TESTS OF TACTUAL SPEECH TRANSMISSION[1]

J. M. PICKETT

The tactual display of speech offers a means of speech transmission under conditions where auditory reception is severely impaired, for example, in teaching the deaf and in receiving speech under very high noise levels. The tests reported here were designed to provide estimates of the transmission of speech sounds by means of a tactual vocoder.[2] The vocoder transformed a representation of the frequency domain of the speech signal into a spatial array on the skin. Ten frequency channels were used and each channel corresponded tactually to one of the ten fingers of the hands of the subject receiving the signals.

The tactual vocoder operated essentially as follows. The speech signal was first compressed in amplitude and then differentiated to emphasize high frequencies (+4 db/oct). The resulting signal was divided by overlapping filters into ten channels having center frequencies of 210, 400, 580, 830, 1050, 1800, 2250, 3320, 5800 and 7700 cps. The response curves of the channel filters were triangular with sides having slopes that approached 12 db per octave. The output signal from each channel was rectified and smoothed to yield a control voltage. Each of the ten control voltages modulated the amplitude of a 300-cps sinusoidal signal. The varying 300-cycle signals were amplified and led to 10 bone-conduction transducers which served as vibrators for stimulating the ventral tips of the subject's fingers. Proceeding from left to right across the dorsal view of the two hands, the frequency channels were presented to the fingers in order beginning with the lowest channel and proceeding to the highest channel.

The speech tests prepared for transmission were of two types: discrimination tests and identification tests. In a discrimination test only one pair of sounds was tested. Various pairs were tested, covering the major speech features that might be dis-

[1] This investigation was carried out during the author's tenure of a Special Research Fellowship from the National Institute of Mental Health, United States Public Health Service.

[2] Levine, Wiesner, *et al* (1) built a 5-channel frequency vocoder, Felix, which presented the channel energies separately to five fingers by means of amplitude-modulated 300-cps vibrations. Short practice on 12 words yielded 91% correct identification. Seven-channel electrical stimulation of the forearm was then substituted for the finger vibrators but further tests of Felix were postponed in favor of informational studies of speech to determine the optimum coding parameters. Rösler (2) reviews the literature on vibratory sensation and reports successful trials in transmitting speech by vibration of the fingers with the 10-channel vocoder of this Laboratory. Guelke and Huyssen (3) describe the development and preliminary testing of a tactual speech analyser which presents eight channels of vibratory stimulation to the fingers of one hand.

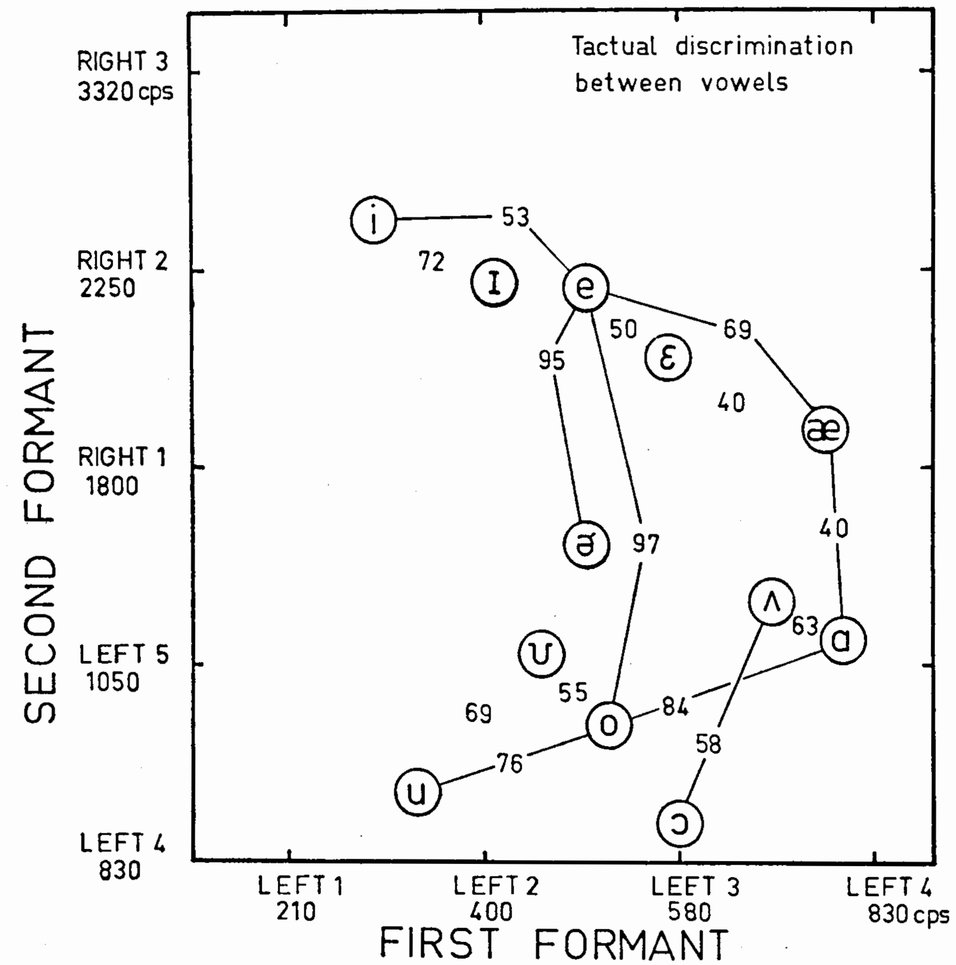

Fig. 1. Results of tactual discrimination tests between pairs of vowels. The vowels are arrayed according to their lower two formant frequencies, F_1 and F_2 in cps, scaled linearly with equal distances between finger vibrators. The abscissa shows the position of F_1 in the series of vibrators across the left hand from LEFT 1, the little finger, through LEFT 4, the index finger. The ordinate shows the position of F_2 in the series of vibrators from LEFT 4, the left index finger through RIGHT 3, the right middle finger. Discrimination is shown between vowels in terms of percentage correct discrimination relative to chance. Each percentage is based on 320 observations: 40 of each vowel were uttered by each of two talkers and judged tactually by each of two subjects. It will be noted that high discrimination is associated with large distances between the vowels. Discrimination between adjacent vowels is lower except for cases where duration differences exist between the vowels (i — ɪ, u — ʊ).

criminated by the tactual patterns. In identification tests, larger sets of test sounds were used to obtain estimates of performance when more than one speech feature was to be transmitted.

Each discrimination test consisted of 80 spoken phrases carrying 40 of each of the two test sounds. The order of the sounds was random with the restriction that each

successive quarter of the test contained ten of each sound. Each identification test contained 15 of each sound to be identified and in a random order.[3]

Results are shown for tactual discrimination of vowel sounds in Fig. 1. In the figure, the vowels are arranged according to their first two formant features as measured on American talkers by Peterson and Barney (3). The diphthongs /eɪ/ and /oʊ/ were tested and they are referred to here as /e/ and /o/ respectively. The scales of the axes of Fig. 1 are arranged with equal distances between the finger vibrators. The band center frequency associated with each vibrator is given to indicate the frequency interval between vibrators. Each interval was scaled linearly in plotting the vowel formant frequencies.

Between pairs of test vowels we have inserted the percentage discrimination of the pair relative to chance performance. It will be noted that discrimination is best between vowels having widely separated values of the second formant, for example /e/ *vs* /o/. Discrimination is also good for some pairs having separated values of the low formant, e.g. /u/ *vs* /o/ and /o/ *vs* /a/. When the separation along one formant dimension is reduced, discrimination is less efficient (cf. /ɛ/ *vs* /e/, /ɛ/ *vs* /æ/, and /æ/ *vs* /a/). However even the lowest values of discrimination represent performance better than chance. In two cases, /i/ *vs* /ɪ/ and /u/ *vs* /ʊ/, there is little formant difference between vowel pairs but good discrimination occurs, apparently based on the

[3] The test utterances were constructed as follows. A standard phrase was adopted to carry each test sound. The phrase was /traCVt/; the /a/ was spoken as in *father*; C represents a test consonant and V represents a test vowel. Consonant and vowel tests were carried out separately. In consonant tests the vowel V was always /a/ and in vowel tests the consonant C was always /p/. Each test phrase was spoken in a fluent manner with the stress on the second syllable.

The tests were recorded on magnetic tape and played back for the tactual perception tests. Each test was recorded twice, once by a male talker, the author, and once by a female talker who had no formal phonetic training. Both talkers were native speakers of American English. The phrases of a test were spoken successively with an interval of 2 to 4 sec. The talker monitored his speech level on a voltmeter with VU damping.

The two talkers also served as the subjects for the perceptual tests. Each test began with a preliminary training series which contained four complete sets of the sounds to be tested; they were spoken in carrier phrases and in a consonant order. The experimenter (E) played the series of four training sets twice, informing the subject (S) as to the test sound before each carrier phrase was played. Then the test followed immediately. S responded by speaking the test sound after each phrase. E then informed S of his success or failure before proceeding to the next phrase, E kept a running score, indicating a total after every 20 test phrases. However most of the learning of the tactual patterns appeared to occur during the first 20 phrases. Slower learning occurred for certain cases where extended practice yielded gradual improvement up to moderate levels of performance (6-vowel identification and discrimination of æ – a, ɛ – æ, e – ɛ, i – e).

To insure that no audible information was available during the tests, the subject wore a tight-fitting headset which introduced a continuous random noise into both ears at a sound level of about 80 db.

Discrimination scores were expressed as percent discrimination relative to chance performance by the following formula:

$$D = 100 \frac{N_P - N_Q}{N_P + N_Q}$$

where D is the discrimination score, N_P is the number of correct responses, and N_Q is the number of wrong responses.

PERCENT DISCRIMINATION

Test Vowels	Variable Utterance	Constant Utterance	Probability on Null Hypothesis
æ — a	35.8	64.2	.0002
æ — ε	40.0	57.5	.0375
ε — e	47.5	66.2	.0212
i — e	57.4	71.2	.0495

TABLE I. Effects of utterance variability on tactual discrimination of vowels. Each entry of percent discrimination is relative to chance and based on 160 observations: 40 of each vowel were spoken by one talker and judged tactually by each of two subjects.

differences in vowel duration for these pairs which are normally found in American speech.

The subjects often reported that only one formant of a vowel was felt, even when the first and second formants were widely separated, as with /i, ɪ, e/. The first formants of /i, ɪ, e/ were too weak to be felt, at least in the presence of the strong vibrations due to the second formants. Thus discrimination between these vowels depended on feeling a range of only one finger interval in the region of RIGHT 1 (1800 cps) and RIGHT 2 (2250 cps). On the other hand discrimination across the series of back vowels /u-a/ was somewhat better and these were felt as a series over nearly four finger intervals of the left hand as indicated by their formant plots. However, two or more adjacent vibrators were felt to vibrate for each vowel. Considerable overlap could be perceived between distributions of the vibrations for /u *vs* o/ and /o *vs* a/.

The procedure used in recording the tests (see fn. 3) allowed variability among the utterances of a given test sound. It was suspected that this utterance variability was an appreciable factor in the poor discrimination of /æ – a/, /æ – ε/, /ε – e/, and /e – i/. A set of eight control tests was constructed for each of these vowel pairs to estimate the effect of utterance variability. The control tests were assembled from the original recordings of one talker. The recordings were modified by inserting, for every 20 original test sounds, a test made up from 10 copies of a single example of each test sound. An original series of 20 sounds will be called a variable test and the 20 copied sounds will be called a constant test. Thus there were four different variable tests from each original recording. Four constant tests of 20 sounds were constructed for each of the vowel pairs, æ – a, æ – ε, ε – e, and e – i. The eight test phrases for constructing the constant tests of a given vowel pair were selected at random from the entire original test. Constant and variable tests alternated with each other in a counterbalanced order. Each test was preceded by 16 practice items. The results are shown in Table I. For every vowel pair, constant utterance was more discriminable than variable utterance. The third column of Table I gives the probabilities of the obtained percentages on the null hypothesis that they are samples from a single population having the mean percentage. Since these probabilities are

Tactual vowel identification

Stimulus vowel spoken (rows) / Response (columns)

	i	e	æ	ɑ	o	u
i	31	11	2	1	3	12
e	12	36	11	1		
æ		8	39	13		
ɑ	1	6	18	33	2	
o	2			7	45	6
u	1				8	51

65 % Correct
54 % Information transmitted

Stimulus vowel spoken (rows) / Response (columns)

	ɪ	ε	ʌ	ɔ	ɚ	ʊ
ɪ	21	5	2			2
ε	2	15	9	2	2	
ʌ	6	9	6	4	5	
ɔ	3	3	3	15	4	2
ɚ	1	1	1	4	14	9
ʊ	3	2	3	3	8	11

46 % Correct
27 % Information transmitted

Fig. 2. Tactual identification of vowels. Each matrix entry gives the frequency of a given vowel identification (columns) as a response to a given spoken vowel (rows). In tests for the left matrix, 15 of each vowel were spoken by each of two talkers and identified tactually by each of two subjects. For the right matrix, 15 of each vowel were spoken by one talker and identified tactually by two subjects. The average information transmitted by a response was calculated from individual test matrices of 90 responses and expressed as the percentage of the information per vowel category, 2.585 bits.

small we conclude that utterance variability had a significant deteriorating effect on tactual discrimination of the difficult vowel pairs.

Vowel identification tests were carried out with sets of 6 vowels. Two sets were used. One set, /i, e, æ, a, o, u/, was chosen to maximize the average distance on the formant plot between members of the set. The other set of 6 vowels consisted of the remaining English vowels. The results of the identification tests are shown as confusion matrices in Fig. 2. In the first set of vowels, /i, e, æ, a, o, u/, three complete repetitions of the two recorded tests were judged on separate days by the two subjects. Results are shown only for the final repetition, but these do not represent a large improvement over performance on the initial test. It will be noted that reasonably good identification was obtained. The percentage of information transmitted, a relatively rigourous measure of performance, indicated that a moderate amount of information was transmitted from the stimulus source to the response distributions.

The results shown in Fig. 2 for the second set of vowels, /ɪ, ε, ʌ, ɚ, ʊ/, represent only two tests of this set on one talker. But it appears that we will not obtain high performance with this set where the vowels are more similar in formant pattern than the vowels of the first set.

The tests of consonant discrimination were performed in the same way as the vowel discrimination tests. The results are given in Table II beginning with the consonant pair that was best discriminated and proceeding to the worst. It will be noted

Test Consonants	Percent Discrimination	Test Consonants	Percent Discrimination
s,t	99.5	f,b	52.0
t,n	84.2	ʃ,ʒ	50.0
z,v	84.2	t,d	50.0
j,l	83.5	d,l	48.5
j,n	82.2	l,n	44.0
s,f	82.2	f,v	40.5
s,ʃ	76.2	w,m	29.7
z,ʒ	63.8	l,r	27.0
s,z	62.3	p,b	22.0
m,b	56.1		

TABLE II. Tactual discrimination between various consonants. Each entry is the percent discrimination relative to chance and based on 320 observations: 40 of each consonant were spoken by each of two talkers and judged tactually by each of two subjects.

from the consonant results that place of consonant articulation is discriminated well for the fricative consonants, even when they are voiced. The glide /j/ was well discriminated from /l/ and /n/, the other continuants which are articulated at about the same place as /j/. Voicing and nasality were discriminated fairly well in some cases, e.g. /m, b/, /ʃ, ʒ/ and /t, d/. The subjects reported that these two features were judged often by feeling for a slower decay and attack in the vowels adjacent to voiced or nasalized consonants than for voiceless or non-nasal consonants.

The results above lead us to believe that tactual speech displays may be of considerable value when used as a substitute for auditory reception. For a suitably restricted context and vocabulary a tactual display may be sufficient for communication. In less restricted situations, tactual speech may be used as an aid to lip-reading or visual presentation of speech.

It is planned to continue tests of the present device on discriminations involving more than one sound per syllable and to obtain estimates of the syllable span of tactual speech perception. Other methods of speech processing will also be considered.

Speech Transmission Laboratory
Royal Institute of Technology
Stockholm

REFERENCES

(1) L. Levine, J. B. Wiesner, O. H. Straus, B. Howland, N. Wiener, and E. E. David, Jr., "Felix" (Sensory Replacement), *Quarterly Progress Reports*, Research Laboratory of Electronics, Massachusetts Institute of Technology, January 1949—April 1951.

(2) G. Rösler, "Über die Vibrationsempfindung. Literaturdurchsicht und Untersuchungen im Tonfrequenzbereich", *Z. f. exper. u. angew. Psych.*, 4, 549–602 (1957).

(3) R. W. Guelke and R. M. J. Huyssen, "Development of Apparatus for the Analysis of Sound by the Sense of Touch", *J. Acoust. Soc. Am.*, 31, 799–809 (1959).

(4) G. Peterson and H. Barney, "Control Methods Used in a Study of Vowels", *J. Acoust. Soc. Am.*, 24, 175–184 (1952).