Signal Processing for Cochlear Prosthesis: A Tutorial Review

Philipos C. Loizou

Department of Applied Science
University of Arkansas at Little Rock
Little Rock, AR 72204-1099, U.S.A.
http://giles.ualr.edu/asd/cimplants/

***Abstract*— Cochlear implants has been very successful in restoring partial hearing to profoundly deaf people. The success of cochlear implants can be attributed to the combined efforts of scientists from various disciplines including bioengineering, speech science and signal processing. This paper will present an overview of various signal processing techniques that have been used for cochlear prosthesis over the years.**

I. Introduction

For centuries, people believed that only a miracle could restore hearing to the deaf. It was not until forty years ago that scientists first attempted to restore normal hearing to the deaf by electrical stimulation of the auditory nerve. Today, a prosthetic device, called *cochlear implant*, can be implanted in the inner ear and can restore partial hearing to profoundly deaf people. Some individuals with implants can now communicate without lipreading or signing, and some can communicate over the telephone. The success of cochlear implants can be attributed to the combined efforts of scientists from various disciplines including bioengineering, physiology, otolaryngology, speech science, and signal processing. Signal processing, in particular, played an important role in the development of different techniques for deriving electrical stimuli from the speech signal. The purpose of this article is to present an overview of various signal processing techniques that have been used for cochlear prosthesis over the past 25 years.

II. Cochlear implants

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 electrically stimulating the remaining auditory neurons directly.

Several cochlear implant devices have been developed over the years [1]. All the implant devices have the following features in common: a microphone that picks up the sound, a signal processor that converts the sound into electrical signals, a transmission system that transmits the electrical signals to the implanted electrodes, and an electrode or an electrode array (consisting of multiple electrodes) that is inserted into the cochlea by a surgeon (Figure 1). In single-channel implants only one electrode is used. In multichannel cochlear implants, an electrode array is inserted in the cochlea so that 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 frequency of the signal. 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 signal processor is responsible for breaking the input signal into different frequency bands or channels and delivering the filtered signals to the appropriate electrodes.

The cochlear implant is based on the idea that there are enough auditory nerve fibers left for stimulation in the vicinity of the electrodes. Once the nerve fibers are stimulated, they fire and propagate neural impulses to the brain. The brain interprets these impulses as sounds. The perceived loudness of the sound may depend on the number of nerve fibers activated and their rates of firing. The loudness of the sound can be controlled by varying the amplitude of the stimulus current. The pitch on the other hand is related to the place in the cochlea that is being stimulated. Low pitch sensations are elicited when electrodes near the apex are stimulated, while high pitch sensations are elicited when electrodes near the base are stimulated. In summary, the implant can effectively transmit information to the brain about the loudness of the sound, which is a function of the amplitude of the stimulus current, and the pitch, which is a function of the place

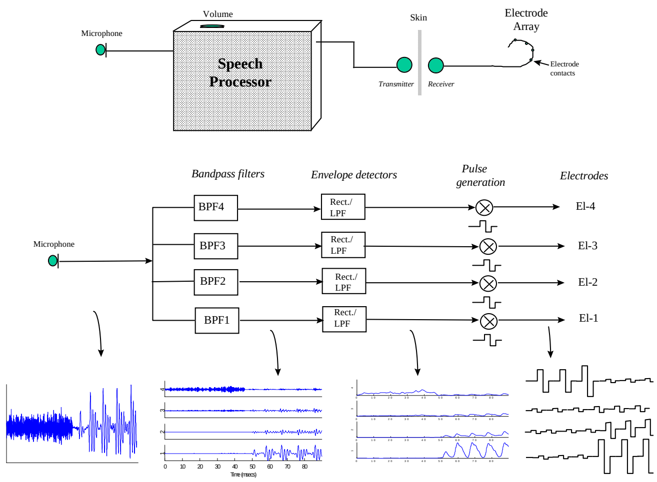

Figure 1: Operation of a 4-channel implant processor.

in the cochlea being stimulated.

Figure 1 shows, as an example, the operation of a four-channel implant. Sound is picked up by a microphone and sent to a speech processor box (the size of a pager) worn by the patient. The sound is then processed through a set of four bandpass filters which divide the acoustic waveform into four channels. Current pulses are generated with amplitudes proportional to the energy of each channel, and transmitted to the four electrodes through a radio-frequency link. The relative amplitudes of the current pulses delivered to the electrodes reflect the spectral content of the input signal (Figure 1).

III. Signal Processing

When multi-channel implants were introduced in the 1980s, several questions were raised regarding multi-channel stimulation. Some of those questions were: "What kind of information should be transmitted to each electrode?" Should it be some type of spectral features or attributes of the speech signal that are known to be important for speech perception (e.g., first and second formants), or some type of waveform derived by filtering the original speech signal into several frequency bands? Depending on how researchers tried to address these questions, different types of signal processing techniques were developed. The various signal processing strategies developed for multi-channel cochlear prosthesis can be divided into three categories: waveform strategies, feature-extraction strategies and "n-of-m" strategies. These strategies differ in the way that information is extracted from the speech signal and presented to the electrodes. A review of these signal processing strategies is given next.

IV. Waveform Strategies

A. *Compressed-Analog (CA) approach*

The compressed-analog (CA) approach was originally used in the Ineraid device manufactured by Symbion, Inc., Utah [2]. The signal is first compressed using an automatic gain control, and then 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 through a percutaneous connection to four intracochlear electrodes. The filtered waveforms are delivered simultaneously to four electrodes in analog form. The CA approach, used in the Ineraid device, was very successful because it enabled many patients to obtain open-set speech understanding. Dorman *et al.* [3] report, for a sample of 50 Ineraid patients a median score of 45% correct for word identification in sentences.

B. *Continuous Interleaved Sampling (CIS) approach*

The CA approach uses analog stimulation that delivers four continuous analog waveforms to four electrodes simultaneously. A major concern associated with simultaneous stimulation is the interaction between channels caused by the summation of electrical fields from individual electrodes. These interactions may distort speech spectrum information and therefore degrade speech understanding.

Researchers at the Research Triangle Institute (RTI) developed the Continuous Interleaved Sampling (CIS) approach [4] which addressed the channel interaction issue by using non-simultaneous, interleaved pulses. Trains of biphasic pulses are delivered to the electrodes in a non-overlapping (non-simultaneous) fashion, that is, in a way such that only one electrode is stimulated at a time. The amplitudes of the pulses are derived by extracting the envelopes of bandpassed waveforms. The signal is first pre-emphasized and passed through a bank of bandpass filters (see Figure 2). The envelopes of the filtered waveforms are then extracted by full-wave rectification and low-pass filtering (typically with 200 or 400 Hz cutoff frequency). The envelope outputs are finally compressed and then used to modulate biphasic pulses. A non-linear compression function (e.g., logarithmic) is used to ensure that the envelope outputs fit the patient's dynamic range of electrically evoked hearing. Trains of balanced biphasic pulses, with amplitudes proportional to the envelopes, are delivered to the six electrodes at a constant rate in a nonoverlapping fashion. The rate at which the pulses are delivered to the electrodes has been found to have a major impact on speech recognition [1]. High pulse-rate stimulation typically yields better performance than low-pulse rate stimulation. Comparison between the CA and CIS approach revealed higher levels of speech recognition with the CIS approach [4].

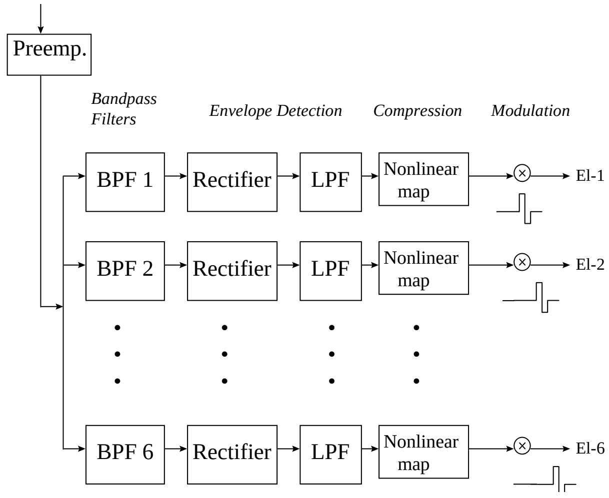

Figure 2: Block diagram of the CIS strategy.

V. Feature Extraction Strategies

Both CA and CIS strategies presented waveform information obtained by filtering the speech signal into a few frequency bands. In contrast, other strategies presented spectral features, such as formants, obtained by formant extraction algorithms. The following sections provide an overview of the feature extraction strategies used in the Nucleus multielectrode implant manufactured by Nucleus Limited, Australia.

A. F0/F2 and F0/F1/F2

The F0/F2 strategy was the first strategy developed for the Nucleus device in the early 1980s [5]. In this strategy, the fundamental frequency (F0) and the second formant (F2) are extracted from the speech signal using zero crossing detectors. One zero-crossing detector is used to estimate F0 from the output of a 270 Hz low-pass filter, and another zero-crossing detector is used to estimate F2 from the output of a 1000-4000 Hz bandpass filter. The amplitude of F2 is estimated with an envelope detector by rectifying and low-pass filtering (at 35 Hz) the bandpassed signal. The F0/F2 processor conveys F2 frequency information by stimulating the appropriate electrode in the 22-electrode array. Voicing information is conveyed with F0 by stimulating the selected electrode at a rate of F0 pulses/sec. During unvoiced segments the selected electrode is stimulated at quasi-random intervals with an average rate of 100 pulses/sec. The amplitude of the pulses are set in proportion to the amplitude of F2. Initial results [6] with the F0/F2 strategy were encouraging as it enabled some patients to obtain open-set speech understanding.

The F0/F2 strategy was later modified to include information about the first formant frequency (F1) [7] and became available in 1985 with the Nucleus wearable speech processor (WSP). In the F0/F1/F2 strategy, an additional zero crossing detector was included to estimate F1 from the output of a 280-1000 Hz bandpass filter.

B. MPEAK

Further improvements to the F0/F1/F2 processor were made in the late 1980s by Cochlear Pty. Limited (a subsidiary of Nucleus Limited) in collaboration with the University of Melbourne [8]. A new coding strategy, called MPEAK, was used that extracted from the speech signal, in addition to formant information, high-frequency information. Similar to the F0/F1/F2 strategy, the extraction of the formant frequencies F1 and F2 was performed using zero crossing detectors, and the amplitudes of F1 and F2 were computed using envelope detectors. High frequency information was extracted, using envelope detectors, from the frequency bands 2000-2800 Hz, 2800-4000 Hz and 4000-6000 Hz. The motivation for using the three additional bandpass filters (> 2 kHz) was twofold: (1) to enhance the representation of the second formant (F2), and (2) to include high-frequency information which is important for the perception of consonants. The estimated envelope amplitudes of the three bandpass filters were delivered to fixed electrodes. Electrodes 7, 4, and 1 were allocated to the outputs of the filters 2-2.8 kHz, 2.8-4 kHz and 4-6 kHz respectively.

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/sec for unvoiced sounds. For voiced sounds, stimulation occurs on the F1 and F2 electrodes and on the high-frequency electrodes 4 (2000-2800 Hz) and 7 (2800-4000 Hz). The high-frequency electrode 1 does not get stimulated because there is generally little energy in the spectrum above 4 kHz in voiced sounds. For unvoiced sounds, stimulation occurs on the high-frequency electrodes 1, 4 and 7 as well as on the electrode corresponding to F2. The electrode corresponding to F1 does not get stimulated because there is generally little energy below 1000 Hz in unvoiced sounds (e.g., /s/).

VI. "N-of-m" strategies

In these strategies, the signal is filtered into m frequency bands, and the processor selects, out of m envelope outputs, the n ($n < m$) envelope outputs with the largest energy. 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 the four channel outputs with the largest amplitudes are selected for stimulation at each cycle. The "n-of-m" strategy can be considered to be a hybrid strategy in that it combines a feature representation with a waveform representation.

A. IP

The Interleaved Processor (IP) developed at the Research Triangle Institute was the first processor to use an "n-of-m" strategy [9]. In this processor, the signal was filtered into six frequency bands and out of the six envelope ouputs, the two envelope outputs with the highest energy were selected at each stimulation cycle. Pulses were then delivered in an interleaved manner to the two electrodes corresponding to the two channels with the highest energy. The pulses were delivered at a maximum pulse rate of 313 pulses/sec per channel.

B. SMSP

The Spectral Maxima Sound Processor (SMSP), developed in the early 1990s for the Nucleus multielectrode cochlear implant, used a 6-of-16 strategy [10]. Unlike previous processors developed for the Nucleus implant, the SMSP processor did not extract any features from the speech waveform. Instead it analyzed the speech signal using a bank of 16 bandpass filters and a spectral maxima detector. The signal was preamplified and then sent through a bank of 16 bandpass filters with center frequencies ranging from 250 to 5,400 Hz. 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 selected, at 4 msec intervals, the six largest filter outputs. The six amplitudes of the spectral maxima were finally logarithmically compressed, to fit the patient's electrical dynamic range, and transmitted to the six selected electrodes through a radio-frequency link. Six biphasic pulses were delivered to the selected electrodes in an interleaved (i.e., non-simultaneous) fashion at a rate of 250 pulses/sec.

VII. State of the Art Processors

There are currently two cochlear implant processors in the United States approved by the Food and Drug Administration (FDA), the Nucleus Spectra 22 processor and the Clarion processor. There is also a processor manufactured by Med-El Corporation, Austria, that is currently in clinical trials. This section provides an overview of these three commercially available implants.

A. Nucleus Spectra 22 processor

The Spectra 22 is the latest speech processor of the Nucleus 22 channel implant system manufactured by Cochlear Pty. Limited, Australia. It can be programmed with either a feature extraction strategy (e.g., F0/F1/F2, MPEAK strategy) or the SPEAK strategy, which is similar to the SMSP strategy. In the SPEAK strategy [11] the incoming signal is sent to a bank of 20 (rather than 16 in SMSP) filters with center frequencies ranging from 250 Hz to 10 kHz. The SPEAK processor continuously estimates the outputs of the 20 filters and selects the ones with the largest amplitude. The number of maxima selected varies from 5 to 10 depending on the spectral composition of the input signal, with an average number of six maxima. The selected electrodes are stimulated at a rate that varies between 180 and 300 Hz depending on: (1) the number of maxima selected, and (2) on patient's individual parameters. Comparison of the SPEAK strategy and the MPEAK strategy showed that the SPEAK strategy performed better on vowel, consonant and sentence recognition [12]. Especially large improvements were found with tests in noise.

B. Clarion processor

The Clarion cochlear implant system [13] is the result of cooperative efforts among the University of California at San Francisco (UCSF), Research Triangle Institute (RTI) and the device manufacturer, Advanced Bionics Corporation. The Clarion implant supports a variety of speech processing options and stimulation patterns. The stimulating waveform can be either analog or pulsatile, the stimulation can be either simultaneous or sequential and the stimulation mode can be either monopolar or bipolar. The Clarion processor can be programmed with either the compressed analog (CA) strategy or the CIS strategy. In the compressed analog mode, the acoustic signal is processed through eight filters, compressed and delivered simultaneously to eight electrode pairs. Analog waveforms are delivered to each electrode at a rate of 13,000 samples/sec per channel. The CA strategy emphasizes detailed temporal information at the expense of reduced spatial selectivity due to simultaneous stimulation. For some patients, use of simultaneous stimulation results in a loss of speech discrimination due to channel interaction. This problem is alleviated in the CIS mode which delivers biphasic pulses to all eight channels in an interleaved manner. In the CIS mode, the signal is first pre-emphasized and passed through a bank of eight bandpass filters. The envelopes of the filtered waveforms are then extracted by full-wave rectification and low-pass filtering. The envelope outputs are finally compressed to fit the patient's dynamic range and then used to modulate biphasic pulses. Pulses are delivered to eight electrodes at a maximum rate of 833 pulses/sec per channel in an interleaved fashion.

C. Med-El processor

The Med-El cochlear implant processor, manufactured by Med-El Corporation, Austria, is currently in clinical trials in the United States. The implant processor [14] is based on the Motorola 56001 DSP, and can be programmed with either a high-rate CIS strategy or a high-rate n-of-m strategy. The Med-El processor has the capability of generating 12,500 pulses/sec for a high-rate implementation of the CIS strategy. The amplitudes of the pulses are derived as follows. The signal is first pre-emphasized, and then applied to a bank of eight (logarithmically-spaced) bandpass filters of Butterworth type and of sixth-order. The bandpass filter outputs are full-wave rectified and low-pass filtered with a cutoff of 400 Hz. The low-pass filter outputs are finally mapped, using a logarithmic-type compression function, to the patient's dynamic range. Biphasic pulses, with amplitudes set to the mapped filter outputs, are delivered in an interleaved fashion to eight monopolar electrodes at a maximum rate of 1,515 pulses/sec per channel. The pulses are transmitted transcutaneously through a radio-frequency link.

VIII. Conclusions

Cochlear implants have been very successful in restoring partial hearing to profoundly deaf people. Many individuals with implants are now able to communicate and understand speech without lipreading, and some are able to talk over the phone. Children with implants can develop spoken language skills and attend normal schools (i.e., schools with normal-hearing children). The greatest benefits with cochlear implantation have occurred in postlingually deafened adults (i.e., adults who acquired speech and language before their hearing loss), and have shorter duration of deafness. Gradual, but steady, improvements in speech production and speech perception have also occured in prelingually deafened adults or children. Much of the success of cochlear implants was due to the advancement of signal processing techniques developed over the years. While this success is very encouraging, there is still a great deal to be learned about electrical stimulation of the auditory nerve, and many questions to be answered.

Acknowledgment

The author would like to thank Blake Wilson and Michael Dorman for providing valuable suggestions on earlier drafts of this manuscript. This work was supported in part by grant No. 98-B-07 from the Arkansas Science and Technology Authority.

References

[1] B. Wilson, "Signal processing," in *Cochlear Implants: Audiological Foundations* (R. Tyler, ed.), pp. 35–86, Singular Publishing Group, Inc, 1993.

[2] D. Eddington, "Speech discrimination in deaf subjects with cochlear implants," *J. Acoust. Soc. Am.*, vol. 68, no. 3, pp. 885–891, 1980.

[3] M. Dorman, M. Hannley, K. Dankowski, L. Smith, and G. McCandless, "Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant," *Ear and Hearing*, vol. 10, pp. 44–49, 1989.

[4] B. Wilson, C. Finley, D. Lawson, R. Wolford, D. Eddington, and W. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236–238, July 1991.

[5] P. Seligman, J. Patrick, Y. Tong, G. Clark, R. Dowell, and P. Crosby, "A signal processor for a multiple-electrode hearing prosthesis," *Acta Otolaryngologica*, pp. 135–139, (Suppl. 411), 1984.

[6] R. Dowell, D. Mecklenburg, and G. Clark, "Speech recognition for 40 patients receiving multichannel cochlear implants," *Archives of Otolaryngology, Head and Neck Surgery*, vol. 112, pp. 1054–1059, 1986.

[7] P. Blamey, R. Dowell, and G. Clark, "Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant," *J. Acoust. Soc. Am.*, vol. 82, pp. 38–47, 1987.

[8] J. Patrick, P. Seligman, D. Money, and J. Kuzma, "Engineering," in *Cochlear prostheses* (G. Clark, Y. Tong, and J. Patrick, eds.), pp. 99–124, Edinburgh: Churchill Livingstone, 1990.

[9] B. Wilson, C. Finley, and et al., "Comparative studies of speech processing strategies for cochlear implants," *Laryngoscope*, vol. 98, pp. 1069–1077, 1988.

[10] H. McDermott, C. McKay, and A. Vandali, "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," *J. Acoust. Soc. Am.*, vol. 91, pp. 3367–3371, 1992.

[11] P. Seligman and H. McDermott, "Architecture of the Spectra 22 speech processor," *Annals of Otology, Rhinology and Laryngology*, pp. 139–141, (Suppl. 166), 1995.

[12] M. Skinner, G. Clark, and et al., "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," *American Journal of Otology*, vol. 15, pp. 15–27, (Suppl. 2), 1994.

[13] D. Kessler and R. Schindler, "Progress with a multi-strategy cochlear implant system: The Clarion," in *Advances in cochlear implants* (I. Hochmair-Desoyer and E. Hochmair, eds.), pp. 354–362, Vienna: Manz, 1994.

[14] C. Zierhofer, O. Peter, S. Bril, P. Pohl, I. Hochmair-Desoyer, and E. Hochmair, "A multichannel cochlear implant system for high-rate pulsatile stimulation strategies," in *Advances in cochlear implants* (I. Hochmair-Desoyer and E. Hochmair, eds.), pp. 204–207, Vienna: Manz, 1994.