SPEECH PROCESSING STRATEGIES IN AN ELECTROTACTILE AID FOR HEARING-IMPAIRED ADULTS AND CHILDREN

R.S.C. Cowan, P.J. Blamey, J.Z. Sarant, K.L. Galvin, and G.M. Clark.

Human Communication Research Centre
University of Melbourne.

ABSTRACT - An electrotactile speech processor (Tickle Talker) for hearing-impaired children and adults has been developed and tested. Estimates of second format frequency, fundamental frequency and speech amplitude are extracted from the speech input, electrically encoded and presented to the user through eight electrodes located over the digital nerve bundles on the fingers of the non-dominant hand. Clinical results with children and adults confirm that tactually-encoded speech features can be recognized, and combined with input from vision or residual audition to improve recognition of words in isolation or in sentences. Psychophysical testing suggests that alternative encoding strategies using multiple-electrode stimuli are feasible. Preliminary results comparing encoding of consonant voiced/voiceless contrasts with new encoding schemes are discussed.

INTRODUCTION

In the simplest tactile devices, a single vibrator positioned on the body presents a representation of the speech amplitude envelope. Efforts to increase the speech information presented through tactile devices have included provision of additional channels (see reviews by Weisenberger, 1989; Lynch, Oller & Eilers, 1989), and presentation of a processed speech signal, providing specific speech feature information such as fundamental frequency (Boothroyd & Hnath-Chisholm, 1988) or vowel formant frequencies (Blamey & Clark, 1985).

A device which combines these two approaches is the multichannel electrotactile speech processor, or Tickle Talker, developed at the University of Melbourne (Blamey & Clark, 1987). The device successfully exploits the fingers as the locus of stimulation, benefiting from the fingers' acknowledged superior tactual sensitivity and larger cerebral representational area compared with other body parts, while not limiting everyday functions of the hand. Results of speech perception testing after training demonstrate that the device can provide significant benefits for profoundly hearing-impaired children and adults (Cowan et al., 1989, 1990, in press). This paper will discuss the approaches to speech processing which have been employed in the device, and future directions to enhance performance of the device.

DESCRIPTION OF THE DEVICE

The Tickle Talker consists of: an ear-level or lapel microphone; a speech processor/stimulator unit; and an electrode handset with associated cabling.

The front-end speech feature extraction hardware is similar to that employed in the University of Melbourne/Cochlear Pty. Ltd. multichannel cochlear prosthesis (Seligman, 1987). An estimate of speech waveform amplitude (EA) is obtained through the output of a peak detector, followed by a 35Hz low-pass filter. An estimate of second formant frequency (EF2) is obtained by a zero crossing detector operating on the output of a 800-4000Hz filter. An estimate of voice fundamental frequency is obtained from a zero crossing detector operating on the output of a 270Hz low-pass filter. Following the zero crossing detector, the output is scaled and used to trigger a pulse generator to initiate a stimulus.

The output of the speech processor circuitry is then electrically encoded in the stimulator circuit (Table 1).

Table 1. Speech processing scheme (e=electrode number).

Speech Feature	Electrical Parameter	Electrical Sensation
EF2	electrode	stimulus position
EF0	pulse rate	stimulus quality
EA	pulse width	stimulus strength
High Frequency Amplitude	e8 pulse width	stimulus strength e8

Amplitude is encoded as pulse width of the 1.5mA constant current biphasic pulses, and sensed by the user as changes in the strength of the stimulus. EF2 is encoded as electrode selected, and sensed by the user as the position of stimulation. Sounds with low EF2 values are felt on the index finger, whereas high frequency sounds such as /s/ are sensed on the little finger. EF0 is electrically encoded as pulse rate, and sensed by the user as differences in the quality of sensation. High EF0 rates are perceived as "smooth", whereas lower EF0 rates are "rougher". Pulse rate is a scaled function of EF0, such that a fundamental frequency of 300Hz resulted in a pulse rate of 150Hz. Pulse rate is scaled on the basis of psychophysical testing which indicated poorer perception of pulse rate differences at rates above 250Hz (Blamey & Clark, 1987). In addition, a 4000-10000Hz band-pass filter was fitted to improve detection of high frequency speech energy. Output from this filter, when present at sufficiently high amplitude, was directed to electrode 8.

The electrode handset consists of eight $50mm^2$ stainless steel electrodes, housed in a support ring, and positioned on the skin surface overlying the digital nerve bundles on the sides of the four fingers of the non-dominant hand. A $9cm^2$ common electrode is located on the underside of the wrist. Electrical stimulation of nerve bundles is a novel feature of the device, providing larger dynamic ranges and a more pleasant quality of sensation than electrocutaneous stimulation of nerve endings at other body sites.

The stimulus used is a constant current (1.5mA) biphasic square waveform, with equal current and duration in both positive and negative phases. A short (100µsec) gap separates the two stimulus phases, during which there is no current flow. The use of biphasic pulses ensures that there is no net current flow to the user. Only one electrode is activated at any given time. Individual threshold and comfortable sensation levels (i.e. pulse widths) are set for each electrode using an interactive computer program. Typical dynamic ranges are from 10µsec to 1msec.

EVALUATIONS OF PRESENT ENCODING STRATEGY

1. ABX Feature Recognition Testing

Tactual perception of speech feature contrasts was evaluated using a test battery consisting of eleven subtests, each containing 24 presentations of two-alternative forced-choice feature contrasts, and one subtest employing 24 three-alternative forced-choice contrasts (Plant, 1989). The test format was ABx (or ABCx), and a variety of phoneme contrasts were included in each subtest. All consonant contrasts were in the initial position. Mean percentage scores for eight adult subjects are shown as Table 2.

Table 2. Mean percentage scores for feature contrast discrimination.

No.	Feature Contrast	Mean Score (%)	No. Above Chance (n=8), ($p<0.05$)
1	syllable number/stress (/b/)	78	8
2	syllable number/stress (/m/)	76	6
3	vowel length (vowel+final stop)	84	8
4	vowel length (vowel+final nasal)	80	8
5	vowel formants (F1/F2)	77	7
6	vowel formants (F2)	75	7
7	initial consonant voicing	60	3
8	consonant manner (m/b, n/d)	70	4
9	consonant manner (p/f, t/s, v/b, z/d)	74	7
10	consonant manner (t/tʃ, d/dʒ)	64	4
11	consonant manner (n/z, m/v)	75	6
12	consonant manner (s/t/st)	69	8

Results for syllable number/stress and vowel length ranged from 76-84%, indicating efficient encoding of time/intensity speech information. Scores for vowel formant contrasts (77%, 75%) suggest that vowel formant frequency information is also efficiently encoded in the tactual signal. Consonant manner of articulation scores varied across subtests, but overall, the results suggest that amplitude envelope and spectral frequency cues to consonant manner are efficiently encoded and perceived. In particular, high frequency consonant fricative information is effectively transmitted, as shown by the scores for subtests 9, 11 and 12. Consonant voicing contrasts were not as well perceived, with mean percentage scores of only 60%.

These results suggest that improvements in encoding of consonant voicing and to some extent consonant manner could be made. The lower scores for initial voicing may reflect the inability of some subjects to perceive differences in pulse rate, as reported in previous psychophysical testing (Blamey & Clark, 1987). Accurate encoding of consonant voicing is especially important to hearing-impaired adults and children since this information is not visible from lipreading.

DEVELOPMENT OF ALTERNATIVE ENCODING STRATEGIES

1. Psychophysics of Multiple Electrode Recognition

Previous psychophysical testing had shown that electrode position was the best-recognized tactual parameter in the encoding scheme (Blamey & Clark, 1987). Therefore, use of additional electrode positions would be the obvious method of improving transmission of voicing information. Initial pilot studies with a matrix-type display employing more than one electrode along the same digital nerve bundle, and with an extended linear display using both hands found these approaches to be impractical. Multiple electrodes located along the same side of the finger were not easily distinguished as separate stimuli, and the logistics of twin sets of cables would reduce practicality of the device for everyday usage. An alternative approach was use of multiple-electrode stimulus patterns delivered through the existing eight electrodes.

Identification of multiple-electrode electrotactile stimuli delivered through the Tickle Talker was evaluated for nine adult subjects. Subjects were presented with four sets of stimuli, consisting of either single electrodes, electrode pairs, electrode triplets or all combinations of single and paired electrodes. Each set consisted of 5 random-order presentations of each possible stimulus combination (i.e. 5x8 single electrodes, 5x28 pairs, 5x56 triplets, 5x36 singles+pairs). Subjects were asked to identify the electrode position(s) activated in each test stimulus. A more detailed description of methodology and results has been reported in Cowan et al. (in press).

Mean percentage correct scores for the nine subjects for each of the stimulus sets is shown in Table 3.

Table 3. Mean percentage and information transmission scores for nine adult subjects using an electrotactile speech processor.

Stimulus Pattern	Stimulus Combinations	Mean Score (%)	Stimulus Information (Bits)	Information Transmitted (Bits)
Single Electrodes	8	98	3.0	2.84
Paired Electrodes	28	62	4.74	2.99
Triple Electrodes	56	32	5.68	2.84
Singles+Pairs	36	74	5.13	3.88

Electrode position identification scores decreased from 98% for single electrodes, to 62% for electrode pairs and 32% for electrode triplets. Analysis of variance showed the difference between mean scores on singles, pairs and triplets to be significant ($F_{(2,24)}=93.63$, $p<0.001$). Mean score for the combined single+paired electrode set was 74%. Information transmission scores for the four stimulus sets are also shown in Table 3. While potential stimulus information increased with additional electrodes, actual information transmitted was 2.84 bits for single electrodes, 2.99 bits for electrode pairs and 2.84 bits for triplets. However, information transmission increased to 3.88 bits in the case of the single+paired electrodes.

While the results suggest that use of two or three electrode multiple stimuli is not a feasible solution, an encoding scheme using combinations of single and paired electrodes could potentially provide additional information to that provided by the current single electrode encoding scheme.

2. Alternative Encoding Strategies

Several feasible encoding strategies using combinations of single and paired electrodes are summarized in Table 4. A common feature of all three strategies is use of the single vs paired electrode dimension to signal the voiced/unvoiced contrast. While the VF2 and UVF2 schemes provide only a voiced/voiceless signal in addition to the information previously encoded (F2 scheme), the F1F2 provides additional vowel first formant frequency and amplitude information by splitting the F1 frequency range between electrodes 1 and 2.

Table 4. Alternative encoding strategies.

Encoder	Speech Feature	Electrical Parameter	Electrical Sensation
VF2	EF2	electrode select (e2-7)	stimulus position (e2-7)
	EA	pulse width (e2-7)	stimulus strength (e2-7)
	EF0	pulse rate	stimulus quality
	high frequency A	pulse width e8	stimulus strength e8
	voiced/voiceless	e1 if voiced	voiced: two stimuli (e1+eF2) unvoiced: single stimulus (eF2)
F1F2	EF2	electrode select (e3-7)	stimulus position (e3-7)
	EA2	pulse width (e3-7)	stimulus strength (e3-7)
	EF1	electrode select (e1-2)	stimulus position (e1-2)
	EA1	pulse width (e1-2)	stimulus strength (e1-2)
	EF0	pulse rate	stimulus quality
	high frequency A	pulse width e8	stimulus strength e8
	voiced/voiceless	e1 or e2 if voiced	voiced: two stimuli (eF1+eF2) unvoiced: single stimulus (eF2)
UF2	EF2	electrode select (E2-7)	stimulus position (e2-7)
	EA	pulse width (E2-7)	stimulus strength (e2-7)
	EF0	pulse rate	stimulus quality
	high frequency A	pulse width e8	stimulus strength e8
	voiced/voiceless	e1 if unvoiced	voiced: single stimulus (eF2) unvoiced: two stimuli (e1+eF2)

Results of pilot studies with four adult subjects on the ABx feature contrast subtests described previously is shown as Table 5. Mean scores for the four adults are shown for syllable length (1), vowel duration (4), vowel F1 (5), vowel F2 (6) and initial consonant voicing (7) subtests.

Table 5. Mean percentage feature contrast scores using four encoding strategies.

Encoder	Feature Contrast Subtest					Overall
	1 (%)	4 (%)	5 (%)	6 (%)	7 (%)	(%)
F2	71	75	74	75	60	71
VF2	89	76	76	64	62	73
F1F2	89	76	89	67	63	77
UF2	96	75	81	69	75	79

As shown, mean percentage scores for syllable length and vowel duration subtests were higher for the three multiple electrode strategies as compared to the previous F2 strategy, suggesting improved encoding of speech time/intensity information. Results for vowel F1 frequency contrasts were also increased for the F1F2 strategy. This would be expected since additional F1 frequency information is presented. Scores for vowel F2 contrasts were however reduced in each of the multiple electrode encoding strategies. This suggests that reducing the spread of EF2 information to fewer electrodes has resulted in some overlap of electrode positions between previously distinct vowels. Future studies will examine whether alterations in electrode frequency boundaries can improve encoding of EF2 in the multiple electrode strategies. Improvements in initial consonant voicing were evident only in the F1F2 and UF2 strategies. Overall mean scores for the F1F2 and UF2 strategies were also higher than for F2, suggesting that improvement in tactual encoding of speech feature contrasts has been realized through use of multiple electrode strategies. Future studies will be required to evaluate whether these improvements in tactual encoding, achieved with multiple electrode strategies, are reflected in improved speech discrimination scores for trained subjects on measures of conversational speech.

REFERENCES

Blamey, P.J. & Clark, G.M. (1985) "A Wearable Multiple-electrode Electrotactile Speech Processor for the Profoundly Deaf", J.Acoust.Soc.Am. 77, 397-405.

Blamey, P.J. & Clark, G.M. (1987) "Psychophysical Studies Relevant to the Design of a Digital Electrotactile Speech Processor", J.Acoust.Soc.Am. 82, 116-125.

Boothroyd, A. & Hnath-Chisholm, T. (1988) "Spatial, Tactile Presentation of Voice Fundamental Frequency as a Supplement to Lipreading: Results of Extended Training with a Single Subject", J.Rehab.Res.& Dev. 25, 51-56.

Cowan, R.S.C., Alcantara, J.I., Whitford, L.A., Blamey, P.J. & Clark, G.M. (1989) "Speech "Perception Studies Using a Multichannel Electrotactile Speech Processor, Residual Hearing, and Lipreading", J.Acoust.Soc.Am. 85, 2593-2607.

Cowan, R.S.C., Blamey, P.J., Galvin, K.L., Sarant, J.Z., Alcantara, J.I. & Clark, G.M. (1990) "Perception of Sentences, Words and Speech Features by Profoundly Hearing-impaired Children Using a Multichannel Electrotactile Speech Processor", J.Acoust.Soc.Am. 88, 1374-1384.

Cowan, R.S.C., Blamey, P.J., Sarant, J.Z., Galvin, K.L., Alcantara, J.I., Whitford, L.A. & Clark, G.M. "Role of a Multi-channel Electrotactile Speech Processor in a Cochlear Implant Program for Profoundly Hearing-impaired Adults", Ear and Hearing (in press).

Cowan, R.S.C., Blamey, P.J., Sarant, J.Z., Galvin, K.L. & Clark, G.M. "Perception of Multiple Electrode Stimulus Patterns: Implications for Design of an Electrotactile Speech Processor", J.Acoust.Soc.Am. (in press).

Lynch, M.P., Oller, D.K. & Eilers, R.E. (1989) "Portable Tactile Aids for Speech Perception", Volta Rev. 91(5), 113-126.

Plant, G.L. (1989) "A Comparison of Five Commercially Available Tactile Aids", Austr.J.Audiol. 11(1), 11-19.

Seligman, P. (1987) "Speech-processing Strategies and their Implementation", Annals Otol.Rhinol. & Laryngol. 96(1), Suppl. 128, 71-73.

Weisenberger, J.M. (1989) "Tactile Aids for Speech Perception and Production by Hearing-impaired People", Volta Rev. 91(5), 79-100.