from old recordings of famous singers of the past. Caruso, for example, had a swing of between 6.8 and 7.7% above and below his pitch frequency; but Lily Pons's swing was sometimes as high as 23.5%. Without vibrato, the voice would be uninteresting; with too much it is unpleasant. But as with all our senses, the sense of hearing responds much more to change than to steady stimuli.

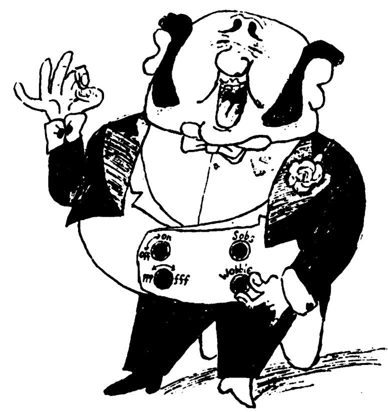

Fig 15 '*Vibrato*'. (*Courtesy Gerard Hoffnung and Dobson Books Ltd*)

A few months ago, in some interesting correspondence with Frank McGuckin, he wrote: 'Now, the complexity of the harmonics of the human voice must be such that the possible moves on the chess-board become simple arithmetic by comparison. How does one recognize Owen Brannigan's voice, or, even given the distortion inevitable in any record, the voice of Paul Robeson, or the unique light tenor of John McCormack – whom I can still hear more than forty years since I last heard him in the flesh? It is not only the voice, or the mere mechanism of the ear; the human brain is the final Arbiter.' This is of course absolutely true; for in the final analysis, hearing is a perceptual process. How much, then, do we know about the central mechanisms of hearing?

Perceptual Processes in Hearing

Perception is the process of identifying and interpreting that information which comes into our consciousness from the world about us, by way of our senses. Held & Richards (1972) wrote:

'Historically, men have thought of the process of perception as the transmission of a copy of an object to a sense organ and then to the seat of consciousness in the brain. But nothing like a replica is maintained in transmission. The processes upon which perception depends [include the] encoding of inputs into neural impulses; [and] in a modern view, perception is regarded as the outcome of the nervous system's processing of the information that comes to it through the senses.'

Held & Richards point out that no organism can process all the external signals which it receives; and experimental psychologists have shown that the capacity of the human brain to deal with the input from the various receptors is limited (in psychological terms) to about 25 'bits' of information per second. It is obvious that the number of external signals passing through the receptors is greatly in excess of this figure; and the main function of neural processing, in its initial stages, is to abstract the most important data and to eliminate redundancy. This abstracting process inevitably gives rise to ambiguities, or (as the psychologists call them) 'illusions'.

In terms of hearing, this means in effect that the bits of information that finally reach the auditory cortex represent a greatly condensed version of the highly complex signals entering the ears; but fortunately, in terms of our hearing for speech, it is possible to dispense with many of these acoustic cues, without detriment to our understanding; as Dr Fourcin has often said, 'Speech is highly redundant'. What we actually hear, when we listen to speech, is not just a conglomeration of constituent cochlear components, but a series of sound symbols; and we fuse all the many ill-assorted sounds of speech into a single sound pattern.

How, then, does the central nervous system deal with these patterns? In the recognition of visual patterns it is known that specialized neurones exist (in the cerebral cortex) which respond exclusively to complex patterns such as the shape of a hand (Gross *et al* 1972), and it has been suggested (Bergeijk *et al.* 1961, p 722) that similar, so-called 'complex auditory neurones' may exist, which respond exclusively to specific complex patterns of sound. The upper limit of such neuronal specialization for particular acoustic patterns is probably, at most, for words; and the recognition of phrases and sentences may depend upon certain temporal and spatial relationships between these specialized neurones.

Ultimately, however, speech recognition is accomplished, not merely by acoustic cues, but largely by linguistic cues (Denes & Pinson 1963). Speaker and listener must share a common language; and there is evidence (Studdert-Kennedy & Shankweiler 1970) that specialization of the dominant cerebral hemisphere in speech perception is due, not to any superior capacity for auditory analysis in general, but rather to its possession of a special linguistic device. This device will surely repay much further study in the future, but at present very little is known about it. I therefore chose to speak about 'The Hearing Ear' – that exquisite little gem without which there would be neither hearing nor music, nor speech, nor language.

Was it not Aristotle who said, 'Nothing is in the mind that did not pass through the senses.'?

REFERENCES

Bachem A
(1955) *Journal of the Acoustical Society of America* **27**, 1180
Ballantyne J C
(1956) An Ear for Music: the Psycho-Physiology of Hearing. Institute of Musical Instrument Technology, London
Békésy G von
(1957) In: Perception: Mechanism and Models. W H Freeman, San Francisco; p 92
(1960) Experiments in Hearing. McGraw-Hill, New York
Bergeijk W A van, Pierce J R & David E E jr
(1961) Waves and the Ear. Heinemann, London
Broadbent D E
(1962) In: Perception: Mechanisms and Models. W H Freeman, San Francisco; pp 22–28
Capps M J
(1967) Thesis, University of Illinois
Corliss Edith L R
(1973) *Journal of the Acoustical Society of America* **53**, 1737
Davis H
(1947) Hearing and Deafness. Murray Hill, New York
(1973) In: Basic Mechanisms of Hearing. Ed. A R Møller. Academic Press, London.
Denes P B & Pinson E N
(1963) The Speech Chain. Williams & Wilkins, Baltimore
Fourcin A J & Abberton Evelyn
(1971) *Medical and Biological Illustration* **21**, 172
Fry D B
(1971) In: Scott-Brown's Diseases of the Ear, Nose and Throat, 3rd edn. Ed. J Ballantyne & J Groves. Butterworths, London; **1**, 379–397.
Gross C G, Rocha-Miranda C E & Bender D B
(1972) *Journal of Neurophysiology* **35**, 96
Held R & Richards W
(1972) In: Perception: Mechanisms and Models. W H Freeman, San Francisco; p 166
Helmholtz H L F
(1863) Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Vieweg, Braunschweig
(1878) *Annalen der Physik und Chemie* **241**, 448–459
(1954) On the Sensations of Tone as a Physiological Basis for the Theory of Music. Trans. A J Ellis of 4th German ed, 1877. Dover Publications, New York
Mills J
(1935) A Fugue in Cycles and Bels. Chapman & Hall, London
Richardson E G
(1954) *Journal of the Acoustical Society of America* **26**, 960
Scholes P
(1950) Oxford Companion to Music, 3rd edn. Oxford University Press, London
Schouten J F, Ritsma R J & Lopes Cardozo B
(1962) *Journal of the Acoustical Society of America* **34**, 1418
Solzhenitsyn A
(1971) The First Circle. Collins and Harvill Press, London
Studdert-Kennedy M & Shankweiler D
(1970) *Journal of the Acoustical Society of America* **48**, 579
Taylor G A
(1976) In: Sounds of Music. BBC, London; p 150
Tonndorf J
(1962) *Journal of the Acoustical Society of America* **34**, 1337
Wever E G
(1949) Theory of Hearing. John Wiley, New York
Winckel F
(1974) In: Ventilatory and Phonatory Control Systems. Ed. Barry Wyle. Oxford Medical Publications, London; pp 248–259