Veterans Administration

Journal of Rehabilitation Research and Development Vol. 25 No. 4
Pages 25–44

Tactile communication of speech: Comparison of two computer-based displays

MARK A. CLEMENTS, LOUIS D. BRAIDA, and NATHANIEL I. DURLACH
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract—Two methods of encoding speech for tactile displays were compared in discrimination experiments using speech segments. One display represented the short-term speech spectrum in time-swept mode and used vibration amplitude to encode spectral amplitude. The other represented the linear predictive coding (LPC)-derived vocal tract shape as a filled bar graph in which the number of active vibrators was used to encode cross sectional area. The displays were applied to the thigh via a matrix of vibrators. The vibrators were driven at 250 Hz during voiced segments, and by random noise during unvoiced segments. Overall results show a slight superiority for the spectral display in vowel discrimination. Detailed results were analyzed in terms of an articulatory description of the speech stimuli, a multidimensional scaling (MDS) analysis of confusions, and an ideal receiver analysis. The results of these analyses suggest that the detailed characteristics of the tactile patterns were only crudely discriminated.

Key words: *area function/spectral display, consonant/vowel discrimination, linear predictive coding, tactile speech communication, vibratory amplitude/waveforms.*

INTRODUCTION

Although studies of the Tadoma Method by Norton *et al.* (11), and Reed *et al.* (15), indicate that tactile reception of speech is possible at reasonable rates and with reasonable accuracy, the performance of artificial tactile speech displays and tactile prostheses for the deaf leaves much to be desired. In principle, Tadoma not only provides display designers with realistic performance standards against which artificial displays can be compared, but also provides suggestions for the design of novel displays. The Tadoma user monitors the actions of the speech articulators by placing a hand on the face of the talker. The tactile speech display provided the user is both directly tied to the articulation process and perceptually rich (including a variety of sensory qualities). By contrast, most artificial tactile speech displays attempt to convey the short-term spectrum of the acoustic speech signal by means of stimulus patterns created by linear or rectangular arrays of vibrating elements.

In previous papers, we discussed the need for controlled experimental comparisons of different schemes for the tactile display of speech and argued that initial comparisons should focus on discrimination tasks rather than identification tasks. Snyder *et al.* (20) compared an Optacon-based spectral display to Tadoma in a consonant discrimination task and found the artificial display to be substantially inferior to Tadoma for voicing and place contrasts, but roughly equal for manner contrasts. Clements *et al.* (6) compared 2 different Optacon-based spectral displays in a vowel discrimination task. They found a time-swept display (in which one dimension of the stimulating matrix is used to present a history of

Address all correspondence and requests for reprints to L.D. Braida, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

M.A. Clements is now at the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

past spectra in a swept mode) that dichotomized spectral amplitude to be slightly superior to the frequency-amplitude display studied by Snyder *et al.* (20), that included no explicit past history of the spectrum.

Although the comparative results obtained for the different display schemes may be informative, the overall results obtained with the Optacon-based display used in these studies are rather unimpressive. Not only are they inferior to Tadoma, but they also appear to be inferior to those reported for other artificial displays (even when differences in stimulus set, experimental design, etc., are taken into account). Moreover, since the comparisons all used the Optacon stimulator, the conclusions are strictly valid only for displays based on this stimulator. The sensory qualities that this stimulating system can convey are limited to those which result from varying the number and locations of active vibrators. In particular, vibration frequency, amplitude, and waveform are fixed parameters of the Optacon stimulator and cannot be varied for encoding purposes. Thus, to achieve a time-swept display, Clements *et al.* were forced to adopt crude amplitude coding. Greater advantages might have been found for the time-swept display, had more refined amplitude coding been possible. In this regard, it is important to note that Ifukube *et al.* (9), who used a transducer matrix which permitted vibration amplitude to encode acoustic intensity, found significant improvement in consonant reception when time-sweeping was incorporated in the display.

A common weakness of many previous studies of tactile speech display systems is a failure to specify the physical properties of the stimuli precisely, and a concomitant failure to determine the effectiveness of the display scheme to convey these properties. Although many attempts have been made to interpret speech reception results in terms of reception of various acoustic and linguistic features, the analyses are insufficient to determine whether the discriminability of speech segments is primarily restricted by the data compression inherent in the acoustic analysis component of the system, or by limitations of the transducers and sensory system.

The purpose of the present study is to compare 2 new methods of encoding speech for tactile displays, using a vibratory matrix that permits manipulation of vibration frequency and waveform, as well as variation of vibration amplitude across the matrix.

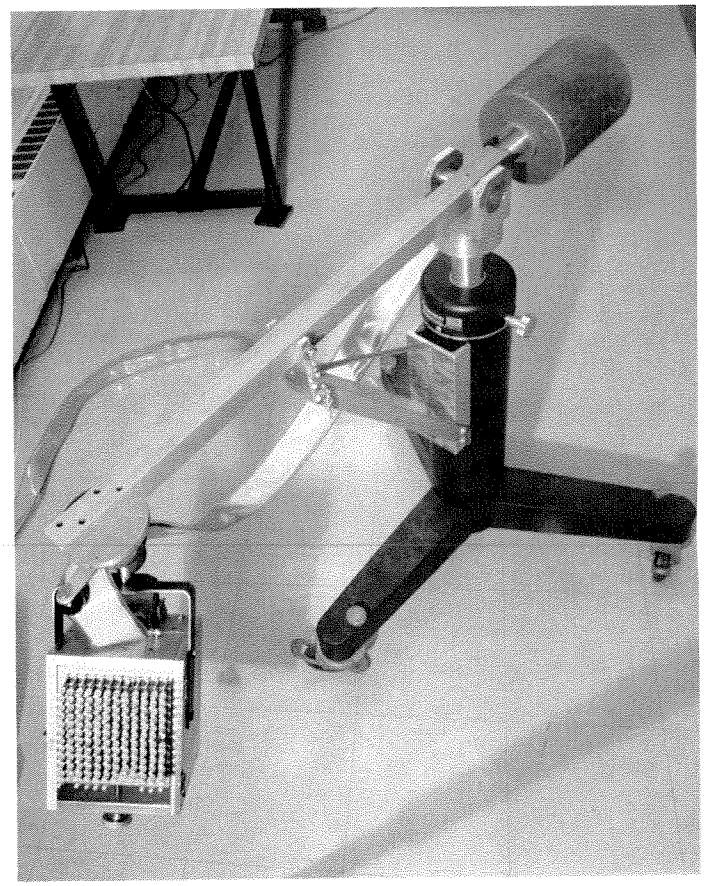

Figure 1.
Tactile stimulating matrix mounted on its supporting structure. The vibrating contactors (nylon screw heads), oriented toward the camera in this illustration, would ordinarily face downward when applied to the thigh.

One display is an adaptation of the conventional time-swept display of the short-term acoustic spectrum studied by Ifukube *et al.* (9). The other is a filled bar-graph representation of an acoustic-tube model of the short-term shape of the vocal-tract shape during speech production. This display represents the articulatory process more directly than the spectral display and permits a large reduction in data rate. Although the representation is derived from short-term power spectra, Shadle and Atal (18) found that it derives shapes that appear to change more smoothly across phoneme transitions than the corresponding spectra. Each of the displays used distinct vibratory waveforms to represent voiced (250 Hz sinewaves) and unvoiced (wideband noise) segments of speech and encoded acoustic intensity in terms of vibratory amplitude.

DISPLAYS

The stimulating system consisted of a 12 × 12 element square matrix (13.3 cm on a side) of piezoelectric bimorph vibrators (16,19) (**Figure 1**), each of which moved in a direction tangential to the skin surface. The surface defined by the bimorph tips was conformable to the shape of the body structure contacted. An arbitrary short-duration "excitation waveform," could be specified during each time *frame*. A set of 144 digital-to-analog converters (DAC's) specified relative vibrator amplitudes (across the array) over a linear range from −128/128 to +127/128 times that of the excitation waveform, while a separate DAC scaled the amplitude of the entire set of vibrators. New amplitude values and a new excitation waveform could be specified for each frame. In the present study, the array was applied to the right thigh of the subject with the tangential vibration aligned along the length of the leg. We use the term *row* to refer to the vibrator coordinate along the leg and *column* to refer to the coordinate across the leg.

Two displays, based on different descriptions of the same acoustic speech waveform, were studied. The *area function* display involved presenting the shape of a model of the vocal tract derived by linear predictive coding (LPC) analysis, and the *spectral* display involved presenting the shape of the short-term speech spectrum.

Area function display

Wakita (22) used linear predictive analysis to derive an estimate of a spatially-sampled vocal tract area function from the acoustic speech waveform. The algorithm used consisted of 4 elements: voice-unvoiced decision, pre-emphasis of voiced segments, vocal tract length and area function estimation, and loudness estimation. The area function display represented the vocal tract area function (computed every 12 msec) in the following manner:

- Each of the 12 rows corresponded to a different acoustic tube segment. The number of rows activated (8,9,10,11, or 12) corresponded to the number of segments in the tube model, which in turn corresponded to estimated vocal tract length.
- The area of each tube section was indicated by the number of vibrators active in the corresponding row of the display. The logarithm of the area was quantized into 1 of 13 values and 0–12 vibrators were activated in the corresponding row centered about the middle column. In preliminary tests, such a filled bar graph representation was found to be superior to an unfilled contour representation.
- Vibration amplitude was varied from frame to frame to make perceived tactile intensity proportional to the loudness of the corresponding acoustic segment. To achieve this, the vibration amplitude for each vibrator was directly proportional to the computed auditory loudness, but inversely proportional to the number of vibrators active during the frame (Cholewiak [4]). Within a given frame, the amplitudes of all activated vibrators were equal. Tactile stimuli were presented at roughly 25 dB sensation level.
- Distinct vibratory waveforms, a 250 Hz sinewave, and a wideband noise were used to represent voiced and unvoiced segments of the speech sounds.

Spectral display

Signal processing for the spectral display consisted of the following steps: filtering of the acoustic waveform by a 12-channel digital filter bank, root mean square (RMS) averaging of the filter outputs over the desired time interval or frame, a voiced-unvoiced decision, and loudness estimation for each filter channel. Filter center frequencies and bandwidths, listed in **Table 1**, approximate the characteristics of auditory critical bands over the 200–4500 Hz range. The envelopes of the filter outputs were displayed in time-swept form, with the vibratory intensity of each channel adjusted to produce perceived intensity roughly proportional to the auditory loudness associated with that channel. Each of 12 rows on the tactile array represented a different frequency band and each of 12 columns represented a different time interval. The duration of each analysis and presentation frame (9 msec) permitted 108 msec of speech to be displayed at any instant, as suggested by Spens (21). (Although we selected this length only after extensive experimentation, its superiority over windows within a factor of 2 in duration was not dramatic.) Voiced and unvoiced intervals received distinct excitation

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 1.
Filter center frequencies and bandwidths (in Hz).

Filter number	Center frequency	Bandwidth
1	250	100
2	350	100
3	450	100
4	574	139
5	731	178
6	932	226
7	1188	289
8	1515	368
9	1931	469
10	2462	598
11	3138	763
12	4000	972

waveforms: 250 Hz sinusoids and wideband noise, as in the area function display.

EXPERIMENTS

To evaluate the ability of users of the display to resolve differences between speech segments, a set of monosyllabic speech materials was used in which most of the American English phonemes were represented. All syllables were spoken 3 times by each of 4 speakers (2 male, 2 female) to minimize artifacts. The speech was recorded in an anechoic environment, using a high-quality microphone and tape recorder. The utterances were segmented by hand, lowpass filtered to 5 kHz, digitized using a 10 kHz sampling rate with 12-bit precision, and then normalized to have equal RMS amplitudes.

The vowel materials consisted of a set of 10 non-diphthongized /b/-V-/t/ syllables, *BAHT, BAT, BEET, BIT, BOUGHT, BOOT, BUHT* (rhymes with "put"), *BERT, BUT,* and *BET.* All 45 possible pairs were used in the discrimination tests. The consonant materials consisted of 24 C-/a/ syllables. Only 55 pairs (selected to test the discriminability of individual articulatory features, and differing by the minimal number of features) were used in the discrimination tests. The specific consonant pairs used are listed in **Table 2.** The same stimulus pairs were tested for both displays and for both subjects.

One male (MD) and 1 female (LD) served as subjects in the main experiments. Both were normal hearing college students; neither was familiar with acoustic phonetics. Two additional male subjects, MC and ND, were used in auxiliary experiments. Subjects were tested individually while seated in a large chair with his or her right thigh exposed. The stimulator matrix was applied to the upper thigh surface with all of the vibrators in contact with the skin. The contact force was roughly 1400 g (10 g per vibrator). Ear plugs and masking noise were used to eliminate possible auditory cues.

To obtain estimates of the discriminability of pairs of tactile patterns with minimum response bias, a 2-interval, 2-alternative forced-choice paradigm was used. The subject was presented with a pair of stimuli, was required to identify the order of presentation, and was then informed of the correct response. The stimuli, each roughly 500 msec long, were presented roughly 500 msec apart. Subjects were allowed to respond when ready. No time limit was imposed, and feedback was presented immediately.

Each discrimination test consisted of 100 trials. The first 10 trials were considered training, and the results discarded. Each trial consisted of the tactile presentation of 1 utterance of each of the 2 words to be discriminated. Each utterance was drawn randomly (with replacement) from a corpus of 12 tokens (4 speakers—2 male, 2 female—3 utterances each).

Training was intended to familiarize the subject with both the test paradigm and the vibratory sensation, and consisted of discrimination tests identical to those used in the testing. Initially, each subject spent roughly 20 hours in such training, using a small subset of the contrasts. Training was considered complete when performance appeared to reach an asymptotic level. When the display scheme or stimulus set changed, several hours were again devoted to training. The testing order of the displays was reversed for the 2 subjects to compensate for possible long-term training effects. A small number of tests were repeated for each condition at the end of the experiments, and indicated that no long-term learning effects were present. In all, each subject

Table 2.
Consonant pairs used in discrimination tests.

A) Voicing contrasts:

1. Plosives: /p-b/, /t-d/, /k-g/
2. Fricatives: /f-v/, /s-z/
3. Affricates: /ch-j/

B) Place contrasts:

1. Plosives: /p-t/, /p-k/, /t-k/, /b-d/, /b-g/, /d-g/
2. Nasals: /m-n/
3. Fricatives: /f-th/, /f-s/, /v-xh/, /v-z/, /s-sh/, /z-zh/
4. Semivowels: /h-wh/, /l-h/, /w-wh/, /w-y/, /l-r/, /r-y/

C) Manner contrasts:

1. Plosive-fricative: /p-f/, /b-xh/, /t-th/, /t-s/, /d-xh/, /k-sh/
2. Plosive-semivowel: /p-h/, /t-l/, /b-w/, /g-y/
3. Plosive-nasal: /b-m/, /d-n/
4. Plosive-affricate: /k-ch/, /d-j/
5. Fricative-nasal: /xh-n/, /z-n/
6. Fricative-affricate: /sh-ch/, /zh-j/
7. Fricative-semivowel: /th-h/, /s-h/, /sh-h/, /r-zh/, /xh-l/, /xh-y/, /z-l/, /zh-y/
8. Nasal-semivowel: /m-w/, /m-r/, /n-l/, /n-y/

All symbols are pronounced as their English equivalent, with the following non-standard cases:

/ch/ as in *chew*	/y/ as in *you*
/xh/ as in *the*	/j/ as in *jaw*
/th/ as in *thaw*	/sh/ as in *she*
/wh/ as in *why*	/zh/ as in *azure*

spent an a total of roughly 100 hours in the evaluations.

RESULTS AND ANALYSIS

The data obtained from the experimental runs were summarized by a 2 × 2 confusion matrix for each syllable pair tested. Each matrix was processed to obtain estimates of sensitivity (d') and bias. The values of sensitivity were then transformed to values of percent correct under the assumption that the response bias was zero. (This correction resulted in an average improvement of only 0.6 percentage points.) In the following discussion, all percent correct scores refer to these *corrected* scores. The average scores for 2 types of display for each of the 2 subjects are presented in **Table 3**. Subject LD achieved higher scores than MD for both displays, and for both stimulus sets. Also, the spectral display led to results superior to those of the area function display for both subjects and stimulus sets. Two-way analyses of variance were employed on results for individual contrasts using arcsine transformed scores to evaluate possible subject and display scheme differences. Although the difference was small, the spectral display scheme was significantly (0.01 level) better than the area function display. The 2 subjects performed at levels that were not significantly different.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 3.
Percent correct for various subsets of data.

Subject	Display/Stimuli	Score	Order
LD	Spectral/Vowels	85.6%	1
LD	Spectral/Consonants	78.3%	2
LD	Area/Vowels	83.6%	3
LD	Area/Consonants	72.0%	4
MD	Spectral/Vowels	85.0%	3
MD	Spectral/Consonants	70.9%	4
MD	Area/Vowels	80.2%	1
MD	Area/Consonants	70.0%	2

Subject/Conditions	Score	Display/Stimuli	Score
LD Overall	79.4	Consonants	72.8
MD Overall	75.9	Vowels	83.6
LD Consonants	75.2	Spectral Overall	79.4
LD Vowels	84.6	Area Overall	75.9
MD Consonants	70.5		
MD Vowels	82.6		
LD Spectral	81.6	Spectral Consonants	74.6
LD Area	77.2	Spectral Vowels	85.3
MD Spectral	77.2	Area Consonants	71.0
MD Area	74.6	Area Vowels	81.9

Vowels

Scores were analyzed to determine: 1) the degree to which classes of vowels were subject to confusion; 2) the characteristics of the vowels used by the subjects to make discriminations; and, 3) how closely simple models of perception could approximate observed performance.

Feature analysis

Vowels were classified by articulatory features using a system suggested by Chomsky and Halle (5), supplemented by the feature retroflexion to differentiate *BERT* from the other vowels, as discussed in the **Appendix.** The reception of individual features was analyzed, and knowledge of the acoustical characteristics of the features used to determine which acoustical differences were important in discrimination. Average discrimination scores for vowel pairs that differed by a given feature, independent of other features, are listed in **Table 4.**

In general, the 2 displays were similar in that the 3 features having greatest influence on F1 (first formant frequency) (*low, high,* and *tense*) were the best perceived. Features affecting F2 and F3 (*round,*

back, and *retroflex*) were less well-perceived than those affecting F1. Unlike the spectral display, in which F1 corresponds rather directly to the location of maximum vibration in the low frequency portion of the display, the area function display represents F1 indirectly in terms of the overall shape of the vocal tract and would be related to the overall "slope" of the area function from glottis to lips. Measurements made on the /b/-V-/t/'s indicated large slope differences for the features *high, low,* and *tense,* but only small differences for the features *back, round,* and *retroflex.* Thus, the average slope of the tract display may have been the perceptual correlate of the acoustical attribute F1.

Table 4.
Performance when vowel pairs differed by at least the listed feature.

Feature	Average Spectral	Average Area
High	89.9%	86.8%
Low	87.8%	89.7%
Back	84.9%	80.0%
Tense	95.9%	87.8%
Round	84.6%	80.6%
Retro	78.8%	75.4%

Multidimensional scaling analysis

A nonmetric multidimensional scaling technique, *KYST*, developed by Kruskal and Wish (10), was applied to the vowel discrimination scores to derive an underlying "perceptual space." In general, 2-dimensional solutions were characterized by sufficiently low values of "stress" that higher order solutions were not considered. **Figures 2, 3, 4, and 5** show the resulting configurations plotted for the 10 vowels. A linear multiple regression technique was employed to interpret the structure of the resulting perceptual space, and to determine how well various simple acoustic and vibratory characteristics of stimuli specify the principal dimensions of the space. Of the 19 characteristics examined, the more salient included duration of the vowel nuclei (DN), word duration (DW), loudness (LS), F1, F2, and vibration amplitude (AM) (area function display only). The measured mutual correlation coefficients of these properties (values greater than 0.63 are significant at the 0.05 level) are listed in **Table 5.**

Table 5.
Correlations between measured properties.

	DW	LS	F2	AM	DN
F1	0.68	0.60	−0.32	0.86	0.65
DW		0.04	−0.23	0.46	0.91
LS			−0.03	0.72	−0.16
F2				−0.41	−0.09
AM					0.26

The correlation coefficients between the projected perceptual distances and the differences in physical parameter values, and the angles of the "best fit" axes for each parameter (where 0 degrees is parallel to the dimension-1 axis), and angles measured counter-clockwise relative to this axis, are listed in **Table 6.** Coefficients with values greater than 0.76 and 0.63 were significant at 0.01 and 0.05 levels respectively.

Table 6.
Correlations and angles of measured properties with 2-dimensional configurations.

	MD AREA		LD AREA	
Property	ρ	Angle	ρ	Angle
F1	0.850	−1.9	0.956	18
DW	0.913	53	0.796	−10
LS	0.787	−71	0.570	−15
AM	0.946	−30	0.814	−13
F2	0.447	−26	0.659	−71

	MD Spectral		LD Spectral	
Property	ρ	Angle	ρ	Angle
F1	0.974	−20	0.850	−34
DW	0.855	−56	0.835	−64
LS	0.912	49	0.736	31
F2	0.324	−28	0.379	59

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

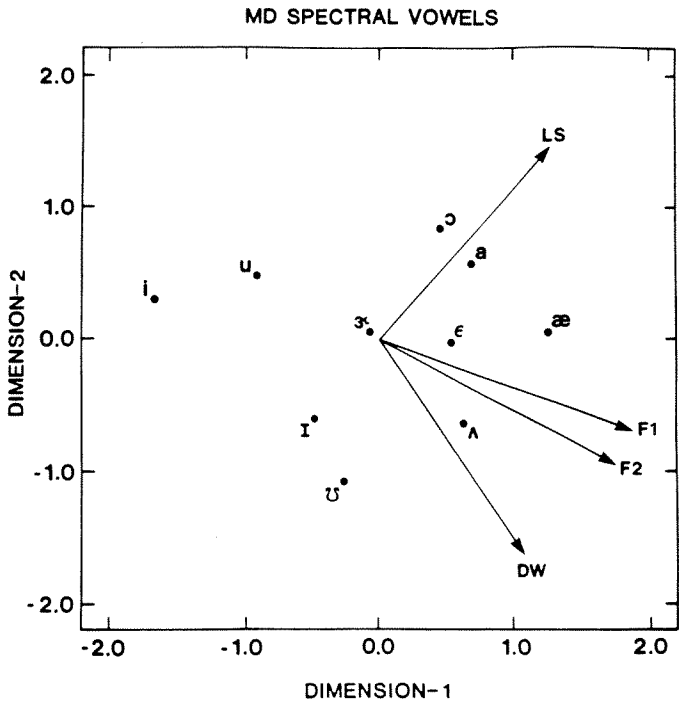

Figure 2.
Multidimensional scaling solution for the vowel discrimination scores of subject MD using the spectral display. Orientations of the axes corresponding to physical stimulus parameters F1, F2, DW and LS are indicated by vectors.

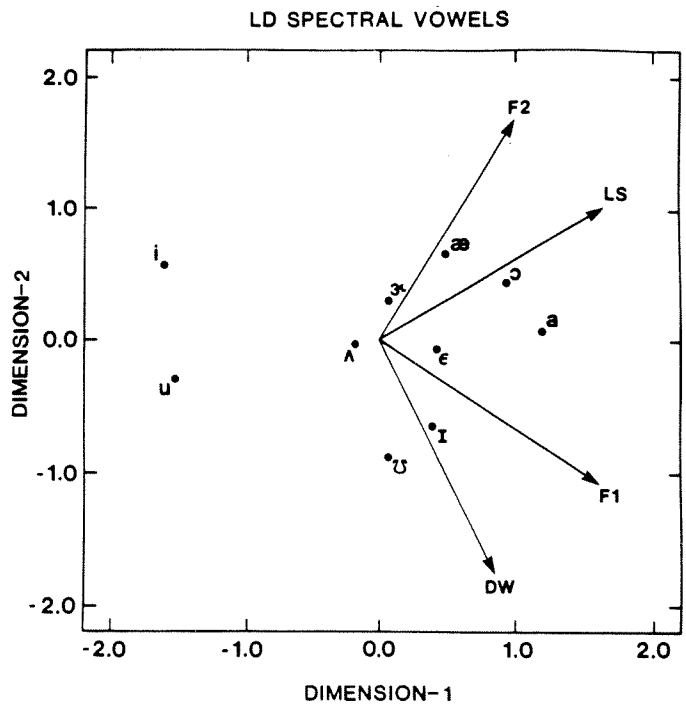

Figure 3.
Multidimensional scaling solution for the vowel discrimination scores of subject LD using the spectral display.

As illustrated in **Figure 2**, for subject MD using the spectral display, F1 resulted in the highest correlation to the MDS configuration of all the tested parameters. An acceptable second dimension may be derived from a linear combination of LS and DW which were found to be roughly orthogonal to F1. For subject LD (**Figure 3**), the correlations were somewhat lower than for MD, but the orientations for properties F1, DW, and LS were similar to those for MD. The stimulus parameters DW and LS (roughly orthogonal to DW) appear to provide the cues for vowel discrimination for LD.

For subject MD using the area function display (**Figure 4**), the MDS configuration was spread roughly twice as much in dimension-1 than in dimension-2, suggesting that only one perceptual quality was very salient. Although F1 is nearly parallel to this dimension, DW could have also been a cue since, on average, syllable pairs differed in duration by roughly 20 percent—roughly the just noticeable difference (JND) in duration for tactile stimuli reported by Geldard (8). AM and DW, as a pair, provide a good description of the perceptual space. It thus appears that the performance of MD on the area function display can be explained in terms of fairly crude stimulus characteristics that have little to do with the detailed shape of the area function. For subject LD (**Figure 5**), F1, which exhibits the highest correlation with the configuration, and F2, which is roughly orthogonal to F1, provide a good description of the perceptual space.

Ideal receiver analysis

An ideal receiver analysis was performed to determine how well the various utterances could be classified on the basis of the tactile signals using simple decision rules. If such a receiver could classify the utterances accurately, poor performance by human subjects would necessarily be linked to limitations associated with the tactile display system or to the perception and interpretation of the display. The signals were first classified by a minimum distance procedure. For each utterance, a vector of the displayed parameter values (the quantized log areas of the 12 tube sections for the tract display; the cube roots of the filter envelopes for the spectral display) averaged in time over the duration of the vowel nuclei was constructed. These data are

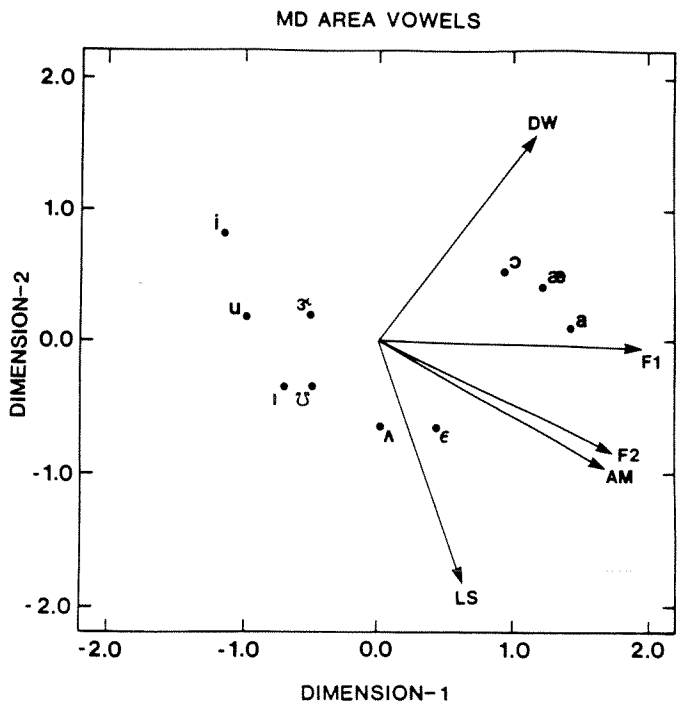

Figure 4.
Multidimensional scaling solution for the vowel discrimination scores of subject MD using the area function display. Orientations of the axes corresponding to physical stimulus parameters F1, F2, AM, DW, and LS are indicated by vectors.

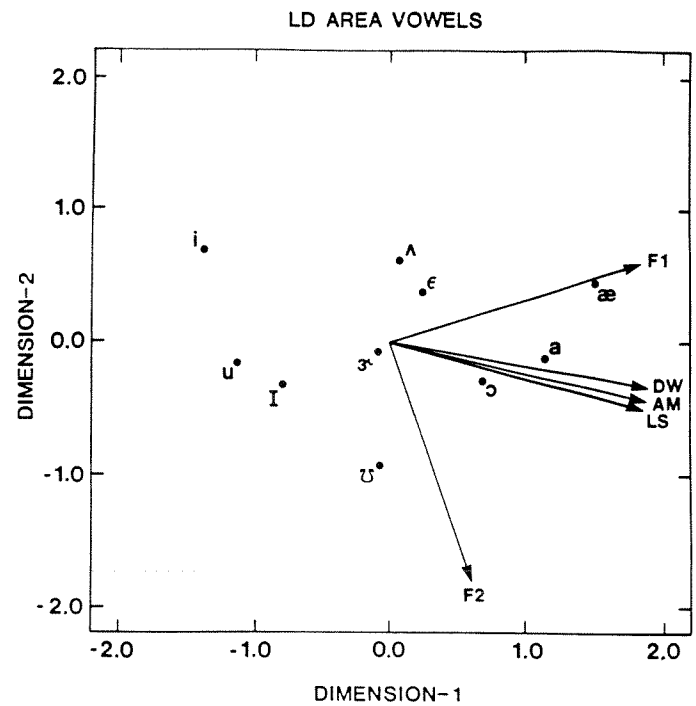

Figure 5.
Multidimensional scaling solution for the vowel discrimination scores of subject LD using the area function display.

simplified versions of the signals applied to the display, and do not include cues such as duration, voicing, and overall amplitude. A mean vector for each syllable was also constructed by averaging the appropriate vectors over the different utterances of the same syllable. A weighted Euclidean distance between each utterance vector and each mean vector was computed, and the utterance was classified as the syllable whose mean vector was closest. The performance of this receiver was perfect for both displays, demonstrating that the area function and spectral parameters of the utterances used in this study were sufficiently consistent to allow the vowel stimuli to be identified, and hence discriminated, without error—even when utterances produced by multiple speakers were included.

To determine the extent to which these physical measures specified the pattern of discrimination performance, we computed the correlation coefficients between the d' data of the discrimination experiments and the measured distances for each vowel pair. These correlations were neither large, nor statistically significant for either display. We also attempted to model certain perceptual effects by degrading and transforming the signals applied to the array, and predicting performance by means of the distance computation. The results observed in the discrimination experiments were then compared with the predicted values to see how well they modeled the perceptual processing of the tactile signals.

We attempted to simulate degraded perception of vibratory signals in the following manner. For the area function display, we further quantized the value of the log-area of each tube segment from 13 to 6, 4, 3, and 2 levels, and spatially averaged the displayed vocal tract shape over 1, 2, 3, 4, and 6 adjacent rows (corresponding to 12, 6, 4, 3, and 2 distinct sections). This smearing considerably improved the correlation between predicted and observed performance. Maximum correlation corresponded to the case of 6 levels of area, and 3 distinct sections. The resulting correlation coefficients were 0.47 and 0.58 (significant at 0.01 level) for MD and LD, respectively. It is interesting to note that the degradation with the highest correlation involved much more smearing in the longitudinal direction than in the transverse direction. Cholewiak (4)

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

previously noted a longitudinal/transverse asymmetry in tasks involving length estimation on the thigh. Also, Boring (2) reported that 2-point discrimination limens are generally smaller in the transverse direction than in the longitudinal direction. The smearing which produced best correlations is roughly consistent with 2-point limens measured by Weinstein (23).

For the spectral display, however, degradations similar to those applied to the area function patterns failed to produce significant correlations with the discrimination data. On the other hand, when the numerical value of the row containing the first formant peak was the sole parameter used, correlations of 0.56 and 0.66 for MD and LD were observed. No significant correlation was observed for the location of the second formant peak. This suggests that the place of maximal excitation was an important cue in vowel discrimination.

DISCUSSION

Articulatory features related to F1 were important determinants of vowel discrimination for both subjects and both displays. However, it is impossible to determine whether discriminations were made on the basis of spectral shape because temporal and intensive properties (e.g., DW, LS) exhibited high correlation with F1 for the stimulus set used. Cues associated with F2 appeared to be important only for LD using the area function display, and even in this case, the salience was not high. For a display to transmit vowels well, it must transmit information concerning the second formant frequency (and perhaps the third) as well as F1. The poor transmission of F2 and related parameters suggests a major shortcoming of both display systems.

Our analyses suggest specific hypotheses about the cues used in vowel discrimination. Feature analysis, MDS, and ideal receiver calculations underscored the relative importance of F1 and unimportance of F2 as determinants of discriminability. For the spectral display, the most salient property of the stimulus appears to be the locus of the most intense excitation, although other factors were probably also involved. For the area function display, the overall tilt of the excitation pattern plays a corresponding role. According to our blurring simulation, the detailed shapes of the area functions were only crudely resolved, with a spatial resolution of 4 cm longitudinally along the thigh, and 2 cm transversely.

Consonants

Consonant scores were analyzed according to articulatory feature distinctions and an ideal receiver model. MDS analysis was not performed, since scores were obtained for only a subset of the possible consonant pairs.

Feature analysis

The features used to classify the consonants are described in the **Appendix**. Since all consonant pairs tested differed in only one feature, the scores were averaged (**Table 7**) to estimate how well the various feature differences were perceived. Since performance followed roughly the same pattern for both subjects, their scores were averaged.

Spectral display. Place of articulation was the least well-discriminated feature with this display. Best place discrimination was found for fricatives (70 percent), with plosives and nasals the poorest. Place is also difficult to perceive auditorily, being cued by spectral differences over brief intervals of time. Voicing distinctions were also poorly perceived (71 percent). Best voicing discrimination was found for plosives, where the distinction is manifest in temporal differences. The poor discrimination of voicing for fricatives indicates that the use of distinct excitation waveforms was ineffective. A set of auxiliary experiments indicated that although in isolation the voiced/unvoiced excitation waveforms could be discriminated tactually with great ease, the intense excitation associated with the succeeding vowels effectively masked these shorter duration differences.

Manner contrasts were well discriminated (80.2 percent). Manner distinctions are associated with a wide variety of acoustic differences: it is not surprising that a wide range of discrimination scores was obtained. Scores were generally better when the acoustical differences were large in terms of both spectral and temporal patterns. For example, very high scores (97 percent), were found between plosives and nasals, and relatively low scores (67 percent) were found for nasals and semivowels. Contrasts between stimuli differing primarily in temporal characteristics (e.g., plosives versus fricatives) or differing primarily in spectral cues (e.g., fricatives versus semivowels) resulted in intermediate scores (roughly 80 percent).

Table 7.
Percent correct scores for the consonant contrasts.

Contrasts	Spectral	Area-function
Voicing		
Overall	71.3	69.0
Plosives	74.8	75.0
Fricatives	68.0	66.3
Affricates	67.5	56.5
Place		
Overall	66.0	64.3
Plosives	63.5	63.7
Nasals	58.5	62.0
Fricatives	70.3	61.8
Semivowels	65.5	67.8
Manner		
Overall	80.2	75.5
Plosive-fricative	80.7	73.7
Plosive-semivowel	82.6	86.6
Plosive-nasal	97.0	88.8
Plosive-affricate	82.3	62.0
Fricative-nasal	75.5	95.8
Fricative-affricate	79.5	61.0
Fricative-semivowel	81.8	78.4
Nasal-semivowel	67.0	59.8
Overall	74.3	70.9

Area function display. The area function display led to results slightly inferior to those for the spectral display, but a similar pattern emerged. Place of articulation was discriminated least well (64 percent) with all manners roughly equivalent in difficulty. Voicing was also poorly perceived, with voicing contrasts for plosives the easiest to discriminate. Even here, the performance was poor (75 percent).

Manner contrasts were better discriminated than those for place and voicing, although some contrasts (nasal-semivowel, fricative-affricate, and plosive-affricate) were not well discriminated. Since the spectral and area function displays were very similar in terms of temporal properties, differences in scores must be attributed to differences in the spectral characteristics of the displays. The spectral display scores can be interpreted plausibly in terms of acoustic phonetic descriptions of the consonants, but no such body of knowledge currently exists for the area function display. For the stimuli used in these experiments, nasals and semivowels were represented by area functions that were closed at the lips; plosives, fricatives, affricates, and unvoiced semivowels by area functions that were open at the lips. This distinction may well have been salient, since manner discrimination scores averaged 89 percent across these 2 groups, but only 67 percent within the groups. Hence, as with the spectral display, only gross temporal and shape differences appear to contribute significantly to discrimination.

Ideal receiver analysis

An ideal receiver analysis was applied to the consonant representations in the same manner as for vowels, except that the data used were derived from a single hand-picked frame for each utterance. For plosives, the frame spanned the burst. For fricatives, nasals, and semivowels, the frame was roughly in the middle of the consonant. For both the area function and spectral displays, perfect classification resulted from the minimum distance decision rule. As with the vowel analysis, no significant correlations between the experimentally-derived and computed d' values was found for either display. In contrast to the case for vowels, predictions were not improved by spatial and/or temporal smearing (by averaging various numbers of frames centered around selected frames).

DISCUSSION

Although both displays led to rather poor consonant discrimination scores, overall performance was better with the spectral display. In place contrasts, the spectral display was appreciably better only for fricatives, and in voicing contrasts, only for affricates. Manner contrasts were generally better discriminated with the spectral display, although certain contrasts were better with the area function

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

display. Most of the differences can be explained by postulating that temporal cues were better conveyed by the time-swept spectral display.

It is difficult to escape the conclusion that displayed differences in shape (whether spectral or area function) were not well-perceived for consonants. The feature analysis strongly suggests that reliable discrimination was possible only when pairs differed significantly both in energy temporal contour and in spatial distribution of stimulation. The poor discrimination of voiced/unvoiced pairs, and of pairs differing only in displayed spectral or area function shapes, may reflect backward masking by the vowel component of the utterances. The ideal receiver analysis, and the failure to find temporal or spatial smearing that could mimic discrimination results, suggests that stimulation pattern differences during the brief consonant portions of the utterances did not greatly contribute to discrimination.

Relation to previous research

In the current study, pairwise discrimination of speech segments was studied for 2 different encoding schemes on the same vibratory matrix using a common pair of subjects. The discrimination results indicate that the 2 displays (one based on the short-term acoustic spectrum, and the other on the short-term shape of the vocal tract) yield roughly equivalent performance. In this section, we discuss these results in light of the previous studies of speech segment discriminability on tactile displays.

Although procedural differences among the studies make precise comparisons tenuous, we believe there is considerable value in such comparisons. The relevant procedural differences include: 1) the tactile stimulator and the body locus employed; 2) the representation of speech used; 3) the specific speech stimuli employed, including the number of talkers and tokens represented in the stimulus set; and, 4) the experimental paradigm employed to measure discriminability. Pertinent differences in such variables are noted in subsequent paragraphs.

Clements *et al.* (6) used the same recorded vowel speech materials and experimental paradigm as the present study; the major procedural differences concerned stimulators and body sites. They also used the tactile stimulation sub-system of the Optacon, a 6 × 24 matrix of piezoelectric bimorphs arranged in a 14 × 30 mm rectangle intended to be applied to the fingertip developed by Bliss *et al.* (1). As discussed previously, vibration frequency (250 Hz) and amplitude are not variable in this system. Speech materials were displayed on the Optacon array in the following manner: 1) the speech waveform was analyzed by a one-third octave analog filter bank; 2) the envelopes of the filter outputs were derived via half-wave rectification and peak detection; and, 3) the logarithms of the envelopes were sampled every 16 msec and used to control the array. Each 16 msec each bimorph was specified as active (excited at 250 Hz at a fixed level) or inactive (not excited). Stimulating patterns differed with respect to the subset of activated bimorphs. In the 2 displays tested, 17 rows of the array represented 17 filter bands (100 to 5000 Hz). Two display schemes were studied, frequency amplitude (FA) and time swept (TS). In the FA display, similar to that used by Yeni-Komshian and Goldstein (26), the 6 columns of the array were used to construct a bar graph of short-term spectral amplitude. The envelope in each band was quantized in steps of 6 dB (total dynamic range of 30 dB). In the TS display, each column of the Optacon matrix represented a different 16 msec time frame (96 msec display window), with rows representing frequency bands. Amplitude was coded dichotomously: a given vibrator was activated only if a specified threshold was exceeded.

Overall vowel discrimination scores of the present study and the Clements *et al.* (6) study (corrected for bias) are presented below.

Present Study		Clements et al. (6)	
Spectral	Area function	Time-swept	Frequency-amplitude
85.3	81.9	81.4	76.1

Two-way analyses of variance indicate that both the spectral and area function displays were superior to the Optacon frequency-amplitude display. Only the spectral display was superior to the time-swept Optacon display. The discrimination scores obtained in the 2 studies were highly correlated ($\rho \approx 0.6$), demonstrating that confusions were similar across the displays.

Pickett and Pickett (12) studied a 10-channel tactile vocoder (200–8000 Hz) in which the envelopes of filter outputs were used to modulate the excita-

Table 8.
Vowels: comparison with Pickett and Pickett (12).

Pair	Pickett & Pickett Raw	Pickett & Pickett Corrected	Area	Spectral	1-Speaker Spectral
BIT-BEET	86	94	89.0	81.0	100.0
BET-BAT	70	77	82.5	64.0	100.0
BAT-BAHT	70	77	70.0	63.0	100.0
BUT-BAHT	82	90	76.0	86.0	90.0
BUT-BOUGHT	79	87	79.0	77.0	85.0
BOOT-BUHT	85	93	88.5	85.0	100.0
Average	78.7	86.3	80.8	76.0	97.5

Raw refers to scores from a 1-interval paradigm.
Corrected refers to scores corrected for a 2-interval paradigm.

tion (300 Hz) of 10 vibrators as applied to the 10 fingers. Vowels were presented in the context /trap/-V-/t/ and consonants were presented in the context /tra/-C-/at/. Recordings were made in an anechoic chamber for 1 male and 1 female speaker. For each vowel or consonant pair, there were 80 1-interval trials with trial-by-trial feedback. Utterances produced by these 2 speakers were not intermixed during testing. When Pickett and Pickett tested vowel discrimination using a fixed pair of utterances, scores increased almost 10 percentage points. Their vowel results (**Table 8**) and estimates of scores that would have obtained in a 2-interval paradigm, indicate discriminability similar to that measured in the present study, with raw and adjusted scores bracketing scores for the area function display. To determine the effect of interspeaker variability on the results of the current study, we conducted auxiliary experiments with utterances produced by only 1 talker at a time using our spectral display. Vowel discrimination was 100 percent for all but 3 pairs, and average consonant discrimination was 89 percent. Thus, the presentation of utterances from multiple speakers in the random fashion used in the present study, appears to influence performance substantially. Our results, using recorded utterances from 1 speaker, are substantially superior to those reported by Pickett and Pickett.

Saunders *et al.* (17) studied a 20-channel vocoder (170–7000 Hz) in which the envelopes of filter outputs were used to control 20 electrocutaneous stimulators applied to the abdomen. Discrimination tests used a 1-interval paradigm, 40 trials per syllable pair, trial-by-trial feedback, and 1 talker. Vowels used an /h/-V-/d/ context and consonants were in the initial position of rhyming 1-syllable words. The scores obtained by Saunders *et al.* (**Table 9**) are similar to those obtained in the present study for common contrasts, and show a rank ordering in degree of difficulty similar to that found with the spectral display. However, as illustrated by the very high scores obtained in single speaker tests of spectral displays, it is inappropriate to take this agreement very seriously.

Weisenberger (24) recently studied a 16-channel vibrotactile spectral display, modeled after the tactile vocoder developed by Brooks and Frost (3), that was applied to the forearm. Stimuli consisted of pairs of words, presented live-voice, that contrasted either the initial consonant or the medial vowel of CVC words. Separate experimental runs tested discrimination for 3 consonant features (voicing, manner, and place), as well as medial vowel. In all the lists, different trials employed different pairs of words, and an ABX procedure was used. Discrimination performance was roughly 65 percent for voicing, 75 percent for manner, 60 percent for place, and 60 percent for vowel contrasts. These scores appear to be somewhat lower than those obtained in the present study, but it is difficult to estimate the effect of live-voice presentation on the scores obtained.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 9.
Vowels: comparison with Saunders *et al.* (17).

	Saunders et al.						
	Naive[a]		Sophist[b]				1-Speaker
Pair	Raw[c]	Cor[d]	Raw[c]	Cor[d]	Area	Spectral	Spectral
BEET-BAT	94	99	98	99	100.0	99.0	100.0
BOOT-BEET	87	94	98	99	69.9	88.0	100.0
BIT-BEET	71	78	90	97	89.0	81.0	
BOOT-BOUGHT	91	97	92	98	96.0	96.9	
BOUGHT-BAT	61	65	77	85	75.0	71.0	100.0
BAT-BET	78	86	75	83	82.5	64.0	100.0
BIT-BET	79	87	57	60	81.5	82.5	100.0
Average	80	86.6	84	88.7	84.7	83.1	100.0

a. Scores achieved by inexperienced subjects.
b. Scores achieved by experienced subjects.
c. Scores refer to 1-interval experiment.
d. Scores refer corrected for a 2-interval experiment.

Table 10.
Vowels: comparison with Tadoma per Reed *et al.* (13).

		Area-function			Spectral			1-Speaker
Pair	Tadoma	Avg	MD	LD	Avg	MD	LD	Spectral
BOUGHT-BAHT	97.5	60.5	60	61	65.0	70	60	90
BOUGHT-BUT	100.0	79.0	91	67	77.0	74	80	85
BUHT-BOOT	95.0	88.5	94	83	85.0	70	100	100
BUT-BIT	70.0	73.0	66	80	78.5	74	83	100
BIT-BEET	55.0	89.0	91	87	81.0	78	84	100
BUT-BAHT	70.0	76.0	66	86	86.0	82	90	90
BAT-BAHT	57.5	70.0	77	63	64.0	54	72	100
BET-BIT	57.5	81.5	77	86	82.5	81	84	100
BUT-BET	45.0	68.5	72	65	72.5	68	77	100
Average	72.0	76.2	77.1	75.3	76.7	81.1	72.3	96.1
Rounded	97.5	76.0			75.7			
Non-rounded	60.8	76.3			77.3			

Finally, the results of the present study may be compared to those reported for Tadoma. Reed *et al.* (13) measured discrimination using an ABX paradigm with feedback for vowel pairs differing by only 1 articulatory feature. Vowels produced by a single speaker were presented live-voice in an /h/-V-/d/ context to normal-hearing subjects (who were trained on Tadoma to an extent comparable to that of the present study). The results (**Table 10**) obtained for Tadoma with a single speaker are similar to those for the artificial displays with multiple speakers, but inferior to those for the spectral display in the single speaker tests. However, in these single-speaker tests, only 3 tokens of each syllable were used; in the live-voice single-speaker tests conducted with Tadoma, a fresh token was produced on each presentation. Discrimination with Tadoma is good only when at least 1 of the syllables contrasted contains a rounded vowel. For the artificial displays, performance was roughly independent of roundedness.

Consonant discrimination with Tadoma was performed as specified for vowels, except that the syllables contrasted on a given trial could differ in any 2 of the tested consonants. The results (**Table 11**) are generally similar to those found in the present study. (Since the same set of contrasts was not tested in both studies, results for a given feature are reported both for the entire set of discriminations, and for the subset that was the same as that used in the Tadoma tests.) Although voicing and manner scores for Tadoma and the artificial displays

Table 11.
Consonants: comparison with Tadoma per Reed *et al.* (13).

		Area-function			Spectral			1-Speaker
Pair	**Tadoma**	**Avg**	**MD**	**LD**	**Avg**	**MD**	**LD**	**Spectral**
Voicing								
p-b	69	75	71	78	73	59	86	90
t-d	60	64	65	62	86	84	87	
k-g	77	87	74	100	67	62	71	
f-v	69	69	67	70	64	62	66	
s-z	73	64	64	64	72	67	77	75
j-ch	63	57	55	58	68	63	72	
Average	68.3	69.0	66.0	72.0	71.4	66.2	76.5	82.3
Place								
p-t	100	59	56	61	68	59	76	100
p-k	100	66	67	65	59	63	54	
t-k	74	68	71	64	71	64	78	
m-n	100	62	59	65	59	59	58	60
f-s	95	56	56	55	67	59	74	
s-sh	65	74	70	77	83	76	90	
l-h	62	96	92	100	63	64	62	
w-wh	68	79	76	81	63	59	66	
r-y	93	55	51	59	68	48	87	
Average	86.8	68.1	66.4	69.7	66.5	61.2	71.7	80.0

(continued on the following page)

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table 11. (*Continued*)
Consonants: comparison with Tadoma per Reed *et al.* (13).

Pair	Tadoma	Avg	Area-function MD	LD	Avg	Spectral MD	LD	1-Speaker Spectral
Manner								
p-f	88	86	81	91	79	73	85	
t-s	80	93	90	96	69	72	66	
k-sh	83	61	58	64	95	90	100	
t-l	80	86	82	89	82	75	89	100
b-w	82	92	90	94	96	94	98	
b-m	78	96	92	100	97	100	93	100
d-n	78	82	73	90	98	95	100	
k-ch	76	64	65	62	88	86	90	
sh-ch	53	65	70	60	88	76	100	
m-v	79	58	64	52	71	76	66	
n-l	62	57	55	59	60	62	57	90
Average	76.0	76.2	74.5	77.9	83.8	81.7	85.8	96.7

were roughly similar, Tadoma scores for place discriminations are much higher. Subject LD was able to achieve substantially better performance on manner and voicing contrasts using the spectral display than the Tadoma subjects.

CONCLUSIONS

Our results suggest that certain of our original goals were not fully attained. In particular, the attempt to present a more refined tactile display than those possible with the Optacon stimulating system, appears in retrospect to have been a failure. The ability of the display system used in the present study to convey distinctions by variations in excitation waveform was largely vitiated as the result of backward masking effects. The ability of the display system to convey spectral shape distinctions by variations in excitation amplitude across the display, likewise appears to have been ineffective: the locus of maximum stimulation appears to have provided the dominant cue for discrimination for the spectral display. Unfortunately, the transducer system used in the present study is not suitable for application to a variety of body loci so that it is not possible to determine whether these conclusions are specific to the body locus used (the thigh), or would generalize to other loci.

By contrast, our attempt to relate the properties of the stimuli used in the discrimination experiments to the patterns of results obtained, appears to have been fruitful. Ideal receiver analysis indicated that perfect discrimination for all stimulus pairs could have been achieved with both displays, using only a small fraction of the tactile stimulus in the absence of transducer or sensory limitations. Thus, the relatively poor performance actually achieved cannot be attributed to any data compression inherent in the acoustic analysis component of the systems studied. Additional modeling suggests that the degradation was rather attributable to a poor match between the properties of the display and the sensory capacity of the skin.

On the other hand, our analysis also points up certain limitations on the ability to specify the stimulus properties that were primarily responsible for the discrimination results. For example, al-

though it is tempting to interpret the performance of the spectral display on vowel discrimination in terms of representation of F1, it is difficult to reject the possibility that the primary cue was, instead, the duration of the utterance. For the specific set of speech sounds used (which included tokens from 4 different talkers), these physical properties were highly correlated, making it impossible to establish unambiguously whether the display was effective at encoding F1 based on these results. In order to do this, it would be necessary to use stimuli (e.g., synthetic speech) which would allow these stimulus properties to be controlled independently.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. 14902, and NSF Grant BNS 84-17817. The assistance of C.M. Reed and W.M. Rabinowitz in all phases of this work is gratefully recognized.

REFERENCES

1. **Bliss JC, Kutcher MH, Rogers CH, Shepard RP:** Optical-to-tactile image conversion for the blind. *IEEE Trans Man-Machine Systems* 11:58-64, 1970.
2. **Boring EG:** *Sensation and Perception in the History of Experimental Psychology.* New York: D. Appleton-Century Company, 1942.
3. **Brooks PL, Frost BJ:** Evaluation of a tactile vocoder for word recognition. *J Acoust Soc Am* 74:34-39, 1983.
4. **Cholewiak RW:** Spatial factors in the perceived intensity of vibrotactile patterns. *Sensory Processes* 3:141-156, 1979.
5. **Chomsky N, Halle M:** *The Sound Pattern of English.* New York: Harper and Row, 1968.
6. **Clements MA, Durlach NI, Braida LD:** Tactile communication of speech II: Comparison of two spectral displays in a vowel discrimination task. *J Acoust Soc Am* 72:1131-1135, 1982.
7. **Clements MA:** A comparison of two tactile speech displays. SM thesis, MIT, 1978.
8. **Geldard FA:** Adventures in tactile literacy. *The American Psychologist* 12:115-124, 1957.
9. **Ifukube T, Yoshimoto C, Shoji H:** A finger-tip tactual vocoder with feature-extracting system. In *Proceedings of Research Conference on Speech Processing Aids for the Deaf,* Gallaudet College, 1977.
10. **Kruskal JB, Wish MJ:** *Multidimensional Scaling.* Beverly Hills: Sage Publications, Inc., 1978.
11. **Norton SJ, Schultz MC, Reed CM, Braida LD, Durlach NI, Rabinowitz WM, Chomsky C:** Analytic study of the Tadoma method: Background and preliminary results. *J Speech Hear Res* 21:625-637, 1977.
12. **Pickett JM, Pickett BH:** Communication of speech sounds by a tactual vocoder. *J Speech Hear Res* 6:207-222, 1963.
13. **Reed CM, Rubin SI, Braida LD, Durlach NI:** Analytic study of the Tadoma method: Discrimination ability of untrained observers. *J Speech Hear Res* 21:625-637, 1978.
14. **Reed CM, Durlach NI, Braida LD, Schultz MC:** Analytic study of the Tadoma method: Identification of consonants and vowels by an experienced Tadoma user. *J Speech Hear Res* 25:108-116, 1982.
15. **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.
16. **Rubin SI:** A programmable tactile display. Unpublished SM thesis, MIT, 1979.
17. **Saunders FA, Hill WA, Simpson CA:** Hearing substitution: A wearable electrotactile vocoder for the deaf. *Sensory Aids for the Hearing Impaired,* H. Levitt, J.M. Pickett, R. Hood (Eds.), IEEE Press, 1980.
18. **Shadle CH, Atal BS:** Speech synthesis by linear interpolation of spectral parameters between dyad boundaries. *J Acoust Soc Am* 66:1325-1332, 1979.

19. **Sherrick CE, Geldard FA:** Princeton Cutaneous Research Project, Report No. 29, 1977.
20. **Snyder JC, Clements MA, Reed CM, Durlach NI, Braida LD:** Tactile communication of speech I: Comparison of Tadoma and frequency-amplitude spectral display in a consonant discrimination task. *J Acoust Soc Am* 71:1249-1254, 1982.
21. **Spens KE:** Is there an optimal time window for tactually conveyed spectral patterns derived from the speech window? In *Proceedings of Research Conference on Speech-Processing Aids for the Deaf,* 119-125. Gallaudet College, 1977.
22. **Wakita H:** Estimation of vocal-tract shapes from acoustical analysis of the speech wave: The state of the art. *IEEE ASSP-27,* No. 3, 1979.
23. **Weinstein S:** Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality. *The Skin Senses,* D.R. Kenshalo (Ed.), Springfield, IL: Charles C. Thomas, 1968.
24. **Weisenberger JM:** Evaluation of single-channel and multichannel tactile aids for the hearing impaired. *J Acoust Soc Am* Suppl. 1:82, S22(A), 1987.
25. **Wicklegren W:** Distinctive features and errors in short-term memory for English consonants. *J Acoust Soc Am* 39:388-398, 1966.
26. **Yeni-Komshian GH, Goldstein MH:** Identification of speech sounds displayed on a vibrotactile vocoder. *J Acoust Soc Am* 62:194-198, 1977.

APPENDIX

Description of Consonant and Vowel Classifications

Vowels are classified by the 6 features: *high, low, back, tense, round,* and, *retro*, and are summarized in **Table A1**.

Table A.1.
Feature attributes of vowels.

Word containing Vowel	Feature					
	Low	Tense	High	Back	Retro	Round
BEET	–	+	+	–	–	–
BIT	–	–	+	–	–	–
BOOT	–	+	+	+	–	+
BUHT	–	–	+	+	–	+
BOUGHT	+	–	–	+	–	+
BAHT	+	–	–	+	–	–
BET	–	–	–	–	–	–
BUT	–	–	–	+	–	–
BERT	–	–	+	+	+	+
BAT	+	–	–	–	–	–

+ signifies presence.
– signifies absence.

The consonants studied were either voiced: /b, d, g, v, xh, z, zh, j, m, n, r, w, l, y/, or unvoiced: /p, t, k, f, th, s, sh, ch, h, wh/.

For place of articulation, we used a 5-level classification suggested by Wicklegren (25):

Place 1: /p,b,f,v,m,w,wh/

Place 2: /t,d,th,xh,n,r/

Place 3: /s,z,1/

Place 4: /sh,zh,ch,j,y/

Place 5: /k,g,h,wh/

Five manners of articulation were used:

1. **Plosives:** /p,t,j,b,d,g/
2. **Fricatives:** /f,th,s,sh,v,xh,z,zh,h,wh/
3. **Affricates:** /ch,j/
4. **Nasals:** /m,n/
5. **Semivowels:** /l,r,y,w,h,wh/

The classifications are summarized in **Table A2**, on the following page.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

Table A.2.
Classification of consonants.

Consonant	Voice	Plosive	Fricative	Affricate	Nasal	SemiVowel	Place
p	–	+	–	–	–	–	1
t	–	+	–	–	–	–	2
k	–	+	–	–	–	–	5
b	+	+	–	–	–	–	1
d	+	+	–	–	–	–	2
g	+	+	–	–	–	–	5
f	–	–	+	–	–	–	1
th	–	–	+	–	–	–	2
s	–	–	+	–	–	–	3
sh	–	–	+	–	–	–	4
ch	–	–	–	+	–	–	4
v	+	–	+	–	–	–	1
xh	+	–	+	–	–	–	2
z	+	–	+	–	–	–	3
zh	+	–	+	–	–	–	4
j	+	–	–	+	–	–	4
m	+	–	–	–	+	–	1
n	+	–	–	–	+	–	2
r	+	–	–	–	–	+	2
w	+	–	–	–	–	+	1
l	+	–	–	–	–	+	3
y	+	–	–	–	–	+	4
h	–	–	+	–	–	+	5
wh	–	–	+	–	–	+	1,5

+ denotes presence.
– denotes absence.