HEARING WITH BIONIC EARS

Speech Processing Strategies for Cochlear Implant Devices

Around 10% of the population in developed countries suffers from hearing impairment. Elderly people (older than 65) comprise the largest group with 45% of people who are hearing impaired. Besides this group, 25- to 45-year-olds make up 42% of the hearing-impaired population. Since these people continuously work in high-risk environments, they are also candidates to incur hearing damage or loss.

Various devices have been invented to improve speech perception of hearing-impaired people. Early versions of hearing-aid devices (before 1900) were often shaped like a tapered horn by which the sound was guided toward a small opening on the side of the ear. In the early 20th century, with the discovery of the diode tube and the triode, the first commercially available hearing aids were introduced. They simply consisted of a microphone, a telephone, and a bulky battery. After the invention of transistors in 1947, the first behind-the-ear hearing aids were introduced at the end of the 1950s. In the 1980s, so-called “in-the-ear” or “in-the-canal” hearing instruments, which could be worn inside the ear, were made possible as a result of rapid advances in high-density IC technologies. With the continuous progress in the IC technology, other features such as volume remote-control, digital programmability, and self-adoptability became possible in hearing-aid instruments.

Efforts in developing an auditory prosthesis can be divided into two major parts: one is concerned about the hearing-aid systems developed for the people who have partial hearing loss, and the other one concerns cochlear implant systems developed for deaf people whose auditory sensors (hair cells in the cochlea) are not functional at all. Several signal processing and acoustic transmission techniques and devices, such as tactile devices, have been proposed and produced over the past 30 years for helping the hearing impaired [1, 2]. In this article, we review the hearing process and types of hearing disorders, and we then review speech processing techniques in early and modern cochlear implant devices.

The Hearing Process

The auditory pathway can be divided into three basic divisions: external, middle, and the inner ear, as shown in Fig. 1. Audition begins with the external ear, which includes...
1. Human auditory pathway.

The pinna, ear channel, and eardrum (tympanic membrane). The external ear collects the sound waves and guides them to the middle ear. The eardrum or tympanic membrane transmits the vibrations in the air to a system of a small bones in the middle ear. The middle ear is a small cavity that contains the smallest three bones of the human body (ossicular chain): malleus, incus, and stapes. Equivalently, the eardrum and the middle-ear bones work as a mechanical transformer that transfers pressure changes in the air to fluid movements within the cochlea. The inner ear is located in a cavity that is protected by the hardest bone of the human body.

The part that deals with hearing is called the cochlea and it consists of a bony tube about 33 mm in length, coiled upon itself for two and a half turns. Two membranes, called Reissner's membrane and the basilar membrane, divide the cochlea into three parallel fluid-filled canals: scala vestibuli (upper), scala media (middle), and scala tympani (lower) as shown in Fig. 2. The basilar membrane supports a structure known as the “Organ of Corti,” on which the hair cells are placed. Displacement of the oval window, which is connected to the last small bone (stapes) of the middle ear, produces fluid displacement within the cochlea, which is from the scala vestibuli to the scala tympani. During this hydraulic displacement, the basilar membrane also moves, and this movement is turned into electrical pulses by hair cells placed in the Organ of Corti. About 30,000 neurons exit the cochlea and these neurons process 1500 to 2500 cochlea inner hair-cell signals.

All kinds of signal processing and manipulation take place in the cochlea. Auditory nerve pulses are transmitted to the cochlear nuclei in the brainstem. The cochlear cells distribute these coded pulses to multiple synaptic points in the brainstem before sending them to the primary and secondary auditory cortex. The auditory signals transfer from the primary and secondary auditory cortex to the Wernicke's area, where the auditory signals are analyzed and interpreted into language-specific, meaningful messages.

2. Cross-section of the cochlea.

Hearing Disorders
Hearing impairments are clinically categorized into two major groups: conductive and sensorineural disorders. The impairment levels are determined with respect to the pure tone threshold average (PTA) in decibels (dB). Problems at the external or
middle ear that block or degrade sound transmission from the external ear to the cochlea are the cause of conductive hearing loss. Today, replacements of the middle ear bones and other sophisticated corrective procedures are available for conductive hearing loss patients [1].

Sensorineural hearing loss involves changes in the inner ear that result in a change in sensitivity to sound. The loss of hair cells in the cochlea due to exposure to loud sound or heavy drug treatment is the most common sensorineural impairment. Damage of hair cells also results in subsequent degeneration of the adjacent auditory neurons. If the hair-cell and auditory nerve damage is excessive, the connection between the central nervous system and the external world is lost and the person who has such level of loss is recognized as being profoundly deaf (PTA > 91 dB). However, some amount of living auditory neurons can still exist in the cochlea, even with extensive loss of hair cells. Direct electrical stimulus of these neurons can create a sound sensation in profoundly deaf people. These electronic neural stimulus systems are called cochlear prostheses. Profoundly deaf people are assumed to be good candidates for cochlear implant devices if they have some degree (< 30%) of open-set speech recognition ability with best-fit hearing aids.

Speech Processors for Cochlear Prostheses

Speech signals include different acoustic features. Peaks in vocal-tract transfer function, (called formant positions), vocal tract excitation rate (pitch), and the energy of the speech signal are three important acoustic features of speech signals. However, not all of the acoustic features are equally important for normal human speech perception. Based on the understanding of human brain functions, different speech processing strategies are proposed and used successfully in cochlear implant devices.

In general, a cochlear implant module contains four components: a microphone, an external speech processor, a transmission link with internal circuitry, and an electrode or array of electrodes, as shown in Fig. 3. Early cochlear implant devices were composed of only a single-channel signal processor and one implanted cochlear electrode.
7. Feature-based speech processor.

3M/House Speech Processor
One of the successful early single-channel cochlear implant devices, shown in Fig. 4 [3], was developed by House and Urban in the early 1970s and manufactured by 3M Corp. The speech processor of the 3M/House device modulates band-limited speech signals with a 16 kHz sinusoidal waveform in nonlinear fashion to achieve dynamic range compression and sends it to the single electrode implanted inside the cochlea. Since speech processing strategy in these devices neglected the temporal response pattern 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/electrode cochlear implants that could enable different electrodes to stimulate different types of auditory nerves with different temporal features have been introduced as a better replacement to the single-channel/electrode cochlear implant devices.

Compressed Analog Speech Processor
One of the widely applied speech processing strategies in multichannel cochlear prostheses in the late 1980s was the compressed analog (CA) design, as shown in Fig. 5. In a CA processor, dynamic range compression of the sound signals is achieved by broadband automatic gain control circuitry. Fundamental and higher formants of the speech information are extracted by the bandpass filters. Depending on electrode threshold levels, the extracted features are applied to channel amplifiers and then sent to the cochlear electrodes. Since sound information is introduced to the auditory neurons simultaneously, better speech discrimination can be achieved [4, 5]. In CA processors, generally four speech formants are extracted. Also, it was found that the CA processing strategy works best for patients with low neural stimulus thresholds and with relatively good nerve survival.

Continuous Interleaved Sampling Processor
Problems of channel interaction inside the cochlea are addressed in the continuous interleaved sampling (CIS) processors by using interleaved nonsimultaneous stimuli as shown in Fig. 6 [4, 6]. Half-wave rectifier and lowpass filters are used to extract the signal envelope in specified channels. The speech envelope of each channel is applied to a nonlinear mapping function in which the dynamic range of the signal is compressed. Channel information is modulated with a biphasic pulse stream, including temporal offset. In the CIS processors, generally four to six channels are used to take advantage of additional electrodes. Clinical studies on human subjects [5] showed that CIS processors provide much better speech perception than CA processors. In some commercial cochlear implant devices, like the Clarion Multi-Strategy Cochlear Implant System from Advanced Bionics Corp., both CIS and CA processing strategies are used.

Feature-Based Speech Processors
The feature-based speech processor (FBSP), as shown in Fig. 7, is one of the modern multielectrode/multichannel cochlear implant speech processors [7-9]. The processor simply estimates five speech parameters: fundamental voice frequency (F0), first formant frequency (F1), first formant amplitude (A1), second formant frequency (F2), and second formant amplitude (A2). The dynamic range of the signal is compressed in the automatic

8. Spectral maxima sound processor.
gain control block. The fundamental frequency is extracted by using an envelope detector, which contains a full-wave rectifier and a 270 Hz lowpass filter. Then pitch information is extracted in the spectral estimator block. The first and second formant frequency, i.e., the pitch and amplitude, are extracted by two bandpass filters and peak detectors.

This processing technique was used in Nucleus Multi-Electrode Cochlear Implant devices, wearable speech processor (WSP), and mini speech processor (MSP) [9]. The MSP was a digital version of the WSP processor with extended and more accurate pitch-extraction capabilities. Notice that a zero-crossing detector, which does not give accurate results for noisy inputs, was used in extracting pitch information in the WSP processor while a repeated peak detection/differentiation function was used in the MSP processor. This processor is used in an intracochlear electrode array, which has 22 platinum rings spaced at 0.75-mm intervals [7] in Nucleus devices.

Spectral Maxima Sound Processor

The spectral maxima sound processor (SMSP) was first successfully used in 1989 [10, 11]. A block diagram of a modified version of the SMSP is shown in Fig. 8 [E]. The processor includes sensitivity control, a microphone preamplifier, and a sound compressor followed by 16 bandpass filters, full-wave rectifiers, and lowpass filters for analog signal processing. A scanning analog-to-digital converter (ADC) is used to convert band signals into digital form with 8-bit resolution, and digitized signals are stored into a first-in first-out (FIFO) memory. Digitized spectral information is processed by a microprocessor and the maximum amplitude or amplitudes of the entire speech spectrum is determined. Depending on the external control parameter values, such as loudness and the implantee's stimulus threshold levels, and the position of the spectral maxima, the microprocessor transfers the electrode numbers with stimulus levels to the data encoder. The data encoder converts data frames into pulse streams and sends them to the RF transmitter.

Speech processing and coding strategies in cochlear implant devices have advanced very quickly and have become commercially available during the last two decades. Performances of different cochlear implant devices are listed in the Fig. 9 [14]. Early modern cochlear implant devices, such as 3M/House and Nucleus WSP devices, were based on analog speech processing techniques. More recent devices started to use sophisticated fully digital or mixed analog/digital techniques to extract the speech features with high accuracy. High-accuracy feature extraction in the speech processor is benefited from the use of off-the-shelf, relatively low-power, low-cost, and high performance DSP chips such as the filterbank chip (NEC uPD7763D), the Hitachi 63A03YF microprocessor, or the TMS320E17 from Texas Instruments. The information listed in Fig. 9 clearly indicates that using mixed-signal speech processing techniques in cochlear implant devices can enhance the intelligent speech perception in hearing impaired people who receive cochlear implants. However, speech processors of current cochlear implant devices are usually worn on a belt because they are not compact enough to be worn behind or in the ear, and they. We have interest in making further contributions in lightweight, low-power speech processors for hearing-impaired people.

### Conclusion

According to the speech perception-rate data listed in Fig. 9, CIS and SMSP techniques are, and probably will be, the best speech processing strategies for multichannel/electrode cochlear implant devices. From packaging and power-consumption viewpoints, today's speech processing systems are very big and are, therefore, worn on the body and consume large electric power. Next-generation cochlear implant devices would be more compact, low-power products that would be worn behind or in the

<table>
<thead>
<tr>
<th>Device</th>
<th>Processor</th>
<th>Year</th>
<th>Sentence Recognition (% correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M / House</td>
<td>Single Channel</td>
<td>1980</td>
<td>0.5</td>
</tr>
<tr>
<td>Nucleus WSP</td>
<td>FBSP</td>
<td>1982</td>
<td>21.5</td>
</tr>
<tr>
<td>Nucleus WSP II</td>
<td>FBSP</td>
<td>1985</td>
<td>30.3</td>
</tr>
<tr>
<td>Nucleus MSP</td>
<td>FBSP</td>
<td>1989</td>
<td>58.0</td>
</tr>
<tr>
<td>Ineraid / MIT</td>
<td>CA</td>
<td>1992</td>
<td>32.3</td>
</tr>
<tr>
<td>Ineraid / RT1</td>
<td>CIS</td>
<td>1993</td>
<td>64.9</td>
</tr>
<tr>
<td>Nucleus Spectra 22</td>
<td>SMSP</td>
<td>1994</td>
<td>77.5</td>
</tr>
<tr>
<td>Clarion / ABC</td>
<td>CA / CIS</td>
<td>1996</td>
<td>77.0</td>
</tr>
<tr>
<td>MedEL / Combi</td>
<td>CIS</td>
<td>1996</td>
<td>83.9</td>
</tr>
</tbody>
</table>

**Fig. 9. Performance of Cochlear Implant Devices [14].**
It is clear that mixed-signal, high-density, and low-power design techniques are required to satisfy compactness, as well as low-power consumption features to realize intelligent speech sensation for the implantees. The especially critical design consideration of power supply lifetime and efficiency might be increased by using new promising technologies like microelectromechanical systems (MEMS).

Suat U. Ay is a Ph.D. student in the Electrical Engineering department at the University of Southern California. Dr. Fan-Gang Zeng is the Director of Auditory Perception Lab in the House Ear Institute. He also teaches graduate-level course at USC. Dr. Bing Sheu teaches in the Electrical Engineering department at USC. He is Chairman of the Colloquium & Publications for the Integrated Media Systems Center and the NSF/Engineering Research Center at USC.

References

BRAINBUSTER

In Bill's sock drawer are eight pairs of white socks and six pairs of red socks. (He's a snappy dresser.) How many socks does Bill have to pick at random to be sure he has a matched pair?

ANSWER TO LAST ISSUE'S BRAINBUSTER
The oarsman's velocity relative to the water is constant. Since he rowed away from the log upstream for an hour, he rowed downstream for an hour before he returned to it. During those two hours the log moved two km, so its velocity relative to the land was 1 km/hr.