MULTICHANNEL COCHLEAR IMPLANT SPEECH PROCESSING: FURTHER VARIATIONS OF THE SPECTRAL MAXIMA SOUND PROCESSOR STRATEGY

A. E. VANDALL, BAENG; J. M. HARRISON, DIPCOMPENG; J. M. HUIGEN, BSC, DAUD; K. PLANT, BSC, DAUD; G. M. CLARK, PhD, FRACS

INTRODUCTION

The spectral maxima sound processor (SMSP) was first developed at the University of Melbourne in 1989. A full description of the SMSP has been given by McDermott et al. In short, the SMSP utilizes an ear-level microphone to measure acoustic sound pressure. A 16-channel band-pass filter bank is used to analyze the sound spectrum at discrete time intervals. Each of the 16 filters is assigned to one of the 16 intracochlear electrodes according to frequency. Within each time interval the six channels with the largest band-pass filter amplitudes are selected and used to stimulate six corresponding electrodes in quick succession. The current implementation of the SMSP differs from the original in that a digital signal processor is used in place of the analog filter bank and the microprocessor. The filter bank has been implemented with a discrete Fourier transform. Also, the input dynamic range has been improved by increasing the resolution of the analog-to-digital converter from 8 to 12 bits.

Studies exploring variations of the standard SMSP strategy have been conducted in the past. These studies looked at the effects on speech perception for variations such as modifying the number of electrodes selected within each time interval from six to a value between four and eight, increasing the constant rate of stimulation from 250 to 400 Hz, and sharpening of spectral peaks prior to selection of the six largest filter bank outputs. The results of these studies showed that all three variations had minimal effect on speech perception in quiet, but that increasing the number of electrodes selected for stimulation to eight, or increasing the rate of stimulation, may have advantages when listening in background noise. More recently, a number of new studies were conducted exploring other variations of the SMSP strategy, the results of which are the topic of this paper.

Three studies were conducted. The first explored the use of different temporal ordering arrangements of the six selected stimuli. Previous psychophysical studies have shown that for electrical stimuli presented in groups in quick succession, the temporal order of the stimuli can affect percepts such as pitch, timbre, and loudness. The purpose of this study was thus to examine whether temporal order could also have a significant effect on speech perception. Three different ordering arrangements were examined. First, the standard SMSP amplitude order (largest to smallest) was used, whereby the stimulus corresponding to the filter bank channel with the largest output amplitude is delivered first to the implant, followed by the next largest, and so on, with the smallest being delivered last in time. Second, a reverse amplitude order (smallest to largest) was examined. Finally, a tonotopic order (basal to apical, or highest to lowest frequency) was also examined. In this case the stimulus corresponding to the highest frequency is delivered first, and the lowest frequency stimulus is delivered last in time.

The second study investigated the use of the speech envelope to control the stimulation rate independently for each channel. It was hypothesized that discrimination of speech from background noise may be improved by employing a stimulation rate that was related to some speech feature, in this case the speech envelope or the voicing pitch for voiced speech. Essentially, the outputs of each filter bank channel were analyzed over time at 1-millisecond (ms) time intervals, and stimuli were generated to coincide with any temporal peaks encountered in the envelope. These temporal peaks usually correspond to the fundamental voicing period for voiced speech. However, because the fundamental frequency for male speakers can be as low as about 80 Hz, the stimulation rate produced by such a scheme could be fairly low. Low stimulation rates can be undesirable, particularly for perception of speech features that change rapidly, such as glides. Therefore, in the intervals between successive stimuli corresponding to temporal peaks, additional stimuli were inserted at a constant rate of 400 pulses per second (pps) per electrode. This produced a fairly constant stimulation rate that was adjusted in phase so that the peaks of the envelope remained locked in phase to the temporal peaks in the signal envelope. For unvoiced speech, a random pattern of temporal peaks exists; thus, this strategy would effectively deliver stimuli at a high random rate (limited by the strategy to 500 pps). This whole process was conducted independently for each channel. Thus, different channels could effectively deliver stimuli at different rates. Note that because stimuli cannot be generated simultaneously, some offset in the stimulation timing can arise. Priority was given to the channel with the largest amplitude.

The final study explored the effect of emphasis of transient speech features such as speech onset, speech transitions, glides, plosives, etc. These transient speech features can be defined by a rapid change of amplitude (over a time period of about 40 ms) of the speech signal in a particular frequency band. It is hypothesized that perception of these transient features and perception of speech may be improved by amplifying the signal during the periods of transition in each frequency band. The strategy implemented was identical to the SMSP strategy up to the point at which the six largest filter bank outputs were selected. At this point, the channel outputs were low-pass filtered by a second-order 35-Hz low-pass filter to remove any fundamental modulation. Any transitions over time in the filtered channel outputs were detected and amplified. The amount of gain applied was proportional to the change in amplitude over time of the detected transition. Emphasis occurred during the time interval of transition, and
Weekly evaluation sessions were conducted over a 3-week period. SMSP strategies (TESM N and TESM T) were evaluated, the standard SMSP strategy. Two versions of transient-emphasis SMSP strategies were then used for selection of the six largest channel outputs. This process was carried out independently for approximately 40 ms was employed to extend the duration of each filter bank channel. The outputs of the transient emphasis were then used for selection of the six largest channel outputs. Processing then continued as it normally would in the standard SMSP strategy. Two versions of transient-emphasis SMSP strategies (TESM N and TESM T) were evaluated, the latter having twice the emphasis gain of the former.

also for some time afterward. A decay time constant of approximately 40 ms was employed to extend the duration of emphasis. This process was carried out independently for each filter bank channel. The outputs of the transient emphasis were then used for selection of the six largest channel outputs. Processing then continued as it normally would in the standard SMSP strategy. Two versions of transient-emphasis SMSP strategies (TESM N and TESM T) were evaluated, the latter having twice the emphasis gain of the former.

**PROCEDURE**

A total of six subjects took part in these studies. All were postlingually deaf adults who were implanted with the Nucleus Minisystem-22 implant, manufactured by Cochlear Pty Limited. All were experienced users of the SMSP processor and experienced research subjects. The speech material used throughout these studies consisted of open-set monosyllabic words and open-set sentences presented in quiet and in background noise (multitalker babble). The speech material was recorded and presented hearing-alone to the subjects in a soundproof room via a loudspeaker at a signal level of 70 dB sound pressure level. No feedback was given during or after any of the tests.

**Study 1.** Four subjects took part in this study. All subjects commenced the study with the standard SMSP processor employing a constant rate of stimulation. Weekly evaluation sessions were conducted over a 3-week period. Subjects 1 and 2 then took home the tonotopically ordered basal to apical (B-A) scheme, and subjects 3 and 4 used the amplitude ordered smallest to largest (S-L) scheme for a period of 6 weeks, with the first week being used for training, leaving 5 weeks for evaluation. The subjects then alternated to the remaining scheme for another 6-week period (1 week of training, 5 weeks of evaluation). All subjects then switched back to the standard amplitude L-S ordered scheme for a final 3-week evaluation period. In each evaluation session, two lists of Consonant-Nucleus-Consonant (CNC) words were presented in quiet and IEEE sentences at three SNRs. B) For study 2, comparing SMSP to speech envelope-derived timing SMSP strategy (SMET). Used CNC words in quiet and IEEE sentences at +10 dB SNR. C) For study 3, comparing SMSP to two versions of transient-emphasis SMSP strategy (TESM N, TESM T). Used CNC words in quiet and BKB sentences at +5 dB SNR.

**Mean results.** See text for abbreviations. A) For three temporal ordering schemes: amplitude order largest to smallest (L-S), amplitude order smallest to largest (S-L), and tonotopic order basal to apical (B-A). Used CNC words in quiet and IEEE sentences at three SNRs. B) For study 2, comparing SMSP to speech envelope-derived timing SMSP strategy (SMET). Used CNC words in quiet and IEEE sentences at +10 dB SNR. C) For study 3, comparing SMSP to two versions of transient-emphasis SMSP strategy (TESM N, TESM T). Used CNC words in quiet and BKB sentences at +5 dB SNR.
list of CNC words was presented in quiet and one list of Bamford-Kowal-Bench (BKB) sentences (50 words per list) was presented in noise at an SNR of +5 dB.

RESULTS

The effect of learning over the duration of the first two studies was examined by comparing the results obtained in the first 3 weeks and the last 3 weeks of evaluation with the standard SMSP strategy. Analysis of variance on these results was conducted for each study and condition. No significant differences were observed; thus, learning effects over the duration of the study were ignored and the results for the first 3 and last 3 weeks with the standard SMSP strategy were pooled together in our analysis.

Study 1. The effect on speech comprehension, for the three temporal ordering schemes evaluated, are shown in the Figure, A. For each subject, the first columns of each panel show the results for the CNC word tests scored as percentage words correct. All three ordering schemes provided similar word comprehension results. One-way and two-way ANOVA of the results across the four subjects showed no significant differences between the three schemes for word results. The next group of columns in each panel shows the results for the sentence tests scored as percentage of key words correct for presentations at three SNRs in order of decreasing SNR. An overall slight improvement in scores for the tonotopic B-A scheme can be observed. One-way and two-way ANOVA across the four subjects at each SNR showed no significant differences. However, when data were pooled across the four subjects and across the five SNRs (three-way ANOVA), a significant difference was observed (amplitude L-S 49.2%, amplitude S-L 50.1%, tonotopic B-A 56.6%, p = .02, F = 3.99, df = 59). Further analysis (three-way ANOVA) between each pair of schemes showed a significant difference between the tonotopic B-A and amplitude L-S schemes (p = .013, F = 6.36, df = 39) and also between the tonotopic B-A and amplitude S-L schemes (p = .022, F = 5.41, df = 39). No significant difference was observed between the amplitude L-S and S-L schemes. No significant interactions were observed between schemes and subjects, or schemes and SNRs; thus, we can be reasonably confident that the difference is a result of scheme alone. All four subjects indicated a preference for the tonotopic B-A scheme. This scheme was thus employed in the standard SMSP strategy (in place of the amplitude L-S ordered scheme) for the remainder of the studies conducted.

Study 2. Results for the study comparing the speech envelope-derived timing SMSP strategy (SMET) and the standard SMSP strategy are shown in the Figure, B. The results for each subject for the CNC word tests in quiet scored as percentage words and phonemes correct are shown in the first columns of each panel. The results for the IEEE sentence tests conducted at a SNR of +10 dB are shown in the next columns. Subject 6 was not evaluated with the sentences because of previous experience with the material. One-way and two-way ANOVA of the results across the three subjects for each test condition showed no significant differences.

Study 3. The results for the study comparing the two transient-emphasis SMSP strategies (TESM N and TESM T) with the standard SMSP strategy are shown in the Figure, C. The mean results for each subject and the mean of the subject means for the CNC word tests in quiet are shown in the first columns of each panel. All three strategies provided similar word comprehension results for subjects 3, 4, and 5. One-way ANOVA of the results for these subjects showed no significant differences among the three strategies for the word tests. Subject 6, however, performed significantly poorer with the TESM T strategy compared to the SMSP strategy. One-way ANOVA of the results for subject 6 showed a significant difference between the SMSP strategy and the TESM T strategy (SMSP 60.7%, TESM T 44.0%, p = .023, F = 12.7, df = 1). Two-way ANOVA of the results across the four subjects showed no significant differences among the three strategies for word results. The mean results for each subject and the mean of the subject means for the BKB sentences presented at an SNR of +5 dB are shown in the next columns of each panel. An improvement of approximately 10% to 20% for both the TESM N and TESM T strategies compared to the SMSP strategy can be observed for all subjects. One-way ANOVA of the results showed no significant difference in individual results. However, two-way ANOVA of the results across the four subjects did show a significant difference (SMSP 44.1%, TESM N 55.5%, TESM T 59.7%, p = .02, F = 4.31, df = 11). Further analysis (two-way ANOVA) between each pair of strategies showed a significant difference between the SMSP and TESM T strategies (p = .16, F = 6.35, df = 7) and between the SMSP and TESM N strategies (p = .05, F = 3.93, df = 7). No significant difference was observed between the TESM N and TESM T strategies. No significant interactions were observed between the strategies and subjects or between the strategies and SNRs.

DISCUSSION

Study 1. The results indicate that the temporal order of stimuli that are presented in groups in quick succession can affect speech comprehension. In our study it was found that the tonotopically ordered scheme, which presented stimuli in order of highest to lowest frequency, provided some perceptual advantage over the amplitude-ordered schemes, at least for the sentence test conducted over a range of SNRs. One possible explanation for this difference is that during voiced passages of speech, the order of the selected stimuli remains fairly constant for the tonotopic B-A scheme, whereas it tends to vary more with the amplitude-ordered schemes because of variations in the amplitudes of each of the selected stimuli over the duration of the vowel. Thus, for the tonotopic B-A-ordered scheme, a voiced passage of speech would tend to sound more periodic or smoother in pitch and thus more distinct from unvoiced speech than it would for the amplitude-ordered schemes. Also, voiced speech may be more distinct from background noise that is unvoiced or has multiple voices. Another possible explanation for the differences stems from the fact that given that the time taken to stimulate the selected stimuli was about 2 to 4 ms, which is on the order of the time taken for sound waves to propagate from the base to the middle turn of the cochlea in normal-hearing people, it would be reasonable to assume that the spatial-temporal stimulus pattern of the tonotopically ordered stimulation scheme would be more like the natural timing of waves arriving at their place of firing in normal hearing than the amplitude-
**INTRODUCTION**

The cochlear implant is a device used to provide the sensation of sound to those who are profoundly deaf. Generally, it consists of a wearable speech processor, a headset transmitter, an implanted receiver-stimulator module, and an electrode array, which together provide an electrical representation of the speech signal to the residual nerve fibers of the peripheral auditory system.

The state-of-the-art speech processors for multichannel cochlear implants perform loudness mapping using microcontroller and look-up tables, and channel selection using spectral maxima strategy. As the input values for the microcontroller are quantified to eight bits and so are the pulse amplitudes and widths, it needs a large amount of memory to store the look-up tables for a multichannel implant. For channel selection, the spectral maxima strategy selects the m

**REFERENCES**


**ARTIFICIAL NEURAL NETWORK-BASED CHANNEL SELECTION AND LOUDNESS MAPPING**

R. J. WANG, MSC; M. A. JABRI, PhD

From the Systems Engineering and Design Automation Laboratory, Sydney University Electrical Engineering, Sydney, Australia.

We present in this paper artificial neural network techniques for implementing loudness mapping and "smart" channel selection for cochlear implant systems.

For loudness mapping, a multilayer perceptron (MLP) is trained to perform the mapping for each channel according to threshold and comfort levels. It is shown that good accuracy mapping can be performed by a very simple MLP architecture. For channel selection, we propose a neural network-based method that can make "smart" selection. We describe and report results for the case in which 6 channels are to be selected from 18. The neural network-based selection system is trained on a multispeaker labeled speech database and tested on a database of different speakers and spoken sentences. Compared with methods used by leading cochlear implant systems, our approach produces significantly better results, and it is easy to implement in the speech processor of the cochlear implant system.

**CONCLUSION**

The results of these three studies have shown that some variations of the SMSP strategy can provide improved speech comprehension, particularly when one is listening to continuous speech in the presence of background noise. In the first study we found that the temporal order of stimuli can affect speech perception. It was observed that frequency ordering of the stimuli, as opposed to amplitude ordering, provided better speech comprehension results for the sentence material presented. As a consequence of this work, the standard SMSP strategy was modified to use a tonotopic B-A stimulus order. The results of the second study, in which the stimulation rate was synchronized to temporal peaks in the speech envelope of each channel, did not show any significant differences. It is difficult to draw any conclusions from this study; however, it can be said that the SMSP strategy appears to be fairly robust to changes in the stimulation rate scheme. In the final study we observed that emphasis (ie, increasing the amplitude) of transient features in speech in a multiband system can provide improved comprehension results for continuous speech presented in background noise at fairly low SNRs. Given the good results of this preliminary study, continued research with this strategy and similar strategies is being conducted.

**REFERENCES**

Author/s: Vandali, A. E.; Harrison, J. M.; Huigen, J. M.; Plant, K.; Clark, Graeme M.

Title: Multichannel cochlear implant speech processing: further variations of the spectral maxima sound processor strategy

Date: 1995


Persistent Link: http://hdl.handle.net/11343/27445

File Description: Multichannel cochlear implant speech processing: further variations of the spectral maxima sound processor strategy

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.