Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, Hew York, Oct 17-20, 1999

A ROBUST SIGNAL PROCESSOR FOR COCHLEAR IMPLANTS

John K. Bates

Time/Space Systems
79 Sarles Lane, Pleasantville, NY 10570
jkbates@ieee.org

ABSTRACT

Cochlear implants are increasingly useful as prosthesis for severe hearing loss. However, improvements are needed in the speech intelligibility and robustness of these implants. The difficulties are related to the Gabor time-frequency limit that inhibits perception of stop consonants and diphones in speech, and induces susceptibility to interference from background sounds. This paper describes and demonstrates a new method of granular waveform decomposition that provides the improved time resolution and dynamic range needed to correct these problems.

1. INTRODUCTION

As aids for the profoundly deaf, cochlear implants are in relatively common use. Surgical implantation techniques have become well developed and their effectiveness has made implants sought after rather than objects of experimentation [1]. However, much remains to be improved, especially in signal processing. While the primary goals have been to improve robustness for speech communication, improvements are also needed in general auditory performance. This paper describes a novel signal processing approach that addresses these problems.

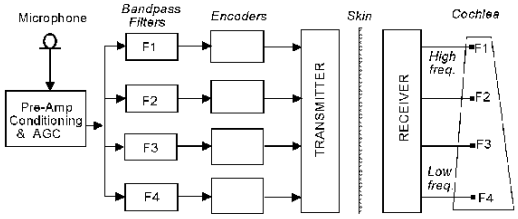

Figure 1. A typical four-channel cochlear implant

Cochlear implants aim to replace the functions of failed hair cells. This is done by splitting the acoustic signal into frequency bands that match neural frequency locations in the cochlea, where they make electrical contact with the cochlear nerve. For example, the functions of a typical cochlear implant having four channels are shown in Figure 1. (Commercially-available implants range from one to 24 channels.) Each filter output is encoded for transmission across the skin barrier by an amplitude-modulated subcarrier. Encoding modifies the signal into a form that creates the desired auditory response. This encoding generally involves ways to convert the waveform into pulsatile form so as to minimize cross-channel interference among the closely spaced wires of the implant. These wires then bypass the failed hair cells and make direct contact with the cochlear nerve. Neural functions behind the cochlea then combine the responses from the frequency-distributed channels into a perceived sound. A difficulty with current implants is that speech comprehension usually requires aid from cues supplied by visual contact. Also, users have little meaningful awareness of the acoustic environment. Thus, important research goals are to improve the comprehension of normally articulated and whispered speech and to provide better awareness of environmental sounds.

It is known that speech intelligibility depends largely on hearing diphones and stop consonants [2]. It is therefore necessary that the time resolution and dynamic range of the implant be sufficient to pass stop consonants and diphones. However, the transient response of conventional frequency-based processors is insufficient for speech transients due to the Gabor time-frequency limit. As a possible solution, a method using granular analysis that models signals using an atomic decomposition technique *w*as proposed by Goodwin and Vetterli [3]. While this method offers a way to improve temporal resolution, its implementation requires windowed time intervals for computing Gabor-type grain paramenters. The problem with this method is that time windows are too arbitrary for dealing with unpredictable events such as transients. Instead, what is needed is a granular method that is synchronous with temporal events within the waveform. The system to be described achieves this goal by decomposing the acoustic waveform into elementary granules having dimensions of time and amplitude that are synchronous with zero crossings. Moreover, each granule is assumed to be an independent event in time, and is therefore processed individually. Over the wide dynamic range of impulsive acoustic sources, this allows instantaneous response and recovery. In this way, the time resolution becomes sufficient for perceiving consonants and diphones and is even good enough for binaural azimuth localization of sources [4].

2. THE GRANULAR PROCESSOR

We now describe the method for applying waveform zeros to obtain high time resolution and dynamic range. It has been known that zeros completely define a waveform, and it has been shown that speech comprehension requires only first and

second-order zeros [5] [6]. The problem with this knowledge, however, has been to find ways to apply it. The solution lies in acknowledging that waveforms are composed of intermixed patterns that are embedded in sequences of zero-based granules. Recognizing these patterns is therefore the key to a successful auditory processor.

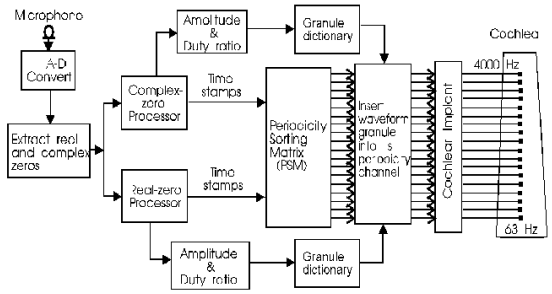

Figure 2. Configuration of granular signal processor

Figure 2 shows a block diagram of the system. The major functions are:

- Extract the time stamps of the real and complex zeros of the waveform.
- Use the time stamps to extract peak amplitudes and duty ratios that identify features of the waveform.
- Collect granule sequences into periodicity bands analogous with conventional filter banks.
- For each periodicity event, use amplitude and duty ratio to select a matching overdamped sinusoid from a dictionary of waveshapes.
- Send each granule waveshape to its assigned channel in the prosthetic.

The first step is to extract the zeros. Real-axis zero crossings identify the waveform periods while the complex (off-axis) zeros identify shape features of the waveform. Complex zeros are obtained from the log derivative of the waveform between real zeros so that the real and complex zeros maintain time synchrony needed for sampling the peak amplitude of each positive half wave. This method allows measuring absolute amplitude of each successive halfwave, thereby eliminating the time-response difficulties of conventional envelope-based automatic gain control. The duty ratio of each halfwave period is used to select a granule shape from a dictionary of overdamped sinusoids.

The next step uses a time-domain method to replace the frequency filter bank as follows: The waveform has now been reduced to a stream of intermixed granule sequences produced by multiple signal sources. A delay line (shift register) stores these sequences, retaining their time relationships. A time-filtering algorithm recognizes patterns of periodicity within the interleaved sequences of granules. Each granule arriving in the delay line is assumed to be an independent event, but *potentially* could be related to a sequence of previous granules in the delay line. We now have the ability to identify intermixed sequences, since we do not depend on knowing only the interval between successive events. Instead, we test

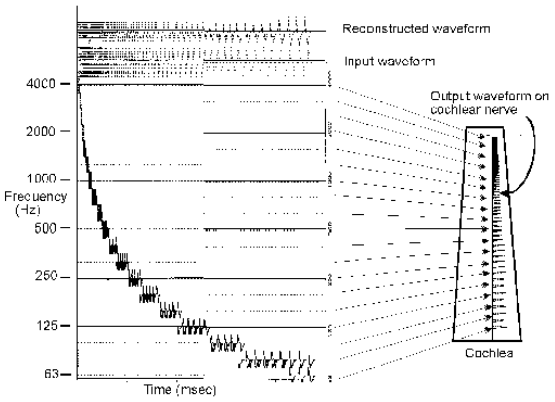

Figure 3. Reconstructed waveform of a stepped-sweep pulse signal from 4000 Hz to 63 Hz

simultaneously for *all* periodic intervals among sequences in the delay line.

The reason for this approach is that the waveform produced by the acoustic environment is *always* composed of multiple signal sources. This means that the stream of granules consists of intermixed unrelated sequences. In fact, even the human voice is the product of multiple unrelated sources [7]. For example, the periodicities of speech formants are only loosely synchronous with glottal pitch, and in whispered speech there is no synchronism at all. This is why the algorithm tests each arriving granule to see if it has any relationship to previous granules, a process not possible in windowed decompositions.

The periodicity sorting matrix (PSM), described in detail elsewhere [8] performs the foregoing functions. Briefly, it is composed of an array of logic gates connected to a tapped shift register delay line that propagates the input stream of granules. The spacings of the taps connected to the gates includes all possible time intervals in which three granules are separated by two equal intervals. As each granule enters the delay line, the matrix tests for a matching event. When such an event occurs, one of the logic gates responds, thereby identifying its periodicity channel. Recognition of false submultiples is prevented by the PSM's time-scaling method. Fortuitously, this time scaling has resulted in a logarithmic scale in which output periodicities correspond closely with the tonotopic frequency distribution of the cochlear nerve. Significantly the periodicity resolution also matches exactly the well-tempered musical scale. For the 19-channel implant configuration the range is seven octaves in one-third octave intervals. Operation of the PSM is illustrated in Figure 3. A test signal of pulses is swept from 4000 Hz to 63 Hz in 19 linear steps corresponding to the one-third octave periodicity channels of the PSM. The logarithmic pattern covers a seven-octave range.

Following the PSM, the periodicity and duty ratio information are combined to select a granule from a dictionary of overdamped sinusoids. These individual granules compose the dispersed waveform passed to the implant electrodes. Notice that each granule is independent in both time and space sequence, thereby providing the desired pulsatile format that minimizes cross-channel interference..

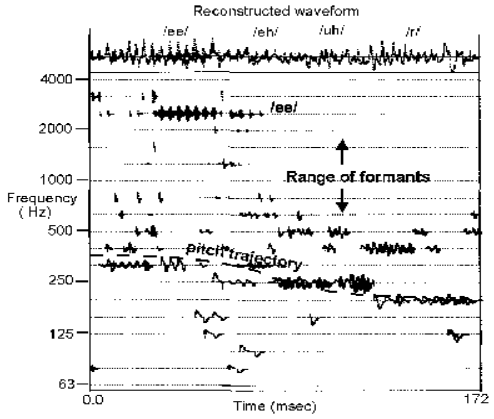

Figure 4. Granular processing for the utterance /ear/

3. EXPERIMENTS

Three requirements for a satisfactory cochlear implant were tested:

- Intelligible perception of both voiced and whispered human speech
- Perception of environmental background sound, especially impulses
- Robust speech perception against background sounds including other voices

To evaluate the processor against these requirements, tests were performed using whispered speech, music, multiple speakers and. some environmental sounds. Space does not permit showing all results here---we present one example of speech analysis and two kinds of acoustic transients. These examples characterize the essential functions needed to address the above requirements.

A difficulty in evaluating a cochlear implant processor is that it is not possible for an experimenter with normal hearing to "hear" exactly the perception that the spatially distributed granules of the implant device will induce in a patient. The best that can be done is to simulate the perception by constructing a waveform derived from the channelized granules.

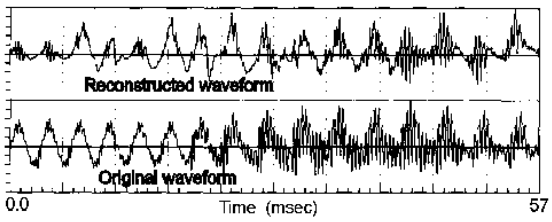

Figure 5. Comparison of the /ee/ phoneme of /ear/ with its original waveform

This reconstructed signal is then compared against the original sound. This is also the method used for evaluating filter-based cochlear implants [1]. Since auditory comparison is not feasible here, performance is evaluated by comparing temporal features.

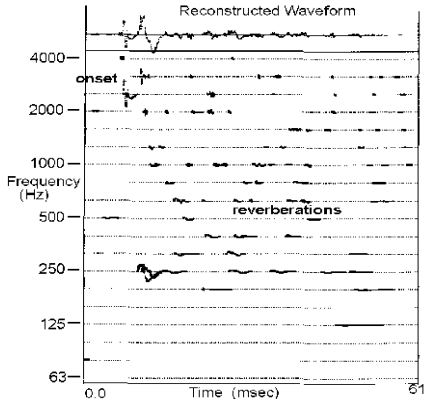

Figure 6. Analysis of acoustic pulse caused by exploding a toy balloon in a reverberant room

Quantitative methods based on mean-squared error are not applicable due to the ear's high tolerance for distortion. Figure 4 shows periodicities of interleaved formants and glottal pitch using the diphthong word, /ear/. Notice that the granules are sorted into periodicity channels in the order of their occurrence within the waveform. The periodicities include both pitch and formants along with fragments of periodicities caused by reverberations and artifacts. The display shows vowel transitions, starting with /ee/ which is characterized by a large frequency ratio between the F3 formant at 2600 Hz and F1 around 500 Hz. Then there is a downward glide of of formant periodicities of the vowels in the diphthong. While the reconstructed sound has a rough timbre due to the ragged envelope, the temporal features are located with accuracy so that the sound has good intelligibility and speaker recognition. Figure 5 compares the original and reconstructed phoneme /ee/ taken from /ear/, showing that the /ee/ formant is properly reproduced.

By analyzing and reconstructing two kinds of acoustic transients the next examples show how transients are localized in time. Figure 6 is a periodicity analysis of the sound of an

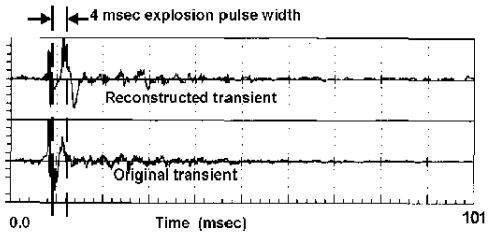

Figure 7. Comparing waveforms of exploding toy balloon

exploded toy balloon in a reverberant room. The distribution pattern of granules over time and periodicity illustrates the events that occur during and after the "pop." The explosion pulse is differentiated by acoustic propagation, leaving two sharp edges. Following the main pulse there is a progressive

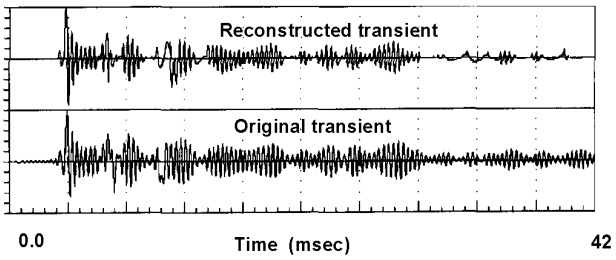

Figure 8. Transient waveform caused by a mechanical click

delay toward the lower periodicities caused by interactions of reverberations as they become increasingly distant from the point of the explosion. Figure 7 compares the shapes of the reconstructed and original waveforms. The essential time relation of the leading and trailing edges is maintained, although the trailing edge shapes are different. The reconstructed sound closely resembles the original. Figure 8 illustrates another type of transient from a mechanically generated click having a 3300 Hz resonance. This shows the decaying ringing waveform with the complex decaying envelope caused by reverberations.

4. DISCUSSION

The foregoing tests have illustrated the basic features of the zero-based granular processor. Analyzing the utterance /ear/ has demonstrated that granule sequences can be sorted into channels that are equivalent in function to a filter bank. However, unlike the band-limited phase-delayed waveforms of a filter bank, they are synchronous with the input signal. The "granulated" periodicity channels are inherently logarithmic on the same scale as that of the basilar membrane. It is significant that the time synchrony of the periodicity channels seems similar to the firing synchrony induced by normal hair cells.

In the examples of the acoustic transients, it is clear (1) that the time relationships of all waveform features are maintained, especially the onsets and offsets, and (2) the dynamic range of each granule is independent, limited only by the range of the input analog-to-digital converter. It is notable too that the transient tests showed that zero crossing detection does not bury the output in a mass of front-end thermal noise, despite operating at the zero axis. Thus, the granular method has good signal-to-noise response.

In the wider range of experiments that were done, there were a number of observations having general significance, including these:

- Rectified (halfwave) sounds are nearly indistinguishable from full wave processing of the reconstructed waveforms.
- Channelization of pitch and timbre seems to have little effect on their perception, and may have implications on improving our understanding of pitch and timbre.
- The perception of interrupted sequences was shown in analysis of whispers, multiple speakers, voice plus a pulse train, a male singer with orchestra, and combination tones. This might help to explain occluded speech, a phenomenon basic to the cocktail party effect.
- Time-difference methods for azimuth direction showed feasibility for binaural source location [4].

In summary, the implant processor reproduces speech that is intelligible under most conditions, even with background noise, for both voiced and whispered utterances. Recognition of speakers and non-speech sound sources is also good.. Perception of emphasis and intonation is normal. The examples of transients demonstrated the instantaneous response to onsets and waveform variations that are essential components of human speech, as well as in most environmental sounds.

Remaining objectives are to reduce the roughness of the reconstructed waveform and to translate code from the APL language into C so as to run in DSP signal processor chips. The next step is to test the system in an actual cochlear implant. Some long term development goals are (1) to explore possibilities of the granular processor for improving conventional hearing aids, and (2) to reduce size and power requirements using the programmable logic array (PLA) instead of the DSP. This could make the implant physically comparable with present hearing aids.

5. REFERENCES

[1] P.C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine*, pp 101-130, September 1998

[2] P. Tallal, Lecture on impairment of speech intelligibility due to deficient temporal perception, New York Academy of Sciences, January 10, 1995

[3] M. Goodwin and M. Vetterli, "Atomic decompositions of audio signals," *WASPAA97 Proceedings, 1997*

[4] J.K. Bates, "Modeling the Haas effect: a first step for solving the CASA problem," *WASPAA97 Proceedings, 1997*

[5] H.B. Voelker, "Toward a unified theory of modulation," Part I, Phase envelope relationships, *Proc. IEEE*, Vol. 63, pp 340-353, March 1966, and Part II, "Zero manipulation," pp735-755, May 1966

[6] J.C.R. Licklider and I. Pollack, "Effects of differentiation, integration, and infinite clipping upon the intelligibility of speech," *J. Acous. Soc. Am.*, Vol. 20, pp42-51, January 1948

[7] H.M. Teager and S.M. Teager, "A phenomenological model for vowel production in the vocal tract," *Speech Science: Recent Advances*, R.G. Daniloff, Editor, College Hill Press, San Diego, California, pp73-109, 1985

[8] J. K. Bates, *U.S. Patent No. 4,559,602*