TEMPORAL CODING OF SPEECH INFORMATION FOR COCHLEAR IMPLANT PATIENTS

J. B. MILLAR, BSc, PhD; L. F. A. MARTIN, BSc, Dip Aud; Y. C. TONG, BE, PhD; G. M. CLARK, PhD, FRACS

A modified speech-processing strategy incorporating the temporal coding of information strongly correlated with the first formant of speech was evaluated in a long-term clinical experiment with a single patient. The aim was to assess whether the patient could learn to extract information from the time domain in addition to the time domain cues for voice excitation frequency already received from the initial strategy. It was found that the patient gained no significant advantage from the modified strategy, but there was no disadvantage either, and the patient expressed a preference for the modified strategy for everyday use.

In a multichannel digitally controlled cochlear prosthesis there are three domains of control of the stimulating signal: the selection of the site of stimulation (spatial domain), the selection of the instant of stimulation (temporal domain), and the selection of the strength of stimulation (energy domain). Responses to stimulus variation in the energy domain are primarily correlated with the psycho-physical dimension of loudness. Those resulting from stimulus variation in the spatial and temporal domains are strongly correlated with the acoustic frequency-related psycho-physical dimensions of pitch and/or sharpness.

In the initial speech-processing strategy proposed for the University of Melbourne prosthesis, a single acoustic frequency-related parameter was assigned to each of the spatial and temporal domains. A measurement strongly correlated with the second formant frequency (F2) was used to select the site of stimulation, and a value inversely proportional to the frequency of glottal excitation (Fo) determined the interval between successive stimulation pulses. Since that initial proposal, two methods of including additional frequency-related information in our speech-processing model have been investigated: first, by coding two formants in the spatial domain, and second, by presenting both formant and voice excitation information in the temporal domain, which is the topic of this paper.

STIMULATION WITH COMPLEX TEMPORAL PATTERNS

Broadly speaking, two methods for structuring the temporal pattern of stimulation are used in cochlear prostheses. In one method, the rate or frequency of stimulation is controlled by an extracted acoustic variable; in the other, the temporal structure of some derivative of the speech pressure wave is replicated. Use of several such derivatives has been reported, ranging from a simple gain-compressed version of the speech wave, through systems that employ a variety of prefiltering, amplitude modulation, and pulse depletion schemes. Although the detailed speech reception performances of patients using these approaches to speech processing vary considerably, it appears that for a few patients temporal information in addition to simple time-varying rate information may be used.

This paper reports on a group of clinical experiments conducted as a prelude to explorations of the utility of temporal patterning in cochlear implant patients. A new speech-processing strategy was evaluated by comparing results of a balanced sequence of tests during which the initial strategy was alternated with the new one. The aim was to assess whether the patient could learn to extract additional information from the time domain, and whether such learning would cause a deterioration of reception of time domain cues already received from the initial strategy.

NEW SPEECH-PROCESSING STRATEGY

The new strategy used the same processes of electrode selection and current level assignment as the initial strategy.4 Each electrode was assigned to a portion of the F2 range. They were stimulated one at a time, according to the instantaneous F2 value, with a current level corresponding to the energy level at the output of the F2 filter. The strategies differed in the temporal pattern of pulses used to stimulate the electrode. In the initial strategy (Fo/F2), the selected electrode was stimulated at a pulse rate determined by an Fo detector. In the new strategy (Fo, F1/F2) electrodes were stimulated by pulses generated from the raw speech wave such that the interpulse interval was dictated by the major peaks of the waveform, which for vowels are strongly correlated with the first formant frequency (F1).

Calibration of Stimuli. Random samples (duration approximately 52 ms) of the pulse stream produced by both strategies were taken during the presentation of the words “heed, heard, and hard” by both speakers. As expected, the pulse rate of the Fo/F2 strategy was substantially constant across samples from the same speaker, but the pulse rate of the Fo, F1/F2 strategy was higher than that for Fo/F2 for each sample tested, and was monotonically related to the first formant of the vowel.

The time spent stimulating each electrode was sampled during the vocalic nucleus of each of 11 [h-vowel-d] words spoken by both speakers. For each speaker, 12 or 13 electrodes were stimulated by the full range of vowels, and for each vowel nucleus between one and seven electrodes were stimulated. The duration-weighted mean electrode position for each vowel nucleus followed the expected pattern of second formant frequency.

Threshold, comfortable loudness levels, and electrode assignments were the same for each strategy. Although higher loudness was reported for the Fo, F1/F2 strategy, it was considered by the patient to be within the range of comfortable loudness. The patient could detect a difference in pitch between the two strategies but was not given any explanation for this. She was told that the purpose of the experiment was to compare and assess two different strategies.

Patient History. The clinical history, preoperative evaluation, and initial postoperative assessment of the patient
involved in this study are reported elsewhere where she is identified as patient HA1 and patient P2. At the beginning of the experimental period she had had several months of daily experience with the initial strategy. During the 2-month experimental period she had similar exposure to the new strategy via a modification to her wearable processor.

This patient was chosen for the current experiment because she had very good pulse rate discrimination for stimulation at rates up to 600 pps (patient P02). This higher than average performance is comparable to other reports of the chosen patient, had been informally tested with the new strategy and seemed to benefit from the additional information when performing a speech-tracking task.

Another patient, previously reported as P03, whose pulse discrimination was good but slightly inferior to that of the chosen patient, had been informally tested with the new strategy and seemed to benefit from the additional information when performing a speech-tracking task.

Assessment Tests. The assessment process used four speech reception tasks. Each session started with a live-voice speech-tracking test with lipreading administered by a female speaker. This was followed by one or two tests administered by either a male or female speaker. These tests were chosen from a vowel identification test, a consonant identification test, and a prosody identification test. Each of these three tests were presented three times by each speaker and in a balanced order across sessions. The assessment was performed in 15 biweekly sessions, in a simulated quiet domestic environment. Speech material was presented via loudspeaker at 75 to 80 dB at a 1.5 m from the patient who received input via an ear-level microphone.

The speech-tracking task was included to provide both training by means of the constant feedback that this method involves, and assessment of the use of auditory cues together with top-down linguistic constraints. The vowel test material comprised a closed set of 11 vowels in an [h-vowel-d] context. The prosody test material comprised six sentences of identical phonetic segmental content. The patient had to detect which of three words was stressed and whether the utterance was a question or a statement. The consonant test material comprised nonsense syllables (consonant-vowel or vowel-consonant) clustered into 11 closed set subtests each of which contained either voiced or voiceless consonants in either initial or final position.

RESULTS

A two-way analysis of variance was performed on an arcsine transformation of the percent correct scores for the vowel test. There was no significant difference between strategies, but there was between speakers (p < 0.01, F = 16.0). Neither the interaction between speakers and strategies nor the session was found to be significant.

No significant difference between processing strategies could be found for either speaker using y²-square analyses on the "totally correct" scores in the prosody test. The "totally correct" scores and "word-stress correct" scores all exceeded the 95% confidence level for nonrandom responses and indicated some learning for the female speaker. However, the nonrandom criterion was exceeded only in one session for one speaker (male) for the "question-statement correct" score.

The reasons for this relatively poor question-statement detection were investigated in two ways. First, the nature and distinctiveness of the signal being received by the patient for this set of material was studied using analytical techniques. Second, the less than optimum, although clinically realistic, listening environment was improved and additional training was given. Improved scores resulted but still without any significant advantage to either strategy.

Percent correct scores for each of the 11 groupings of consonants were averaged over the three presentations, arcsine transformed, and subjected to a three-way analysis of variance. No significant difference was found for different speakers or strategies or any interactions between factors, but the consonant groupings were significantly different (p < 0.01, F = 5.062). This result occurred because consonant-vowel and vowel-consonant transitions that involve more movement of stimulation across the electrode array are easier to perceive correctly.

The mean words per minute (wpm) scores for the speech-tracking test were graphed separately for the first five sessions (when the strategy was known to the talker), and the last nine (when the talker did not know the strategy; Figure). The overall mean scores reveal a 3 wpm superiority of the Fo, F1/F2 strategy over the Fo/F2 strategy. This difference was found not to be significant using a Wilcoxon matched-pairs signed-ranks test. However, a mildly significant trend (p < 0.05) was found for the Fo, F1/F2 scores, which increased over all 14 sessions at a rate of 0.66 wpm/session. No significant trend with respect to the overall mean was found for the Fo/F2 scores.

DISCUSSION

The aim of this experiment was to discover whether the inclusion of F1 information in the time domain would lead to a superior speech processor when tested using standard clinical tests. During the course of this experiment, it was found that the patient gained no significant advantage from the new strategy, but that there was no disadvantage either. The patient expressed a preference for the new strategy for everyday use. An information transfer analysis of the vowel test showed that a similar amount of information was conveyed by each strategy, with only 9% to 21% of F1 information being conveyed compared to 33% to 52% of F2 information. The confusion matrices for both
strategies indicate a similar error pattern. It therefore appears that the patient was not aware of the additional information being presented, or of how to use it. The slight learning observed in the speech-tracking results may indicate that slow learning of relevant information for phonetic distinctions was occurring for this patient when higher order linguistic cues were available.

The overall result of this experiment directs attention to the need for specific "transition" training when adding new information to a familiar speech-processing strategy. The patient was aware of the difference in pitch of the two strategies and was able to use the F0 information just as effectively in the new strategy as in the old. Hence there exists the potential for at least limited use of F1 information presented by temporal pattern if the patient can be trained to use for speech perception the information she is aware of at a psychophysical level.

The ability of patients to detect pulse patterns rather than simply pulse rates is a central issue. Psychophysical studies to this end have recently been designed and we hope to be able to report our preliminary results soon.

ACKNOWLEDGMENT — We thank Dr Peter Seligman of Nucleus Ltd for his hardware design for the real-time presentation of F0 and F1 information in the time domain.

REFERENCES
Author/s:
Millar, J. B.; Martin, L. F. A.; Tong, Y. C.; Clark, Graeme M.

Title:
Temporal coding of speech information for cochlear implant patients

Date:
1987

Citation:

Persistent Link:
http://hdl.handle.net/11343/27254

File Description:
Temporal coding of speech information for cochlear implant patients

Terms and Conditions:
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.