The Multi-channel Cochlear Implant: Psychoacoustics and Speech Perception

Graeme Clark

The Graeme Clark Centre for Bionic Ear and Neurosensory Research,
La Trobe University, Melbourne, Australia

ABSTRACT

The multi-channel cochlear implant is the first clinically successful interface between the world of sound and human consciousness, and the first means of giving severely deaf people hearing and speech understanding, and children spoken language. It has arisen from multi-disciplinary research in neurophysiology; communication, electronic, mechanical and bioengineering; neurobiology; anatomy and pathology; surgery; psychophysics; speech science; audiology and education. Physiological research showed that brainstem nerve cells could only respond to electrical stimuli up to 500pulses/, but they fired deterministically rather than stochastically which is the case with sound. Behavioural research in the experimental animal confirmed this limitation and showed temporal coding could occur when stimulating different sites along the cochlea. Electrophysiological and mathematical studies demonstrated how to stimulate the cochlea for limited place coding. After a series of anatomical, pathological, and biological studies we established it to be safe to implant deaf people. Psychophysical findings from the first patient demonstrated how simple and complex electrical stimuli were perceived. There was a relation between timbre and site of stimulation that could be scaled and they were perceived as vowels. This led to the first successful speech coding strategy that extracted the second formant frequency for place coding. A second patient had similar results and demonstrated that the memory for speech sounds could be retained for many years. Later research also established that the strategy was appropriate for tonal languages. The extraction of other formant and band-pass filtered frequencies for place coding has shown a steady improvement in results. The research is also providing an understanding of neural processing in the auditory pathways, and how this underlies the conscious experience of speech. The research directions to achieve high fidelity sound for hearing in noise and appreciating music are now more clearly defined. The speech processing strategy is very effective especially for young deaf children diagnosed under 12 months of age. With good education they can achieve near normal spoken language. Bilateral implants can provide sound localization and some improved hearing in noise. It has been developed industrially through close collaboration between the University of Melbourne's research group and the company Cochlear Limited created to take university research to the clinic and market place. The bionic ear has also paved the way for a new discipline in Medical Physics and Biomedical Engineering, I have called Medical Bionics.

INTRODUCTION

This plenary address is entitled: "The Multi-channel Cochlear Implant: "Psychophysics and Speech Perception"; because our research on the perception of sounds with electrical stimulation, has been crucial in discovering speech coding strategies that enable severely- to- profoundly deaf people to understand speech. This has resulted in the first clinically successful interface between the world of sound and human consciousness. It is also the first means for deaf children to develop spoken language.

This address summarises research that is the culmination of our initial studies presented on Psychophysics and Speech Processing at the 10th World Congress on Acoustics in Sydney 30 years ago.

Our psychophysical and speech discoveries at the University of Melbourne have followed on from electrophysiological research I commenced at the University of Sydney in 1967. I aimed to find out how the neural responses in the brain were the basis for the conscious experience of sound and speech, and how this could lead to a hearing prosthesis for severely deaf people. It was later referred to as a cochlear implant or bionic ear.

When I commenced my research in 1967 most scientists said that the use of electrical stimulation to relieve severe deafness was nearly impossible. For example, Merle Lawrence a leading auditory neuroscientist said in 1964, "Direct stimulation of the auditory nerve fibres with resultant perception of speech is not feasible" (Lawrence, 1964; Clark, 2003a). He claimed that the nerve supply of the cochlea was too complex for electrical stimulation to code speech sounds. In addition, Schuknecht (1964) had evidence there would be too few fibres left in deafness for speech understanding.

However, in science I considered that it is even more important to examine why an outcome can occur, rather than why it cannot. Furthermore, I was not convinced the evidence had conclusively shown that speech understanding was near to impossible, and needed to satisfy myself there was a way around the difficulties.
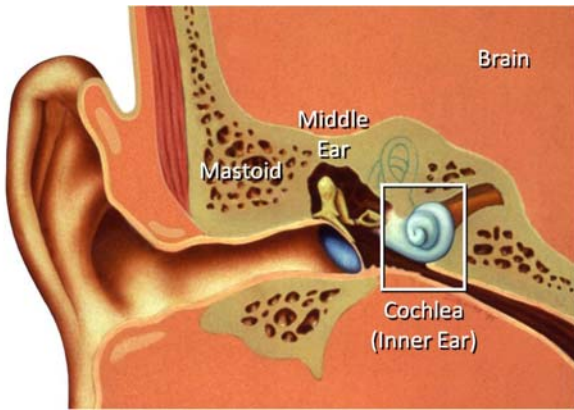
Figure 1a. The outer, middle and inner ears. (Cochlear Limited)

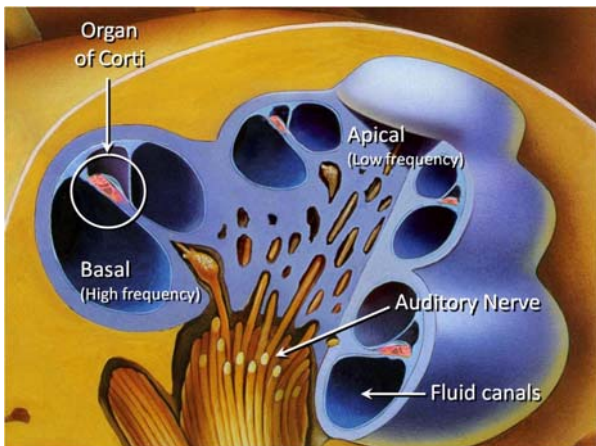**Figure 2.** Temporal coding. Top: sound waves; Bottom: nerve responses. (Clark, 2003b; Clark, 2008).

Figure 1b. The inner ear or cochlea with sense organ of hearing on the vibrating basilar membrane and connected to the auditory nerve. (Cochlear Limited)

Figure 3. Place coding of frequency. Left: tuned responses of the auditory nerve; Right: spatial ordering of frequency responses. (Clark, 1996)

So in 1967 I commenced research in auditory neurophysiology at the University of Sydney to study the neural responses to sound and electrical stimulation.

However, in understanding how the cochlear implant speech coding strategy was discovered it is first necessary to be reminded of the way sound is coded normally. The sound vibrations are transmitted along the ear canal via the small ossicles in the middle ear to the sense organ of hearing (organ of Corti) in the cochlea or inner ear (Figure 1a). In the organ of Corti, sound is first coded by hair cells that convert the vibrations into electrical signals for further coding by the higher centres of the brain (Figure 1b).

Frequency is coded by temporal and/or place coding, and intensity by the rate and/or the population of nerve responses.

The temporal coding of frequency, shown in Figure 2, occurs when individual neurons fire in time and phase with the sound waves. The time interval between the action potentials could also be an important parameter in temporal coding.

But not all individual neurons fire every cycle or at precisely the same phase. So a group response could account for the temporal coding of frequency. On the other hand, with place coding the inner ear filters sound frequencies, as shown on the left of Figure 3. The different frequency regions are connected spatially or tonotopically to all the centres in the brain as shown on the right so that a frequency scale is preserved. In other words, we recognise the pitch of a sound according to the place of stimulation.
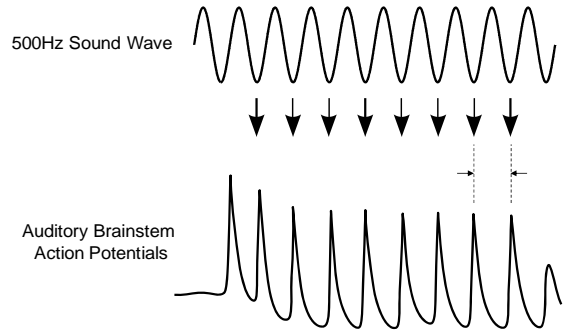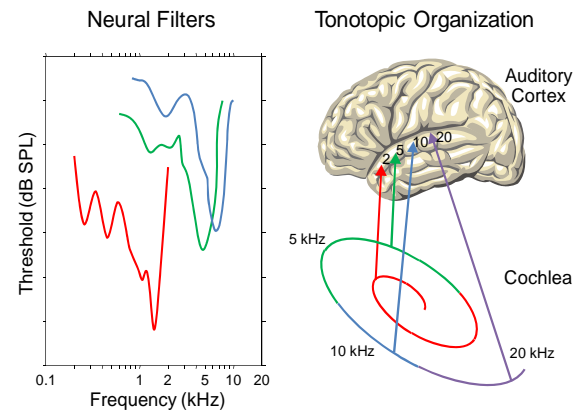
With sound, temporal and place coding are combined. The sound frequencies lead to waves that travel along the basilar membrane as illustrated in Figure 4. The temporal information is preserved in the vibrating membrane. Place information is preserved in the site of maximal vibration.

But at the time, the relative importance of the timing and place codes for frequency, as well as the codes for intensity, were not clear, let alone how well they could be reproduced with electrical stimulation.

With severe-to-profound deafness there is a marked loss of hair cells, and hence amplifying sound with a hearing aid will not lead to speech comprehension. So the challenge was, could electrical stimulation of the residual cochlear nerve fibres, thus by-passing the cochlea, reproduce the coding of sounds so speech could be understood?
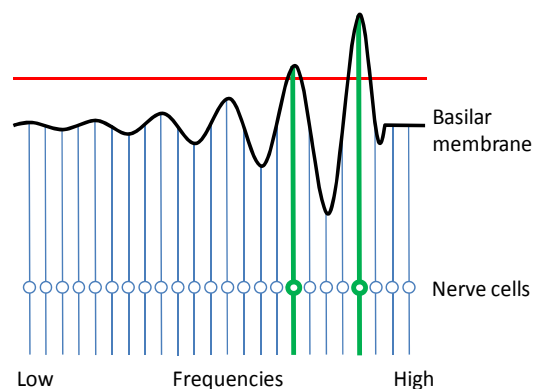

Figure 4. The basilar membrane travelling wave and patterns of cochlear nerve responses. (D. Grayden)

ELECTROPHYSIOLOGY & BIOLOGY

Our psychophysical discoveries have followed on from electrophysiological and biological research I commenced at the University of Sydney in 1967. I aimed to find out if electrical stimulation of the auditory nerve could reproduce the coding of speech sounds for deaf people.

I first undertook research on experimental animals to see the effect of rate of electrical simulation on single cells and groups of cells (i.e. field potentials). I considered this would help show to what extent we could reproduce speech frequencies using a temporal code.
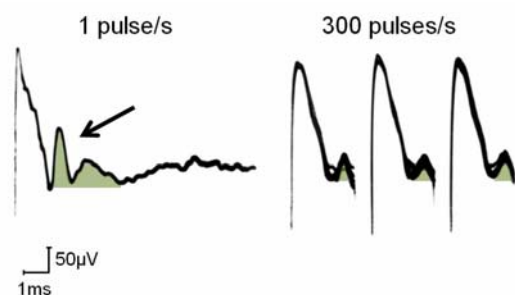

Figure 5. The summed neural activity (arrow) in the auditory brainstem of the experimental animal for stimuli of 1 pulse/s and 300 pulses/s (1967-1969). (Clark, 2003b; Clark, 2006; Clark, 2010)

Figure 5 shows the summed neural responses from the auditory brainstem cells. Note the marked fall in neural response from 1 pulse/s to 300pulses/s. This lack of neural response indicated that temporal or rate coding would not be adequate for speech frequencies above approximately 300 pulses/s.

I concluded from my physiological studies that "if pure tone reproduction is not perfect meaningful speech may still be perceived if speech can be analysed into its important components, and these used for electrical stimulation. More work is required to decide which signals are of greatest importance in speech perception" (Clark, 1969; 2003b).

But first the neurophysiological findings needed to be confirmed with behavioural studies on conditioned animals, and to help establish the correlation between brain activity and an animal's perception. The studies were undertaken using a plug and socket, but the sockets became infected. For that reason I embarked on the engineering of an implantable receiver-stimulator for use in deaf people, where information and power were transmitted through intact skin by radio waves. This was a much more difficult and expensive task, requiring judgment about the circuitry to provide as much flexibility and transparency with the stimuli for a possible speech coding strategy.

In the experimental animal behavioural studies it was discovered that low rates of electrical stimulation could be discriminated nearly as well as sounds of the same frequency, but only up to 500 pulses/s. Higher rates could not be discriminated. Discrimination for low rates of stimulation also occurred when stimulating the low frequency or apical region of the cochlea, as well as in the basal or high frequency regions. This helped establish that a temporal code was important for the perception of low frequencies. This also indicated that with a cochlear implant speech processing strategy, the low frequencies of voiced sounds could be presented as a rate of stimulation across all electrodes.

The poor discrimination for rates above 500 pulse/s meant that the important higher frequencies for consonants could not be conveyed by rate of stimulation, and a place code would need to be explored to discover if this was a possible mechanism.

So the next challenge was to determine how to localise current to discrete groups of nerve fibres in the cochlea for effective place coding. But would the fluid in the inner ear canals short circuit the current away from the nerve fibres?

So we mathematically modelled the resistances of the structures in the cochlea and discovered that with monopolar stimulation from a voltage between the cochlea and a distant ground, a current could be partially localised to separate groups of nerve fibres (Figure 6). Electrophysiological research in the experimental animal also showed that bipolar stimulation, where current passes between two nearby electrodes, could also partially achieve place coding of frequency.

Thus as frequencies could only be partially reproduced with temporal and place coding, important speech frequencies would need to be selected and transmitted through an electro-neural coding "bottle-neck" between the world of sound and the brain.

So the important scientific question was: could a speech coding strategy be discovered for deaf people so they could understand speech with electrical stimulation of the brain?
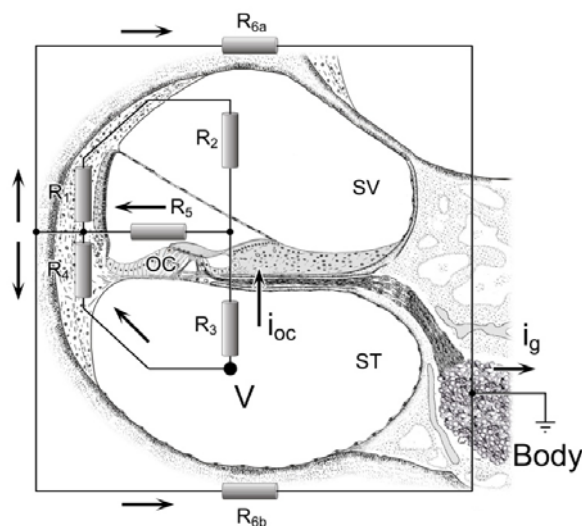

Figure 6. The model of resistances (R_1-R_6) across structures in the cochlea (Black and Clark, 1980). The arrows show the current through the organ of Corti (i_{oc}) and to ground (i_g), for a voltage V between the scala tympani and the auditory nerve. (Clark, 2008)

Prior to the perceptual research on deaf people biological safety studies were undertaken on the experimental animal. This ensured that surgical trauma, electrical stimulus parameters, materials toxicity, and inner ear infection did not lead to the loss of neural elements or other adverse effects such as meningitis.

The perceptual research required an implantable receiver-stimulator (Figure 7). The electronic circuit had to be designed to allow flexible use of stimulus parameters, as all information as well as power had to be transmitted inductively via radio waves. This made it less transparent than a plug and socket. It had to be water-tight to prevent the entry of body fluids as they would corrode the circuitry.
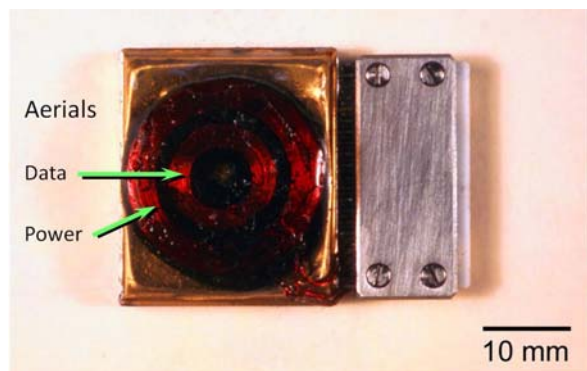
Figure 7. The inaugural University of Melbourne prototype receiver-stimulator with connector at the right. (Clark, 2003)

It was used with a connector between the package and electrode until our later biological studies showed that a replacement package and electrode bundle could be easily reinserted if they failed.

The electrode bundle needed to pass around the cochlea to allow multi-channel stimulation of the speech frequency region. The required mechanical properties of the electrode for it to pass around the tightening spiral of the cochlea were discovered by passing small branches and grass blades around a "Turban" shell (Ninella torquata) that had a similar shape to that of the human cochlea (Figure 8). This demonstrated the electrode needed to be flexible at the tip, and stiffer at the base to pass upwards into the cochlea to reach the speech frequency region. A suitable pre-operative evaluation procedure had to be developed to ensure the patient had no useful hearing.

In 1978 I selected my first patient Rod, who had lost all hearing after a car accident 18 months earlier, and badly wanted to hear speech. He said it is a "nightmare being deaf".

I also selected Rod as I reasoned he should remember the quality of the sounds he had previously heard after only 18 months of deafness. Furthermore, as he was totally deaf there would be no electrophonic hearing from residual hair cells to contaminate the findings. But I was worried that the fracture at the base of his skull may have cut the auditory nerves, and then neural responses from electrical stimuli would not reach the higher centres of the brain. At the time polytome X-rays gave poor resolution, so I electrically stimulated the promontory of his middle ears so the current could spread to the auditory nerves. As responses were present from both sides this suggested the auditory nerves were intact, and I was free to operate on either the right or left ear.

Figure 8. A grass blade with flexible tip and stiffer proximal end passing around a turban shell with a shape similar to the human cochlea. (Clark, 2000; Nystrand, 2006)

Figure 9. The first Multi-channel cochlear implant operation on Rod by Graeme Clark and Brian Pyman on 1st August 1978. (The Graeme Clark Collection, National Film & Sound Archives, Australia)
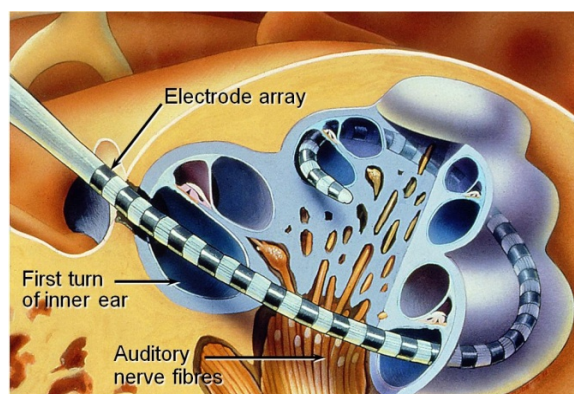

Figure 10. Banded electrode bundle inserted around the basal turn of the cochlea. (Cochlear Limited)

On the 1st August 1978 I operated with assistant Brian Pyman to implant the receiver-stimulator package in Rod's left temporal bone. The details of the surgery are described elsewhere (Clark et al., 1979).

The electrode bundle was guided around the first turn of the cochlea by a specially designed micro-claw ("Clark Claw") (Clark, 2000). It went far enough to lie opposite the speech frequency region (Figure 10).

Six weeks later, on the 18th September, we set out to see if Rod could recognise tunes. Their amplitudes and frequencies were converted to electrical intensities and stimulus rates on a single electrode. He recognised our alternative national anthem "Waltzing Matilda" immediately. He said he identified the song from the rhythm rather than the melody, but the pitch was too high (To be hears in The Graeme Collection in the National Film & Sound Archives, Australia).

PSYCHOPHYSICS

Psychophysical studies were then carried out using simple electrical stimuli to discover how pitch was perceived for rate and place of electrical stimulation. First rate of stimulation was perceived as a "true" pitch sensation. However, changes in rate could only be discriminated at low rates.

Figure 11 shows pitch on the vertical axis versus site of stimulation on the horizontal axis, for rates of stimulation of 50, 100, 200pulses/s, and 500pulsese/s. Note the pitch only increased with rate up to 200pulses/s, and then reached a plateau with no further increase from 200pulses/s. This was

consistent with the behavioural findings on the experimental animal. Later studies on additional subjects showed some variation in rate discrimination, but it reached a plateau at 600pulses/s. This was a clear demonstration that rate of stimulation or frequency of neural excitation was the essential code for low frequencies.
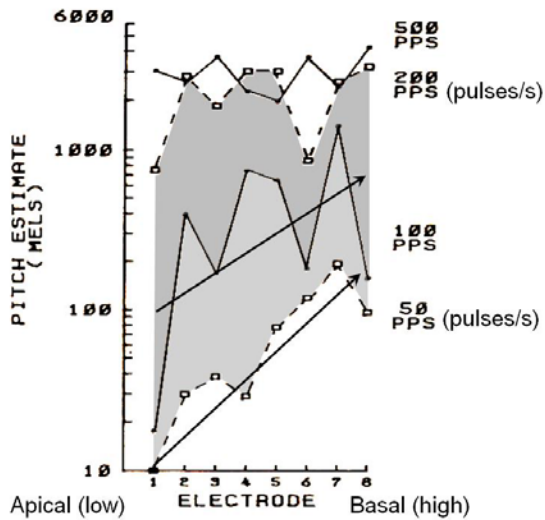

Figure 11. Pitch for rate and place of stimulation. (Tong et al., 1979)
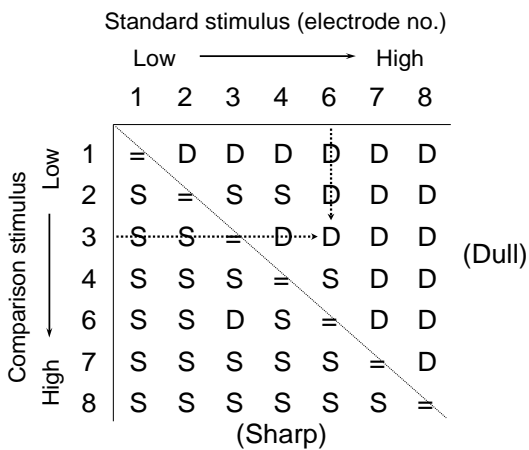

Figure 12. Ranking of timbre versus electrode place. (Tong et al., 1982)

In addition, pitch with place of stimulation, probably had a different pitch dimension, and varied from the low frequency apical electrodes to high frequency basal electrodes. It was discovered too that with place of electrical stimulation, the sensation was perceived as timbre rather than "true" pitch (timbre is the quality of the sound that distinguishes two musical instruments playing the same note, at the same intensity). Timbre varied from dull at the low frequency end of the cochlea, to sharp at the high frequency end. This suggested that place of stimulation made a significant contribution to timbre.

Timbre could also be scaled for place of stimulation. In Figure 12, standard and comparison stimuli were ranked from a low frequency location at #1 to high at #8. When for example a standard stimulus at #6, a moderately high frequency location, was compared with a stimulus #3 at a lower frequency location, the comparison stimulus was described as duller and is marked "D". With good coding of timbre on a place basis the "S" for sharp and "D" for dull should lie along the diagonal, which is essentially the case.

In addition, with the coding of intensity using electrical stimulation, the loudness, rose rapidly over a 6dB range in current amplitude. This was a narrower range than the 60 dB for acoustic clicks.

Then for the next 10 weeks till the end of November 1978, Rod came regularly so we could explore in detail the sensations experienced with the electrical stimuli. I was surprised when he heard vowels that also varied with place of stimulation, and corresponded with electrodes that gave sensations that varied from sharp to dull (National Film & Sound Archive).

When we examined the relationship between the vowels and the sharpness or dullness of the sound, we realised that the vowels corresponded to those experienced by normal hearing listeners when a similar site of the cochlea was excited by a single or second formant frequency.

This is illustrated in Figure 13. The /e/, /I/, /i/ sounds were sharp and had high single or second formant frequencies, and were produced by the electrodes in the basal high frequency turn. The /ɒ/ and /ɔ/ were from electrodes that had slipped outside the cochlea, and I realised from the anatomy that when the current passed around the cochlea it would stimulate the low frequency nerves on the outside of the nerve, and they were also perceived as dull.

However, by mid October 1978 I was concerned that we had only discovered how to code vowels, but not consonants, which are more important for intelligibility. So my next hope was could we code transitions in the second formant frequencies, as they are some of the important cues for the plosive consonants /ba/, /da/, and /ga/, as well as other consonants. For example with /ba/ the second formant is rising, with /da/ it is flat, and with /ga/ rapidly falling.
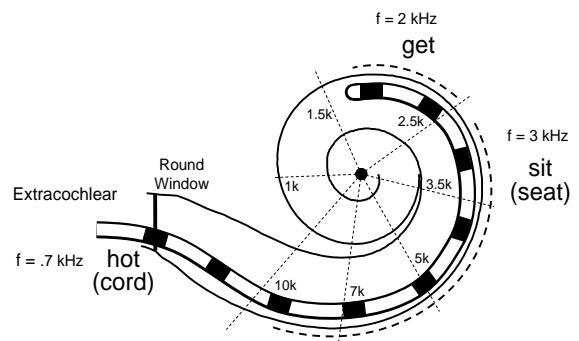

Figure 13. Perception of vowels versus site of stimulation in first patient's cochlea. (Clark, 2010)
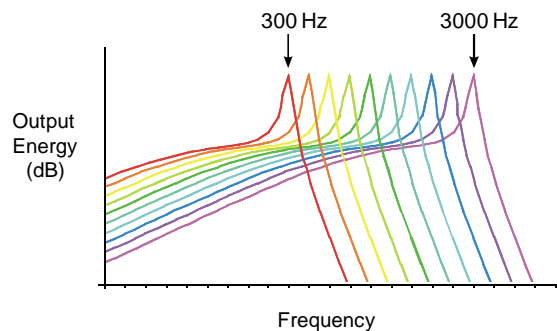

Figure 14. Physiological speech processor with filter bandwidths similar to those for auditory nerve fibres. (Clark, 2010)

But by the middle of November 1978 we still could not produce consonants reliably, so for one week we trialled a hard-wired physiological speech processor to produce the spectral changes required. This had 8 frequency filters, as shown in Figure 14, with responses similar to the filters in the auditory nerve fibres (Laird, 1979). Overall speech understanding was very limited because the electrical fields around each electrode overlapped, and with simultaneous stimulation it was difficult to predict the loudness. This also demonstrated that speech processing strategies should only provide non-simultaneous stimulation.

SPEECH PROCESSING

As my prior research had indicated that the greatest chance of success in understanding speech would be to transmit the key speech information through the electro-neural bottle-neck we selected the second formant frequency (F2) for site of electrical stimulation, and the sound pressure (A0) as current level. The voicing frequency (F0) was presented as rate of stimulation across the electrode sites, as our earlier physiological and behavioural findings had shown the brain processed rate and place information along two separate channels. This F0/F2 strategy was first implemented in 1978 on a laboratory computer.

Then I was most anxious to see if my patient could recognise any unpractised monosyllabic words using electrical stimulation alone. That would establish whether the formant-based speech processor would be better than a lipreading aid in providing speech understanding for deaf people.

In December 1978 I asked my audiologist if she would present open-set words to Rod. I dared to hope that he would get one or two right. When this happened I realised that we had achieved what most said was impossible, and was overcome with relief and joy. My joy was dampened by the knowledge that I only had two month's salary left if my application to the federal government failed. But at the beginning of 1979 there was the good news: our research would continue, and the government would also fund the commercial development of what had become known as the bionic ear. This funding allowed me to validate the speech perception results with a battery of audiological tests.

The coding strategy was found to give good open-set speech understanding when combined with lipreading, and some limited understanding of open-sets of words with electrical stimulation alone, as the patterns of excitation in his brain were not yet a good match for those previously imprinted by sound. His understanding of sentences was better as there were more contextual cues. I then had to address other important scientific questions: 1) Would the formant extraction strategy benefit all people, and languages?; 2) Would the memory for speech sounds be retained in other people after long periods of deafness?

To see whether we had discovered a speech strategy that was specific for one person's brain I operated on a second patient in July 1979. I only did so when the development of a wearable speech processor was well under way so that he could get help in his daily life, and not only be a research subject.

As the second patient had been deaf for 17 years it made it possible to show whether the memory for speech sounds persisted after long periods of deafness. At his second test session he was read to from the newspaper without help from lipreading (National Film and Sound Archive Australia). His

ability to remember speech sounds and language was evidence that the coding of speech and language can lie dormant in the retained synaptic connections and/or proteins of brain cells for many years.

The speech perception scores for the first two University of Melbourne patients using the F0/F2 processor are shown in Figure 15. There was a marked improvement in speech perception when electrical stimulation was combined with lipreading compared to lipreading alone, and the patients could understand up to 30% of words using electrical stimulation alone with 'live voice' presentations.
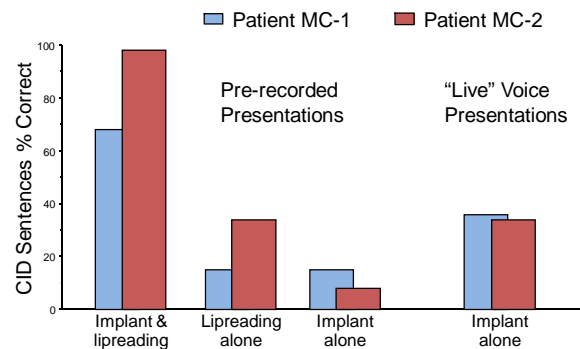

Figure 15. Speech perception for the first two patients with the F0/F2 speech processor. (Clark, 2010)

Figure 16. First patient uses his wearable speech processor to talk with his wife. (Clark, 2010)

The speech coding strategies, discovered using software on the laboratory computer, were implemented in a hardware – based wearable processor the size of a binocular case (Figure 16) by engineers Peter Seligman and Jim Patrick, and it performed to expectations.

It was then necessary to establish a partnership with industry so the discovery could be commercialised, and the benefits given to as many people as possible. A Public Interest Grant from the Australian government helped establish in 1981 a new firm Cochlear Pty Ltd to do this. The Chairman Paul Trainor created a "tiger team" of talented engineers. The implant, speech processor and transmitting coil and head band, shown in Figure 17, were ready for the clinical trial for the US FDA that was commenced by the University of Melbourne surgical team in 1982.

It provided the same speech coding strategy as the University prototype, but there was more precise control of the stimuli. As a result a significant proportion of the patients were able to achieve open-set speech understanding with electrical stimulation alone.

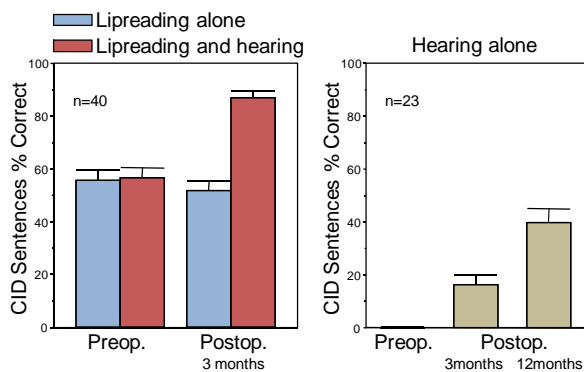**Figure 17.** The first Nucleus cochlear implant system. (Clark, 2010)

Figure 18. F0/F2 speech results for FDA trial. (Clark, 2010)

After a trial on 40 people in the US, and Australia, from 1982-1985, the US Food and Drug Administration approved the implant as safe and effective for speech understanding in adults with hearing before going deaf. It was the first multi-channel cochlear implant to be approved by any world health regulatory body. There was a 70% improvement when electrical stimulation was used in conjunction with lipreading compared to lipreading alone, and with electrical stimulation alone the open-set scores rose from 16% at 3 months to 40% at 12 months, in a group of 23 patients.

Although the strategy initially benefitted English speakers (Figure 18), we were also able to demonstrate its value for other formant-based languages on people in Melbourne, and later internationally. It was also a great help to people with tonal languages.

But as we had not achieved normal speech understanding for most deaf people I felt the need to press on with further research. So the next key question was: what additional speech information should be selected and transmitted through the electro-neural "bottle-neck" for improved speech coding? The initial good F0/F2 results but poor physiological model results had demonstrated that selecting information rather than presenting unprocessed information to the nervous system was for us the preferred option.

To determine the speech features to select, the vowel and consonant scores for the F0/F2 processor were first compared. The consonant scores at 36% were significantly worse than the vowels at 52%.

Therefore, I considered it essential to determine how well the different consonant speech features were being transmitted. Many were analysed, and the information transmission for voicing, nasality, and affrication was moderately good, but it was much poorer for place of articulation.
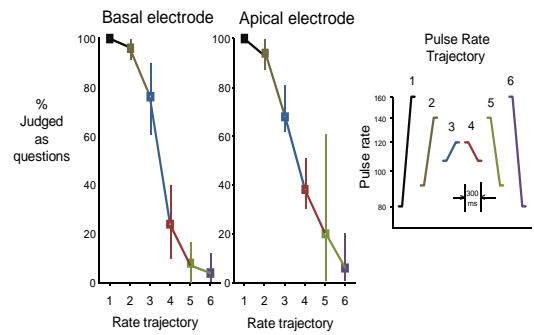

Figure 19. Voicing categorization for rate trajectories. (Tong and Clark, 1982)

So our research then focussed on discovering the electrical stimuli that maximised the transmission of speech features through the electro-neural "bottle–neck", and especially for place of articulation.

To code for the conscious experiences of speech sounds the next question asked: was rate of electrical stimulation perceived as voicing, and not just pitch (Figure 19)? A rising voicing frequency for the question "Mary had a little lamb?" was conveyed by increasing the rate of stimulation. On the other hand, a flat or falling frequency was experienced as a statement. The results were the same for both low frequency and high frequency electrodes. These data thus demonstrated that rate of stimulation was perceived as supra-segmental information for voicing in recognizing questions and statements, and was processed by the brain independently from spatial stimulation.

It was also important to discover whether rate or place coding was important for the conscious experience of segmental speech information. How well could they convey important frequency transitions for the perception of consonant speech features, and especially those of short duration in speech? This applies in particular to plosives, and the coding of second formant transitions.

On the left in Figure 20 perceptual discrimination is plotted against changes in rate of stimulation. The percentage of responses judged different on the vertical axis is plotted against increasing rate of stimulation on the horizontal axis, for stimulus durations from 25 to 100ms. Note the percentage responses judged different fall at a duration of 25ms. Thus rate of stimulation was not appropriate for coding consonants.
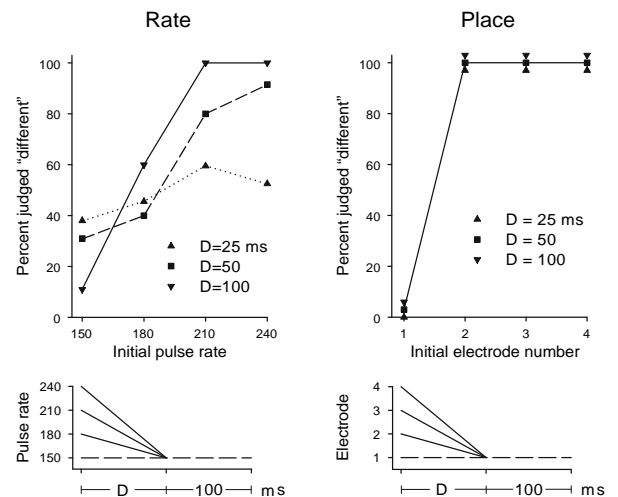

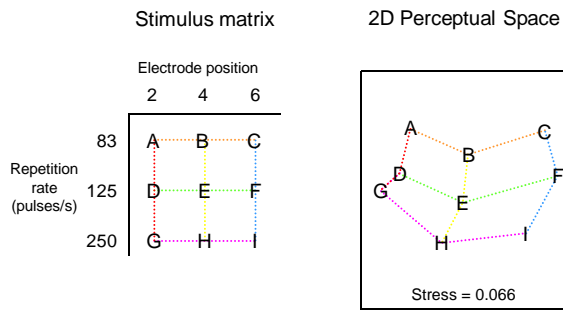**Figure 20.** Rate and place discrimination vs duration. (Tong et al., 1982)

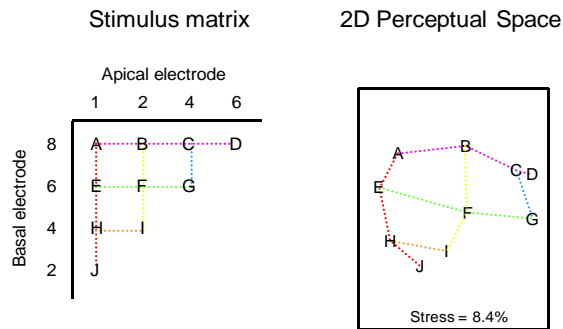**Figure 21.** Rate & place pitch: perceptual space dimensions. (Tong and Clark, 1982)

Figure 22. Electrode pairs: perceptual space dimensions. (Tong et al., 1983)

On the right perceptual discrimination is plotted against changes in site of stimulation, over stimulus durations from 25 to 100ms. The discrimination was the same. Thus the formant transitions of consonants, which are in the order of 20ms, could be recognised through place of stimulation.

It was next important to establish the perceptual independence of temporal and spatial information. This was evaluated using multi-dimensional scaling.

Figure 21 shows that the perceptual space on the right for combined rate and place stimulation compares favourably with the stimulus matrix on the left. The two dimensional solution was the best for both the stimuli and perceptual data indicating, that there are separate perceptual spaces for rate and place of stimulation. This suggests that the conscious experience of complex speech sounds is at upper levels in the auditory pathways where the two brain responses for place and rate can be integrated.

The next question was: could electrical stimulation of two separate frequency sites in the brain be perceived as two separate sensations? This would support the development of speech coding strategies for two or more vowel or consonant formants on a place coding basis.

Figure 22 shows the two electrode stimulus matrix on the left, and as there is a reasonable fit between the stimuli and perceptual spaces on the right, this indicated a 2-dimension perceptual space for electrode place gave the best solution. Furthermore, it was discovered that although place of stimulation gave two perceptual dimensions they were fused into the conscious experience of one vowel.

Then an acoustic model of electrical stimulation was developed for better indirect evidence of how electrical stimuli were perceived. The model, illustrated in Figure 23, used seven bands of pseudo-random white noise with centre frequencies corresponding to the electrode sites to represent different speech frequencies. The psychophysical results

were similar to those we had obtained with multi-channel electrical stimulation indicating the perceptual similarity of electrical stimuli to filtered noise.

Having discovered how complex patterns of sound were perceived and how to transmit them through the electro-neural "bottle-neck" the next challenge was to discover how to code the patterns for speech understanding.

Figure 24 shows the mean vowel and consonant scores for the F2/F0 strategy as an acoustic model of electrical stimulation, and electrical stimulation itself. The results are similar suggesting the model provided a good representation of the electrical stimuli. The results for the acoustic model of adding the first formant or F1 to achieve the F0/F1/F2 processor are in dark blue, and show an improvement in vowels and consonants. Later it was found that with an F0/F1/F2 processor for electrical stimulation there were similar improvements to those obtained with the acoustic model. Thus the acoustic model for electrical stimulation provided similar spatial patterns of stimulation to those with electrical stimulation.
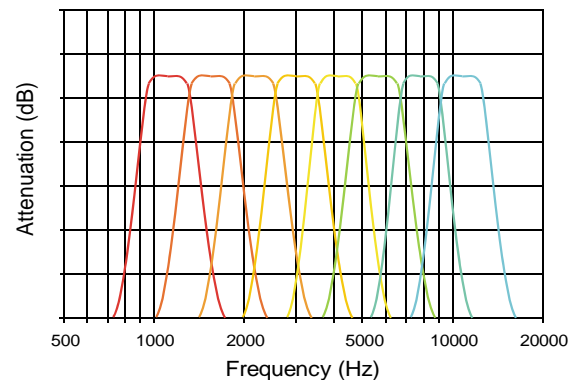

Figure 23. Acoustic model: electrical stimulation of the auditory nerve.
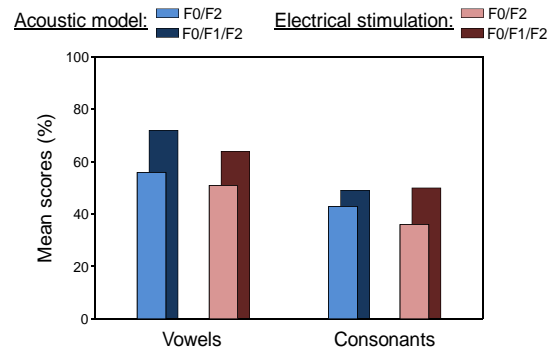

Figure 24. Acoustic and electrical stimulation for vowels and consonants. (Blamey et al., 1984)

In addition, the acoustic model was used to study not only the place coding of speech features with electrical stimulation, but also the processing of temporal information for electrical stimulation. With the acoustic model with filtered noise the fine temporo-spatial patterns of stimulation seen with sound were not present. Normally with acoustic stimuli the phase relations in a small population of neurons are preserved. This was not the case with the acoustic model where voicing was simply the amplitude modulation of the noise burst at the voicing frequency.

The information transmission for voicing using rate of stimulation with the F0/F2 acoustic model is shown in Figure 25, and with electrical stimulation. It can be seen they are similar. This showed electrical stimulation was not transmitting the fine temporo-spatial patterns for sound frequency coding.
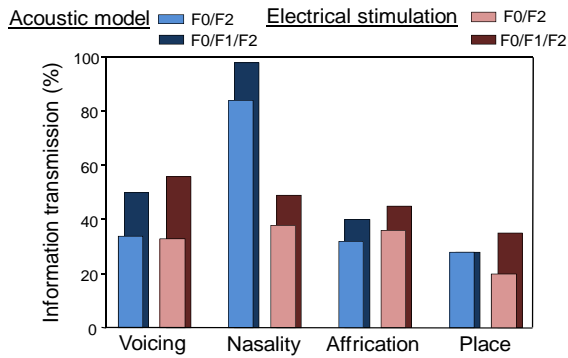
Figure 25. Acoustic model of electrical stimulation and speech features. (Blamey et al., 1984)
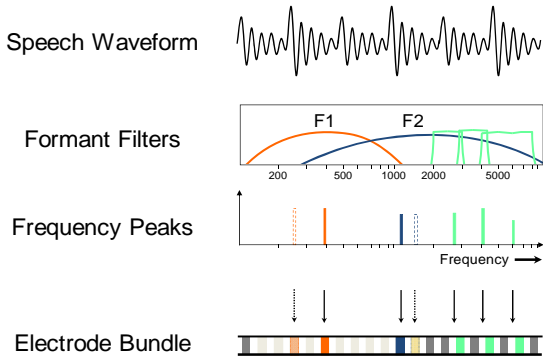

Figure 26. The F0/F1/F2 and F0/F1/F2 + high fixed filter "Multipeak" strategies. (Clark, 2010)

After the US FDA approved the F0/F2 strategy as safe and effective in 1985, further advances were achieved by extracting additional speech frequencies, and coding them on a place basis, as illustrated in Figure 26. This shows the extraction and coding of the first formant (F1) as well as the second formant (F2) with voicing still as rate of stimulation across electrodes. Then rather than extract the third formant, a hybrid scheme (Multipeak) was evaluated in which the outputs of three band-pass filters for the high frequency bands (2000-2800 Hz, 2800- 4000 Hz and > 4000 Hz) were coded on a place basis as well as the F1 and F2 formant frequencies.

Then the extraction of 6 to 8 maximal filter outputs, and coding them as place of stimulation at a constant low rate of stimulation produced improved speech perception especially in noise. With this strategy, referred to as SPEAK, the fundamental or voicing frequency was transmitted as amplitude variations in the speech wave at a constant rate across electrodes, rather than a rate proportional to the voicing frequency. An improved version was this strategy at a high rate of constant stimulation, and referred to as ACE.
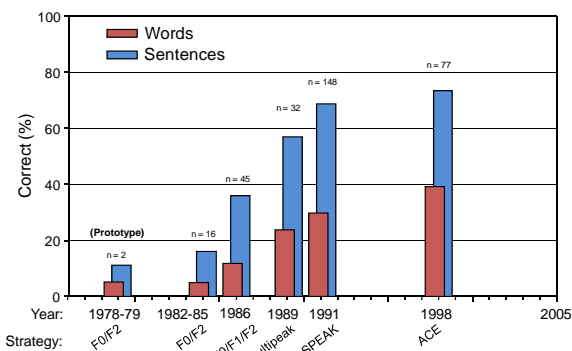

Figure 27. Open-set word and sentence scores for the F0/F1/F2, Multipeak, SPEAK and ACE strategies. (Clark, 2003b)

The progressive improvements in speech understanding for the F0/F2, F0/F1/F2, Multipeak, SPEAK and ACE strategies are shown in Figure 27. Although there was a steady increase in scores, they have reached a plateau, and new strategies and electrodes are needed to provide a fine temporo-spatial pattern of stimuli. These stimuli should also help people hear in noise as this is difficult at a 0dB SNR with the present implant strategies, but with electrical stimulation they are more similar to normal scores at 5-10dB SNRs. In addition, music does not have high fidelity, and could be improved with fine temporal spatial patterns of excitation.

The fine temporo-spatial coding of frequency was studied by constructing an interval histogram by summing the intervals between the nerve spikes. If the time bins collecting the intervals are small any jitter in nerve firing will also be detected.
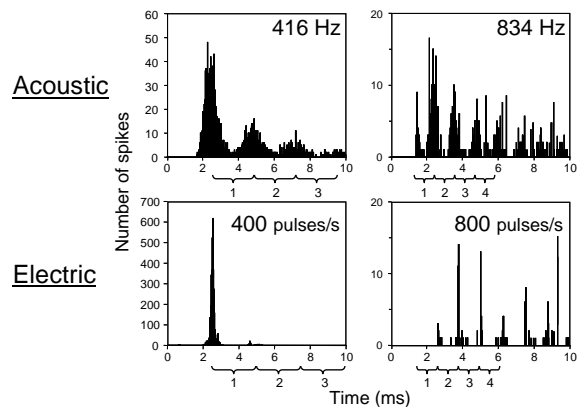

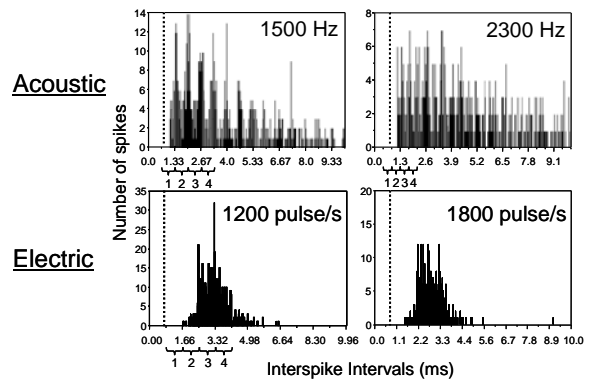**Figure 28.** Interspike interval histograms. (Paolini and Clark)

Figure 29. Inter-spike interval histograms for high rates of acoustic and electric stimulation. (Paolini and Clark)

The inter-spike interval histograms for acoustic and electrical stimuli at approximately 400 Hz and 800 Hz, are shown in Figure 28. Interspike interval histograms. (Paolini and Clark). The interval histograms for acoustic stimulation are on the top. The intervals are multiples of the periods of the sound waves. Note too that the responses do not always occur at precisely the same interval, i.e. they are jittered or are stochastic. On the other hand, for electrical stimulation at a low rate of 400 Hz, on the bottom left, the responses are more precisely timed or deterministic. They also only fire every 2.5 ms, which is the period of the stimuli. On the other hand, on the bottom right, for a higher rate of 800 Hz, the interval peaks were multiples of the periods of the stimulus, and there was stochastic firing as seen for sound.

Furthermore, with higher rates (Figure 29) with acoustic stimuli, there was time locking at 2,300 Hz, but for high rates

of electrical stimulation of 1200 pulses/s, there was very little time locking, and it was lost at 1800 pulses/s. So there was a big difference between the temporal responses to acoustic and electrical stimuli.

The next question was: to what extent did the temporal responses of single neurons to electrical stimulation correlate with the discrimination of pitch in implanted patients. As discussed above the discrimination of changes in pitch reached a plateau with a stimulus rate of 300pulses/s and above. Thus the firing properties of individual neurons to electrical stimulation did not account for temporal pitch discrimination at all well.

Therefore, the temporal response properties in a group of neurons rather than in a single neuron were examined and thought to be important. As shown in Figure 30, with sound not all fibres fire each sine wave, but the population as a whole does. Thus the brain may decode frequency as a population response. In this case, nerve action potentials converge on higher brain cells, and firing occurs if they arrive within a certain time window corresponding to the period of the frequency.

Furthermore, our mathematical modelling studies have shown that the probability of neighbouring cells firing depends whether their input reflects the phase changes in the basilar membrane travelling wave at its maxima or along its length. This is seen for acoustic stimuli (Figure 31, Figure 32), but not an electric stimulus where the phase changes are absent.

In addition, our modelling studies have demonstrated, as shown on the left of Figure 33, that with excitation from a very localised region for the cochlea, the responses from the brainstem cells would be synchronised with the sound frequency. But if they came from widespread regions, they would not be synchronised, as seen on the right.
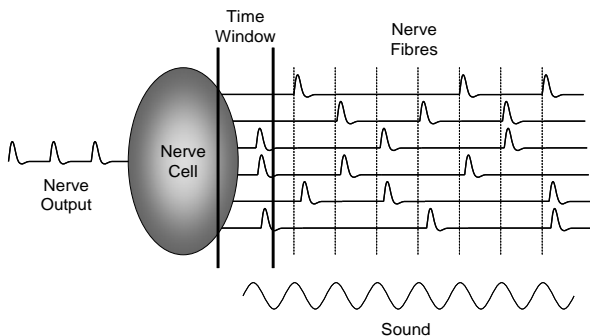

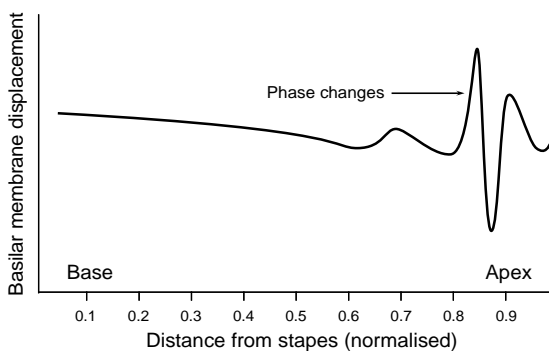**Figure 30** Temporal properties in a group of neurons.

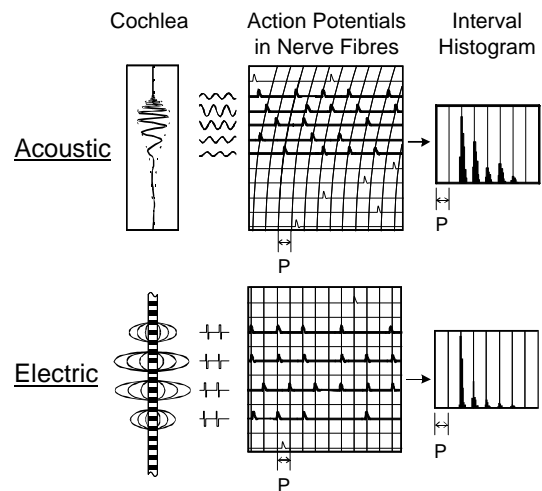**Figure 31.** The basilar membrane travelling waves and phase changes at these maxima.

Figure 32. Temporo-spatial firing patterns. (Clark, 2001)
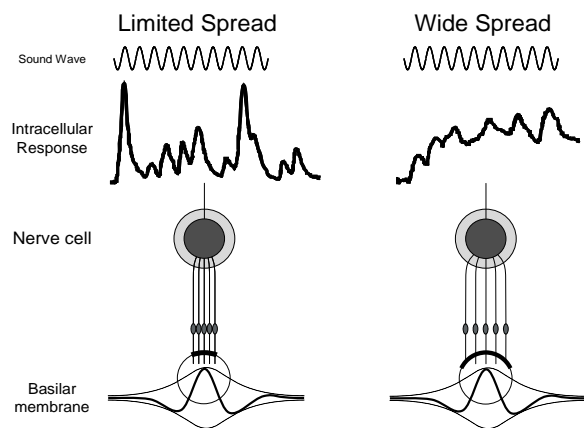

Figure 33. Synchronization and spatial spread for electrical stimulation of neurons. (Paolini et al., 2001)

As the present bundle of electrodes has an electrode separation of 0.75 mm it is not ideal for fine temporo-spatial coding, but a speech processing strategy has been developed to incorporate the longitudinal phase changes.

PRELINGUISTICALLY DEAF CHILDREN

Another key question was would children born deaf be able to develop the right neural connections to be able to understand speech with electrical stimulation using the strategies developed for adults who had hearing before going deaf?

The implant was modified by Cochlear Limited in collaboration with The Melbourne Research team so that it could be used on children. In particular magnets were placed in the implant and transmitting coil to allow ease of alignment (Figure 34).

The first three children to receive the multiple-channel cochlear implant became deaf early in life and were operated on in 1985 and 1986. Following improved speech perception results on these children in Melbourne, a world trial was undertaken for the US FDA on 142 children from two to 17 years of age.

In this trial there was a greatly increased percentage of children achieving maximal performance in either closed or open-set speech understanding after implantation.
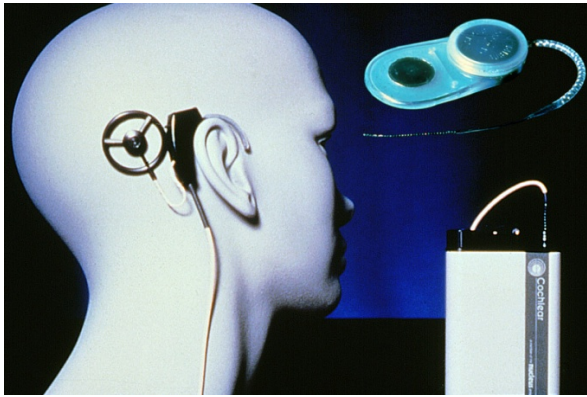
Figure 34. The first Cochlear Pty Ltd "Mini" cochlear implant for children. (Clark, 2003b)

In 1990 the FDA announced that the 22-channel cochlear implant was safe and effective in enabling deaf children from ages two through 17 years to understand speech both with and without lipreading. It was the first cochlear implant to be approved by any world regulatory body for deaf children.

It was next important to analyse the results and determine whether electrical stimulation of the auditory nerve during the early plastic phase of the child's brain development led to better speech perception

The speech results for children under 4 years of age are plotted against those for adults with prior hearing (Figure 35). The results were comparable for phonemes, words and sentences. These suggest that the younger the child at surgery the better the results are likely to be.
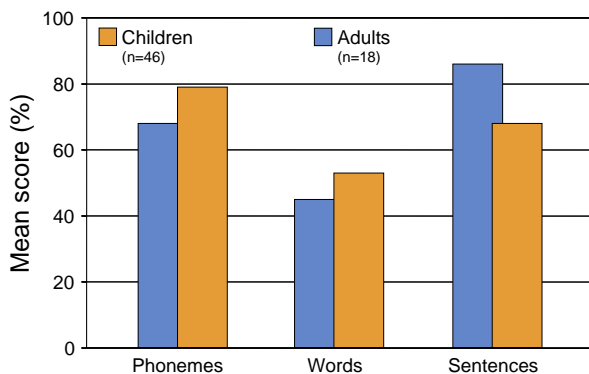

Figure 35. Improvement and variation in results versus age at implantation. (Clark, 2003b)
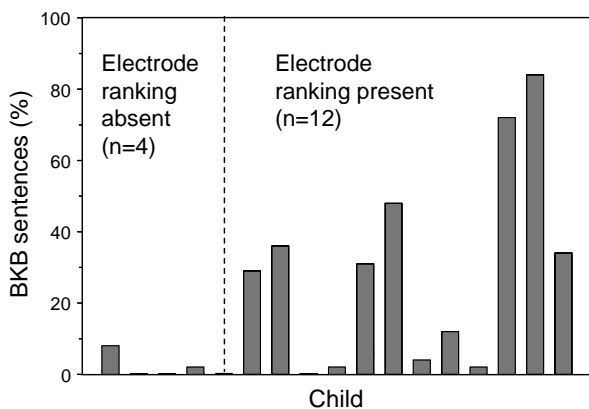

Figure 36. The factors for speech perception. (Clark, 2003b)

In addition, it was shown that with early exposure to stimuli good speech perception was associated with good place pitch discrimination, but this was not the only factor for good speech understanding. In Figure 36 the ability of 16 children to rank electrodes according to whether pitch changed monotonically with electrode place is shown. Speech scores are on the vertical axis. The left sector shows that 4 children could not rank electrodes, and they all had poor speech perception. The right shows 12 children could rank the pitch of the electrodes through early exposure to sound, but not all the children who could rank pitch had good speech perception. This suggests that apart from the ability to rank electrodes there were other factors such as language contributing to good speech perception.

In view of the better results the younger the child at operation, further biological studies were undertaken to address important safety issues in this young age group, namely the effect of head growth, and prevention of a Pneumococcal middle ear infection spreading to the cochlea with the risk of meningitis.

It was only after the safety studies were complete, and showed minimal risk that operations were carried out on young children. We have now demonstrated that children operated on under 4 years of age, and trained in listening skills and spoken language can achieve near normal spoken and receptive language.
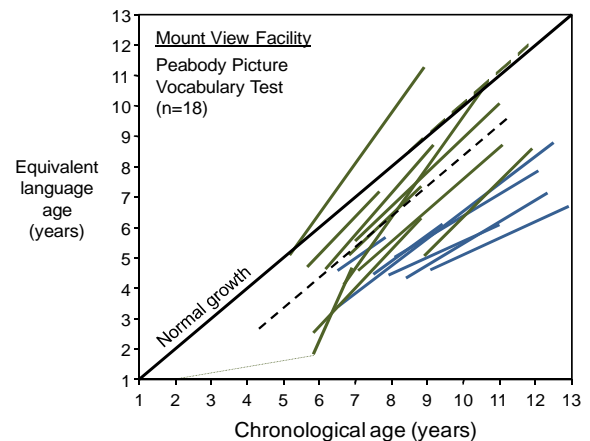

Figure 37. Receptive language at Mount View. (Clark, 2010)

For example, in 18 children at a primary school which provides "Auditory-verbal" education when their receptive language measured as PPVT when plotted against chronological age 61% had a slope >1, i.e. they were developing language faster than children with normal hearing (Figure 37). In addition, 40% of the 18 children showed normal or above normal receptive language. The remaining children had language levels comparable to those expected for children with hearing aids and a severe hearing loss.

CONCLUSION

This research has established the clinical value of electrical stimulation of the central auditory pathways for hearing and speech understanding in deaf people. At the same time it has provided an essential understanding of brain coding, and its relation to perception and the conscious experience of sound. It has also led to establishing a new and successful field of medical prosthetic devices developed by Cochlear and the other companies. Finally it has created a new medical field in neuroprosthetics, not only for the bionic ear, but more generally in Medical Bionics.

REFERENCES

Blamey, P. J., Dowell, R. C., Tong, Y. C. and Clark, G. M. (1984). An acoustic model of a multiple-channel cochlear implant. Journal of the Acoustical Society of America. **76:** 97-103.

Clark, G. M. (1969). Middle ear and neural mechanisms in hearing and the management of deafness. PhD thesis.

Clark, G. M. (1996). Electrical stimulation of the auditory nerve: the coding of frequency, the perception of pitch and the development of cochlear implant speech processing strategies for profoundly deaf people. Clinical and Experimental Pharmacology and Physiology. **23:** 766-776.

Clark, G. M. (2000). Sounds from Silence. Sydney, Allen and Unwin.

Clark, G. M. (2001). Editorial Cochlear implants: climbing new mountains. The Graham Fraser Memorial Lecture 2001. Cochlear Implants International. **2:** 75-97.

Clark, G. M. (2003a). Cochlear implants. Springer Handbook of Auditory Research. Speech Processing in the Auditory System, Springer Verlag.

Clark, G. M. (2003b). Cochlear Implants: Fundamentals and Applications. New York, Springer-Verlag.

Clark, G. M. (2006). "The multiple-channel cochlear implant: the interface between sound and the central nervous system for hearing, speech, and language in deaf people-a personal perspective." Philosophical Transactions of The Royal Society **361**: 791-810.

Clark, G. M. (2008). "Personal reflections on the multichannel cochlear implant and a view of the future." Journal of Rehabilitation Research & Development **45**(5): 651-694.

Clark, G. M. (2010). "The Multi-channel Cochlear Implant and the Mitigation of Severe-to-Profound Deafness." ACTO DE INVESTIDURA DEL GRADO DE DOCTOR HONORIS CAUSA Prenas Universitarias de Zaragoza.

Clark, G. M., Pyman, B. C. and Bailey, Q. R. (1979). The surgery for multiple-electrode cochlear implantations. Journal of Laryngology and Otology. **93:** 215-223.

Laird, R. K. (1979). The Bioengineering development of a sound encoder for an implantable hearing prosthesis for the profoundly deaf. Master of Engineering Science Thesis. The University of Melbourne.

Lawrence, M. (1964). Direct stimulation of auditory nerve fibers. Archives of Otolaryngology. **80:** 367-368.

Nystrand, A. (2006). "Kokleaimplantat ger allt battre resultat." Lakartidningen **103**(37): 2616-2619.

Paolini, A. G., Fitzgerald, J. V., Burkitt, A. N. and Clark, G. M. (2001). Temporal processing from the auditory nerve to the medial nucleus of the trapezoid body in the rat. Hearing Research. **159:** 101-116.

Schuknecht, H. F. (1964). Further observations on the pathology of presbycusis. Archives of Otolaryngology. **80**.

Tong, Y. C., Black, R. C., Clark, G. M., Forster, I. C., Millar, J. B., O'Loughlin, B. J. and Patrick, J. F. (1979). A preliminary report on a multiple-channel cochlear implant operation. Journal of Laryngology and Otology. **93:** 679-695.

Tong, Y. C. and Clark, G. M. (1982). Percepts produced by electrical stimulation of the human cochlea. Proceedings of the Australian Physiological and Pharmacological Society. **13:** 150P.

Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A. and Dowell, R. C. (1982). Psychophysical studies for two multiple-channel cochlear implant patients. Journal of the Acoustical Society of America. **71:** 153-160.

Tong, Y. C., Dowell, R. C., Blamey, P. J. and Clark, G. M. (1983). Two-component hearing sensations produced by two-electrode stimulation in the cochlea of a deaf patient. Science. **219:** 993-994.

ACKNOWLEDGEMENTS