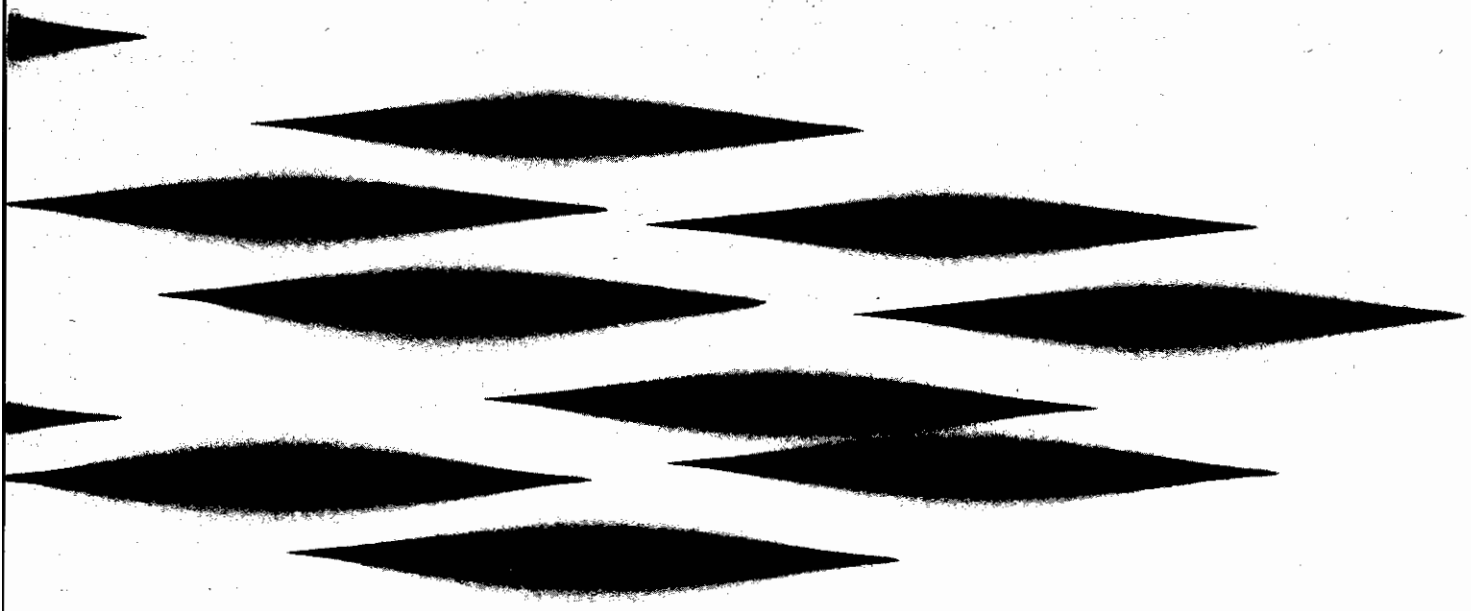
ISSN 0066-071X

RESEARCH ON TACTILE COMMUNICATION OF SPEECH: A REVIEW



ASHA MONOGRAPHS

NUMBER 20 A PUBLICATION OF THE AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION

RESEARCH ON TACTILE COMMUNICATION OF SPEECH: A REVIEW

RESEARCH ON TACTILE COMMUNICATION OF SPEECH: A REVIEW

Charlotte M. Reed
Nathaniel I. Durlach
Louis D. Braida

Massachusetts Institute of Technology
Cambridge, Massachusetts 82139

ASHA Monographs Number 20 (ISSN 0066-071X)

AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION

Rockville, Maryland 20852

May 1982

Contents

Research on Tactile Communication of Speech: A Review

Charlotte M. Reed Nathaniel I. Durlach Louis D. Braida
Massachusetts Institute of Technology, Cambridge

Abstract

Experimental results on the tactile communication of speech are reviewed from the viewpoint of comparative analysis. The review includes studies of Tadoma and spectral displays, studies of tactile supplements to lipreading, and a comparison of Tadoma with lipreading. In addition, comments are made on general research strategy and directions for future research.

Introduction

Recent work has demonstrated that it is possible to communicate speech through the tactile sense. More specifically, it has been shown that highly trained deaf-blind subjects, some of whom lost their hearing and sight at a very early age, can understand running speech by placing a hand on the talker's face and monitoring actions associated with the speech production process (the "Tadoma method"). Detailed experimental results on speech perception performance via this method are available in Norton, Schultz, Reed, Braida, Durlach, Rabinowitz, and Chomsky (1977), and Reed, Durlach, Braida, and Schultz (1982). As a result of this demonstration, investigators can now focus on the development of tactile encoding and display schemes that function at a distance, secure in the knowledge that the tactile sense is not fundamentally inadequate for speech reception.

To evolve an optimally successful scheme, it is important to determine and understand the advantages and disadvantages of each scheme considered. Although the results obtained with Tadoma appear to be superior to those obtained with various laboratory devices, the reasons underlying this apparent superiority are not yet understood. In particular, no one yet knows how to weigh the following factors: (a) the overall richness of the Tadoma display (according to display theory, Tadoma should be superior because the display is more "multidimensional"); (b) the articulatory nature of the display (according to certain theories of speech perception, Tadoma should be superior because it is directly tied to articulation); (c) the use of the hand to sense the display (consideration of both natural tactile processes and tactile physiology suggest that the hand should be an exceptionally good body site); and (d) the extensive training received by the deaf-blind Tadoma users (which is many orders of magnitude greater than the training received with any laboratory device). Similarly, for laboratory systems that use, like the ear, a frequency-to-place transformation (i.e., that analyze the acoustic signal into frequency bands and apply the outputs of different bands to different regions of skin), little is known about the relative merits of different encoding methods, different transducer systems, different body locations, and so on. Examples of studies that have used frequency-place transformations (often referred to as "spectral displays") and that, taken together, represent significant variation over some of these variables can be found in Pickett and Pickett (1963), Kirman (1974), Engelmann and Rosov (1975), Spens (1976), Ifukube, Yoshimoto, and Shoji (in press), Yeni-Komshian and Goldstein (1977), Sparks, Kuhl, Edmonds, and Gray (1978), Saunders, Hill, and Simpson (1980), Scilley (1980), and Clements, Durlach, and Braida (in press). Laboratory systems that use articulatory rather than spectral displays (e.g., displays of vocal-tract shape) will soon be available and will provide further material for comparative analysis and evaluation.

The purpose of this paper is to review recent experimental results on the tactile communication of speech from the viewpoint of comparative analysis. Aside from its focus on comparative analysis and the inclusion of substantial amounts of recent data, the present review differs from those of Kirman (1973, in press) in its concern with detailed experimental results, in the factors stressed in the interpretation of these results, and in the ideas expressed about the constraints that must be satisfied for a tactile communication system to be effective. The ideal studies from the viewpoint of comparative analysis are those in which different schemes, systems, or training experiences are tested under identical conditions. Unfortunately, relatively few studies of this type have been performed; therefore, other studies are included as well. In general, experiments in which the speech reader's inputs are confined solely to the tactile

sense are considered separately from those in which tactile input supplements visual lipreading.

This paper is divided into six sections. Section 1 is concerned exclusively with Tadoma and compares different types of subjects and training experiences. Section 2 is concerned exclusively with spectral displays and compares different encoding schemes and transducer systems. With Sections 1 and 2 as background, Section 3 then presents results comparing spectral displays and Tadoma. Section 4 compares Tadoma and lipreading. Section 5 presents results of studies in which tactile input is used to supplement lipreading. Finally, Section 6 presents some comments regarding general research strategy and future work.

SECTION 1

TADOMA

Tadoma is a method of tactile speech communication based on monitoring the actions present on the face and neck during articulation. The hand of the person receiving speech is placed on the face and neck of the talker. Typically, the thumb rests lightly on the lips and the fingers fan out over the face and neck. Although it is possible that significant information is derived in this method from kinesthetic cues associated with passive movements of muscles and joints generated by movements of the lip and jaw against the thumb and fingers, it appears that the primary perceptual cues in Tadoma are based on sensations of "surface touch." Consequently, throughout this report we refer to Tadoma, as well as any method based on an artificial vibratory or electrocutaneous display, as a method of "tactile" speech communication. The method originated in the education of a deaf-blind adolescent in Norway in the 1890s by a teacher named Hofgaard (Hansen, 1930). It was introduced in the United States by a teacher of the deaf named Sophia Alcorn, and it was subsequently used in the education of deaf and deaf-blind children in various schools throughout the country (e.g., see Alcorn, 1932; Gruver, 1955; van Adestine, 1932; Vivian, 1966).

As already indicated, it is unclear to what extent the superior speech reception performance achieved by highly experienced deaf-blind Tadoma users is due to the method itself and to what extent it is due to the extensive training incorporated with the method. To help answer this question, and to permit comparisons between Tadoma and laboratory systems on comparable subjects with comparable amounts of training, efforts are underway to study Tadoma with relatively inexperienced subjects as well as highly experienced users.

At present, speech reception data using the Tadoma method are being accumulated on three classes of subjects: (Class A) deaf-blind subjects who lost their hearing and sight at an early age and have used Tadoma extensively for many years (and, in certain cases, did most of their language learning through Tadoma); (Class B) subjects who have normal hearing and vision and have received little previous training in Tadoma; and (Class C) deaf-blind subjects who have received little previous training in Tadoma.

Detailed comparisons between one subject in Class A and two subjects in Class B for tests of discrimination and identification of closed sets of nonsense monosyllables were reported previously (Reed, Rubin, Braida, & Durlach, 1978; Reed et al., 1982; Reed, Doherty, Braida, & Durlach, in press). According to these results, performance in these tasks is roughly comparable for the two classes of subjects. In the discrimination task, in which all that is required of the subject is the ability to respond differentially to two different tactile patterns (and not to associate a specific pattern with a specific monosyllable), the Class B subjects were able to perform as well as the Class A subject with very little training (i.e., less than 2 hours). In the identification task, in which the subject is required to identify the monosyllable corresponding to the tactile pattern, the performance of the Class B subjects after approximately 100 hours of training was comparable to (or better than) that of the Class A subjects. Further data obtained on eight additional Class A subjects (Reed, Conway-Fithian, Braida, Durlach, & Schultz, 1980) indicate that the single Class A subject used in these comparisons is one of the most proficient Class A subjects available (one of the top three studied).

Given the apparent equivalence between the Class A and Class B subjects in these discrimination and identification tests (after only modest amounts of training for the Class B subjects), the overall performance for these subjects in these tests can be summarized as follows: In discrimination tests using an ABX procedure and a variety of monosyllabic materials, scores were in the range of 70–100% correct ($p = .50$). The highest scores were obtained for the easy task of discriminating between randomly selected pairs of monosyllabic words (W-22 lists). The lowest scores were obtained for discriminating pairs of vowels in a CVC context (consonants fixed) where most of the pairs were selected to differ in only one articulatory feature (e.g., round, tense, back, high, low) and for discriminating between consonant clusters and a single consonant from the cluster. Intermediate scores were obtained for word pairs selected from the *Modified Rhyme Test* (each pair was selected from the six alternatives specified in the subtests) and for pairs of consonants in CV or VC syllables (vowel constant) contrasting single articulatory features (voicing, manner, and place). In identification tests involving a set of 24 consonants in CV syllables (vowel fixed), scores were in the range of 50–75%. In identification tests involving a set of 15 vowels and diphthongs in CVC syllables (consonants fixed), scores were in the range of 55–80%. In general, the higher scores in these tests were obtained by the Class B subjects. An analysis of confusions made in the identification tests in terms of various articulatory/phonological features indicates that similar features were well-received by the two classes of subjects. The features of *voicing, place, frication,* and *round* were well-received for consonants, as were the features *round, tense,* and *vertical lip separation* for vowels.

The results of the discrimination and identification tests, together with basic knowledge about the speech production process, our own experience as laboratory subjects, and measurements of various face and neck actions (e.g., Hansen, 1964), have led us to conjecture that the primary physical actions sensed by the Tadoma reader are lip movements, jaw movements, laryngeal vi-

bration, and oral air flow. Muscle tension and nasal air flow are thought to play a secondary role. The characteristics of the cue system employed are being explored further through experiments that measure the speech reception effects of limitations in hand position (e.g., thumb removed from the lips, fingers removed from the cheek and neck, etc.).

Although identification experiments with very large sets (such as all monosyllabic words) have not been performed on Class B subjects, it is obvious that these subjects would perform very poorly unless they were trained extensively. Scores for Class A subjects on such a test (randomly selected W-22 words) were in the range of 26–56%. It is also clear that Class B subjects could not begin to approach the percent-correct scores obtained by Class A subjects on sentence tests (e.g., 70–85% on the CID sentence test at slow-to-normal speaking rates) unless they received much more training. For the more proficient and experienced deaf-blind Tadoma users, the relation between performance on sentences and performance on isolated speech segments shows substantial use of context; roughly speaking, these subjects achieved speech-reception results for both segments and sentences comparable to those obtained in audition when listening with a signal-to-noise ratio in the range 0–6 dB.

A preliminary probe of the abilities of Class B subjects to learn to comprehend sentence material was reported by Reed et al. (in press). The training procedure involved the gradual acquisition of a 43-word vocabulary that, although presented initially in isolation, was used primarily in sentence context. The vocabulary was made up of various parts of speech (including nouns, verbs, pronouns, adjectives, prepositions, etc.) and was selected so that it could be used to generate a very large set of sentences with a variety of grammatical forms. Two subjects spent 36 hours each in training and testing sessions in which they responded to sentences that were repeated until a correct word-for-word response was achieved. Tests on identification of sentences that were formed from the 43-word vocabulary and averaged 7 words in length indicated that the subjects perceived 20–45% of the words correctly on the first presentation. The average number of presentations required for a correct repetition of a sentence was 3 for one subject and 4.5 for the other. These results indicate that, within the confines of a small, fixed vocabulary, Class B subjects can learn to perceive connected speech with a fair degree of accuracy following only modest amounts of training.

A special group of Class B subjects that would be interesting to test is the small set of subjects identified by Craig (1977) who showed a remarkable ability to read text with the Optacon (Linvill & Bliss, 1966). The Optacon is a reading aid for the blind in which printed text is converted to vibrotactile patterns and applied to the finger. It consists of two major components, a camera with a 24 × 6 array of photosensitive elements and a transducer array composed of 24 × 6 piezoelectric vibrators. As the camera scans the printed material, the vibrators are activated in a pattern corresponding to the visual images of successive alphanumeric characters. The two extraordinary observers identified by Craig were able to read 70–100 words per minute (wpm) on the Optacon following only several hours of practice, compared to an average reading rate of 20 wpm for blind users of the Optacon after a 50-hour training period. Although Craig attempted to relate the ability of these subjects for Optacon reading to underlying perceptual and cognitive abilities, no one has yet examined how quickly these subjects could learn to understand speech through Tadoma.

Informal training was conducted with one Class C subject who is congenitally deaf and became blind in early adulthood. This subject initially received training on discrimination and identification of consonants and vowels. Her performance on consonant and vowel identification after a short period of training (70% and 55%, respectively) was comparable to that of the Class A and Class B subjects just described. This subject also received training on perception of continuous speech through Tadoma (including practice on sentences, children's stories, and simple conversations). Although a formal evaluation of the subject's performance on connected speech has not been conducted, it appears that after about 50 hours of practice she is able to conduct simple conversations with only a modest number of errors.

In general, the results of the experimental studies of Tadoma thus far indicate that (a) there are deaf-blind individuals who can understand continuous speech at reasonable rates and with reasonable accuracy via Tadoma; (b) some of these individuals lost hearing and sight in infancy and have learned language and how to speak, as well as how to receive speech, primarily through Tadoma; (c) the ability of relatively inexperienced observers to discriminate and identify speech segments is comparable to that of experienced subjects (i.e., the basic tactile sensitivity of Class A subjects is no different from that of Class B and C subjects); and (d) the results of preliminary studies involving training of Class B and C subjects to comprehend continuous speech, although incomplete, indicate that significant progress towards this goal can be achieved with only modest amounts of training.

SECTION 2

SPECTRAL DISPLAYS

Spectral displays may be thought of as rough analogues of the spatial frequency analysis performed by the peripheral auditory system. In these displays, which are often referred to as "tactile vocoders," the acoustical signal is analyzed into frequency bands and the envelopes of the outputs of these bands are used to drive an array of stimulators in such a way that frequency is transformed into place of stimulation. The specific way in which information concerning frequency, amplitude, and time is encoded varies among the systems. Among the factors that vary over the different systems tested are the geometry of the array (linear vs. planar), the number of stimulators (a few to a few hundred), the type of stimulation (mechanical vs. electrical), the portion of the body stimulated (finger, hand, arm, thigh, stomach), the type of encoding scheme (e.g., how amplitude is represented), and a variety of parameters related to the detailed signal processing (choice of filters, time constants, sampling rates, etc.). In most linear displays, amplitude is encoded directly as vibration intensity. In planar arrays, amplitude is encoded either as vibration intensity or, using the second dimension of the array, as place of stimulation. If the second dimension is not used to encode amplitude, it is often used to "time-sweep" the pattern of vibration across the skin (simulating certain aspects of the motion associated with natural tactile exploration). In general, in the following discussion, these factors are specified for each of the studies considered.

Our review focuses on those studies that have attempted to evaluate performance on speech materials in a detailed manner, and it is organized primarily according to the geometry of the array. One major class of transducer systems is characterized by a unidimensional array of stimulators in which *location of stimulation* corresponds to acoustic frequency region and *intensity of stimulation* at a given location corresponds to acoustic amplitude for that frequency region. Such devices were used in some of the early research on tactile speech communication (Gault & Crane, 1928; Wiener, Wiesner, David, & Levine, 1949-1951), as well as in several more recent studies (Biber, 1961; Engelmann & Rosov, 1975; Guelke & Huyssen, 1959; Oller, Payne, & Gavin, 1980; Pickett & Pickett, 1963; Saunders et al., 1980; Scilley, 1980).

The second major class of transducer systems is characterized by a two-dimensional array of stimulators in which frequency is encoded in one dimension and either amplitude or time in the other. In most of these systems, the vibrators are activated at some constant intensity. Amplitude is encoded by defining a dynamic range for each channel, dividing this range into a prescribed number of steps, and presenting either a bar graph or contour display of this number in the second dimension of the array. When time is encoded in the second dimension, the spectral values from a given instant in time are presented along with previous values and the most current value. The total "time window" of the display depends on the sampling rate and the size of the dimension used for encoding time. Studies of frequency-amplitude displays include those of Yeni-Komshian and Goldstein (1977), Mook (1978), Snyder, Clements, Reed, Durlach, and Braida (in press), Sparks et al. (1978), and Clements et al. (in press). Studies of time-swept frequency displays include those of Kirman (1974), Spens (1976; in press), Ifukube et al. (in press), and Clements et al. (in press). Following a review of these studies, which are summarized in Table 1, various encoding schemes and transducer systems are compared.

REVIEW OF SELECTED STUDIES OF SPECTRAL DISPLAYS

Guelke and Huyssen (1959) tested a device in which the output of each of eight 300-Hz bands in the region 410–2880 Hz was lowered into the 100–400-Hz region (by heterodyning) and applied to electromagnets with a number of mechanically tuned reeds. The electromagnets were applied to eight sites on the fingers and palm of one hand. The evaluation of the device was concerned with transmission of segmental speech elements. Informal observation indicated that vowels were easily discriminable and that absolute identification of five vowels approached 80–90% correct after a short training period. Consonants were less easily discriminated than vowels. Voicing and manner distinctions among consonants were better perceived than place distinctions. Poorer performance on consonants compared to vowels was attributed to their lower power, shorter duration, wider frequency distribution, and greater dependence on context, and to the greater importance of frequencies above 3000 Hz for consonants.

Biber (1961) presented speech to the forearm using the tactile cochlear model designed by von Bekesy (1955). This device consists of a plastic tube cast around a brass tube, forming a membrane along a slit in the inner tube. The width of the membrane is constant along the length of the fluid-filled tube although its stiffness varies with a gradient of 10:1. Traveling waves are produced when a bellows inserted into the tube is compressed and dilated by moving an attached piston. Biber stimulated the device using speech that was slowed down by factors of 2, 4, and 8 (to bring the frequencies into a range more compatible with the frequency response of the skin). Three groups of words, each composed of a different set of phonemes, were used for training and testing. The first group consisted of phonemes with spectral energy concentrated in the high-frequency region (/i, e, ä, st, s, z/), the second group consisted of mid-frequency

TABLE 1. Summary of studies of spectral displays of speech.

Study	Type of stimulation	Geometry of array	Body site	Encoding scheme[a]	Amplitude encoding	Test materials[b]	Task[c]	Duration of training[d]	Number of subjects	Average score[e]
Guelke & Huyssen (1959)	Mechanical	Linear	Fingers & palm	Static	Vibration intensity	V Pairs	D	–	1	"Easy"
						C Pairs	D	–	1	"Harder"
						5 V	I	6 hrs	1	80–90%
Biber (1961)	Mechanical	Linear	Forearm	Cochlear	–	Words	I	12-32 hrs	4	75–100%[f]
Pickett & Pickett (1963)	Mechanical	Linear	Fingers	Static	Vibration intensity	14 V Pairs	D, 1-I	–	2	83%
						19 C Pairs	D, 1-I	–	2	80%
						6 V	I	12 hrs	2	60%
						6 V	I	8 hrs	2	42%
Engelmann & Rosov (1975)	Mechanical	Linear	Fingers, forearm, or thigh	Static	Vibration intensity	60 Words	I	70-80 hrs	2	90%
						165 Words	I	170 hrs	1	80%
Saunders et al. (1980)	Electrical	Linear	Abdomen	Static	Biphasic pulse rate	10 V Pairs	D, 1-I	–	8	80%
						11 C Pairs	D, 1-I	–	8	72%
Scilley (1980)	Mechanical	Linear	Forearm	Static	Vibration intensity	70 Words	I	40 hrs	1	80%
						150 Words	I	55 hrs	1	80%
Oller et al. (1980)	Mechanical	Linear	Forearm	Static	Vibration intensity	6 Word Pairs	D, 1-I	–	8	73%
Yeni-Komshian & Goldstein (1977)	Mechanical	Planar	Fingers or palm	Static	Bar graph	3 V	I		10	75%
						4 Spondees	I	15 hrs	10	71%
						4 Durations	I	70 hrs	10	94%
Mook (1978)	Mechanical	Planar	Finger	Static	Bar graph	12 C	I	70 hrs	2	30–40%
Snyder et al. (in press)	Mechanical	Planar	Finger	Static	Bar graph	32 C Pairs	D, 2-I	–	2	68%
Sparks et al. (1978)	Electrical	Planar	Abdomen	Static	Unfilled contour	8 V	I	15 hrs	1	95%
						8 Plo/Nas	I	10 hrs	2	50%
						9 Fric.	I	8 hrs	1	70%
Kirman (1974)	Mechanical	Planar	Hand	Swept	None	15 Words	I	40 hrs	6	80%
Spens (1976)	Mechanical	Planar	Finger	Swept	Vibrator pattern	10 Words	I	6 hrs	1	94%
Ifukube & Yoshimoto (1974)	Mechanical	Planar[g]	Fingers	Static	Vibration intensity	5 V	I	–	4	91%
						5 C	I	–	4	91%
Ifukube et al. (in press)	Mechanical	Planar	Fingers	Swept	Vibration intensity	5 C	I	–	4	76%
Clements et al. (in press)	Mechanical	Planar	Finger	Static	Bar graph	45 V Pairs	D, 2-I	–	2	83%
				Swept	Dichotomous	45 V Pairs	D, 2-I	–	2	87%

[a]Static or Swept. Static refers to those encoding schemes in which the value of the display at a given time is dependent solely on the value of the most recent sampling of the signal. Swept refers to those displays in which memory of previous sampling values is incorporated into the display.
[b]V = Vowel, C = Consonant, Plo = Plosive, Nas = Nasal, Fric = Fricative.
[c]D = Discrimination, I = Identification, 1-I = One-Interval-Forced-Choice Procedure, 2-I = Two-Interval-Forced-Choice Procedure.
[d]Hours of training are listed only for those studies in which at least 5–6 hours training was conducted prior to testing.
[e]The percent-correct scores obtained in the particular task employed by the authors of a given study.
[f]Slow playback by a factor of 4.
[g]Although Ifukube and Yoshimoto (1974) used a two-dimensional array of vibrators (16 × 3), identical spectral information was presented across the three columns associated with a given row.

phonemes (/a, au, eu, l, g, p/), and the third group consisted of low-frequency phonemes (/o, u, m, n, w/). Although the details of the experiments are not clear, it appears that seven monosyllabic words formed from each of the three groups of sounds were used in the training sessions; the tests consisted of these 21 words plus other words (perhaps an additional 20) formed from the phonemes within each of the three groups. Four subjects participated in the experiments. Performance was highest for slow playback by a factor of 4, and results are summarized for this condition. For one subject, performance reached 100% on recognition of the phonemes in the 21 familiar words after about 12 hours of training. For the other three subjects, correct phoneme recogni-

tion for the familiar words reached 75–90% after roughly 32 hours of training. Scores for phoneme recognition in the unfamiliar words, reported for two subjects, averaged 80% following approximately 30 hours of practice.

Pickett and Pickett (1963) evaluated a tactile vocoder consisting of 10 filters with center frequencies in the range of 210–7700 Hz. The envelope of each band modulated the amplitude of a 300-Hz sine wave that activated a bone-conduction vibrator. These vibrators stimulated the fingertips of both hands. Tests of pairwise discriminability of consonants and vowels were conducted (using a one-interval-two-alternative-forced-choice procedure) as well as additional tests on identification of small sets of vowels. The discrimination tests were conducted following a short period of familiarization with the two stimuli for each test, whereas more extensive training was provided for the identification tests. The average score for discriminating pairs of vowels was 83%.[1] Discriminability was related to separation in the F1-F2 space, with higher scores obtained on pairs of vowels with greater separation along the two dimensions. Additional tests of vowel identification used two sets of six vowels each: One set was chosen to maximize the average distance in the formant space and the other set was composed of the remaining vowels. For the set chosen to be "easy," performance averaged 60% after four training sessions. Lower scores were obtained with the second set of vowels, with scores averaging roughly 40%. The average score for discriminating pairs of consonants that differed by voicing, nasality, manner, or place of articulation was 80%. Subjects reported that voicing and nasality distinctions were cued by differences in attack and decay characteristics. The stop/continuant distinction was better perceived for voiceless than voiced sounds. Place of articulation for fricatives was well-perceived.

Engelmann and Rosov (1975) examined the effects of prolonged training on speech perception through a 23-channel vocoder whose filters covered the frequency range 85–10,000 Hz and whose bandwidths increased with center frequency. The envelopes of the outputs of the various bands modulated the amplitude of 60-Hz solenoid vibrators that were applied to the fingers, forearms, or thighs. The normal-hearing subjects were trained to identify words from a 60-item list. Two subjects achieved scores above 90% on words randomly selected from the complete list after 70–80 hours of training. The rate of vocabulary acquisition appeared to be inversely related to the strictness of the criterion used for introducing new words (as would be expected), but performance was roughly independent of stimulation site and was also observed to transfer readily when the solenoids were moved from the forearm to the leg. Learning rates appeared to accelerate during the tests, and sentences constructed from words in the list were perceived easily. The four deaf children, aged 8–14 years, were trained for approximately 50–170 hours primarily to recognize isolated words presented through solenoids placed on their thighs. The most proficient subject achieved a score of 80% correct on a test involving 165 words after about 170 hours of training. For this subject, vocabulary acquisition rates increased dramatically during the training period from .5 words per hour during the first two-thirds of the program to 2 words per hour during the final third. The other three deaf subjects achieved far less impressive results either in terms of vocabulary acquisition or learning rates.

Saunders et al. (1980) performed preliminary tests using a 20-channel electrotactile vocoder in which the envelope of the output of each of 20 filters controlled the number of biphasic, constant-current pulses applied to electrodes mounted in a belt worn around the waist. The filters had center frequencies ranging from 190 Hz to 6200 Hz and bandwidths of approximately ⅓ octave. Discrimination of vowels (averaged across 8 subjects and 10 vowel pairs) was 80% and discrimination of consonants (averaged across 8 subjects and 11 consonant pairs) was 72%. One subject was also trained to identify a vocabulary of approximately 33 nouns, verbs, and pronouns in isolation (with scores ranging 72–87% after 1 hour of training on each class of words), and in sentences (where informal tests indicated that most sentences were identified correctly within the first three presentations).

Scilley (1980) studied the effects of prolonged training on vocabulary acquisition using a 16-channel tactile vocoder that activated solenoids at a 100-Hz rate applied to the forearm. The ⅓-octave filters had center frequencies in the range of 200–8000 Hz. Two normal subjects participated in the study, which involved exposure to tactile representations of their own speech as well as that of a variety of other speakers. One subject achieved 80% identification scores on a list of 70 words after 40 hours of training. A second subject achieved this level of performance on a 150-word list after 55 hours. Average vocabulary acquisition rates for the two subjects were 1.7 and 2.7 words/hr., with no indication of rate acceleration as training progressed. The subjects were able to generalize readily to different speakers and to changes in stimulation site. Scilley analyzed the perception of speech through the tactile vocoder in considerable detail. Pairwise discrimination tests with words confused in the identification tests and with rhyming words contrasting single phonetic elements indicated that discriminability was high in most instances. An information theoretic analysis of the confusions indicated that information transfer reached plateaus for both subjects (4.3 and 5.3 bits) after roughly 20 hours of training. Beyond this point information transfer remained fairly constant; although identification scores of 80% continued to be achieved as vocabulary size increased, error patterns became increasingly random. Thus, while the discrimination tests are encouraging with respect to the basic reso-

[1]We modified the scores reported by Pickett and Pickett by undoing their "correction for guessing" to make their results more directly comparable to those reported by other investigators. Our modified score was obtained by the following computation: Percent Correct = (Reported Score/2) + 50%.

lution of the display, the information measures suggest that it may not be possible to achieve further increases in vocabulary size while maintaining the 80% performance criterion.

Scilley also reported some results of tests conducted with a 13-year-old postlingually deaf subject. This subject was able to use the vocoder to identify 50 environmental sounds fairly accurately and to identify small sets of words whose contrasting elements were highly confused through lipreading. A pilot study indicated that the use of the vocoder improved the intelligibility of the subject's productions of short phrases.

Oller et al. (1980), using the same device as Engelmann and Rosov (and applying it to the thigh), tested the discriminability of six pairs of words in which the contrasted phonetic elements were difficult to distinguish through lipreading. After a brief training period, eight hearing-impaired subjects received a one-interval two-choice, 40-trial test (with presentations evenly divided between a male and a female speaker) with each pair. The average score across the six word pairs for individual subjects was in the range of 58–86%, with a mean of 73%. The average score on individual word pairs ranged from 60% (for three pairs of words that contrasted one phonetic element each) to 70% (for one pair which also contrasted one phonetic element) to 95% (for two word pairs each of which contrasted two phonetic elements per word). Although the authors attributed the differences in scores across word pairs to a factor related to the percentage of vibrators activated by different words, this factor appears to be confounded with the number of phonetic elements by which a pair of words differed.

Yeni-Komshian and Goldstein (1977) presented a two-dimensional display of frequency and amplitude through the Optacon transducer (Linvill & Bliss, 1960), a finger-sized 24 × 6 array of piezoelectric vibrators. Frequency was encoded in 18 of the 24 available rows of the Optacon and intensity was encoded by means of a bar-graph display in the six columns. The filters covered the frequency range of 250–7700 Hz and their bandwidths increased with frequency; intensity covered a 20–25 dB range in each channel. Subjects, using their fingertips or palm to sense the display, were trained in three tasks: identification of the three cardinal vowels, four spondaic words, and four durations of a synthetic vowel. At the end of 30 half-hr. training sessions, subjects averaged 75% correct on vowels, 71% correct on spondaic words, and 94% correct on durations.

Mook (1978) studied consonant identification using an Optacon-based frequency-amplitude display (similar to the display used by Yeni-Komshian & Goldstein) applied to a finger. The center frequencies of the ⅓-octave filters were in the range of 160–5000 Hz, and intensity covered a range of 30 dB in each channel. Subjects were trained to identify 12 consonants in CV syllables with two vowels recorded three times each by each of four talkers. After approximately 40 hours of training, consonant identification scores were roughly 30–40% correct. Subjective reports indicated the use of perceptual cues related to the features of voicing, nasality, frication, and place. Analysis of confusion matrices in terms of these features suggested that voicing and nasality were well-perceived, whereas frication was well-perceived only when voicing was absent. Comparable tests performed with an analogous visual display led to scores of approximately 60–70%.

Snyder et al. (in press) studied consonant discriminability using the same system as that studied by Mook. Thirty-two pairs of consonants contrasting the features of voicing, manner, and place were tested in "live-utterance" CV syllables with three vowels using a two-interval paradigm. The consonant contrast was held fixed throughout a run, but the vowel was varied randomly from trial to trial. The average discriminability of the consonant pairs was 68%. The manner feature was best perceived (78%), but discriminability of voicing and place was poor (55% and 56%, respectively). The comparative results obtained on Tadoma in this study (using the same materials, procedures, and subjects) are discussed next.

Sparks et al. (1978) evaluated the MESA electrotactile display (worn as a belt) in which 36 horizontal channels are used to code frequency and 8 vertical channels to code intensity. The filters had center frequencies in the range of 85 Hz–10,500 Hz and bandwidths that increased with center frequency (from 19 Hz for the lowest channel to 2350 Hz for the highest channel). For each frequency channel, a 40-dB dynamic range was divided into eight 5-dB steps, each of which corresponded to one of the eight vertical channels. Only one electrode in each of the 36 columns was active at a given time, producing an unfilled amplitude contour (rather than the entire area beneath the contour). Subjects were trained to identify sets of consonants and vowels (using "live-voice" utterances) under conditions of the tactile aid alone, lipreading alone, and the combined mode. Only the results obtained on the tactile aid alone are discussed here. Recognition of eight vowels in /b/-V-/d/ context presented by one talker over 35 training sessions reached 95% for the one subject tested. In another test with three talkers (two unfamiliar to the subject) the same subject obtained a score of 76%, indicating a reasonable ability to generalize across talkers. The major confusions for all talkers were among the vowels /ɛ, æ, ɑ, ʌ/. Average recognition scores for eight plosive and nasal stimuli following roughly 20 training sessions were 50% for the two subjects tested (whose training protocols were slightly different). Recognition of a set of nine fricatives after eight training sessions levelled off at 70% for the one subject tested. Perception of voicing was better than that of place on both sets of consonants, although place perception was higher for fricatives than for plosives (presumably because of the inherently longer durations of fricatives).

Kirman (1974) conducted tests of vowel perception using a 15 × 15 array of vibrotactile stimulators that covered an area of 2.8 × 2.8 in. (7.1 × 7.1 cm) and was applied to the hand. Information about the speech signal

was presented in the form of a time-swept display of formant frequencies. Computer-derived formant values were entered vertically on the left side of the display and then shifted one column to the right every succeeding 10 msec, yielding a 150-msec time window for a full display. Six subjects were trained to identify 15 words (composed entirely of vowels and vowel-like sounds) that were recorded twice by one talker. Mean asymptotic performance in the training task was roughly 80%. Scores dropped to 54% for tests with fresh utterances by the talker used in the training paradigm and to 35% for utterances produced by three new talkers. In all cases, the front vowels /ɛ–ɪ/ were confused as were the back vowels (ɑ–ɔ/; many errors occurred on words ending in /l/ and /r/.

Spens (1976) encoded spectral, amplitude, and temporal parameters of speech using an Optacon transducer applied to a finger. Computer-derived measurements of samples of the speech signal taken every 16 msec included the "center of gravity" (comparable to formant frequency) and the intensity in each of two frequency regions, 50–800 Hz and 700–3600 Hz. The frequency measurements were encoded in the rows of the Optacon. Four ranges of intensity were also encoded in the rows by varying the number and position of active vibrators assigned to a given frequency. Varying time windows (16 through 96 msec) were achieved by using a constant sampling interval of 16 msec and controlling the number of columns that received parallel information. For example, for a 16-msec time window, all six columns displayed the same information, whereas for the 96-msec window each of the columns contained separate information from six adjacent time samples. Spens studied the effect of the time window size on the ability of a subject to identify 10 Swedish numerals. For words presented in isolation, scores were similar for the four time windows studied (roughly 94%) following a 6-hour training period. When the numerals were presented in continuous strings of 2–9 words and the subject was asked to identify the next-to-last word, performance increased with the size of the time window (up to 96 msec), but was always lower than that observed for isolated words. Performance dropped for time windows greater than 96 msec (attained by increasing the sampling interval).

Ifukube and Yoshimoto (1974) and Ifukube et al. (in press) experimented with a 16-channel tactile vocoder in which the analysis filters had center frequencies in the 250–4000 Hz range and bandwidths that increased with frequency. The display consisted of a 16 × 3 array of reeds (activated by piezoelectric vibrators) applied to the fingertips. The vibrators were driven by 200-Hz square waves amplitude-modulated by the filter output envelopes. Two types of encoding schemes were studied: one in which frequency was encoded along the long axis of the display with identical information presented in the three columns, and the other in which frequency information was swept through the display as a function of time. In both schemes, amplitude was encoded as intensity of vibration. Results with the nonswept display indicated that after ½ hour of training, four normal subjects were able to identify 91% of five vowels (/i a u e o/) and 66% of five consonants (/k s n h r/). Vowels with wide separation of the first two formants were easily distinguished. Most of the consonant errors occurred for /n h r/. Further results with this display indicated that after ½ hour of training, profoundly deaf children identified the five vowels with 85% accuracy. Three sets of consonants were also tested with the deaf subjects, including five consonants that were difficult to lipread (65%), five plosives (35%), and five nasal and glide phonemes (50%). Results for stimuli played back at slower-than-real times indicated that peak performance in consonant recognition occurred for playback that was slowed by a factor of 4 (where improvements of 5–25% were obtained for consonant recognition). Results using the time-swept spectral display (with a 15-msec time window) indicated that the recognition score for five CV syllables was 10% higher than that obtained with the nonswept display.

Clements et al. (in press) compared two Optacon-based spectral displays applied to a finger in a vowel discrimination task: the frequency amplitude display used by Mook (1978) and Snyder et al. (in press) and a time-swept spectral display (roughly similar to that used by Spens, 1976, and Kirman, 1974). In the time-swept display, rows corresponded to frequency bands, columns corresponded to swept time, and amplitude was coded dichotomously. The 45 pairs of vowels tested were produced both naturally (four speakers, 3 utterances per speaker per vowel) and synthetically (one waveform per vowel). The average discrimination score (obtained using a two-interval paradigm) was roughly the same for both the frequency-amplitude (83%) and the time-swept (87%) displays. Averaged across the two displays, performance on the synthetic stimuli was roughly 13% better than on the natural stimuli. Similar error patterns were obtained for all four experimental conditions, however. Performance was highly dependent on the number of features by which the vowels differed and the feature best discriminated was tenseness, followed by high and low, round and back, and finally retroflexion. An examination of the relation of discrimination performance to physical cues available for discrimination (amplitude, duration, F1, and F2/F1) showed no pronounced correlation with any single cue.

Comments on Spectral Displays

General characteristics. It is difficult to provide a unified summary of the results of the various studies just discussed (and summarized in Table 1) because of a variety of differences across studies. These differences include different transducer systems, body sites, encoding schemes, type of stimulation, type of speech material, the manner in which the material was represented in the utterance set, the general nature of the task, the detailed characteristics of the paradigm, and the type and amount of training. Some comments can be made, however, concerning several characteristics that emerge from the re-

sults of the various studies. These characteristics are grouped under three categories discussed in the paragraphs below: *performance on vowels relative to that on consonants*, the *effects of long-term training*, and the *ability to generalize* from the specific conditions encountered in training.

1. Performance for vowels appears to be better than for consonants through most spectral displays. Results on discrimination or identification of both consonants and vowels through the same spectral display are available in several cases.[2] Clements et al. (in press) used an Optacon-based frequency-amplitude bar-graph display and obtained an average score of 83% for vowel discriminability. Snyder et al. (in press), using the same display, obtained an average score of 68% for consonant discriminability. Saunders et al. (1980) used a linear electrotactile display and obtained average discrimination scores of 88% for vowels and 79% for consonants. Pickett and Pickett (1963), however, obtained roughly equal scores for consonant and vowel discriminability (approximately 90%). In the two studies that examined identification of both vowels and consonants, the vowel scores again tended to be superior. Ifukube and Yoshimoto (1974) reported 91% accuracy in identifying 5 vowels compared to 66% accuracy in identifying 5 consonants following a brief period of familiarization with the display. Sparks et al. (1978) found that, after training, performance on a set of 8 vowels was nearly perfect, whereas performance on consonants plateaued at 50% for a set of 8 plosives and nasals and at 70% for a set of 9 fricatives. As Sparks et al. (1978) noted, one component contributing to the superiority of performance on vowels relative to consonants may be that the relatively steady-state spectral information important to vowel identification is more readily adaptable to presentation through spectral displays (as currently realized) than is the rapidly changing spectral information necessary for identification of certain classes of consonants.

2. In examining the effects of training through tactile displays, we focus on two studies in which large amounts of time (50–200 hours) were spent in training on word recognition (Engelmann & Rosov, 1975; Scilley, 1980). The two studies used similar tactile displays (linear vibrotactile arrays), but they differed with respect to the types of subjects tested (prelingually deaf vs. normal-hearing). The results of the two studies differed with respect to the overall average rate of vocabulary acquisition and the manner in which the rate of acquisition varied as a function of training. Words were acquired at an average rate of 2 words/hr. by Scilley's normal subjects compared to 1 word/hr. for the best deaf subject tested by Engelmann and Rosov. The word-acquisition rate for the deaf subject, however, increased as training progressed, rising from .5 words/hr. for the first 36 weeks to 2 words/hr. thereafter. The learning curves for Scilley's subjects, on the other hand, do not exhibit any such acceleration in rate. In addition, the results of an information-theoretic analysis indicate that information transfer levelled off after roughly 20 hours of training in Scilley's study.

3. Many of the spectral display studies reviewed include tests of the subjects' abilities to transfer learning to conditions more general than those included in the original training sessions. Engelmann and Rosov reported immediate transfer of learning in a word-recognition task when the display was moved from the forearm to the thigh, as did Scilley when changing stimulation from the right arm to the left arm. A number of studies have examined the effect of generalizing speech materials by increasing the number of tokens produced by a given talker or by including utterances from new talkers. Pickett and Pickett found, as one would expect, that significantly better discrimination results were obtained when a single utterance was used to represent a speech element than when a fresh utterance was employed on each trial. Similarly, Clements et al. (in press) observed much better performance on discrimination tests using single-token synthetic vowels than on tests using multiple-token natural utterances. Kirman (1974) tested the extent to which learning transferred to new utterances by the speaker used in the training sessions as well as to three new speakers. For new utterances by the familiar speaker, scores dropped from 80% to 50%. For the same words produced by new speakers, scores dropped from 80% to 30%. Spens (1976) reported good transfer of learning to a new talker (94% to 84%) and to faster utterances by the original talker (94% to 80%). Scilley reported good transfer of learning to new talkers as indicated by an average score of 56% (compared to a score of 80% with the familiar speaker) on first-time tests with new talkers. In addition, fewer training sessions were required to reach a criterion of 80% with each succeeding new talker. Mook observed that performance continued to improve when multiple talkers and tokens were introduced into the set of utterances presented to subjects in training sessions. Sparks et al. (1978), who conducted training and testing with live-voice utterances, found that performance on vowel identification dropped from 95% to 75% for a test using three talkers, two of whom were unfamiliar to the subject. They also found that a subject trained in a single vowel context did as well on a test using CVs with eight vowels as did a subject who received training in all eight contexts.

Comparisons among systems. Within the general category of spectral displays and for tasks involving the discrimination and/or identification of speech segments, it appears that relatively good results were obtained by Pickett and Pickett (1963), Ifukube et al. (in press), and Sparks et al. (1978), whereas relatively poor results have

[2]All discrimination scores cited that were obtained with a one-interval paradigm have been modified to be consistent with scores obtained with a two-interval paradigm, that is, the one-interval percent-correct scores were converted to d' assuming unbiased observation, d' was then multiplied by a factor of $\sqrt{2}$, and, finally, $\sqrt{2}d'$ was reconverted to a percent-correct score assuming unbiased observation.

been obtained by Yeni-Komshian and Goldstein (1977), Snyder et al. (in press), and Clements et al. (in press).

In considering these results, one cannot differentiate between the good and poor systems on the basis of the geometric dimensionality of the array (linear vs. planar), site of stimulation (finger vs. stomach), or mode of encoding (frequency-amplitude vs. time-swept). Although the MESA system used by Sparks et al. (1978) involves a two-dimensional array, the system used by Pickett and Pickett was linear and the system used by Ifukube et al. was equivalent to a linear system when time-sweeping was not used (since all three columns were treated identically in this condition). Similarly, although good results were obtained using portions of the hand, some of the poor results also occurred with portions of the hand. Finally, in one case in which a time-swept display was observed to have an advantage (discrimination of consonants in the study by Ifukube et al., in press), the comparison was made between a display using two dimensions (one to encode frequency and the other to encode time) and an essentially linear reference system (because the second dimension in this system was used only to supply redundant frequency information). If the reference system had made use of the second dimension to encode some other parameter (e.g., amplitude), it is unclear whether this advantage would have been maintained. Furthermore, Clements et al. (in press) found no distinct advantage for time sweeping in the discrimination of vowels.

In contrast to the above factors, some factors that may differentiate between the good and poor systems are those relating to the use of the Optacon transducer system and the way in which spectral amplitude was encoded. Whereas all the poor systems used the Optacon and encoded amplitude as a bar graph, none of the good systems followed such a procedure; Pickett and Pickett and Ifukube et al. encoded acoustic amplitude as vibration amplitude, and Sparks et al. (1978) used a contour graph, not a bar graph. Although Sparks et al. did not report any formal results using a bar-graph display, the results of pilot studies led them to believe that such a display would have degraded their results. Clearly, further experiments are required to answer these questions.

SECTION 3

COMPARISON OF SPECTRAL DISPLAYS WITH TADOMA

Snyder et al. (in press) have provided the only direct comparison between Tadoma and a spectral display in their study of consonant discriminability. Test stimuli were 32 pairs of consonants presented in live-voice utterances of CV syllables with one of three vowels. The consonant contrast was held constant throughout a run, but the vowel was varied randomly from trial to trial. Two male subjects alternated as the talker and the observer. For tests with Tadoma, auditory and visual cues were eliminated by using blindfolds and masking noise. The results of these discrimination tests showed Tadoma to be significantly superior to a frequency-amplitude bar-graph display on the Optacon: 81% correct for Tadoma and 68% correct for the frequency-amplitude display. Moreover, the primary advantage of Tadoma occurred for contrasts of voicing and place (Tadoma, 90% and 82%; frequency-amplitude display, 55% and 56%); for contrasts of manner, the two displays were roughly equivalent (78%).

We cannot conclude that spectral displays generally are inferior to Tadoma from the results of Snyder et al. (in press). For example, the results Pickett and Pickett obtained for consonant discriminability are better overall than those Snyder et al. obtained through Tadoma (88% vs. 81%). (Pickett and Pickett's scores were modified by undoing their "correction for guessing" and compensating for the difference between one-interval and two-interval paradigms.) The scores of Pickett and Pickett for place and manner were superior to Tadoma (93% vs. 82% for place and 89% vs. 78% for manner), but their voicing score was inferior (80% vs. 90%). It is clear that for consonant discriminability the spectral display Snyder et al. used was inferior to that which Pickett and Pickett used.

Results obtained by Reed et al. (1978) for vowel discriminability through Tadoma have been compared to results obtained by Clements et al. (in press) and Pickett and Pickett (1963). The results of Clements et al., averaged across frequency-amplitude and time-swept displays of natural utterances, indicated a score of 79% compared to a score of 73% through Tadoma averaged across constant-stimulus and roving-stimulus lists. Whereas the Tadoma study involved trial-by-trial fresh utterances from one talker, the study of Clements et al. involved recordings of four talkers and three utterances per talker for each syllable; discriminations were made across talkers as well as across utterances. The average score of Pickett and Pickett on vowel discriminability (modified in the manner just described) was 91% compared to an average score of 76% through Tadoma on constant-stimulus lists. Thus, spectral displays appear to have an advantage over Tadoma for discrimination of vowels.

A comparison of performance on identification of speech elements has been made for Tadoma (using results of Reed et al., 1982, in press) and for the MESA frequency-amplitude spectral display (Sparks et al., 1978). This study of Sparks et al. was chosen for the comparison because it represents the most systematic evaluation of both consonant and vowel identification performed to date on a spectral display. Procedural differences between the two studies have been noted and attempts made (where possible) to control for these differences.

Because the stimuli used by Sparks et al. (1978) were a subset of those used in the Tadoma studies, differences in set size were taken into account by forming subsets from the Tadoma matrices that were equivalent to the sets used by Sparks et al. Normalized identification scores for the elements in these subsets were then calculated by using the constant-ratio rule formulated by Clarke (1957).[3] To obtain an indication of the applicability of the rule to the Tadoma data, we applied the rule to data obtained on an experienced Tadoma user from two tests of vowel identification that differed only in the sets used: a set of 15 vowels and a subset of 8 vowels. Performance on the subset predicted from the larger matrix using the constant-ratio rule was within 5% of the performance obtained on the 8-vowel test. This result offers some support for the use of the constant-ratio rule in the present context.

For the set of plosives and nasals, normalized scores for the Tadoma users (one class A subject and two class B subjects) in C/ɑ/ context exceeded 90%. Asymptotic performance for the subject in the Sparks et al. (1978) study who was trained in C/ɑ/ context was roughly 50%. For the set of fricatives in C/ɑ/ context, normalized scores in the Tadoma study were 69% for the one Class A subject and 94% and 76% for the two Class B subjects. Asymptotic performance in the Sparks et al. study was 70% for the subject who was trained and tested on fricatives in C/ɑ/ context with one talker. For the set of vowels tested by Sparks et al., normalized Tadoma scores were 64% for the Class A subject and 76% for both Class B subjects for vowels presented in /g/-V-/d/ context. Performance on the spectral display for vowels presented in /b/-V-/d/ context by one talker approached 95%. The results of these comparisons indicate superior performance with Tadoma for the set of plosives and nasals, comparable performance levels on fricatives, and superior performance with the spectral display on vowels. Until

[3]The constant-ratio rule states that the ratio between any two entries in a row of a submatrix is equal to the ratio between the corresponding two entries in the full matrix.

training and testing with full sets of consonants and vowels are performed using a spectral display, we cannot be certain how performance on these full sets will compare to that obtained through Tadoma.

The comparisons presented here suggest that (a) the performance of spectral displays (at least of the frequency-amplitude type) relative to that of Tadoma are much better for vowels than for consonants; (b) the MESA is substantially more effective than Optacon-based bar-graph displays; and (c) overall, the MESA may not be significantly inferior to Tadoma for the transmission of speech segments. To date, the amount of training that has been employed with spectral displays is too limited to permit serious comparative analyses of these displays and Tadoma in the context of open-set vocabularies of words and/or continuous speech.

SECTION 4

COMPARISON OF TADOMA WITH LIPREADING

Tadoma and lipreading will be compared for performance on identification of consonant and vowel stimuli. The lipreading studies used in the comparison contained fairly extensive identification data on relatively well-trained subjects. Data on segmental identification through Tadoma are available for the Class A subject and the two Class B subjects mentioned previously in Section 1.

Three studies of consonant confusions through lipreading were selected (Erber, 1974; Heider & Heider, 1940; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). The first two studies are similar in that they both involved identifying a set of 20 consonants, using children in schools for the deaf as subjects, and using teachers familiar to the subjects as speakers. The syllabic contexts and the number of observations per subject varied between the two studies. Heider and Heider used CV syllables with V = /ɔɪ/ or /ɪ/, and each of 39 subjects contributed one response to each consonant in each context. Erber used VCV syllables with V = /i/, /ɑ/, or /u/, and each of 6 subjects contributed 10 responses to each consonant in each context. The study of Walden et al. tested adults with sensorineural hearing loss who participated in a training program for the acquisition of lipreading skills. Twenty consonants were presented in C/ɑ/ syllables, and each of 31 subjects contributed 20 responses per item. The studies of vowel identification through lipreading examined here were those of Heider and Heider (1940) and Berger (1970). Heider and Heider studied the identification of 16 vowels and diphthongs in CVC syllables (/p/-V-/p/ and /f/-V-/f/). A teacher presented the stimuli to 37 deaf students, who each contributed one response to each vowel in each context. Berger's study involved the lipreading of 12 vowels in VC and CV contexts (with the consonants /b/, /n/, and /g/) by 45 normal-hearing subjects. Each subject contributed one response per vowel in each context.

In the study of Tadoma with a Class A subject (Reed et al., 1982), 24 consonants were presented in CV syllables where V = /ɑ/, /i/, /u/. Approximately 25 responses per consonant were obtained in each context. Fifteen vowels and diphthongs were presented in three contexts (/h/-V-/d/, /g/-V-/d/, and /b/-V-/d/), and roughly 25 responses per vowel were obtained in each context. In the study of the Class B subjects (Reed et al., in press), 24 consonants were presented in CV syllables where V = /ʌ/, /ɑ/, /i/, and /u/. For each subject, approximately 40 responses per consonant were obtained in C/ʌ/ context and 20 responses in each of the remaining contexts. Each subject also responded to 50 presentations of each vowel in one context, /g/-V-/d/. The set sizes, items in the sets, and syllabic contexts used in the Tadoma and lipreading studies are sufficiently similar to warrant a comparison of scores.

Despite the previously mentioned procedural differences in the lipreading studies, similar average scores (in the range of 44–50%) were obtained across the three studies of consonant identification. The average score for consonant identification for the Class A Tadoma subject (55%) was roughly 5–10% higher than the average lipreading score; the average score of the Class B subjects (75%) was roughly 25–30% higher. Performance on vowel identification was similar for the two lipreading studies (scores of 53% and 60%). The average score for the Class A Tadoma subject (56%) was similar to the lipreading scores, whereas scores for the Class B subjects were roughly 20% higher.

The confusion matrices obtained through lipreading and Tadoma were examined further in terms of various articulatory/phonological features. The features described by Miller and Nicely (1955) were used for analysis of consonant confusions, and those described by Chomsky and Halle (1968) were used for the vowel analysis. The percentage of unconditional transmitted information was calculated for each feature. These percentages are presented in Table 2 for the lipreading studies and for the Class A and Class B Tadoma subjects.

For consonants, the features voicing and nasality were transmitted better through Tadoma than through lipread-

TABLE 2. Proportion of unconditional information transfer on various features through lipreading and through Tadoma.

	Consonants				
	Lipreading			*Tadoma*	
Feature	*Heider & Heider (1940)*	*Erber (1974)*	*Walden et al. (1977)*	*Class A*	*Class B*
Voicing	.16	.06	.02	.84	.79
Nasality	.16	.12	.02	.43	.41
Frication	.78	.57	.80	.49	.59
Duration	.28	.63	.75	.26	.46
Place	.81	.80	.68	.47	.73

	Vowels			
	Lipreading		*Tadoma*	
Feature	*Heider & Heider (1940)*	*Berger (1974)*	*Class A*	*Class B*
High	.54	.54	.36	.68
Low	.43	.36	.36	.73
Back	.47	.61	.37	.62
Tense	.41	.22	.51	.73
Round	.65	.74	.46	.97

ing. For voicing, a score of approximately 80% was obtained for both classes of Tadoma subjects, compared to an 8% average across the lipreading studies; for nasality the score was roughly 40% for Tadoma compared to 10% for lipreading. The features of duration, place, and frication were better perceived through lipreading than through Tadoma. For these three features the scores for the Class B subjects were 5–15% lower than those for lipreading, whereas those of the experienced Tadoma user were 25–30% lower than for lipreading. The results shown here for lipreading concur with most summaries of consonant confusions through lipreading, which indicate that nasality and voicing are poorly perceived whereas features related to place and manner are better perceived. For vowels, the scores for lipreading and for the Class B Tadoma subjects were generally similar and were higher than those for the Class A Tadoma subject. One feature, *tense*, was better perceived through Tadoma than through lipreading, presumably on the basis of the Tadoma users' ability to detect a tightening of the neck musculature for tense vowels. The similarity of the performance on most of the vowel features for lipreading and for Tadoma indicates that similar attributes of vowel perception may be used for the two methods.

SECTION 5

TACTILE INPUT AS A SUPPLEMENT TO LIPREADING

The most simple tactile devices that have been used to supplement lipreading are single-channel devices that translate the speech signal directly to the skin. Many of the gross temporal aspects of speech available through such devices may not be easily perceived through lipreading alone. Among recent studies of such single-channel devices are those of Erber and Cramer (1974), Zeiser and Erber (1977), and Beguesse (1976).

Erber and Cramer (1974) studied the ability of six normal-hearing subjects to learn to identify sentences presented through a bone-conduction vibrator applied to the finger. Sentences presented in groups of 10 differed in overall duration, total number of syllables, and other prosodic characteristics. The sentences were first presented in fixed order for training, and then a test score was obtained using random presentation. Each list (10 in all) was presented 10 times in this manner. By the fifth presentation, scores approached 100%. The subjects reported that they relied on the following cues in identifying sentences: (a) total duration; (b) number of stressed syllables; (c) changes in intensity and duration; (d) pauses and bursts; and (e) onset and offset characteristics of each sentence.

Additional unpublished work by Erber and Cramer[4] indicated that subjects were able to identify emotional emphases and syntactic structures of sentences presented through the vibrator at levels well above chance. Zeiser and Erber (1977) investigated the ability of profoundly hearing-impaired children and normal-hearing adults to determine the number of syllables in words that could contain one, two, or three syllables. The stimuli were presented under two modes: auditorily and through a bone-conduction vibrator. Performance was very similar for the hearing-impaired subjects under both methods of presentation and for the normal subjects under the tactile condition. Results could be predicted from the time-intensity envelopes of the stimuli as observed visually on an oscilloscope. In particular, multisyllabic words containing short, unstressed syllables were frequently perceived to contain fewer syllables than they actually contained.

Beguesse (1976) compared the conventional microphone-amplifier-vibrator circuit to a modified single-channel device in which the envelope of the speech signal was used to modulate the amplitude of a 250-Hz tone applied to a small bone vibrator worn on the back of the wrist. Threshold measurements, as expected, showed a relatively flat response across the frequency range for the modified aid compared to the frequency-dependent thresholds obtained for the simple aid. Maximal sensitivity for the simple aid, observed at 250 Hz, was at the same absolute level as thresholds for the 250-Hz modulated vibrator. Following training in identifying 27 environmental sounds, scores on the simple aid were roughly a third lower than those obtained on the modified aid (for which scores were in the range of 25–40% among subjects). Results on a closed-set sentence identification test (Erber & Cramer, 1974) indicated scores of 70–90% on the modified aid, compared to 50–80% on the simple aid following ½ hour of practice on each system. Beguesse concluded that his aid was at least as good as the simple amplifying system.

Other studies have explored the benefits of extracting specific types of information from the speech signal and presenting this information through the tactile sense. In particular, tactile input has been used to code information in isolated frequency bands of speech (e.g., De Filippo & Scott, 1978; Ling & Sofin, 1975; Scott & De Filippo, 1976; Scott, De Filippo, Sachs, & Miller, in press), to code fundamental frequency (e.g., Grant, 1980; Rothenberg & Molitor, 1979; Spens, 1975; Stratton, 1974; Willemain & Lee, 1972), to code phonological features (Martony, 1974), and to code articulatory signals (Miller, Engebretson, & De Filippo, 1974a, 1974b).

Ling and Sofin (1975) used a single-channel bone-conduction vibrator to provide tactile coding of high-frequency signals for profoundly deaf children. Signal components above 5000 Hz activated an 800-Hz oscillator connected to a bone-conduction vibrator held between the subject's thumb and first two fingers. Ten students from a school for the deaf were trained and tested using 4-word sets in which all four words had the same vowel and two of the words contained /f/, /s/, /ʃ/, or /z/. Scores were obtained for two conditions, one in which auditory, lipreading, and tactile cues were presented together and the other involving auditory and lipreading cues alone. Fewer errors were made when the tactile cues were available, both before and after training. Pretraining scores (66% with the tactile cues and 60% without) were improved by 15% with training under both conditions.

De Filippo and Scott (1978) evaluated a 2-channel device in which one channel presented information for frequencies below 1000 Hz to the palm through a mechanical vibrator and the other channel coded frequencies above 4000 Hz through electrotactile pulses presented to the back of the hand. Two normal-hearing subjects participated in the study, which consisted of 10–13 hours of training on tracking of continuous speech, followed by tests on the perception of nonsense syllables, words, and sentences under conditions of lipreading alone and lipreading plus the tactile aid. Test scores for the condition of lipreading plus the tactile aid were 6–29 percentage

[4]For details, contact Norman P. Erber, Associate Professor, Washington University, St. Louis, MO 63110.

points higher for the various test materials than those obtained for lipreading alone. Data were also obtained on the subjects' ability to repeat material using the tracking procedure. In this method, the talker reads a passage of text phrase by phrase, the subject is required to repeat the phrase exactly, the talker repeats elements until the responses are correct, and performance is evaluated in terms of the rate at which the passage is conveyed. After 8 hours of experience with this task, the better of the two subjects was tracking at a rate of 75 wpm with lipreading plus the aid compared to 50 wpm with lipreading alone. For the material used, the comprehension rate for normal listening was 110 wpm. These results suggest that tactile cues can be integrated with lipreading to improve comprehension of continuous discourse.

Scott and De Filippo (1976) and Scott et al. (in press) reported results obtained with a modified version of the aid described. In this version, two electrotactile channels were used to code high-frequency information [one electrode for frequencies above 8000 Hz and two other (identical) electrodes for frequencies in a band centered around 2400 Hz], and a vibratory channel was used to code first-formant information. The same subjects served in evaluating this aid as the 2-channel aid, and their performance on the tracking procedure was higher with the 3-channel aid. For the better of the two subjects, comprehension of continuous discourse with lipreading plus the aid was 75% that of the normal auditory rate, whereas comprehension for lipreading alone was 56% that of the normal auditory rate.

Tactile displays of vocal fundamental frequency obtained by the use of a throat microphone to detect voiced speech were presented to deaf children by Willemain and Lee (1972) and Stratton (1974) primarily for improving pitch control in speech. In the study by Willemain and Lee, high or low pitch was indicated by the vibration of one of two solenoids worn on the fingers. Pilot studies indicated that some deaf subjects were able to learn to lower their voice pitch, particularly at the start of a new phrase. The display Stratton used involved activation of one of 10 solenoids to represent the value of fundamental frequency in the range 80–600 Hz. Subjects were trained to imitate intonation contours in short phrases and sentences. Recordings were made of subjects' utterances before and after training with the device. Listeners' judgments of the recordings indicated that the training improved the subjects' ability to control intonation but did not improve the pleasantness or expressiveness of their speech.

Spens (1975) presented a time-swept tactile display of the fundamental frequency contour of sentences through an Optacon transducer. The frequency scale, chosen to suit a male talker, was divided into four 20-Hz intervals between 80 and 160 Hz, and one interval for frequencies above 160 Hz. Each interval was coded by two rows of the display. Data points were shifted through the six columns of the Optacon at intervals of 4, 8, 16, or 24 msec, giving total time windows on the display of 24 through 144 msec. Stimuli were closed sets of 15 sentences (all approximately the same length and containing no labial or labio-dental sounds) video-taped by a male talker. Nine normal-hearing subjects who had little or no experience with tactile displays lipread the sentences supplemented by the tactile display of fundamental frequency. Subjects received no training prior to testing but were given immediate correct-answer feedback after choosing a response. Five of the subjects had peak performance for the 96-msec time window, two had peak scores for the 24-msec window and next-highest scores for the 96-msec window, and two subjects performed essentially the same for all four windows. On the basis of these data, Spens hypothesized that tactile pitch information can supplement lipreading at two different levels: first, at the segmental level for short time windows (i.e., less than 90 msec); and second, at a suprasegmental level for time windows of roughly 90 msec.

Rothenberg and Molitor (1979) examined the ability of eight normal and five hearing-impaired subjects to identify stress/intonation patterns of a short sentence based on vibrotactile encoding of fundamental frequency. Various transformations of fundamental frequency to vibrotactile frequency were achieved primarily by the use of different scale factors and by shifts in center frequency and frequency range covered. The vibrotactile stimuli were delivered through an Electrodyne vibrator applied to the forearm at roughly 15 dB SL. The six stimuli were constructed from a fixed 3-word sentence *(Ron will win)* by variations in placement of primary stress and by delivery of the sentence as either a question or statement. The subject's task was to identify which of the six stimuli had been presented. The results indicated that the best performance (averaging roughly 50% correct where $p = .167$) was obtained when the fundamental frequency range was expanded by a factor of 2 or shifted down to a center frequency of 50 Hz. Scores were similar for the normal and hearing-impaired subjects. Results of extensive training of one subject on one scheme in which the initial score was roughly 40% correct indicated that the score appeared to reach a plateau at 65%. An analysis of errors indicated an inability to determine the location of stress, whereas questions and statements were easily distinguished.

Grant (1980) designed a tactile aid to convey information about fundamental frequency and tested its ability to supplement lipreading of connected discourse. The aid was a 10-electrode array applied to the forearm, in which frequency was coded as a function of place. Fundamental frequency was extracted from the speech materials and displayed on the array. Two subjects (one normal-hearing and the other hearing-impaired) participated in the study. In one test, 10 sentences were each read as questions or as statements. Using the tactile aid alone, the subjects were able to determine quite accurately whether a question or statement had been presented. Their ability to identify which of the 10 sentences was presented averaged 25% correct for the tactile aid alone compared to 98% correct for aided lipreading. Results using the continuous-discourse tracking

procedure (De Filippo & Scott, 1978) indicated that tracking rates for aided lipreading were approximately 15 wpm higher than for lipreading alone. The average tracking rates for aided lipreading for the two subjects were 57 and 70 wpm compared to a normal auditory tracking rate of 110 wpm for the materials used. Additional tests on small, closed sets of sentences indicated that the subjects were able to detect locations of phrase boundaries and of stress quite accurately through the use of the tactile aid alone.

Martony (1974) provided subjects with either visual or tactile cueing of two of the features displayed visually in the Upton eyeglass (Upton, 1968). In Upton's device, a set of miniature lights mounted on the lens of an eyeglass is used to inform the lipreader whether a sound has the property of voicing, frication, or plosiveness. In Martony's experiments, the voiced/voiceless and continuant/stop binary features, derived from computer analysis of recorded speech, were indicated by activating lights or vibrators corresponding to these features. One set of tests examined the ability of normal-hearing subjects to identify CV syllables with lipreading alone or with lipreading plus the visual or tactile cues of voicing/nonvoicing. In the visual case, a red or blue lamp was lit to indicate voiced or voiceless sounds, respectively; in the tactile case, one of two bone-conduction vibrators was activated. Before each test, a short training session with feedback was administered. Detection of the voicing feature rose from 70% for lipreading alone to roughly 90% with the addition of either visual or tactile cues; however, the overall identification score was the same (35–40%) for all three conditions. Seven children in a school for the deaf were also tested with the tactile device. These children were first given a short training period and then wore the aid in the classroom for four weeks. Results indicated that the aid was more helpful in understanding isolated words than in understanding normal communication. Generally, the aid was better liked by the more profoundly deaf children.

Cueing of voiced/voiceless and stop/continuant dimensions was also tested with 10 normal subjects. Three lights or three vibrators were used to cue voiced, voiceless, or stop consonants. The perception of CV syllables rose from 29% to 44% with the addition of tactile cues and to 59% with the addition of visual cues. Normal subjects were also trained and tested on sentence identification with the visual and tactile feature-detecting schemes. Subjects received 6 hours of training on closed sets of sentences containing four key words. For testing, the response set was open, and each sentence was preceded by a picture of one of its key words. Identification of words in sentences was 30% for lipreading alone. Although this score increased by only a few points with the use of the visual feature detector, it was improved by 45% when tactile cues were available. On the other hand, identification of isolated words, drawn from the sentences, was 50% for lipreading alone and close to 70% with the addition of either aid. The results indicate that identification of isolated words was better than identification of these same words in sentences for both types of aids. For sentence identification, however, the results indicate that visual presentation may interfere with lipreading, whereas tactile presentation permits the same cues to be integrated more easily with those provided by lipreading.

Miller et al. (1974a, 1974b) experimented with a 3-channel tactile display based on the following articulatory signals: nasal vibrations, laryngeal vibrations, and the acoustic signal at the mouth. The envelopes of these signals were used to modulate 100-Hz square waves applied to piezoelectric vibrators sensed by the fingers. Experiments with normal subjects involved identification of nonsense syllables and words under conditions of lipreading alone and lipreading plus the tactile display. For tests using alveolar and palatal consonants in CV and VC syllables, significant increases in perception of the voiced/voiceless, stop/continuant, and nasal/oral distinctions occurred when the tactile display was used. An overall advantage of 30 percentage points on the nonsense syllable tests was observed with the tactile cues. Word tests were selected from a known list of 240 words containing equal numbers of monosyllables, trochees, and spondees. The use of tactile cues improved the perception of monosyllables and trochees, but not spondees (because of the high scores obtained on these words with lipreading). An average score of 86% was obtained for the combined condition, compared to 74% for lipreading alone. When the set of words was increased to 1,000 monosyllables, an 18%-gain in recognition was obtained in the combined condition (52% correct) over lipreading alone. The binary features voiced/voiceless, stop/continuant, and nasal/oral are difficult to perceive through lipreading alone, but they can be discriminated through the tactile display.

Several of the tactile vocoder systems described in Section 2 have also been tested in conjunction with lipreading (Pickett, 1963; Risberg, Galyas, & Franzen, 1965; Sparks, Ardell, Bourgeois, Wiedmer, & Kuhl, 1979; Sparks et al., 1978).

Pickett (1963) continued evaluation of the device used by Pickett and Pickett (1963) (described in Section 3). Tests conducted on 16 students in a Swedish school for the deaf compared performance for lipreading alone, for the tactile device alone, and for lipreading plus the tactile device. On tests of discrimination of pairs of vowels or consonants, the average score for the tactile device (80%) was similar to that obtained through lipreading (84%), although on individual pairs performance was sometimes higher under one condition than the other. Tests comparing lipreading alone to the combination of lipreading and the tactile vocoder involved identification of closed sets of words. For the largest set tested (12 words), the scores for the combined condition for two different speakers (85% and 75%, respectively) were 25% and 10% higher than for lipreading alone.

Risberg et al. (1965) tested essentially the same tactile device used by Pickett and Pickett (1963) and Pickett (1963) on 12 deaf children. Three conditions were

tested: lipreading alone, lipreading plus group hearing aid, and lipreading plus tactile vocoder. The tests involved distinguishing between /m b/, /p m/, or /p b/ in 1-, 2-, and 3-syllable words and identifying initial /t d n s/ in words. After several hours of training, improvements of roughly 10–20% over lipreading alone were obtained for distinguishing /m p/ and /p b/ (with essentially equal scores for the conditions involving the hearing aid and the tactile device) and for identification of /t d n s/ (with better scores achieved for the condition with tactile cues). In several cases, scores on the pretraining tests with lipreading plus the tactile device were better than for lipreading alone. This finding is similar to that of Ling and Sofin, which indicates that some of the tactile information may have been put to immediate use.

Sparks et al. (1978) tested identification of consonants and vowels through the MESA tactile array alone, lipreading alone, and lipreading plus the tactile device. A description of the tactile device, stimulus sets, and experimental procedure is provided in the discussion of Sparks et al. (1978) in Section 2. The following scores were obtained on posttraining tests with vowels: 98% for the combined mode, 90% for lipreading alone, and 95% for the tactile aid alone. For the set of plosive and nasal consonants, performance was 92% for the combined mode, 46–65% for lipreading alone, and 54–58% for the tactile aid alone. For fricatives, perfect performance was obtained under the combined mode, compared to scores of 75% for lipreading alone and 70% for tactile alone. An evaluation of the MESA as an adjunct to lipreading in the perception of continuous speech was conducted by Sparks et al. (1979) using the tracking procedure of De Filippo and Scott (1978). Three normal-hearing adults (one of whom had previous experience in syllable identification through the MESA) received 15 hours of training in tracking continuous discourse under conditions of lipreading and lipreading plus the aid. At the end of training, the rate of tracking was similar for the two conditions. The best subject achieved a rate of roughly 65 wpm under both conditions. These results differ from those obtained by De Filippo and Scott who found that the use of a 2-channel tactile aid in conjunction with lipreading increased the tracking rate by as much as 25 wpm over that obtained through lipreading alone. Among possible reasons offered by Sparks et al. for the apparent failure of the MESA to supplement lipreading of continuous discourse are that the MESA "overloaded" the tactile sense in continuous speech, that the transmission of suprasegmental information may have been inadequate, or that the selection of subjects was inappropriate (e.g., better results might have been obtained with hearing-impaired children).

Several studies have been performed at Johns Hopkins concerning the effects of vibrotactile stimulation (in addition to visual lipreading) on the development of speech and language in prelingually deaf children. Goldstein and Stark (1976) experimented with the Optacon-based frequency-amplitude display used by Yeni-Komshian and Goldstein (1977), which is described in Section 2. The subjects were children 2 to 4 years old with hearing loss greater than 80 dB in the speech range, none of whom produced intelligible speech. The children participated in seven 20-min training sessions designed to encourage the production of CV syllables. An experimental group of four children used the tactile aid (in addition to lipreading) in the third to seventh training sessions, whereas a control group of an additional four children did not receive exposure to the aid. The ratio of CV utterances to the total set of utterances produced was calculated for the first two training sessions (before introduction of the tactile aid to the experimental subjects) and for the last two sessions. The children trained with the tactile display showed a significantly greater increase in the ratio of CV utterances to total utterances as a function of training than did the children in the control group.

Proctor and Goldstein (1980) applied a single-channel vibrotactile device to a profoundly deaf child who wore the aid for "at least 15 hours per week." The aid was similar to that studied by Beguesse. Observation of the child, in addition to speech and language therapy sessions, occurred between the ages of 33 and 43 months. At the outset of the program, the child had a receptive vocabulary of approximately 5 words through lipreading. This vocabulary remained constant throughout the first 3 months of training. Following the introduction of the tactile device at 36 months of age, the child's understanding of single words increased from 5 to 400 words in 7 months (similar to the rate observed in normal-hearing children during the early stages of vocabulary acquisition).

Many of the above studies offer substantial evidence that tactile cues can serve as an effective supplement to lipreading for various types of speech materials, including connected discourse. Positive results on connected speech were obtained with several types of tactile aids, including aids that encoded information in isolated frequency bands of speech (De Filippo & Scott, 1978; Scott & De Filippo, 1976; Scott et al., in press), aids that provided fundamental frequency information (Grant, 1980), or aids that encoded phonological features (Martony, 1974). The results of the studies with young deaf children (Goldstein & Stark, 1976; Proctor & Goldstein, 1980) suggest that the application of tactile information (whether single- or multi-channel) may benefit the development of speech and language abilities in such children.

SECTION 6

CONCLUDING REMARKS

With the advantage of hindsight, early work on the tactile communication of speech (e.g., most studies conducted prior to those reviewed in this paper) appears naive in two important senses: (a) the schemes selected for study took little account of general perceptual principles, the characteristics of speech, and the limitations of the tactile sense; and (b) the testing procedures used to evaluate the schemes almost totally ignored the importance of training. Because of these factors, the results appeared to be highly negative, the possibilities for tactile communication of speech appeared to be minimal, and the initial optimism in this field was replaced by excessive pessimism.

Current thought on the tactile communication of speech recognizes that such communication is possible provided the scheme is "appropriate" and adequate training is provided (e.g., as evidenced by Tadoma users). The basic problems now facing investigators in this field are to (a) determine the ultimate limits of the tactile sense for purposes of speech communication, (b) clarify the nature of the constraints required to achieve effective displays, and (c) develop practical systems. Comments on each of these problems follow.

ULTIMATE LIMITS

The performance exhibited by experienced, deaf-blind Tadoma users, although remarkable, probably does not represent the ultimate potential of the tactile sense for speech communication. Not only do these users appear limited in the rates at which they can comprehend continuous speech, but significant numbers of errors occur in their identification of speech segments. Furthermore, and as one might expect, many of these errors appear to be related to inadequate knowledge of tongue position. Consequently, there is reason to believe that if the usual information sensed through Tadoma were augmented by information on tongue position (e.g., as derived from a palatograph and presented on an auxiliary tactile display), performance could be substantially improved. This notion, combined with the results indicating that certain spectral displays are superior to Tadoma on certain types of tasks (e.g., vowel identification), suggests that the best tactile speech reception performance which has been measured to date does not represent the best performance that can ultimately be achieved.

CONSTRAINTS REQUIRED FOR EFFECTIVE DISPLAYS

Two obvious fundamental psychophysical constraints concern basic resolution and memory. First, the scheme must be sufficiently fine to permit discrimination of basic speech elements. Second, the scheme must be sufficiently multidimensional (in the sense of the "7 ± 2 phenomenon" discussed by Miller, 1956) to permit good identification performance with large sets of such elements. Clearly, many of the schemes tested by the early investigators failed to satisfy either of these constraints.

Further, "deeper" constraints can be derived from general perceptual theory, from consideration of the types of tasks the tactile sense should be capable of performing in terms of natural tasks and evolution, and from current theories of speech perception. Examples of factors that have been considered in this context are (a) the extent to which the scheme is consistent with the natural act of manual tactile exploration of objects (e.g., active vs. passive, use of the hands vs. use of other body parts), and (b) the extent to which the scheme permits direct access to the speech production process and is thereby consistent with the motor theory of speech perception.

The extent to which such deeper constraints are really important, however, is still unclear. The Tadoma method clearly satisfies all such constraints with the possible exception of those associated with the notion of active exploration; superficially, at least, the hand appears to be relatively passive in Tadoma. On the other hand, the spectral display studied by Sparks et al. (1978) satisfies very few of them. Aside from the fact that this display is not strongly multidimensional, it does not involve active exploration with the hand (or even passive sensing by the hand) nor does it provide direct access to the articulation process. Despite these characteristics, the results on the identification of small sets of speech segments with this scheme do not appear to be substantially inferior to those obtained with Tadoma. Furthermore, the results on continuous speech with this system, regarded very negatively by the investigators who obtained them (Sparks et al., 1979), must be interpreted with caution. Aside from any questions related to training, it seems possible that relatively minor changes in the scheme could lead to radical improvements. For example, merely using different stimulating frequencies for voiced and unvoiced portions of the speech stream might have improved the results on syllable counting.

In estimating the relevance of these deeper constraints, it is also important to consider the existence of methods of tactile communication other than Tadoma currently used by the deaf-blind population, such as tactile fingerspelling and tactile signing. Although the input in these cases is not spoken speech, these cases do provide forms of tactile communication that have been learned by many deaf-blind individuals (many more than

have learned Tadoma) and that appear to function with considerable efficiency. Furthermore, these forms generally evolved in the context of visual communication with only minor adaptations for use with the tactile sense.

Perhaps the real problem in this field is not finding a method that works, but finding a method that satisfies the aforementioned psychophysical constraints and that, together with substantial training (e.g., comparable to that received by the deaf-blind), doesn't work! It is worth noting that no such method has yet been reported in the literature; all failures can be ascribed to a violation of at least one of the basic psychophysical constraints and/or to a lack of training. Whether these constraints are truly sufficient, or whether further constraints concerned with higher level processing and evident only in the context of continuous speech must be added, can only be determined by further experiments that incorporate substantial training.

For displays that are intended to supplement visual lipreading, an important constraint (apart from those just discussed) concerns the integration of tactile and visual input. Independent of the type of information that one chooses to display in a lipreading aid, the way in which this information is displayed should be optimally integrable with the visual input. The detailed implications for display design of this constraint, however, are still far from clear.

A second possible constraint concerns the type of information to be displayed in the lipreading aid. In most research projects on lipreading aids, investigators have focused on the display of information that is not available through lipreading. Although this approach is clearly a sound one, and many positive results have been achieved with it, some questions can be raised. For example, in certain cases it might be worthwhile to use a portion of the display "resources" for redundant encoding of information that is available through lipreading but not always well-perceived. Because the ultimate effectiveness of the lipreading aid depends on the subject's use of the information rather than on the available information itself, and because redundant encoding is known to be very beneficial in certain situations, the trade-off between supplementary information and redundant presentation of the same information is not always clear. Furthermore, to the extent that the tactile display is designed only to provide information not available through lipreading, its value is likely to decrease rather sharply as the lipreading information is degraded (because of degraded visual input). In fact, it is conceivable that in the long run the best tactile display for supplementing lipreading will be closer to the best tactile display for the no-visual-input case than is currently realized. An alternative strategy that would be interesting to pursue in this connection would be to begin with the best system available for the no-visual-input case and then to explore how performance with lipreading varied as the system was modified (e.g., information that was redundant with the lipreading information was traded off for more robust encoding of other information).

Development of Practical Systems

Although Tadoma is used successfully by several deaf-blind individuals, the number of users is small (e.g., about 20 in the United States). The method requires direct physical contact, and its successful use requires extensive, specialized training. Ideally, a practical aid would (a) be applicable to a wide range of individuals; (b) function at a distance (and without visual input); (c) require no special training; and (d) be useful for all acoustic inputs, not only speech.

The constraint that the aid must function at a distance implies that the input must be acoustical, not articulatory. It does not necessarily imply, however, that the display cannot be structured to represent articulatory actions (e.g., as in a display of vocal-tract shape and excitation). If such a display were to prove beneficial for speech reception in a laboratory environment, it would of course then be necessary to explore its performance in real environments. Such exploration would have to include study not only of the effects of background noise and reverberation, but also of the ability to interpret meaningful, nonspeech environmental sounds using such a display. Offhand, it would seem that spectral displays would be substantially superior for this purpose.

The constraint that no special training is required could be satisfied only if the aid were applied continuously and at a very early age (perhaps infancy). Under such conditions, speech reception, speech production, and language competence might develop in a fashion similar to that which occurs when normal hearing is available. This procedure creates a substantial additional burden with respect to the actual device to be used (re size, weight, reliability, safety, etc.) and with respect to evaluation procedures (since measures of speech and language development in infants are still relatively crude). Nevertheless, such infant field evaluations must eventually be conducted if the ultimate goal is to be achieved.

In the more immediate future, it seems possible to extend further the concept of the speech reception performance that can be achieved with the tactile sense, to define more clearly the nature of the system that should be used in an eventual tactile substitute for hearing, and to develop and evaluate simpler systems designed for special situations (e.g., lipreading aids for use in schools for the deaf).

Perhaps the ultimate system for use as a substitute for hearing will incorporate two modes or subsystems. One would be addressed to the problem of interpreting the general acoustic environment including speech) on a continuous basis and would not require use of the hands (for sensing) or the eyes (for lipreading). The other would be designed to optimize speechreading and could involve use of the hands and eyes. Such a system might

have substantial advantages over systems that attempt to address all problems in a unitary fashion.

REFERENCES

ALCORN, S. The Tadoma method. *Volta Review*, 1932, *34*, 195-198.

BEGUESSE, I. M. *A single-channel tactile aid for the deaf.* Unpublished master's thesis, Massachusetts Institute of Technology, 1976.

BERGER, K. W. Vowel confusions in speechreading. *Ohio Journal of Speech and Hearing*, 1970, *5*, 123-128.

BIBER, K. W. *Ein neues Verfahren zur Sprachkommunikation über die menschliche Haut.* Doctoral dissertation, University of Erlangen, Erlangen, Germany, 1961.

CHOMSKY, N., & HALLE, M. *The sound pattern of English.* New York: Harper & Row, 1968.

CLARKE, F. R. Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 1957, *29*, 715-720.

CLEMENTS, M. A., DURLACH, N. I., & BRAIDA, L. D. Tactile communication of speech: Comparison of two spectral displays in a vowel discrimination task. *Journal of the Acoustical Society of America*, in press.

CRAIG, J. C. Vibrotactile pattern perception: Extraordinary observers. *Science*, 1977, *196*, 450-452.

DE FILIPPO, C. L., & SCOTT, B. L. A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America*, 1970, *63*, 1186-1192.

ENGELMANN, S., & ROSOV, R. Tactile hearing experiment with deaf and hearing subjects. *Journal of Exceptional Children*, 1975, *41*, 243-253.

ERBER, N. P. Sensory capabilities in normal and hearing-impaired children: Discussion of lip-reading skills. In R. E. Stark (Ed.), *Sensory capabilities of hearing-impaired children.* Baltimore: University Park Press, 1974.

ERBER, N. P., & CRAMER, K. D. Vibrotactile recognition of sentences. *American Annals of the Deaf*, 1974, *119*, 716-720.

GAULT, R. H., & CRANE, G. W. Tactual patterns from certain vowel qualities instrumentally communicated from a speaker to a subject's fingers. *Journal of General Psychology*, 1928, *1*, 353-359.

GOLDSTEIN, M. H., & STARK, R. E. Modification of vocalizations of preschool deaf children by vibrotactile and visual displays. *Journal of the Acoustical Society of America*, 1976, 59, 1477-1481.

GRANT, K. *Investigating a tactile speechreading aid: The transmission of prosodic information in connected discourse and sentences.* Unpublished master's thesis, University of Washington, 1980.

GRUVER, M. The Tadoma method. *Volta Review*, 1955, *57*, 17-19.

GUELKE, R. W., & HUYSSEN, R. M. J. Development of apparatus for the analysis of sound by the sense of touch. *Journal of the Acoustical Society of America*, 1959, *31*, 799-809.

HANSEN, A. The first case in the world: Miss Petra Heiberg's report. *Volta Review*, 1930, *32*, 223.

HANSEN, R. J. *Characterization of speech by external articulatory cues as a basis for a speech-to-tactile communication system for use by the deaf-blind.* Unpublished master's thesis, Massachusetts Institute of Technology, 1964.

HEIDER, F. K., & HEIDER, G. M. An experimental investigation of lipreading. *Psychological Monographs*, 1940, *52*, 124-153.

IFUKUBE, T., & YOSHIMOTO, C. A sono-tactile deaf-aid made of piezoelectric vibrator array. *Journal of the Acoustical Society of Japan*, 1974, *30*, 461-462.

IFUKUBE, T., YOSHIMOTO, C., & SHOJI, H. A finger-tip tactual vocoder with feature-extracting system. *Proceedings of the Research Conference on Speech Processing Aids for the Deaf.* Gallaudet College, in press.

KIRMAN, J. H. Tactile communication of speech: A review and analysis. *Psychological Bulletin*, 1973, *80*, 54-74.

KIRMAN, J. H. Tactile perception of computer-derived formants from voiced speech. *Journal of the Acoustical Society of America*, 1974, *55*, 163-169.

KIRMAN, J. H. Current developments in tactile communication of speech. In W. Schiff & E. Foulke (Eds.), *A sourcebook on haptic perception.* In press.

LING, D., & SOFIN, B. Discrimination of fricatives by hearing-impaired children using a vibrotactile cue. *British Journal of Audiology*, 1975, *9*, 14-18.

LINVILL, J. G., & BLISS, J. C. A direct translation reading aid for the blind. *Proceedings on the Institute of Electrical and Electronics Engineers*, 1966, *54*, 40-51.

MARTONY, J. Some experiments with electronic speech-reading aids. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1974, *2-3*, 34-56.

MILLER, G. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81-97.

MILLER, G., & NICELY, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 338-352.

MILLER, J. D., ENGEBRETSON, A. M., & DE FILIPPO, C. L. Tactile speech-reception aids for the hearing-impaired. *Journal of the Acoustical Society of America*, 1974, *56*, S47. (a)

MILLER, J. D., ENGEBRETSON, A. M., & DE FILIPPO, C. L. *Preliminary research with a three-channel vibrotactile speech-reception aid for the deaf.* Talk presented at Speech Communication Seminar, Stockholm, August 1974. (b)

MOOK, D. *Evaluation of a two-dimensional spectral tactile display for speech.* Unpublished master's thesis, Massachusetts Institute of Technology, 1978.

NORTON, S. J., SCHULTZ, M. C., REED, C. M., BRAIDA, L. D., DURLACH, N. I., RABINOWITZ, W. M., & CHOMSKY, C. Analytic study of the Tadoma method: Background and preliminary results. *Journal of Speech and Hearing Research*, 1977, *20*, 574-595.

OLLER, D. K., PAYNE, S. L., & GAVIN, W. J. Tactual speech perception by minimally trained deaf subjects. *Journal of Speech and Hearing Research*, 1980, *23*, 769-778.

PICKETT, J. M. Tactual communication of speech sounds to the deaf: Comparison with lipreading. *Journal of Speech and Hearing Disorders*, 1963, *28*, 315-330.

PICKETT, J. M., & PICKETT, B. H. Communication of speech sounds by a tactual vocoder. *Journal of Speech and Hearing Research*, 1963, *6*, 207-222.

PROCTOR, A., & GOLDSTEIN, M. H. *Lexical comprehension in a totally deaf child learning to talk.* Paper presented at National Convention of the American Speech-Language-Hearing Association, Detroit, November 1980.

REED, C. M., RUBIN, S. I., BRAIDA, L. D., & DURLACH, N. I. Analytic study of the Tadoma method: Discrimination ability of untrained observers. *Journal of Speech and Hearing Research*, 1978, *21*, 625-637.

REED, C. M., CONWAY-FITHIAN, S., BRAIDA, L. D., DURLACH, N. I., & SCHULTZ, M. C. Further results on the Tadoma method of communication. *Journal of the Acoustical Society of America*, 1980, *67*(Suppl. 1), S79.

REED, C. M., DURLACH, N. I., BRAIDA, L. D., & SCHULTZ, M. C. Analytic study of the Tadoma method: Identification of consonants and vowels by an experienced Tadoma user. *Journal of Speech and Hearing Research*, 1982, *25*, 108-116.

REED, C. M., DOHERTY, M. J., BRAIDA, L. D., & DURLACH, N. I. Analytic study of the Tadoma method: Further experiments with inexperienced observers. *Journal of Speech and Hearing Research*, in press.

RISBERG, A., GALYAS, K., & FRANZEN, O. Phonemic identification with lip-reading alone and lipreading supplemented by residual hearing or tactual communication. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1965, *1*, 14-21.

ROTHENBERG, M., & MOLITOR, R. D. Encoding voice fundamental frequency into vibrotactile frequency. *Journal of the Acoustical Society of America*, 1979, 66, 1029-1038.

SAUNDERS, F. A., HILL, W. A., & SIMPSON, C. A. Hearing substitution: A wearable electrotactile vocoder for the deaf. In H. Levitt, J. M. Pickett, & R. Hood (Eds.), *Sensory aids for the hearing impaired*. New York: IEEE Press, 1980.

SCILLEY, P. L. *Evaluation of an auditory prosthetic device for the profoundly deaf*. Unpublished master's thesis, Queen's University, Kingston, Ontario, Canada, 1980.

SCOTT, B. L., & DE FILIPPO, C. L. Evaluating a two-channel lipreading aid. *Journal of the Acoustical Society of America*, 1976, *60*, S124.

SCOTT, B. L., DE FILIPPO, C. L., SACHS, R. M., & MILLER, J. D. Evaluating with spoken text a hybrid vibrotactile-electrotactile aid to lipreading. *Proceedings of the Research Conference on Speech-Processing Aids for the Deaf*. Gallaudet College, in press.

SNYDER, J. C., CLEMENTS, M. A., REED, C. M., DURLACH, N. I., & BRAIDA, L. D. Tactile communication of speech: Comparison of Tadoma and a frequency-amplitude spectral display in a consonant discrimination task. *Journal of the Acoustical Society of America*, in press.

SPARKS, D. W., KUHL, P. K., EDMONDS, A. A., & GRAY, G. P. Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of segmental features. *Journal of the Acoustical Society of America*, 1978, 63, 246-257.

SPARKS, D. W., ARDELL, L. A., BOURGEOIS, M., WIEDMER, B., & KUHL, P. K. Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of connected discourse. *Journal of the Acoustical Society of America*, 1979, 65, 810-815.

SPENS, K. E. Pitch information displayed on a vibrator matrix as a speech-reading aid: Some preliminary results. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1975, 2-3, 34-39.

SPENS, K. E. Preliminary results from an experiment on recognition of spectral patterns on a vibrator matrix with different time windows. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1976, 2-3, 59-65.

SPENS, K. E. Is there an optimal time window for tactually conveyed spectral patterns derived from the speech window? *Proceedings of the Research Conference on Speech-Processing Aids for the Deaf*. Gallaudet College, in press.

STRATTON, W. D. Intonation feedback for the deaf through a tactile display. *Volta Review*, 1974, 76, 26-35.

UPTON, H. W. Wearable eyeglass speechreading aid. *American Annals of the Deaf*, 1968, *113*, 222-229.

VAN ADESTINE, G. An evaluation of the Tadoma method. *Volta Review*, 1932, *34*, 199.

VIVIAN, R. The Tadoma method: A tactual approach to speech and speechreading. *Volta Review*, 1966, 68, 733-737.

VON BEKESY, G. Human skin perception of traveling waves similar to those on the cochlea. *Journal of the Acoustical Society of America*, 1955, 27, 830-841.

WALDEN, B. E., PROSEK, R. A., MONTGOMERY, A. A., SCHERR, C. K., & JONES, C. J. Effects of training on the recognition of consonants. *Journal of Speech and Hearing Research*, 1977, *20*, 130-145.

WIENER, N., WIESNER, J. B., DAVID, E. E., & LEVINE, L. FELIX (Sensory Replacement Project). *Quarterly Progress Reports*. Research Laboratory of Electronics, MIT, 1949-1951.

WILLEMAIN, T. R., & LEE, F. F. Tactile pitch displays for the deaf. *IEEE Transactions on Audio and Electroacoustics*, 1972, *AU-20*, 9-16.

YENI-KOMSHIAN, G. H., & GOLDSTEIN, M. H. Identification of speech sounds displayed on a vibrotactile vocoder. *Journal of the Acoustical Society of America*, 1977, 62, 194-198.

ZEISER, M. L., & ERBER, N. P. Auditory/vibratory perception of syllabic structure in words by profoundly hearing-impaired children. *Journal of Speech and Hearing Research*, 1977, *20*, 430-436.