Veterans Administration

Journal of Rehabilitation Research and Development Vol. 26 No. 1
Pages 15–24

Speech perception using combinations of auditory, visual, and tactile information

Peter J. Blamey, PhD, R.S.C. Cowan, MSc, DipAud, Joseph I. Alcantara, BSc, DipAud, Lesley A. Whitford, BSc, DipAud, and G.M. Clark, PhD, FRACS
Department of Otolaryngology, University of Melbourne, The Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria 3002, Australia

Abstract—Four normally-hearing subjects were trained and tested with all combinations of a highly-degraded auditory input, a visual input via lipreading, and a tactile input using a multichannel electrotactile speech processor. The speech perception of the subjects was assessed with closed sets of vowels, consonants, and multisyllabic words; with open sets of words and sentences, and with speech tracking. When the visual input was added to any combination of other inputs, a significant improvement occurred for every test. Similarly, the auditory input produced a significant improvement for all tests except closed-set vowel recognition. The tactile input produced scores that were significantly greater than chance in isolation, but combined less effectively with the other modalities. The addition of the tactile input did produce significant improvements for vowel recognition in the auditory-tactile condition, for consonant recognition in the auditory-tactile and visual-tactile conditions, and in open-set word recognition in the visual-tactile condition. Information transmission analysis of the features of vowels and consonants indicated that the information from auditory and visual inputs were integrated much more effectively than information from the tactile input. The less effective combination might be due to lack of training with the tactile input, or to more fundamental limitations in the processing of multimodal stimuli.

Key words: *auditory/tactile/visual input, lipreading, multichannel electro-tactile speech processor (Tickle Talker), multimodal stimuli, normally-hearing subjects, speech perception.*

Address all correspondence and reprint requests to: Peter J. Blamey, Ph.D., The University of Melbourne, Department of Otolaryngology, The Royal Victorian Eye and Ear Hospital, 32 Gisborne Street, East Melbourne, Victoria 3002, Australia.

INTRODUCTION

The purpose of this study was to investigate combinations of auditory, visual, and tactile modalities for speech recognition. The auditory and visual modalities have been studied extensively, individually, and in combination. The auditory input is the natural primary source of speech information, and it is well known that lipreading can be used to achieve a limited level of speech recognition: see, for example, Jeffers and Barley (16). The combined auditory-visual input has been studied for its basic psychological interest (11,19) and for its practical importance to normally-hearing listeners in noise (28) and to hearing-impaired listeners (29). The tactile modality differs from the others because speech information is not normally available in a tangible form. A small number of highly-trained persons have demonstrated that speech can be recognized by the Tadoma method in which the "listener's" hand is placed on the speaker's face to feel the movements and flow of air during the production of speech (24). More usually, a speech processor has been used to convert the acoustic speech waveform to a tangible form. Tactile speech processors have been investigated since the pioneering work of Gault in 1924 (15) as possible information sources for totally deaf people who do not

benefit from conventional hearing aids. A recent review of this research has been published by Sherrick (26). At least one device has been studied as a unimodal input (7), but usually tactile devices have been used as a supplement to lipreading for totally deaf people (9,12,13,17,18,22,23,25,27). Thus the visual (V), auditory (A), and tactile (T) senses have been used individually, and in the combinations AV and VT, to enhance the speech communication potentials of hearing-impaired people.

There are two further combinations, AT and AVT that are just beginning to be studied (12,17,18). The clinical situations in which these combinations would be relevant, are the cases of severely and profoundly hearing-impaired people who gain some benefit from conventional hearing aids, but not enough to achieve a high level of comprehension. The present investigation was designed as an initial evaluation of the usefulness of tactile information in the AT and ATV conditions, and included equivalent investigations of the A, V, T, AV, and VT conditions for comparative purposes. These combinations of sensory inputs have seldom been studied in controlled circumstances with the same set of subjects. The specific questions addressed by the study are whether the combined modalities present more information than the individual modalities, and whether the A, V, and T information combine equally effectively.

There are two major obstacles to be overcome before an investigation of this type can be undertaken. The first is the provision of a tactile speech processor that is capable of presenting useful information. It is possible that redundant information may improve performance in a combined condition, but it is more likely that information not already provided by the auditory input will have a greater effect. Single-channel tactile devices provide mostly amplitude envelope and fundamental frequency information that is likely to be present in the auditory signal for a person with residual low-frequency hearing (23). Multiple-channel tactile devices, which are capable of providing spectral information, have been designed and tested (9,13,22,25,27). The first-referenced device, the "Tickle Talker," was used in the present investigation. The second problem is the availability of experienced users of the tactile device, since it is obvious that any person will need an extensive period of training before the newly-presented tactile information can become associated with the meaningful perception of speech. This contrasts strongly with the considerable experience most persons have with auditory and visual speech perception. In the present study, four normally-hearing listeners, who had been trained with the tactile device in a previous study, participated as subjects. This situation was far from ideal because of the limited experience of the subjects. Further studies are in progress with hearing-impaired users of the Tickle Talker who will be able to wear the device for everyday communication outside the laboratory. These studies are more difficult to control with regard to the hearing loss and experience of the subjects than the investigation reported here.

METHODS

Subjects and training

Four normally-hearing subjects took part in this experiment. Each was a female tertiary-level student who was paid for her participation. Their ages ranged from 20 to 27. Each subject reported that she had normal vision, but no formal tests were carried out. Each subject had electrotactile thresholds and comfortable stimulation levels that fell within the normal range.

All four had previously been trained with the Tickle Talker over a 6-month period, using lipreading, but no auditory signal (9). In the earlier study, each subject was trained for a total of 70 hours, using the speech-tracking procedure of De Filippo and Scott (10) in the V and VT conditions, and closed sets of nonsense syllables and words in the V, T, and VT conditions. At the conclusion of this 6-month period, the subjects showed significant differences between the V and VT conditions on open-set word and sentence recognition tests, on closed-set vowel and consonant tests, and also in speech tracking rates. Scores for recognition of closed sets of vowels and consonants were also well above chance in the T condition. **Table 1** summarizes the mean scores obtained by these four normally-hearing subjects in the earlier study.

A gap of two months occurred between the end of the earlier study and the start of the present one. The subjects did not use the Tickle Talker at all during this time. No specific training was given to the subjects for the present study, but some im-

provement in scores was observed for those tests that were repeated.

Evaluation methods and materials

The four subjects were tested in sessions lasting one hour or two hours with a short break in the middle. Each subject attended 15 to 20 sessions in a 2-month period. In the majority of the sessions, the subjects were tested using different sensory modalities with 10 minutes of speech tracking, followed by one or two closed-set recognition tasks. The modalities A, V, T, AV, AT, VT, and AVT, were tested in rotation, in a different order for each subject. Speech tracking was not done in the T condition because of the difficulty of this task.

Three different closed-set tasks were used: vowel recognition, using the words *hid, head, had, hud, hod, hood, heed, heard, hard, who'd, hoard*; consonant recognition, using the consonants /p,b,m,f,v,s,z,n,g,k,d,t/ in an /a/-consonant-/a/ context; and a set of 12 words containing monosyllables, trochees, spondees, and polysyllabic words (MTSP): *fish, ball, shoe, table, pencil, water, airplane, toothbrush, popcorn, elephant, Santa Claus,* and *butterfly,* proposed by Erber (14). Each closed-set task consisted of a block containing four of each stimulus in a randomized order. Results were obtained for three blocks of vowels or consonants in each condition. Each subject scored 100 percent for the MTSP tests in the V condition, so two blocks of results were collected for the A, T, and AT conditions, omitting those that included a visual component. The tests were presented with live voice, and feedback was given after each item. The speaker for all of the above tests was an Australian male, previously unknown to the subjects, who had not been involved in the training or testing for the earlier study. The speaker was aware of the condition being tested. Results were obtained for six 10-minute sessions of speech tracking in each condition except T.

In the final sessions, the subjects were tested with the open-set Bench, Kowal, and Bamford (BKB) sentence test (1), and the open-set Consonant-Nucleus-Consonant (CNC) word test (21) in each condition. These tests were recorded on videotape, using another Australian male speaker. A different test list was used for each condition, and the order of testing the conditions was balanced across the four subjects.

Table 1.
Mean scores[a] for the 4 subjects in an earlier study (9) of speech recognition in the visual, tactile, and visual-tactile conditions.

Test	Condition		
	V	VT	T
11 vowels (%)	89	99	75
12 consonants (%)	66	95	40
CNC words (%)[b]	50	63	[c]
CID Sentences (%)	47	63	[c]
Speech Tracking (wpm)	34	51	[c]

[a]Percent correct or words per minute.
[b]The CNC words were scored phonemically.
[c]Not tested.

Auditory input

A degraded auditory input was provided to the subjects who were seated in a sound-attenuating booth and could not hear the direct signal from the speaker's voice. In the live-voice testing situation, the speech signal was picked up by microphone about 20 cm from the speaker's lips. The signal was amplified, and then filtered with a digital elliptic filter with 7 poles and 6 zeroes, and a cut-off frequency of 400 Hz. The slope of the filter skirt was greater than 70 dB per octave. The filtered signal was then amplified again to 80 dBA peak level and mixed with white noise at a level of 70 dBA. The signal was presented binaurally to the subjects through headphones. The white noise (without the filtered speech) was presented in the V, T, and VT conditions to mask the direct voice signal, which was reduced by the sound-attenuating booth. The measured attenuation was 45 dBA. In the A, AT, AV, and AVT conditions, the white noise also had the effect of masking quiet sections of the filtered speech signal and high-frequency components that were not completely removed in the filtering process. The acoustic signal was chosen to provide a very crude simulation of a severe hearing loss and the filter frequency was adjusted to obtain a score of 40 to 50 percent in a prior test of the vowels and consonants with a listener who was not a subject in the study.

In the recorded tests, amplified microphone signal was replaced by audio output of the video recorder.

Visual input

The visual signal was provided via a double-glazed window in the sound-attenuating room for the live-voice testing. The speaker's face was well lit by lamps from both sides of the face, and lighting was turned off on the listener's side of the window to avoid reflections in the glass. The total distance between speaker and listener was approximately 1 m.

In the recorded tests, the visual signal was presented with a 48 cm color television monitor at a distance of about 1.5 m. The speaker's head was shown completely, and occupied about 90 percent of the vertical extent of the screen.

Tactile input

The tactile signal was provided via the Tickle Talker, a multiple-channel electrotactile speech processor. This device has been described in detail previously (3). The Tickle Talker used a speech coding scheme that was originally devised for a multiple-channel cochlear implant (6,8). The wearable speech processor produced estimates of the fundamental frequency, EF_0, the second formant frequency, EF_2, and the amplitude, EA, of the speech signal. The method used to estimate F_0 was full wave rectification, followed by low-pass filtering at 270 Hz and a zero crossing detector. The value of EF_0 was then scaled linearly, so that a frequency of 250 Hz produced a stimulation pulse rate of 150 pps. In the case of unvoiced sounds, this circuit produced a series of pulses with random time intervals between them. EF_2 was derived by a zero crossing detector following a high-pass filter that was shaped so that the second formant predominated over the other formants for most vowels. The EF_2 range was divided into 8 regions, corresponding to the 8 electrodes worn by the subjects. There was one electrode on each side of each finger (excluding the thumb) of the left hand, and a common electrode on the left wrist. Each electrical pulse (at the scaled EF_0 rate) was applied between the wrist electrode and one of the finger electrodes, according to the value of EF_2 at the time the pulse was applied. The frequency boundaries between the electrodes were 900, 1125, 1350, 1575, 1800, 2400, and 3300 Hz. Although the second formant does not usually extend as high as 3300 Hz, the output of the EF_2 circuit could exceed this value for sounds such as /s/ and /z/, which include intense high-frequency components. These components are not second formant resonances, but still provide useful information to the subjects. The amplitude estimate, EA, controlled the duration of the 1.5 mA biphasic constant current pulses that were applied between the selected finger electrode and the wrist electrode. A 30 dB range of EA was compressed into the range from threshold to "maximum comfortable level" (MCL) for each electrode.

The electrode array was designed to produce stimulation of the digital nerve bundles by using electrodes on the surface of the fingers immediately over the bundles. Earlier empirical investigation had suggested that stimulation of nerve bundles produced a more comfortable sensation than stimulation of nerve endings or transducers of the somatosensory system (3). The wrist electrode was much larger in area than the finger electrodes, so that electric current densities were not sufficient to produce stimulation at the wrist.

Detailed psychophysical measures of the electrotactile stimuli produced by the Tickle Talker, including thresholds, maximum comfortable levels, intensity discrimination, pulse rate discrimination, and electrode recognition, have been published previously (3). Thresholds and MCLs were measured for each electrode at regular intervals throughout the present study, by adjustment of the pulse duration, with a knob that the subject controlled directly. Subjects were instructed to choose an MCL that they would be quite comfortable with if stimulation continued for several minutes. When initial values had been measured, a stimulus that swept across the electrodes was used to achieve an equal subjective intensity at the MCL, and at a point midway between threshold and MCL on each electrode. The allowable range of pulse duration for the Tickle Talker was 10 μs to 1 ms. There was some variation between subjects within this range, and typical dynamic ranges were 5-10 dB between threshold and MCL. The sensations produced by the electrical stimulus, controlled by the speech processor, were such that loud sounds produced stronger sensations. Higher-pitched voices produced higher pulse rates, which were perceived as smoother sensations. The highest EF_2 values caused stimulation by electrode 8 on the little finger, while the lowest EF_2 values caused stimulation by electrode 1 on the index finger. The subject's task was to interpret these dynamic patterns of stimulation as phonemes and

words. Because this is not a naturally-acquired skill, the proficiency of the subjects can be expected to be determined partly by their experience and previous training.

RESULTS

The mean scores obtained for the different tests in each condition are shown in **Table 2**. For each test, a 2-factor analysis of variance was carried out, using condition and subject as the factors. For the vowel, consonant, and MTSP tests, the scores for separate blocks of data were used as repeated measures in the analysis. Similarly, word-per-minute scores for separate 10-minute sessions were used as repeated measures in the analysis of the tracking data. Since each subject was tested only once in each modality with the BKB sentences, the mean square term for the interaction of subject and modality was used as the error term for the analysis. Every analysis of variance (ANOVA) showed a highly significant variation among the mean scores for the different modalities and their combinations: $F(6,56)=61.5$, $p<.001$ for vowels; $F(6,56)=146.9$, $p<.001$ for consonants; $F(2,12)=51.8$, $p<.001$ for MTSP; $F(6,18)=110.3$, $p<.001$ for CNC words; $F(6,18)=68.1$, $p<.001$ for BKB sentences; and, $F(5,120)=457.7$, $p<.001$ for tracking. There were highly significant differences between the subjects for the vowels: $F(3,56)=7.92$, $p<.001$; consonants: $F(3,56)=12.9$, $p<.001$; and, speech tracking: $F(3,120)=68.9$, $p<.001$. The difference between subjects for the CNC words was less significant: $F(3,18)=3.97$, $p<.05$. The analyses showed no significant differences between subjects for the BKB sentences and MTSP test. The interaction between subject and condition was significant for tracking: $F(15,120)=5.4$, $p<.001$; and consonants: $F(18,56)=1.87$, $p<.05$; but not for vowels or the MTSP tests.

For each test, the mean scores for individual modalities and their combinations were compared using the Newman-Keuls procedure with a 95 percent confidence level for the criterion level. This led to the following relations between the scores:

For vowels,

$$A < T < AT < (V,VT)$$
$$V < (AV,AVT) \quad [1]$$

For consonants

$$T < (V,A) < (AT,VT) < (AV,AVT) \quad [2]$$

For MTSP,

$$T < (A,AT) < V \quad [3]$$

For CNC words,

$$T < (A,AT) < V < VT < (AV,AVT) \quad [4]$$

For BKB sentences,

$$T < (A,AT,V,VT) < (AV,AVT) \quad [5]$$

For tracking,

$$(A,AT) < (V,VT) < (AV,AVT) \quad [6]$$

In order to take a more detailed look at the combinations of cues provided by the three modali-

Table 2.
Average scores[a] for the 4 subjects in the auditory, visual, tactile, and combined conditions.

Test	Condition						
	T	A	V	AT	VT	AV	AVT
11 Vowels (%)	54	42	80	63	87	90	93
12 Consonants (%)	29	48	44	56	56	91	91
MTSP (%)	63	94	100	95	[b]	[b]	[b]
CNC words (%)[c]	12	30	50	36	59	84	84
BKB sentences (%)	1	27	34	27	39	93	92
Tracking (wpm)	[b]	11	21	12	22	75	77

[a]Percent correct or words per minute.
[b]Not tested.
[c]The CNC words were scored phonemically.

ties, the vowel results were analyzed in terms of the percentage of information transmitted for a number of features. This method was described in detail by Miller and Nicely (20), who applied it to a study of auditory consonant confusions. **Table 3** shows the classification of vowels according to the three features: duration, F_1 and F_2. The vowels were classified on the basis of data for average male speakers given by Bernard (2). The "total" column indicates that consideration of each vowel as a group by itself yields a value for the total information transmitted in each condition. The feature groupings allow the evaluation of more specific types of speech information in each condition. **Table 4** shows the percentage of information transmitted for each vowel feature in each condition, calculated from the confusion matrix for the four subjects together.

Table 3.
Groups of vowels used in the information transmission analysis of vowel confusions. In each column, vowels with the same number are grouped together for the analysis of the feature indicated.

Vowel	Feature			
	Total	Duration	F_1 Frequency	F_2 Frequency
heed	1	1	1	1
heard	2	1	2	2
hard	3	1	3	2
who'd	4	1	1	2
hoard	5	1	2	3
hid	6	2	1	1
head	7	2	2	1
had	8	2	3	1
hud	9	2	3	2
hod	10	2	3	3
hood	11	2	2	3

In brackets after each percentage for a combined modality is the value predicted from the values for individual modalities. The prediction was made assuming that each modality independently contributed a proportion of the information, and that an error occurred in the combined modality only if the speech feature was incorrectly perceived in *both* of the individual modalities. For example:

$$1 - P_{AV} = (1 - P_A)(1 - P_V) \qquad [7]$$

was used to predict the proportion of information transmitted in the AV condition, P_{AV}, from the proportion transmitted in the A and V conditions, P_A and P_V. In this case, the proportion of information incorrectly perceived is $(1\text{-}P_{AV})$ in the combined modality. In the individual modalities, the proportions of information incorrectly perceived are $(1\text{-}P_A)$ and $(1\text{-}P_V)$. Since the information provided by each modality is assumed to be statistically independent, the probability of an error in *both*

Table 4.
Percentage of information transmitted for the vowels in the auditory, visual, tactile, and combined conditions. The values in brackets for the combined modalities are predicted from the values for individual modalities using Equation [7].

Condition	Feature			
	Total	Duration	F_1 frequency	F_2 frequency
A	43	91	33	24
V	75	71	76	75
T	43	49	28	51
AV	86 (86)	100 (97)	85 (84)	82 (81)
AT	56 (68)	96 (95)	40 (52)	50 (63)
VT	82 (86)	81 (85)	81 (83)	80 (88)
AVT	91 (90)	100 (99)	89 (88)	90 (89)

Table 5.
Groups of consonants used in the information transmission analysis of consonant confusions. In each row, consonants with the same number are grouped together for the analysis of the indicated feature.

Feature	Consonant											
	b	p	m	v	f	d	t	n	z	s	g	k
Voicing	1	2	1	1	2	1	2	1	1	2	1	2
Nasality	1	1	2	1	1	1	1	2	1	1	1	1
Affrication	1	1	1	2	2	1	1	1	2	2	1	1
Duration	1	1	1	1	1	1	1	1	2	2	1	1
Place	1	1	1	1	1	2	2	2	2	2	3	3
Visibility	1	1	1	2	2	3	3	3	3	3	3	3
Amplitude envelope	1	2	3	1	4	1	2	3	1	4	1	2
High F_2	1	1	1	1	1	1	2	1	2	2	1	2

modalities is the product $(1\text{-}P_A)$ $(1\text{-}P_V)$. This formula has been shown to provide a good description of combined auditory-visual perception of nonsense syllables by cochlear implant users (4). Similar equations were used to calculate predicted values for the AT, VT, and AVT combinations.

Note that the values in **Table 4** observed for the AV and AVT conditions were all greater than, or equal to, the predicted values. The non-parametric Wilcoxon signed-rank test was applied to these differences, indicating that the observed values were significantly greater than the predicted values ($n=8$, Wilcoxon rank sum $=0$, $p<.01$). For the AT and VT conditions, the observed values were all less than the predicted values, with the exception of the proportion of duration information in the AT condition. The Wilcoxon test indicated that observed values were less than predicted values ($n=8$, rank sum $=1$, $p<.01$).

A similar analysis of the consonant results was

Table 6.
Percentage of information transmitted for the consonants in the auditory, visual, tactile, and combined conditions. Values in brackets for combined modalities are predicted from the values for individual modalities using Equation [7].

Feature	Condition						
	A	V	T	AV	AT	VT	AVT
Total	54	53	29	91 (78)	59 (67)	61 (67)	91 (85)
Voicing	79	7	11	91 (80)	78 (81)	15 (17)	94 (83)
Nasality	91	27	43	100 (93)	100 (95)	48 (58)	100 (96)
Affrication	66	70	29	95 (90)	70 (76)	87 (79)	96 (93)
Duration	38	58	80	94 (74)	74 (88)	81 (92)	95 (95)
Place	15	80	20	84 (83)	23 (32)	80 (84)	84 (86)
Visibility	24	100	12	100 (100)	29 (33)	100 (100)	100 (100)
Amplitude envelope	84	24	25	96 (88)	84 (88)	36 (43)	96 (91)
High F_2	27	74	59	84 (81)	52 (70)	83 (89)	86 (92)

Journal of Rehabilitation Research and Development Vol. 26 No. 1 Winter 1989

carried out. **Table 5** shows the classification of the consonants by feature. The first five features were used by Miller and Nicely (20) and are based mainly on articulation of the consonants. The visibility feature is based on the three groups of consonants commonly distinguished by lipreaders (16). The last two features were used by Blamey *et al.* (4,5) to describe the information available to cochlear implant users in a similar experiment. They are based on the amplitude and F_2 frequency parameters estimated by the speech processor.

Table 6 shows the percentage of information transmitted for each consonant feature in each modality, together with values for the combined modalities predicted from Equation [7]. Observed scores for AV and AVT were greater than predicted scores, with the exception of place for AVT ($n=9$, Wilcoxon rank sum $=0$, $p<.01$ for AV; $n=9$, rank sum $=3$, $p<.01$ for AVT). With the exception of affrication, observed VT scores were less than predicted ($n=9$, rank sum $=6$, $p<.025$). Observed AT scores were also less than predicted, except for the nasality feature ($n=9$, rank sum $=4$, $p<.025$).

DISCUSSION

First, it should be noted that each of the individual modalities A, V, and T produced scores that were well above chance for the closed-set vowel, consonant, and MTSP tests. In the case of the CNC words and BKB sentence tests with open response sets, non-zero scores were obtained for each modality. The scores for the T modality might be attributable to chance for these two tests. For vowel recognition, the T modality produced a higher score than the A modality, but for all other tests, T was the lowest scoring modality. V was the highest scoring unimodal condition for all tests except consonant recognition. The unimodal scores show that each modality was capable of conveying useful speech information, although the subjects were unable to use the tactile information effectively for open-set tasks.

The scores on the vowel and consonant tests for the V, VT, and T modalities were considerably lower than those found in the previous study (9), using the same subjects and the same evaluation methods. The difference may be attributable to the use of a different test speaker, but more probably to the period of two months when the subjects were not trained at all. The results of **Table 1** were obtained immediately after a 70-hour training program. In contrast, the results of **Table 2** were obtained during an evaluation period spread over two months. It appears likely that the subjects lost some of the skills that were trained in the earlier study. As mentioned above, some improvement in scores was observed in cases where repeated measures were obtained, but these improvements were ignored in the analyses. These improvements increased the residual sum of squares attributed to random sampling in the analyses of variance, thus making the comparisons of modalities more conservative than they would otherwise have been.

Equations [1] to [6] show that the visual input was an effective supplement in every situation since VT > T, AV > A, and AVT > AT for every test. The auditory input also produced a significant increase, since AV > V, AT > T, and AVT > VT for every test except the vowels. In the case of the vowels, the third inequality did not reach the criterion at the 95 percent confidence level. This may have been a consequence of the limiting effect caused by the high score for VT (87 percent). The tactile input produced only four significant increases in score: AT > A for vowels; AT > A and VT > V for consonants; VT > V for CNC words. Despite the fact that the tactile input conveyed useful information in isolation, it did not seem to have as great an effect as A or V information when combined. This was especially true in the open-set tests.

Comparison of the observed and predicted information transmission values in **Table 4** and **Table 6** also suggests that the AT and VT combinations were less effective than the AV combination in the closed-set tasks. The very effective combination of A and V may be a consequence of long experience with these modalities during the normal development of speech and language, especially in situations where the auditory signal is degraded by background noise. Alternatively, there may be specialized neural mechanisms for the combination of A and V information that are used in the AV mode of speech perception. The less-effective combination of T information with either A or V may, therefore, be a consequence of lack of experience and appropriate training, or lack of appropriate neural structures and functions to carry out the necessary

combination. The data presented here are not sufficient to distinguish between these situations, although it is clear that training (or lack of it) has had a large effect on the results in **Table 1** and **Table 2** for the V, VT, and T modalities. It should also be noted that the subjects had no prior experience in the combined AT and ATV modalities.

If it is assumed that there are no physiological limitations preventing the use of tactile information as effectively as auditory and visual information, the present study leads to two conclusions of practical significance. To be useful, a tactile aid must be capable of providing information in the T condition, and this information must be combined effectively with information from other modalities. The present study showed that it is possible to obtain a good score in the T modality, without obtaining the full benefit of the tactile information in combined modalities. This implies that training in the combined modalities will be necessary, as well as training in the T condition. Secondly, the study showed that tactile information could supplement limited auditory information in closed-set tasks. Provided that training can extend this performance to open-set tasks, tactile devices may be beneficial to hearing aid and cochlear implant users with limited auditory function.

ACKNOWLEDGMENTS

Ms. I. Havriluk, Dr. A. Blunden, Dr. P. Seligman, and Mr. R. Millard provided technical support and advice that was essential to the evaluations carried out. Mrs. A. Brock provided administrative support and typed the manuscript. At different stages of this research, financial support has been obtained from: the Deafness Foundation of Victoria, the National Health and Medical Research Council of Australia, the George Hicks Foundation, the Department of Industry, Technology and Commerce of the Australian Commonwealth Government, and the Australian Bionic Ear Institute.

REFERENCES

1. **Bench J, Bamford J (Eds.):** *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children.* London: Academic Press, 1979.
2. **Bernard JRL:** Toward the acoustic specification of Australian English. *Zeit Phonetik* 23:113-128, 1970.
3. **Blamey PJ, Clark GM:** Psychophysical studies relevant to the design of a digital electrotactile speech processor. *J Acoust Soc Am* 82:116-125, 1987.
4. **Blamey PJ, Dowell RC, Brown AM, Clark GM, Seligman PM:** Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *J Acoust Soc Am* 82:48-57, 1987.
5. **Blamey PJ, Martin LFA, Clark GM:** A comparison of three speech coding strategies using an acoustic model of a cochlear implant. *J Acoust Soc Am* 77:209-217, 1985.
6. **Blamey PJ, Seligman PM, Dowell RC, Clark GM:** Acoustic parameters measured by a formant-based speech processor for a multiple-channel cochlear implant. *J Acoust Soc Am* 82:38-47, 1987.
7. **Brooks PL, Frost BJ, Mason JL, Chung K:** Acquisition of a 250-word vocabulary through a tactile vocoder. *J Acoust Soc Am* 77:1576-1579, 1985.
8. **Clark GM, Tong YC, Patrick JF, Seligman PM, Crosby PA, Kuzma JA, Money DK:** A multi-channel hearing prosthesis for profound-to-total hearing loss. *J Med Eng Technol* 8:3-8, 1984.
9. **Cowan RSC, Alcantara JI, Blamey PJ, Clark GM:** Preliminary evaluation of a multichannel electrotactile speech processor. *J Acoust Soc Am* 83:2328-2338, 1988.
10. **De Filippo CL, Scott BL:** A method for training and evaluating the reception of ongoing speech. *J Acoust Soc Am* 63:1186-1192, 1978.
11. **Dodd B, Campbell R (Eds.):** *Hearing by Eye: The Psychology of Lip-Reading.* London: Lawrence Erlbaum, 1987.
12. **Eilers RE, Widen JE, Oller DK:** Assessment techniques to evaluate tactual aids for hearing-impaired subjects. *J Rehabil Res Dev* 25(2):33-46, 1988.
13. **Engelmann S, Rosov R:** Tactual hearing experiment with deaf and hearing subjects. *Except Child* 41:243-253, 1975.
14. **Erber NP:** *Auditory Training.* Washington, DC: A.G. Bell Assn., 1982.
15. **Gault RH:** Progress in experiments on tactual interpretation of oral speech. *J Abnormal Soc Psych* 14:155-159, 1924.
16. **Jeffers J, Barley M:** *Speechreading (Lipreading).* Springfield, IL: Charles C. Thomas, 1971.
17. **Kozma-Spytek L, Weisenberger JM:** Evaluation of a multichannel electrotactile device for the hearing-impaired: A case study. Presented at *The 114th Meeting of the Acoustical Society of America,* Miami, FL, 1988.
18. **Lynch MP, Eilers RE, Oller DK:** Speech perception

through the sense of touch by profoundly deaf adults and children. Presented at *The 114th Meeting of the Acoustical Society of America,* Miami, FL, 1988.

19. **Massaro D:** *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale, NJ: Lawrence Erlbaum, 1987.
20. **Miller GA, Nicely PE:** An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27:338-352, 1955.
21. **Petersen GE, Lehiste I:** Revised CNC lists for auditory tests. *J Speech Hear Disord* 27:62-70, 1962.
22. **Pickett JM, Pickett BH:** Communication of speech sounds by a tactual vocoder. *J Speech Hear Res* 6:207-222, 1963.
23. **Plant G:** A single-transducer vibrotactile aid to lipreading. *Speech Transmission Laboratories Quarterly Progress and Status Reports* STL-QPSR 1:41-63, 1986.
24. **Reed CM, Rabinowitz WM, Durlach NI, Braida LD, Conway-Fithian S, Schultz MC:** Research on the Tadoma method of speech communication. *J Acoust Soc Am* 77:247-257, 1985.
25. **Saunders FA:** Information transmission across the skin: High resolution tactile sensory aids for the deaf and the blind. *Int J Neuroscience* 19:21-28, 1983.
26. **Sherrick CE:** Basic and applied research on tactile aids for deaf people: Progress and prospects. *J Acoust Soc Am* 75:1325-1342, 1984.
27. **Sparks DW, Kuhl PK, Edmonds AE, Gray GP:** Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of segmental features of speech. *J Acoust Soc Am* 63:246-257, 1978.
28. **Sumby WH, Pollack I:** Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212-215, 1954.
29. **Walden BE, Prosek RA, Worthington DW:** Auditory and audiovisual feature transmission in hearing-impaired adults. *J Speech Hear Res* 18:272-280, 1975.

Author/s:
Blamey, Peter J.; Cowan, Robert S. C.; Alcantara, Joseph I.; Whitford, Lesley A.; Clark, Graeme M.

Title:
Speech perception using combinations of auditory, visual, and tactile information

Date:
1989

Citation:
Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., & Clark, G. M. (1989). Speech perception using combinations of auditory, visual, and tactile information. Journal of Rehabilitation Research and Development, 26(1), 15-24.

Persistent Link:
http://hdl.handle.net/11343/27275