A Cochlear Implant Speech Processing Strategy Based On An Auditory Model

David B Grayden [1], Anthony N Burkitt [1], Owen P Kenny [2],
Janine C Clarey [1], Antonio G Paolini [3], Graeme M Clark [1]
[1] *The Bionic Ear Institute*
384-388 Albert St, East Melbourne VIC 3002, Australia, # dgrayden@bionicear.org
[2] *Dept. of Electrical and Electronic Engineering, James Cook University*
Townsville, Australia
[3] *School of Psychological Science, La Trobe University*
Melbourne, Australia

Abstract

The use of auditory models in cochlear implant sound processing strategies aims to improve cochlear implant users' perception of speech, particularly in noisy environments. To date, cochlear implant sound processing strategies have been designed using simple envelope extraction techniques. Our new strategy simulates the behaviour of the cochlea and the auditory nerve to give a stimulation sequence that produces responses in the auditory nerve that are closer to those in normal hearing. The development and evaluation of the strategy will also help to answer the fundamental question of how much benefit is provided by fine-grained timing information to the perception of different sounds that make up speech.

1. Introduction

The first multi-channel cochlear implant that provided open-set speech perception to adult users was implanted in 1978 and has proven to be effective for people with severe-to-profound hearing loss. However, there are a large number of cochlear implant users who achieve considerably less benefit from the device than others. In addition, background noise has a serious detrimental effect on speech perception for all cochlear implant users. Consequently there has been considerable effort recently to improve both cochlear implant hardware and speech processing strategies in order to overcome these shortcomings.

The effectiveness of a cochlear implant speech processing strategy is principally determined by the amount of information about the original speech signal that is encoded in the electrical pulses that stimulate the auditory nerve. The development of such strategies builds upon the known properties of how the auditory pathway encodes speech. This involves the encoding of information in both a place code (i.e., which neurons respond to the sound) and a temporal code (i.e., the timing of the action potentials that are generated in response to the sound), as summarized briefly in the next section.

Conventional cochlear implant speech processing strategies filter incoming sound using a filterbank or fast Fourier transform, with one filter for each electrode of the implant. The output of each filter is low-pass filtered and the amplitude is used to determine the amount of stimulation to apply to each electrode. The electrodes are then stimulated sequentially with current pulses at a fixed rate of stimulation. Thus the place information is conveyed by which electrode is stimulated and envelope temporal information is provided in the stimulus level. For reviews see [3] and [21].

The Spike-based Temporal Auditory Representation (*STAR*) strategy is a new cochlear implant speech processing strategy that is based upon an auditory model that uses fine-grained temporal information [13]. The auditory model provides precise timing information that is used to control the times that the electrodes are stimulated. This preserves the envelope temporal information and but also controls the rate of stimulation on each electrode and the precise time of stimulation of each electrode across the implanted array.

A large number of auditory models have been developed to model the behaviour of the inner ear (cochlea) and auditory nerve (for an overview of auditory models, see [15]). In particular, Ghitza developed a model that could be applied to Automatic Speech Recognition (ASR) [9], [10], [11], [12]. This model, and those further developed by Kim *et al.* [18], used linear and non-linear filters to replicate the mechanical behaviour of the basilar membrane. The behaviour of the hair cells and auditory nerve was then modelled using thresholds or zero-crossings to obtain spikes (action potentials). Thus the models were able to produce both a rate representation and a temporal representation similar to that observed in the auditory nerve. The models were shown to improve ASR performance in the presence of different types of background noise by providing a robust frequency representation from threshold crossing intervals [9], [12], [18].

The STAR speech processing strategy uses an auditory model similar to that of Kim *et al.* [18]. Extra information is provided to the listener in the temporal sequence of stimulation. Fine-grained temporal coding provided by the auditory model gives rise to formant locking in the temporal representation so that the temporal code provides additional information about the formant frequency across multiple electrodes. In addition, the precise control of stimulation time ensures that onset events, such as plosive bursts and glottal pulses are well represented across the electrodes in a way that more precisely resembles normal hearing compared to conventional cochlear implant strategies. The lower-frequency electrodes also exhibit locking to the fundamental frequency, helping to perceive the pitch of the incoming speech. These temporal properties of the stimulation sequences are combined to improve speech perception in noisy environments, particularly in the presence of other talkers (multitalker babble or cocktail party noise).

This paper will provide a description of the STAR strategy and the basis of its development. In the next section is a short description of the mammalian auditory system, demonstrating the development of place-based models and auditory models. This is followed by a summary of existing cochlear implant processing schemes and then a more detailed description of auditory model-based cochlear implant processing.

2. The Mammalian Auditory System

A. Place-based Coding of Speech in the Auditory System

Sound received by the auditory canal of the ear undergoes a series of transformations as it travels through the outer, middle and inner ear. The acoustic pressure waves captured by the outer ear are converted into mechanical vibrations by the small bones of the middle ear. The signal is then transformed into vibrations of the fluid in the inner ear (the cochlea), which is a snail shaped cavity filled with fluid. In the cochlea there is a flexible membrane, the basilar membrane, which has varying stiffness along its length; it is more rigid at the basal part of the cochlea and more flexible at the apical part of the cochlea. This systematic variation in stiffness of the basilar membrane results in different stimulus frequencies producing vibration displacements of different sections of the membrane. Attached to the basilar membrane are the hair cells, whose motion causes ionic channels in the cell membrane to open that in turn generates action potentials in the auditory nerve fibres to which they are connected. This arrangement produces a tonotopic encoding of the stimulus, i.e., the signal is decomposed into its component frequencies that generate an place code in the pattern of auditory nerve activity. This tonotopic organization of neural activity is maintained at each successive stage of the auditory brainstem and in the early stages of the processing carried out in the auditory cortex.

B. Place-based Models

The pioneering work of von Békésy [36] established that the cochlea is capable of analysing an input acoustic signal into its different frequencies by the vibration of the basilar membrane. A pure tone of a particular frequency causes maximal vibration of the basilar membrane at the “characteristic place” of that frequency, which is turn causes the hair cells attached to that part of the basilar membrane to be maximally activated at that frequency (called the “characteristic frequency” of the particular nerve fibres). Consequently the cochlea acts very much like a spectrum analyzer, decomposing complex sounds into their frequency components.

This place code plays a central role in the neural representation of elements of speech, an early demonstration of which was the way in which the discharge rates of auditory nerve fibres with different characteristic frequencies provides a good spectral representation of steady-state vowels [31]. This place-rate representation of the formant structure of steady-state vowels has subsequently been observed at higher centers of the auditory pathway [28], [32].

ASR systems extract features for training and recognition using the place code (for reviews see [5] and [27]). A filterbank analysis of the incoming speech is performed and than modified using a psychoacoustic frequency scale, such as the mel-scale, and possibly other transformation functions. This results in *frames* of data, typically with durations of 10 msec, with a number of features, or dimensions, of information in each frame derived from the spectral distribution of energy. The task of the ASR engine is then to classify speech according to the features within each frame.

C. Temporal Coding of Speech in the Auditory System

In addition to the frequency information carried by the place code described above, nerve fibres with characteristic frequencies below a couple of kiloHertz can also carry frequency information in their temporal response properties, i.e., in the exact timing of the action potentials that they produce. At these lower frequencies action potentials tend to be produced at the same phase of each cycle, a property that is called “phase-locking”. Individual auditory nerve fibres have been shown to lock to the phase of signals up to 2 kHz and then the phase-locking ability decreases until it is virtually non-existent at 5 kHz [30]. Although individual auditory neurons do not respond to each cycle above $\sim$600 Hz due to refractory effects, the population response of a group of such neurons are encoded through the volley principle [37]. This temporal code is manifested both by phase-locking to the pure-tone acoustical stimulus and by the intervals between the successive action potentials (as plotted in the interspike interval histogram – ISIH) being multiples of the period of the pure tone.

The temporal code has been found to play an important role in the neural representation of the formant structure of vowels, where electrophysiological studies have found that it provides a more accurate and robust neural representation than the place code both in the auditory nerve and in higher stages of the auditory pathway [32]. Subsequent studies showed that neurons in the auditory pathway tend to phase-lock to the nearest formant, i.e., neurons with characteristic frequencies over a range of frequencies near the most dominant formant will tend to phase-lock to the formant frequency [28]. This phenomena is called “formant-locking” and plays a key role in understanding the benefit of the new cochlear implant speech processing algorithm described below.

In addition, one cell type within the cochlear nucleus (the bushy cells that have primary-like responses) preserves the phase-locking of AN fibres through specialised, so-called “secure” synapses (end-bulbs of Held) [2], so that spike times are precisely conveyed to other brainstem areas. Indeed, these bushy cells of the anteroventral cochlear nucleus have been found to phase-lock in a more precise manner than auditory nerve fibers [17]. Within the cochlear nucleus there is also a cell type (octopus cells) that receives convergent inputs from AN fibres with a wide range of best frequencies. These cells respond to the onset of a sound, as well as transient, broadband events within a complex acoustic stimulus, with great precision [26].

D. Auditory Models

Auditory models are designed to produce output that resembles that of the auditory nerve, although further processing is usually performed to allow the model output to be used in automatic speech recognition or other tasks. To demonstrate the working of an auditory model, the model developed by Kim, Lee and Kil [18] is outlined below.

The first stage of an auditory model, as illustrated in Fig. 1, is a filter bank that models the cochlear filters. Different auditory models have used from 16 filters [18] to 192 filters [10]. The filters may be non-linear filters designed to model the non-linear behaviour of the basilar membrane and hair cells, or may be linear filters that capture the essential band pass function of the basilar membrane.

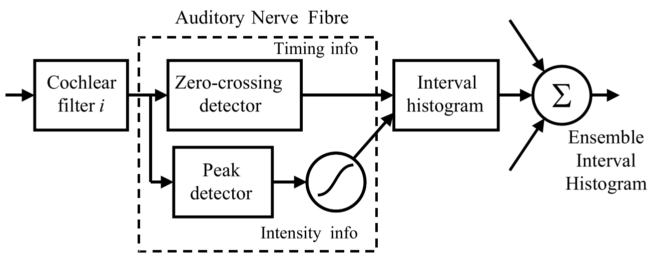

Fig. 1: Schematic view of the auditory model showing how the outputs of each cochlear filter generates a “spike” at each zero crossing. The amplitude of the spike is determined by the maximum of the signal between subsequent spikes. The inverse of the inter-spike interval times gives an instantaneous frequency that is binned to form the interval histogram. A “temporal spectrum” is then created by combining the interval histograms from all filters. After [18].

The interaction between the basilar membrane and the hair cell-auditory nerve fibre components is then obtained from zero crossings of the filter outputs. Each zero crossing generates a “spike”. The spikes may have a fixed amplitude, as they are in the auditory system, by using multiple thresholds for each filter output instead of zero crossings [10]. The Kim *et al.* model instead uses spikes with amplitudes that vary with the amplitude of the filtered waveform after each zero crossing. This is illustrated in Fig. 2. The spike amplitude may then be modified by a compressive nonlinearity that represents the amplitude nonlinearities of the hair cell-auditory nerve interface.

For use in conventional ASR programs, the spike-based representation of the auditory model must be converted into a frame-based representation. This is achieved by creating a frequency histogram by binning the inverse times between adjacent spikes (the inter-spike intervals) for each filter. This results in an *interval histogram* for each modelled auditory

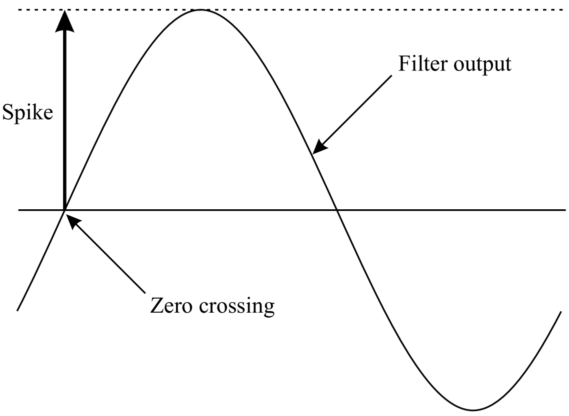

Fig. 2: A spike is generated from the filtered signal representation at the time of a zero crossing. The amplitude that is assigned to the spike is the peak amplitude after the zero crossing.

nerve fibre. The interval histograms are then combined across all filters to obtain the *Ensemble Interval Histogram (EIH)* [10]. An example of an EIH is shown in Fig. 3, which is a frequency representation of a portion of the vowel /ɛ/ as in “bet”. A spectrum derived from LPC analysis is also shown for comparison. The EIH may then be transformed using techniques such as mel-scale and cepstral analysis to create features for ASR.

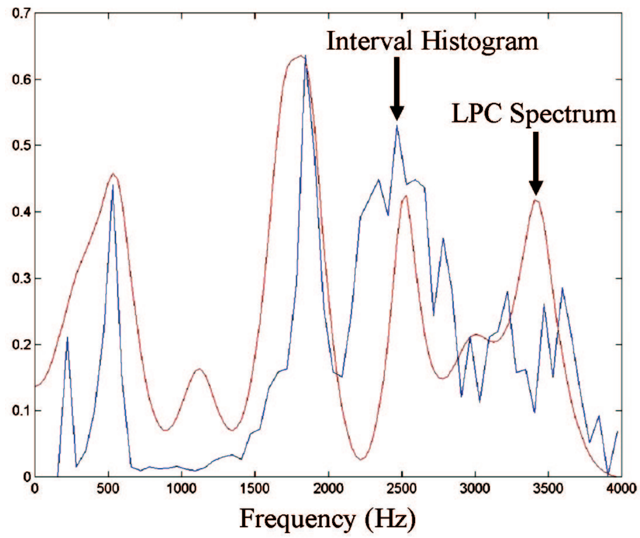

Fig. 3: An example of an interval histogram generated by the auditory model for a portion of the sound /ɛ/ in the word “bet” spoken by a male speaker. Also shown is the LPC spectrum. The locations of the four formants can clearly be seen by the largest peaks in the LPC spectrum. Note that the interval spectrum detects the voicing frequency (the peak just below F1) and has narrow peaks for the formants F1 and F2. The ability to detect high frequency formants is diminished.

The benefit provided by the EIH is that the inter-spike intervals carry information about the frequencies of peaks in the spectrum of a signal that highlights the locations of lower formant frequencies, especially the most important ones, F1 and F2 [10]. This can be seen in Fig. 3 where the interval histogram has much sharper peaks in the first and second formant positions. This provides more robust features that improve ASR performance in the presence of high levels of background noise [12], [18].

3. Existing Cochlear Implant Processing Schemes

There are two types of deafness: conductive and sensorineural. Conductive deafness results from middle ear damage that reduces the efficiency of transmission of the mechanical vibration due to sound. This can usually be rectified by the use of acoustic amplification by hearing aids. In sensorineural deafness the hair cells are typically damaged (or have died), which results in reduced (or non-existent) activation of the auditory nerve fibres. This hair cell damage can have many causes such as: disease (e.g., meningitis, Meniere's disease), congenital disorders, ototoxic drugs, and physical damage. Acoustic amplification can help to give sufficient hearing to understand speech provided the amount of hearing loss is not too severe or profound.

Those with severe to profound hearing loss may benefit from the use of a cochlear implant. Unlike hearing aids, which amplify sounds to the ear, the cochlear implant electrically stimulates the auditory nerve directly, bypassing the ear completely. An electrode array is surgically inserted into the cochlea, usually through the round window and along the scala tympani, below the basilar membrane. This places the electrodes at different positions along the cochlea so that electric current will stimulate auditory nerve fibres with different characteristic frequencies: higher frequencies at the basal end of the implant and lower frequencies towards the tip or apical end. The implant thus conveys information about the amplitude and frequency of the acoustic frequency using sequences of stimulating electrical pulses.

Early signal processing designs extracted the second formant (F2) and fundamental frequency (F0) to control electrode stimulation [4]. The frequency of F2 controlled the location of electrode stimulation, and F0 controlled the rate of stimulation. Improvements were made by also extracting the first formant (F1) and adding a second stimulated electrode for each pitch period [6]. The MULTIPEAK stimulation strategy added stimulation of a number of fixed electrodes towards the base of the cochlea to better represent high-frequency information and improve consonant perception [7]. The next stages of development were the Spectral Maxima Sound Processor (SMSP) strategy [22] and SPEAK strategy [33]. These were a departure from the others as they used a fixed stimulation rate and stimulated electrodes that corresponded to maxima in the sound spectra. Perception of the fundamental frequency of speech now relied on perception of the amplitude modulation across the electrodes. Another fixed-rate strategy, the Continuous Interleaved Sampling (CIS) strategy, consecutively stimulated all of a small number of electrodes to represent the sound spectra [38]. Each of the developments in cochlear implant sound processing demonstrated a significant improvement in speech perception from its predecessor. Since the development of these strategies, the only differences made to commercially available strategies has been increasing the rate of stimulation of the electrodes and application of various noise reduction and gain control algorithms.

Most clinically available strategies employ pulsatile non-simultaneous stimulation of the implanted electrodes, necessitated by the need to reduce current interactions between electrodes that will occur if electrodes are stimulated simultaneously [20]. In each case, the incoming sound is filtered into a series of frequency bands. Each of these filters is assigned to one of the electrodes of the implant, with high frequencies assigned to electrodes at the basal end of the cochlea and low frequencies assigned to the apical end. The amplitude envelopes of the filter outputs are sampled at intervals equal to a chosen fixed stimulation rate. Then some or all of the implanted electrodes are stimulated with current pulses that are modulated by the extracted envelopes.

In addition, some manufacturers provide analogue strategies that sample the filter outputs continuously and provide continuous currents to all of a small number of electrodes [8]. However, the presence of current interactions makes this strategy unsuitable for most users and also cannot be employed in many of the cochlear implants that are available.

Apart from the analogue strategies, all of the clinically available sound-processing strategies use fixed-rate electrical stimulation. Clinical evaluations have shown that fixed-rate stimulation is superior to strategies that employ the extracted fundamental frequency as the stimulation rate because of difficulty in extracting this information in the presence of background noise [22], [33]. The fixed-rate strategies rely heavily on place-pitch perception where the location of stimulation within the cochlea provides information about the frequency components of the acoustic signal. This is particularly effective for vowels and sonorant consonants (e.g., /r/ and /m/) because they are identified by the locations of lower formants, which are peaks in the intensity of the spectrum. However, the perception of voicing (i.e., when the vocal folds are vibrating) and plosive consonant sounds (e.g., /p/ and /b/) requires detection of precise timing information and rapid changes in amplitude that the envelope sampling procedures are inadequate to represent. Perception of voicing, in particular, is important for separation of speech from competing noise, especially other voices [1]. Good perception of these sounds requires high stimulation rates, but conventional high-rate strategies typically use the same width analysis windows as low-rate strategies so short duration transient events are smoothed anyway and there is no overall improvement in speech perception [14], [35].

4. Auditory Model-Based Cochlear Implant Processing

Kitazawa *et al.* describe a strategy for the spectra/CI-22 system, an early model implant and sound processor, using an auditory model for choosing electrodes to stimulate and for extracting fundamental frequency to control rate of stimulation [19]. This strategy was tested by simulation with normal hearing listeners. Stimulation was restricted to low rates and required explicit extraction of pitch information. Performance was reduced for word recognition compared to the Wearable Speech Processor (WSP) running MULTIPEAK, probably because of the low rate restriction.

Meyer-Bäse *et al.* developed a neural model with the aim of extracting inter-spike intervals for control of cochlear implant stimulation [23], [24], [25]. However, this work was on the development of hardware for implementing auditory models and the papers give no details of how the cochlear implant strategy would be implemented.

A. The STAR Sound Processing Strategy

The Spike-based Temporal Auditory Representation (STAR) sound processing strategy uses a model of the cochlea and auditory nerve to create a spike-based representation of the acoustic signal much like that observed in the normal-hearing auditory nerve. In essence, the strategy simulates the behaviour of the hair cells of the cochlea and of the auditory nerve to create a spike-based representation modelled on action potentials generated in the auditory neurons. The cochlea is electrically stimulated using a sequence of electrical pulses that replicates this spike-based representation as closely as possible within device limitations.

The advantage of the spike-based representation and stimulation method is that this strategy provides fine-grained timing information in addition to place-based spectral information. The STAR strategy provides stimuli that more closely match the input to these cells than those provided by conventional strategies. Improved detection of transient and onset events is important for speech perception in the presence of noise, especially pseudo-stationary random noise and multi-talker babble. Voicing information is conveyed by temporal information that is observed across the whole frequency range (called “voicing striations” in spectrograms) and this is conserved by the spike-based representation, but lost in existing fixed-rate strategies. In addition, noise tolerance is increased by providing an improved representation of fine-grained temporal information, especially phase locking to formants across multiple stimulation channels.

The basis of the strategy is an auditory model that was developed by Ghitza [10] for improving the performance of automatic speech recognition in noisy environments. This strategy was further refined and simplified by Kim *et al.* [18] for ASR and is in a form that may be adapted to cochlear implant signal processing.

B. Filtering

The acoustic signal is first passed through an auditory filterbank whose centre-frequency distribution is based on psychophysical and physiological measures [34], [16], similar to those used with existing implant strategies. The filters’ bandwidths are based on equivalent rectangular bandwidth and so have significant overlap. In the initial development of the STAR strategy, linear filters have been used. Up to 22 filters are used with the Nucleus™ CI-24 devices; the number used depends on the number of electrodes available to the cochlear implant user.

C. Spike Times and Amplitudes

For each filter output a series of “spikes” is extracted relating to the instances at which the filter’s signal output crosses the zero axis (see Fig. 2). The amplitude of the spike is determined by the following peak of the signal, though the stimulation level is obtained by a logarithmic loudness growth function. The cochlea is electrically stimulated using a sequence of electrical pulses that replicates this spike-based representation as closely as possible within cochlear implant device and safety limits.

An example of an electrical stimulation sequence is given in Fig. 4. This figure illustrates a stimulation pattern for the word “bet”. The horizontal lines represent electrodes of the implant distributed across different frequency bands on a logarithmic scale. The vertical bars represent stimulation pulses. The height of the vertical bars gives a representation of the amount of current.

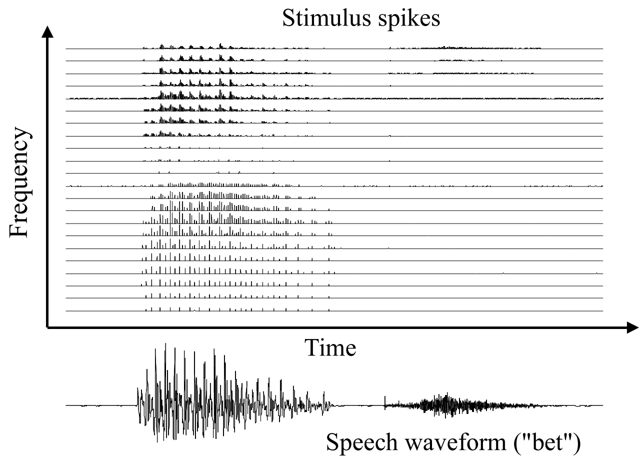

Fig. 4: An example of an electrical stimulation sequence for the word “bet” using the STAR speech processing strategy. The vertical bars represent stimulation pulses on each of the 22 electrodes.

D. Dealing with Spike Timing Contention

Electrodes are stimulated at threshold crossing times (delayed across all electrodes by up to 10 msec to allow peak detection). Owing to limitations in existing cochlear implants, a limited number of electrodes are chosen for stimulation by selecting channels with the highest amplitude. Temporal contention, which is when two spikes occur at the same time, is resolved by systematically shifting spikes to different time instances around the zero crossing location, with preference for the correct timing given to spikes with higher amplitude. Average rate of stimulation are limited on each electrode by ensuring that spikes in high frequency channels are not stimulated above a fixed maximum average rate, currently set at 1800 Hz. This is performed because phase-locking to electrical stimuli is lost at this rate of stimulation [29].

5. Conclusion

The aims of current cochlear implant research are:

1) to provide a better representation of speech by electrical current in the cochlea to improve speech perception in noise;
2) to produce a more natural perception of sound, particularly of speech and music;
3) to learn more about the properties of speech and function of the auditory system through the analyses of speech feature perception and psychophysics.

The Spike-based Temporal Auditory Representation (STAR) strategy was designed to benefit each of the aims. The primary difference between STAR and existing commercial cochlear implant strategies is that it provides a fine-grained temporal representation. This temporal representation is expected to improve speech perception in noise and make sounds appear more natural compared to a fixed-rate stimulation strategy. In addition, the temporal coding allows the investigation of the impact of the temporal properties of sound on the perception of speech.

Acknowledgments

This work was supported by NHMRC Development Grant No. 248500 and The Bionic Ear Institute. Part of this work was carried out under a collaborative research agreement with Telstra.

References

[1] J.P.L. Brookx and S.G. Nooteboom, "Intonation and the perception of simultaneous voices", *J. Phonet.*, 1982, Vol. 19, pp. 23-26.

[2] N.B. Cant and D.K. Morest, "The bushy cells in the anteroventral cochlear nucleus of the cat. A study with the electron microscope", *Neuroscience*, Vol. 4, 1979, pp.1925-1945.

[3] G. Clark, *Cochlear Implants: Fundamentals and Applications*, Springer, New York, 2003, pp. 381-453.

[4] G.M. Clark, Y.C. Tong, Q.R. Bailey, et al., "A multiple-channel cochlear implant", *J. Otolaryng. Soc. Aust.*, Vol. 4, 1978, pp.208-212.

[5] J. Deller, J. Proakis, and J. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, New York, 1993.

[6] R.C. Dowell, P.M. Seligman, P.J. Blamey, and G.M. Clark, "Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise", *Acta Oto-Laryngologica*, Vol. 104, 1987, pp.439-446.

[7] R.C. Dowell, L.A. Whitford, P.M.Seligman, B. Franz, and G.M. Clark, "Preliminary results with a miniature speech processor for the 22-electrode/Cochlear hearing prosthesis", In: T. Sacristan, ed., *Otorhinolaryngology, head and neck surgery*, Kugler and Ghednin, Amsterdam, , 1990, pp.1167-1173.

[8] D.K. Eddington, "Speech discrimination in deaf subjects with cochlear implants", *J. Acoust. Soc. Am.*, 1980, Vol. 68, pp. 885-891.

[9] O. Ghitza, "Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment", *J. Phonetics*, Vol. 16, 1988, pp.109-123.

[10] O. Ghitza, "Auditory nerve representations as a basis for speech processing", In: *Advances in Speech Signal Processing*, S. Furui & M.M. Sondhi (Eds), Marcel Dekker, New York, pp. 453-485.

[11] O. Ghitza, "Adequacy of auditory models to predict human internal representation of speech sounds", *J. Acoust. Soc. Am.*, Vol. 93, 1993, pp.2160-2171.

[12] O. Ghitza, "Auditory models and human performances in tasks related to speech coding and speech recognition", *IEEE Trans. Speech Audio Proc.*, Vol. 2, 1994, pp.115-132.

[13] D.B. Grayden, A.N. Burkitt, O.P. Kenny, J.C. Clarey, A.G. Paolini, P.F. Duke, and G.M. Clark, "Generation of Electrical Stimuli for Application to a Cochlea" (2003) Patent PCT/AU03/00639 (International Publication Number WO 03/099179 A1).

[14] D.B. Grayden and G.M. Clark, (2000) "The effect of rate stimulation of the auditory nerve on phoneme recognition", In: *Proc. Eighth Aust. Int. Conf. Speech Science and Tech.*, Canberra, 2000, pp. 356-361.

[15] S. Greenberg (Ed.) Theme issue "Representation of Speech in the Auditory Periphery", Journal of Phonetics, Vol. 16, No. 1, January 1988.

[16] D.D. Greenwood, "A cochlear frequency-position function for several species - 29 years later", *J. Acoust. Soc. Am.*, 1990, Vol. 87, pp. 2592-2605.

[17] P.X. Joris, L.H. Carney, P.H. Smith, and T.C.T. Yin, "Enhancement of neural synchronization in the anteroventral cochlear nucleus: I. Responses to tones at the characteristic frequency", *J. Neurophysiol.*, Vol. 71, 1994, pp.1022-1036.

[18] D.-S. Kim, S.-Y. Lee, and R. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy environments", *IEEE Trans. Speech Audio Proc.*, Vol. 7, 1999, pp.55-69.

[19] S. Kitazawa, K. Muramoto, J. Ito, "Acoustic simulation of auditory model based speech processor for cochlear implant system", In *Proc. International Conference on Spoken Language Processing, ICSLP'94*, Japan, 1994, pp. 2043-2046.

[20] R.K. Laird, *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, 1979.

[21] P.C. Loizou, "Mimicking the Human Ear", *IEEE Signal Processing Magazine*, 1998 (Sept), pp. 101-130.

[22] C.M. McKay, H.J. McDermott, A. Vandali, and G.M. Clark, "A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (Multipeak) processor", *Acta Oto-Laryng.*, 1992, Vol. 112, pp. 752-761.

[23] U. Meyer-Bäse, A. Meyer-Bäse, H. Scheich, "Auditory Neuron Models for Cochlea Implants", *SPIE AeroSense*, Orlando, Florida, UAS, 1997, Vol. 3077, pp. 582-593.

[24] U. Meyer-Bäse, "An interspike interval method to compute speech signals from neural firing", *SPIE AeroSense*, Orlando, Florida, UAS, 1998, Vol. 3390, pp. 560-571.

[25] U. Meyer-Bäse, A. Meyer-Bäse, H. Scheich, "An inter-spike interval method for computing phase locking from neural firing", *Biological Cybernetics*, Vol. 82, 2000, pp. 283-290.

[26] D. Oertal, R. Bal, S.M. Gardner, P.H. Smith, and P.X. Joris, "Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus", *Proc. Nat. Acad. Sci.*, Vol. 97, 2000, pp.11773-11779.

[27] D. O'Shaughnessy, *Speech Communication*, Addison-Wesley, Reading, MA, 1987.

[28] A.R. Palmer, I.M. Winter, and C.J. Darwin, "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons", *J. Acoust. Soc. Am.*, Vol. 79, Jan. 1979, pp.100-113.

[29] A. Paolini, J. Fitzgerald, A. Burkitt, and G. Clark, "Temporal processing from the auditory nerve to the medial nucleus of the trapezoid body in the rat", *Hearing Research*, 2001, Vol. 159, pp. 101-116.

[30] J.E. Rose, J.F. Brugge, D.J. Anderson, and J.E. Hind, "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey", *J. Neurophysiol.*, Vol. 30, 1967, pp.769-793.

[31] M.B. Sachs and E.D. Young, "Encoding of steady-state vowels in the auditory nerve: Representations in terms of discharge rate", *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp.470-479.

[32] M.B. Sachs, C.C. Blackburn, and E.D. Young, "Rate-place and temporal-place representations of vowels in the auditory nerve and anteroventral cochlear nucleus", *J. Phonet.*, Vol. 16, 1988, pp.37-53.

[33] M.W. Skinner, G.M. Clark, L.A. Whitford, P.M. Seligman, *et al.*, "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channels cochlear implant system", *Am. J. Otology*, 1994, Vol. 15, pp. 15-27.

[34] S.S. Stevens, "The relation of pitch to frequency: A revised scale", *Am. J. Psych.*, 1940, Vol. 53, pp. 329-353.

[35] A.E. Vandali, L.A. Whitford, K.L. Plant, and G.M. Clark, "Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system", *Ear Hear.*, 2000, Vol. 21, pp. 608-624.

[36] G. von Békésy, *Experiments in Hearing*, Wiley, New York, 1960.

[37] E.G. Wever, *Theory of hearing*, Wiley, New York, 1949.

[38] B.S. Wilson, D.T. Lawson, M. Zerbi, and C.C. Finley, "Speech processors for auditory prostheses", *Twelfth Quarterly Progress Report, NIH contract No. 1-DC-9-2401*, Research Triangle Institute, Apr. 1992.