SPEECH PROCESSING STRATEGIES FOR COCHLEAR PROSTHESES-THE PAST, PRESENT AND FUTURE: A TUTORIAL REVIEW

P Mahalakshmi¹, M R Reddy²
Biomedical Engineering Group, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai-600 036, Tamil Nadu, India.
¹Email: maha_50@yahoo.com  ²Email: rsreddy@iitm.ac.in

ABSTRACT

Cochlear Implants are widely accepted prosthetic devices that improve the hearing ability of people with profound hearing loss. The cochlear implant speech processor is responsible for decomposing the input sound into different frequency bands and delivering the most appropriate pattern to the electrodes. The performance of the cochlear prostheses depends on various parameters such as number of channels, number of electrodes, type of stimulation, rate of stimulation and compression function. The objective of this paper is to review how the sound signals are coded in cochlear implants, using different speech processing strategies focusing on waveform, feature-extraction and hybrid. The review describes the coding of sound for single and multi-channel implants based on the type (analog and pulsatile), rate of stimulation and compression function. Also, speech processing strategies used in currently available commercial cochlear speech processors are presented. Results of several investigations show that the strategies based on spectral signal analysis allow for better speech understanding than speech feature extraction. Though the performance of different strategies is variable from patient to patient, certain strategies have variable features which improve the speech perception.

Keywords - Auditory prostheses, cochlear implants, electric stimulation, loudness, signal processing, speech perception.

I. INTRODUCTION

Sensori-neural deafness affects a large number of people throughout the world. This can be caused either by cochlear damage or by damage within the auditory nerve or the neurons of the central auditory system. A profoundly deaf ear is typically one which the majority of sensory receptors in the inner ear, called hair cells, are damaged or diminished. The hair cells are the sensory cells that transduce mechanical motion into electrical signals. The auditory neurons carry information from the hair cells to the brain. Research has shown that the most common cause of deafness is the loss of hair cells, rather than the loss of auditory neurons.
This was very encouraging for cochlear implants because the remaining neurons could be excited directly through electrical stimulation. A cochlear prosthesis is therefore based on the idea of bypassing the normal hearing mechanism (outer, middle and part of the inner ear including the hair cells) and directly stimulates the inner ear sensory cells of the auditory nerve by delivering electrical signals to an electrode array implanted inside the cochlea. These electrical signals are derived from the external sound acquired from a microphone.

The type of signal processing used for coding speech signals is defined as the speech coding/processing strategy. These strategies play a major role in extracting various parameters from the acoustic signals and converting them into electrical signals. Various speech processing strategies have been developed and reported in literature [1-3] over time for cochlear prostheses which include Compressed Analog (CA), Continuous Interleaved Sampling (CIS), Feature based strategy, Multipeak (MPEAK), Spectral Peak (SPEAK), Advanced Combination Encoder (ACE) and Spectral Maxima Sound Processing Strategy (SMSP). The purpose of this article is to present an overview of various speech processing strategies that have been used for cochlear prostheses over the past four decades. Section II gives information about the principle of functioning of cochlear implants and stimulation parameters required for the implant. Section III discusses the classification of implants and various speech processing strategies that are commonly used. Section IV presents the strategies used in commercial cochlear processors. Section V summarizes the paper with the concluding remarks.

II. COCHLEAR IMPLANTS

A. Cochlear Implant Functioning

A cochlear implant is an electronic system that is used to provide hearing to subjects affected by severe or profound hearing loss. The system consists of two main elements, an external processor and an internal element that is implanted into the patient by means of a surgical operation. The implanted element has an electrode array which is placed in the cochlea, in order to provide stimulation of the auditory nerve by means of electrical stimuli. A block diagram of a general cochlear implant is shown in Figure 1 [2].

The basic functioning of the cochlear implant is as follows: The microphone (Mic) acquires the sound signal, transforms it into an electrical signal, and sends it to the amplifier. The signal-processing circuit contains filters or feature-extraction electronics to decompose the electrical signal. It analyzes the sound and determines the stimulation level to be sent at each electrode.

![Fig.1 Block diagram of cochlear prostheses](image)

The stimulation pattern is sent to the internal part (Receiver/Stimulator) of the system by radio transmission, and the internal part generates the electrical pulses, that are presented at each intra-cochlear electrode of the implant. The pulses at each electrode cause the activation of the neural ends of the auditory nerve providing a hearing sensation.

B. Stimulation parameters

The stimulation technique depends on (a) filter bank frequencies (b) stimulation/pulse rate and order (c) dynamic range of compression. The filter bank must represent the auditory system with a non linear frequency distribution along the electrode array. The
stimulation/pulse rate defines the number of pulses per second delivered to each electrode. Pulse rates can vary between 100 and 2500 pulses per second (pps). Loizou reported that some patients obtain a maximum recognition performance with a pulse rate of 833 pps and a biphasic pulse width of 33µs/phase [6]. This can be interpreted that high-pulse rate stimulation can represent fine temporal variations in a better way. The stimulation order can be varied to minimize possible interaction between channels. The stimulation order refers to the sequence with which the electrodes are stimulated. One possibility is to stimulate the electrodes in an apex-to-base order. In this way, signals in the low frequencies (apex) are stimulated first, and signals in the high frequencies (base) are stimulated last. Transforming acoustic amplitudes into electrical amplitudes is done through compression function. This transformation is essential because the range in acoustic amplitudes in conversational speech is considerably larger than the implant patient’s dynamic range [3]. The range in electrical amplitudes between threshold and loudness levels is said to be the dynamic range.

III. SIGNAL PROCESSING FOR COCHLEAR PROSTHESES

Speech signals comprise of three important acoustic features like formant, pitch and energy of the speech signal. Based on the understanding of human brain functions, different speech processing strategies are proposed and used successfully in cochlear implant devices. These speech processors extract the parameters that are essential for intelligibility and then encode them for electrical stimulation of the auditory nerve. Early cochlear devices used single channel implants and later in 1980s, multi-channel implants were introduced.

A. SINGLE CHANNEL IMPLANTS

Single-channel implants were commonly used in 1970s. They provide electrical stimulation at a single site in the cochlea using a single electrode [3]. The device was composed of only a single-channel processor and one implanted cochlear electrode.

a) 3M/House Speech Processor

One of the successful single-channel cochlear implant devices was developed by House and Urban in the early 1970s and manufactured by 3M Company [4]. Figure 2 shows the block diagram of this device. The acoustic signal is picked up by a microphone, amplified, and then processed through a 340-2700Hz band pass filter (BPF). The band passed signal is then used to modulate a 16 KHz carrier signal. The modulated signal goes through an output amplifier and is applied to an external coil. The output of the implanted coil is finally sent to the implanted active electrode in the scala tympani.

Fig.2 House single-channel cochlear implant processor

Since speech processing strategy in these devices has neglected the temporal details of the cochlea nerves that differ from place to place, single-channel devices have not been successful in providing accurate speech perception for implantees. Thus, multichannel cochlear implants that could enable different electrodes to stimulate the auditory nerve fibers with different temporal features have been introduced as a better replacement to the single-channel/electrode cochlear implant devices.
B. MULTI-CHANNEL IMPLANTS

Multichannel implants provide electrical stimulation at multiple sites in the cochlea by using an array of electrodes. Thus, different auditory nerve fibers can be stimulated at different places in the cochlea thereby exploiting the place mechanism for coding frequencies. Different electrodes are stimulated depending on the signal frequencies. Electrodes near the base of the cochlea are stimulated with high-frequency signals, while electrodes near the apex are stimulated with low-frequency signals. The digital speech processing used in cochlear prostheses is built around a filter bank model. The number of filter bands depends on the number of stimulation channels to be considered in the implant system [5].

The various signal-processing strategies developed for multichannel cochlear prostheses can be divided into three categories: Waveform, Feature-extraction and Hybrid. These strategies differ in the way information is extracted from the speech signal and presented to the electrodes. Waveform strategies present some type of waveform derived by filtering the speech signal into different frequency bands. Feature-extraction strategies present some type of spectral features, such as formants, derived using feature extraction algorithms. Hybrid strategy is one that combines features and waveform representation.

A. WAVEFORM STRATEGIES

a) Compressed Analog Strategy

The Ineraid device, manufactured by Symbion Inc. used the Compressed Analog (CA) design in its speech processor as shown in Figure 3 [6]. The CA design uses analog stimulation that delivers four continuous analog waveforms to four electrodes simultaneously. The signal is compressed by the automatic gain control (AGC) circuit in a logarithmic fashion so that the signal amplitude is within the dynamic hearing range (just perceived sound to maximum comfort level). It is then band pass filtered into four contiguous frequency bands with center frequencies at 0.5, 1, 2 and 3.4 KHz. The filtered waveforms go through adjustable gain controls, and then sent directly to four electrodes (El-1 to El-4) through a percutaneous connector (one that pierces the skin).

![Fig.3 Four-channel compressed analog strategy](image)

Superior speech recognition performance was obtained using CA approach over the single-channel approach. But with simultaneous stimulation, interactions between channels caused by the summation of electrical fields from individual electrodes arise, which may distort the speech spectrum information and degrade speech understanding.

b) Continuous Interleaved Sampling Strategy

Problems of channel interaction inside the cochlea are addressed in the continuous interleaved sampling (CIS) strategy through the use of interleaved non-simultaneous stimuli. Researchers at the Research Triangle Institute (RTI) developed the CIS approach by using
non-simultaneous, interleaved pulses [6]. It emphasizes timing information of speech. This strategy uses a fixed set of electrodes and offers new stimulation rates. The block diagram of CIS strategy is shown in Figure 4.

The speech signal is first pre-emphasized and is divided into sub-bands using a bank of band-pass filters and interleaved biphasic pulses are generated from the envelopes of the band-pass filter outputs. The envelope detection (ED) block performs full-wave rectification and low-pass filtering (LPF-cut off frequency: 200-400Hz) that are used to extract the envelopes of the filtered waveforms [7]. The amplitude of each stimulus pulse is determined by a logarithmic function (non-linear mapping) that compresses the signal to fit the patient’s dynamic hearing range. The electrodes are activated with biphasic pulses sequentially at a relatively high stimulation rate. The rate of stimulation on each channel usually exceeds 800 pulses per second [8].

Clinical studies on human subjects showed that CIS processors provide much better speech perception than CA processors. The CIS strategy is currently implemented in several multichannel cochlear implant systems with slight variations in the stimulation rates and number of channels.

**B. Feature Extraction Strategies**

The CA and CIS strategies presented waveform information obtained by filtering the speech signal into a few frequency bands. The feature-based speech processor operates by extracting the spectral information from the input speech signal and using this information to generate the stimulus to the electrodes. For proper perception of speech, it is important to present the fundamental and formant frequencies [9]. The lowest frequency of a periodic waveform is called the fundamental frequency (F0). The peaks that are observed in the spectral envelope are called formants, the first peak being the first formant frequency (F1) and the second peak being the second formant frequency (F2). The information that humans require to distinguish between vowels can be represented purely quantitatively by the frequency content of the vowel sounds. Therefore, formant frequencies are extremely important features and formant extraction is thus an important aspect of speech processing. The Nucleus implant manufactured by Cochlear Corporation and developed at the University of Melbourne in the early 1980s, used these techniques. Some of the techniques used in this device are discussed in the following sections. The Nucleus cochlear implant was a 24-electrode device.
a) **F0/F2 Strategy**

This strategy uses zero-crossing and envelope detectors. Zero-crossing rate is an important parameter for voiced/unvoiced classification. It is a measure of number of times in a given time interval that the amplitude of the speech signals passes through a value of zero. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. Voiced and unvoiced speech usually shows low and high zero-crossing count respectively.

The fundamental frequency F0 is estimated using a zero-crossing detector at the output of a 270Hz low pass filter. The second formant frequency F2 is estimated using a zero-crossing detector (ZCD) at the output of a 1000-4000Hz band pass filter. The estimated energy value in the frequency region of F2 is used to select the electrode to be stimulated. The amplitude of the F2 formant is obtained after rectification and low pass filtering the band passed output. Voicing information is conveyed with F0 by stimulating the appropriate electrode at a rate of F0 pulses/sec [10]. During unvoiced segments, the selected electrode is stimulated at quasi-random intervals at an average rate of 100 pulses/sec.

b) **F0/F1/F2 Strategy**

This strategy was an improvement on previous F0/F2 technique since it included the first formant frequency F1 also. The block diagram of F0/F1/F2 strategy [11] is shown in Figure 5.

![Block diagram of F0/F1/F2 strategy](image)

In the F0/F2 strategy, an additional zero-crossing detector was included to estimate F1 at the output of a 280-1000 Hz band pass filter. Two sets of electrodes were stimulated, one with F1 formant information and the other with F2 formant information. The F1 information was used to stimulate the apical electrodes and the F2 formant information for the basal electrodes. However, the selection of a particular electrode for stimulation in F1 and F2 is not understood. 200 µsec pulses were used with a separation of 800 µsec to avoid channel interaction. The pulse amplitudes were proportional to the amplitudes A1 and A2 of the F1 and F2 formants and the stimulation rate was still F0 pulses per second for voiced segments and at an average rate of 100 pulses per second for unvoiced segments.

The addition of F1 in the F0/F2 strategy improved the speech recognition performance of patients using the Nucleus cochlear implant. This strategy emphasizes low frequency information, which is required for vowel recognition and it did not yield significant improvements on consonant-recognition [6]. The majority of the consonants contain high-
frequency information and this has motivated the refinement of the F0/F1/F2 strategy to the MPEAK strategy.

c) **MPEAK**

A further improvement over the F0/F1/F2 scheme was the Multipeak (MPEAK) that extracted and used high-frequency information [12] from the speech signal to stimulate the electrodes as shown in Figure 6.

A 800-4000Hz band pass filter was used to extract F2. High frequency information was extracted using three additional band pass filters (2000-2800Hz, 2800-4000Hz, 4000-6000Hz) [13]. The motivation for using the three additional band pass filters is to include high-frequency information which is important for the perception of consonants. The estimated envelope amplitudes of the three band pass filters were delivered to fixed electrodes. The MPEAK strategy stimulates four electrodes at a rate of F0 pulses/sec for voiced sounds, and at quasi-random intervals with an average rate of 250 pulses per second for unvoiced sounds.

For voiced sounds, stimulation occurs on the F1 and F2 electrodes and on the high frequency electrodes 4 (2800-4000Hz) and 7 (2000-2800Hz). Due to the less energy in the spectrum above 4 KHz for voiced sounds, electrode 1 was not stimulated. For unvoiced sounds, stimulation occurs on the high-frequency electrodes 1(4000-6000Hz), 4, and 7, as well as on the electrode corresponding to F2. As there is generally less energy present in the spectrum below 1 KHz for unvoiced sounds, electrode corresponding to F1 was not stimulated.

![Fig.6 Block diagram of MPEAK strategy](image_url)

Due to the availability of the high frequency information, improved consonant identification was observed. Though this strategy has proven to be an efficient strategy for consonant identification, it has one major limitation. It tends to make errors in formant extraction when the speech signal is embedded in noise. This limitation in feature-extraction algorithms motivated the development of Spectral Maxima Sound Processing Strategy (SMSP).
C. HYBRID/N-OF-M STRATEGIES

a) Number-of-Maxima (n-of-m) Strategy

In “n-of-m” strategy, the signal was filtered into \( m \) frequency bands, and the processor selects the \( n \) (\( n<m \)) envelope outputs with higher amplitudes [14]. Only the electrodes corresponding to the \( n \) selected outputs are stimulated at each cycle. For example, in a 4-of-8 strategy, from a maximum of eight channel outputs, only four channel outputs with higher amplitudes are selected for stimulation at each cycle [15]. The implementation of this type of processing in the Nucleus implant is referred to as the spectral peak extraction (SPEAK) strategy. SPEAK analyzes the incoming sound to identify the filters that have the larger amounts of energy, selects a subset of filters, and then stimulates the selected electrodes. The stimuli are pulsatile and non-simultaneous. With SPEAK, 6 to 10 electrodes are activated sequentially at a rate that averages approximately 250 pulses per second on each activated electrode [16]. The Advanced Combination Encoder (ACE), offered in the Nucleus system, combines the spectral maxima detection of SPEAK with a higher stimulation rate.

b) SMSP

The Spectral Maxima Sound Processing Strategy (SMSP), developed in the early 1990s for the Nucleus multi-electrode cochlear implant, used a 6-of-16 strategy. Unlike previous strategies developed for the Nucleus implant, the SMSP strategy did not extract any features like F0, F1, from the speech waveform. The speech signal was analyzed using a bank of 16 band pass filters (with center frequencies ranging from 250 to 5400 Hz) and a spectral maxima detector. The output of each filter was rectified and low-pass filtered with a cutoff frequency of 200 Hz. After computing all 16 filter outputs, the SMSP processor selects at 4 msec intervals, the six larger filter outputs [17]. The six amplitudes of the spectral maxima were finally logarithmically compressed, to fit the patient’s electrical dynamic range, and are transmitted to the six selected electrodes through radio transmission. Typical clinical rates of stimulation range from 250 pps to 1800 pps [18].

IV. PROCESSING STRATEGIES USED IN COMMERCIAL PROCESSORS

Presently, there are three major cochlear implant processors: the Nucleus 24, manufactured by Cochlear Corporation, (www.cochlear.com), Australia the Clarion, manufactured by Advanced Bionics Corporation, USA (www.advancedbionics.com), and the Med-El by Medel Corporation, Austria (www.medel.com), with cochlear being the dominant company. This section gives an overview of speech processing strategies used in these commercially available implant processors.

(i) Nucleus -24 Processor

The Nucleus-24 device is equipped with an array of 22 intra-cochlear electrodes and 2 extra-cochlear electrodes. One of the extra-cochlear electrodes is a small platinum ball electrode placed under the temporalis muscle. The second one is a platinum plate on the body of the receiver/stimulator. The extra-cochlear electrodes are used as reference electrodes.

This processor can be programmed with the ACE and CIS strategies [19]. Both strategies estimate the input signal spectrum using FFT. In CIS strategy, a fixed number of amplitudes are used for stimulation based on processing the signal through 10-12 bands. In the ACE strategy, 8-12 maximum amplitudes are selected for stimulation. The remaining electrodes are inactivated. Electrodes corresponding to the selected bands are then stimulated from basal to apical order. The stimulation rate can be chosen from a range of 250 to 2400 pulses per sec per channel and is limited by a maximum rate of 14,400 pulses per sec across all channels.
(ii) **Clarion Processor**

The Advanced Bionics Corporation’s (ABC’s) implant has undergone a number of changes in the past decade. ABC’s first generation implant (Clarion S-Series) used an electrode array with 8 contacts and supported simultaneous stimulation strategy. Its second generation device (Clarion CII) used 16 electrodes and supported a high-rate CIS strategy. Clarion’s CIS strategy, called HiRes, differs from the traditional CIS strategy in the way it estimates the envelope. It uses half-wave rectification rather than full-wave rectification, and it does not use a low-pass filter. HiRes operate at a stimulation rate of 2800 pulses per sec using a pulse width of 11 µsecs/phase.

(iii) **Med-El Processor**

The Med-El cochlear implant processor, manufactured by Med-El Corporation, can be programmed with either a high-rate CIS strategy or a high-rate spectral maxima strategy. The Med-El processor has the capability of generating 18,180 pulses/sec for a high-rate implementation of the CIS strategy in the 12-channel implant. The amplitudes of the pulses are derived as follows. The signal is first pre-emphasized, and then applied to a bank of 12 logarithmically spaced band pass filters. The envelopes of the band pass filter outputs are extracted. Biphasic pulses, with amplitudes set to the mapped filter outputs, are delivered in an interleaved fashion to 12 electrodes at a rate of 1515 pulses per sec per channel [19]. The latest Med-El device supports simultaneous stimulation of 12 electrodes.

V. **Conclusion**

Cochlear implants have been very successful in restoring partial hearing to profoundly deaf people. This paper provides an overview of various speech processing strategies developed for the cochlear implants since the early 1970s. Single electrode implant systems were quite commonly used in 1970s, but the advances achieved in multi-electrode systems subsequently made them more common. The first generation multi-channel Nucleus 24 device extracted the fundamental frequency (F0), which is a source information reflecting the voice pitch, and the second formant frequency (F2). In later versions of the implant, the first formant was added, followed by additional three spectral peaks between 2000 and 6000 Hz. In the late 1980s and early 1990s, extraction of temporal information was given importance as it supported obtaining a high level of speech recognition. CIS strategy avoided channel interactions and preserved the temporal envelope. Then, the n-of-m strategy was introduced and in that, a total of $m$ frequency bands are analyzed and the $n$ electrodes corresponding to the $n$ highest energy bands are stimulated on a given processing cycle. The SPEAK strategy selects 6 to 8 largest peaks of the band pass filtered output. The ACE strategy has a larger range of peak selection and higher rate than the SPEAK strategy. If $n=m$, then the SPEAK and ACE strategies are essentially same as the CIS strategy. This paper also presented an overview of the processing strategies used in currently available commercial processors. Current cochlear implants do not adequately reproduce several aspects of the neural coding of sound in the normal auditory system. Improved electrode arrays and coding systems may lead to improved coding and it is hoped for a better performance.

**References**


