First European Conference on
Speech Communication and Technology
EUROSPEECH '89
Paris, France, September 27-29, 1989

DEVELOPMENT OF A SINGLE-CHANNEL TACTILE AID FOR THE PROFOUNDLY DEAF

Ian R Summers and Judith Farr

Department of Physics, University of Exeter, Stocker Road, EX4 4QL

10.21437/Eurospeech.1989-330

ABSTRACT

In this study of lipreading with supplementary tactile input, normally hearing subjects with noise masking were presented with tactile information relating to the voiced/unvoiced contrast. Considerable training was required before subjects were able to make use of this information so as to gain higher scores for lipreading with tactile supplement than for lipreading alone. These results are in contrast with those from earlier experiments which showed that relatively little training was necessary for successful tactile perception of voice stress.

INTRODUCTION

This paper describes the latest stage in a project to develop a tactile aid for the profoundly deaf. The main aim throughout the project has been to produce a device which gives significant benefit to the user whilst remaining simple enough to be made available at low cost and in a cosmetically acceptable form. With this in mind, the investigation has been restricted to the transmission of information via a single vibrator on the wrist or hand. The significance of our recent test results is most apparent when these are compared with results from our previous investigations. There follows a brief summary of this earlier work.

THE TAM

The Tactile Acoustic Monitor (TAM) is primarily intended as an environmental aid for the profoundly deaf. It consists of a piezoelectric vibrator, worn on the wrist, and a body-worn unit containing microphone and electronics. The vibrator is activated (at constant amplitude and constant frequency, ~ 200 Hz) when the sound level at the microphone rises above a preset level [1]. The vibrotactile output is thus a series of tonebursts coincident with the 'loud' sections of sounds in the acoustic environment. The TAM user can detect sounds and, to some extent, distinguish between different sounds on the basis of distinctive temporal patterns in the tactile output [2]. The TAM was developed by the Royal National Institute for the Deaf (London) in conjunction with the Medical Physics Group at Exeter University. It went into commercial production in 1986. Several hundred are now in use.

TOWARDS AN AID TO LIPREADING

The TAM was designed to be as simple as possible so that it might avoid the fate of many other tacile aids which have never progressed beyond the laboratory bench. It was always the intention to incorporate more sophisticated circuitry once a basic device was in production. Some progress has now been made in this direction with investigations of the tactile transmission of speech features.

The capacity of a single tactile channel is very limited [3,4], and there is no realistic possibility of such a channel conveying all the significant information in a speech signal. However, most profoundly deaf people make use of lipreading and in this context a tactile channel can convey sufficient information to usefully supplement the visual input. So as not to overload the tactile channel, it is necessary to extract from the speech signal an appropriate slowly varying parameter for tactile coding. (By 'slowly' we mean on a timescale of, say, ~ 10ms.) Such parameters include gross spectral features (such as zero-crossing rate or high-frequency/low-frequency energy balance, giving some phonemic information) and features such as voice pitch and voice intensity (giving information on stress, inflection, etc.).

TACTILE TRANSMISSION OF STRESS PATTERNS

Modulations of voice pitch and voice intensity convey information to the normal listener which is not easily available to the lipreader. It has been shown [5,6] that supplementing visual input with information on either or both of these speech features can significantly improve lipreading performance. With this in mind, tests were carried out to establish if such information could be transmitted via a trandsucer systen similar to that of the TAM (in this case not on the wrist but on the distal pad of the finger). This work has been described in detail in a recent paper [7].

Various coding strategies were compared, using normally hearing subjects with noise masking via headphones. After approximately 20 minutes of explanation and demonstrations, subjects were presented with a sequence of six short tests, each with a different tactile coding.

Using tactile information only, subjects attempted to identify the stressed word (from three alternatives) in various simple sentences of three to six words. The most effective coding strategies proved to be (i) voice frequency presented as a continuously variable stimulus frequency over the range 40-220 Hz, with a correlated modulation of stimulus amplitude, and (ii) speech amplitude presented as two discrete levels of stimulus amplitude. In these cases scores of approximately twice the chance score of 33% were obtained.

With hindsight (that is after the more recent

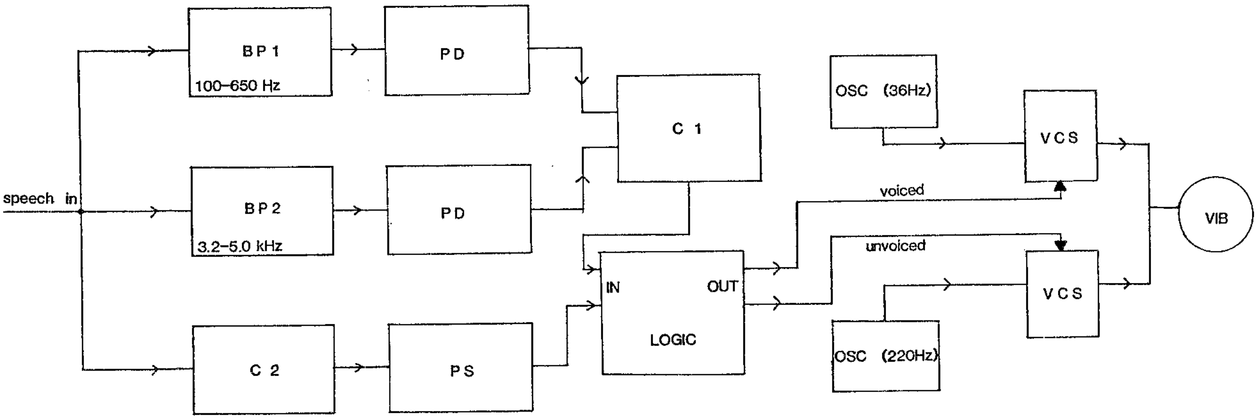

Figure 1. The speech processing circuitry. *Key:* BP1, 4th-order bandpass filter (low-frequency channel); BP2, 4th-order bandpass filter (high-frequency channel); PD, peak detector, 10ms time constant; C1, comparator (voiced/unvoiced decision); C2, comparator (compares input signal level with a d.c. level just higher than signal noise); PS, 8ms pulse stretcher; OSC, square-wave oscillator; VCS, voltage controlled switch; VIB, vibrator, (in practice a tape link was used in the line to the vibrator).

experiments detailed below), a particular feature of the results from these tests was the ease with which the subjects were able to come to terms with the range of tactile codes. Not only were good results obtained with minimal training, but subjects were able to cope with changing from one coding to another when given only a short period of adjustment.

CURRENT WORK

After the experiments on stress identification described above, attention moved to the transmission of phonemic features by tactile presentation of the high-frequency/low-frequency energy balance in the speech signal. This work is described in the remainder of the paper.

APPARATUS

Figure 1 shows a block diagram of the electronic circuitry used to generate the tactile stimuli. The input speech signal is bandpass filtered into two channels - low frequency and high frequency, the signal level in each being controlled by potentiometers which are set to compensate for the preponderance of signal at low frequencies, i.e. so that the long-term mean signal levels in the two channels are approximately equal. A comparator determines which channel contains the larger signal at any instant. If this is the high frequency channel, a 220 Hz square wave is applied to the vibrator; if the low frequency channel, a 36Hz square wave is applied to the vibrator. A second comparator which operates on the broad-band input signal, in conjunction with a simple logic circuit, ensures that no output is produced unless the input signal is significantly above the noise background.

The piezoelectric vibrator has a marked resonance at 220 Hz; hence driving it with this frequency produces a strong sinusodial output. When driven with 36 Hz it produces an output considerably reduced in intensity. This has subjectively a pulsating quality which corresponds to the series of damped tonebursts (at the vibrator's resonant frequency) which are generated by the rising and falling edges of the square-wave drive signal (see Figure 2). These 'high' and 'low' output signals are clearly distinguishable (having significant differences in both subjective intensity and perceived 'quality').

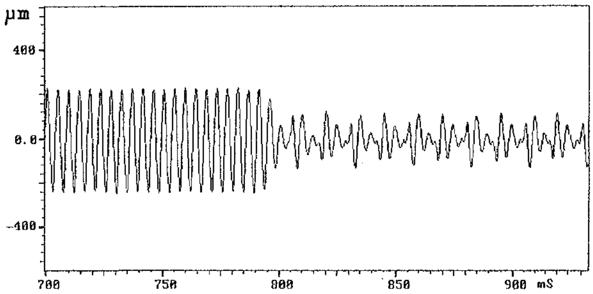

Figure 2. Typical vibrator output. A section of 'high' output, followed by a section of 'low' output. Displacement is measured using a Knowles type BU-1771 subminiature vibration sensor.

At any time the speech signal is classed as either having predominantly low-frequency or high-frequency content. Since unvoiced speech sounds have spectra biased towards high frequencies and voiced sounds have spectra biased towards low frequencies, the tactile output from this circuitry conveys the distinction between these classes of phonemes. Many speech sounds which may be indistinguishable to the lipreader because they look similar (visemes) are expected to be distinguishable on the basis of this voiced/unvoiced contrast.

TEST MATERIAL

The material used during testing is designed for the assessment of lipreading ability. The Intervocalic Consonant Test (University College London) measures subjects' ability to recognize 12 consonants spoken within the context of the vowel 'ah'. Test items containing the consonants B, D, F, G, K, M, N, P, S, T, V, Z are spoken by a female speaker and recorded on video tape. In our tests, normally hearing subjects with noise masking are required to identify and write down each consonant.

Figure 3 shows the vibrator output for various test

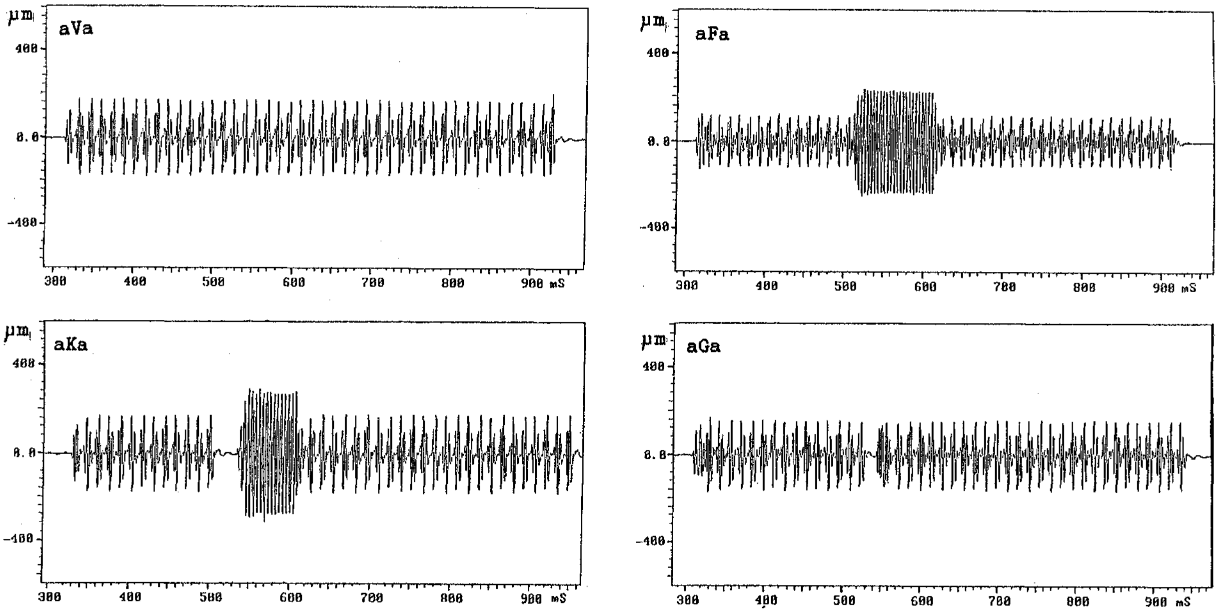

Figure 3. Tactile outputs for the items aVa, aFa, aKa and aGa.

items. Significant cues to help lipreading are provided by the high-frequency/low-frequency information and also by the indication of short periods of silence within the syllables.

EARLY RESULTS

A preliminary evaluation of this coding strategy as an aid to lipreading was carried out by the authors using themselves as subjects. Neither were practised prior to the test but both were familiar with the test material and the type of output produced by the vibrator when using this particular coding strategy. The number of correctly identified consonants using tactile information plus lipreading were scored, along with those from lipreading alone. Subjects IRS and JF obtained scores (with / without tactile information) of 66.7% / 39.6% , and 79.2% / 47.9% respectively.

Unfortunately, the ease with which the authors were able to use the tactile information as an aid to lipreading was not shared by naive subjects who had no previous experience of such tactile stimuli (even though some of these subjects were familiar with the test material as they had participated in an earlier test using acoustic stimuli). When presented with the 12-consonant test after a short training period, these subjects derived no benefit from the tactile information. In contrast to our findings in the earlier stress identification tests, it was clear that more extensive training was required before subjects could make use of tactile information which related to phoneme distinctions.

Accordingly, a sequence of training / testing was devised in which subjects were first familiarised with the tactile stimuli and then taught to make use of stimuli in progressively more complex lipreading tasks.

TRAINING PROGRAMME AND RESULTS OF FURTHER LIPREADING TESTS

Note: The number of subjects for whom results are quoted varies throughout the test sequence. (See Table 1). This is due to some subjects becoming unavailable, or being ineligible to do a test because they had prior experience of a similar one. All results given are for subjects who were at an equivalent stage of training and had spent the same amount of time participating in lipreading tests.

In order to familiarise subjects with the types of tactile output from the vibrator, a microcomputer was used to generate stimuli of a similar type to those produced by the speech processing circuitry. (See Figure 3). The four types of stimuli were:
(i) uninterrupted ~ 650 ms stimulus, all 'low' (as for aVa)
(ii) as (i), but interrupted by a ~ 50 ms 'high' stimulus (as for aFa)
(iii) as (i), but interrupted by a ~ 30 ms silence (as for aGa)
(iv) as (i), but interrupted by a ~ 60 ms silence followed by a ~ 50 ms 'high' stimulus (as for aKa)

Subjects, after a small amount of familiarisation (~15mins), took part in a test in which they were required to classify the stimuli played to them, (this test involved no lipreading). Results for eight subjects (who had previously done one lipreading test only), showed an (87 ± 4) % correct identification of the stimulus type.

Encouraged by this performance, we then began lipreading tests which required a combination of visual and tactile recognition. Consonant items were generally introduced to the subjects as part of a viseme group. Data obtained previously from lipreading tests on 48 subjects enabled classificaion of the test items into these groups. Clearly aMa, aPa and aBa look similar to a lipreader, as do aFa and aVa; aKa and aGa are mutually confuseable, and aNa may be confused with both these items; aNa also tends to be mistaken for aDa, which groups with aTa. Also aSa and aZa form a viseme group; (some subjects confused aSa with aTa).

Lipreading tests, both with and without tactile information, were carried out in what was expected to be increasing order of difficulty, raising the number of items in each test from 2 (aFa and aVa)

TEST	CHOICE OF TEST ITEMS	LENGTH OF EACH TEST	NUMBER OF SUBJECTS
A	2: aFa, aVa	16 items	10
B	4: aFa, aVa aBa, aPa	16 items	12
C	6: aTa, aDa aMa, aNa aKa, aGa	18 items	8
D	12: aMa, aPa, aBa, aNa aGa, aKa aFa, aVa aDa, aTa aSa, aZa	24 items	8

Table 1: Details of the test procedure.

through to the full 12, (See Table 1). Each test was preceeded by a training video which informed subjects of the type of tactile stimulus they would feel for each item, with examples, and then gave a mock test in which subjects were told afterwards which item had been played.

The duration of training prior to each test was ~15 minutes, varying to accomodate the number of new items in the test. All subjects performed the 'lipreading plus tactile' test before the 'lipreading only' test because it was assumed that any learning effect during the test would then contribute favourably to the lipreading only score, hence not giving an unfair bias to the performance of the tactile aid. No feedback as to the subjects' performance was given duing the course of testing (approximately 1 month).

Table 2 shows the scores for each of the tests described above.

DISCUSSION AND CONCLUSION

Although naive subjects appear unable to make use of tactile information in the 12-consonant test, suitably trained subjecs show some benefit from the tactile presentation. However, results for the 12-item test, D, (after training) are still rather disappointing when compared to those from the 2-item, 4-item and 6-item tests (A, B and C), and when compared to the preliminary results obtained with the authors as subjects. This may indicate that even more training is needed if subjects are to fully come to terms with this 12-item test.

It seems clear that, at least in our test formats, the task of making use of tactile stimuli relating to phonemic information is considerably more difficult than the equivalent task with stress information. This result is important for the development of tactile aids since this work often involves the comparison of different tactile codings of speech information. Insufficient subject training may give an unwarranted bias away from those coding schemes which convey phonemic information.

TEST	LIPREADING + TACTILE (%)	LIPREADING ALONE (%)	L+T /L.A (averaged over all subjects)
A	80.6 ± 3.0	54.4 ± 5.0	1.57 ± 0.12
B	82.3 ± 3.4	56.7 ± 3.8	1.55 ± 0.16
C	71.7 ± 4.3	42.4 ± 4.3	1.82 ± 0.21
D	52.1 ± 3.2	38.3 ± 3.7	1.48 ± 0.18

Table 2: Scores for tests described in Table 1.

ACKNOWLEDGEMENTS

This work is supported by The National Deaf Children's Society and the British Technology Group. Vibration transducers were kindly provided by Knowles Electronics Company.

REFERENCES

[1] I R Summers and M C Martin, "A tactile sound level monitor for the profoundly deaf", British Journal of Audiology, Vol 14, pp 30-33: 1980.

[2] I R Summers, M A Peake and M C Martin, "Field trials of a Tactile Acoustic Monitor for the profoundly deaf", British Journal of Audiology, Vol 15, pp 195-199: 1981.

[3] W M Rabinowitz, A J M Houtsma, N I Durlach and L A Delhorne, "Multidimensional tactile displays: identification of vibratory intensity, frequency and contactor area", Journal of the Acoustical Society of America, Vol 82, pp 1243-1252: 1987.

[4] M Rothenberg, R T Verillo, S A Zahorian, M L Brachman and S J Bolanowski, "Vibrotactile frequency for encoding a speech parameter", Journal of the Acoustical Society of America, Vol 62, pp 1003-1012: 1977.

[5] S M Rosen, A J Fourcin and B C J Moore, "Voice pitch as an aid to lipreading", Nature (London), Vol 291, pp 150-153: 1981.

[6] K W Grant, L H Ardell, P K Kuhl and D W Sparks, "The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speech reading in normal-hearing subjects", Journal of the Acoustical Society of America, Vol 77, pp 671-677: 1985.

[7] I R Summers and J Farr, "Coding strategies for a single-channel tactile aid", British Journal of Audiology, Vol 23, in press 1989.