Journal of Rehabilitation Research and Development, Vol. 23 No. 1 1986
pages 119–128

Continuing Evaluation of the Queen's University Tactile Vocoder. I: Identification of Open Set Words

P. L. BROOKS, Ph. D.
B. J. FROST, Ph. D.
J. L. MASON, Ph. D.
D. M. GIBSON, M. Sc.

Departments of Psychology* and Electrical Engineering*
Queen's University
Kingston, Ontario, Canada K7L 3N6

Abstract — Identification of open set words, by an experienced normal hearing subject using the tactile vocoder developed at Queen's University, was examined. The tactile vocoder filters and processes the acoustic waveform into 16 filter channels, each of which controls a vibrator on the skin surface. After acquiring a 250-word vocabulary through the tactile vocoder, the subject was presented with three sets of 1000 different open set words in three reception conditions. The percentages of words correctly identified in the tactile vocoder (TV), lipreading (L), and lipreading plus tactile vocoder (L + TV) conditions were 8.8, 39.4, and 68.7 percent respectively. Phonemic analysis of stimulus/response pairs revealed that 36.7, 64.9 and 85.5 percent of the phonemes were correctly identified in TV, L, and L + TV conditions, respectively, indicating that incorrect-response words often contained many correct phonemes. Also, syllabic stress of stimulus and response words was identical 88 percent of the time in the TV condition. Important information about speech was transmitted through the tactile vocoder.

INTRODUCTION

There have been many attempts to provide tactile sensory communication for the deaf through tactile vocoders. These devices filter the acoustic waveform and transduce it into vibratory or electrical patterns of skin stimulation: see references (1) through (14). A review of these systems and their evaluation can be found in Reed et al., 1982 (15) and Sherrick, 1984 (16).

The experiments described in this paper are part of a continuing evaluation of the tactile vocoder developed at Queen's University. In this system the acoustic waveform is filtered and the output activates a series of vibrators that are in contact with the skin. The goal of this research is to provide the profoundly deaf with a wearable tactile vocoder that will aid in identification of speech and environmental sounds, and provide feedback for improvement of vocalizations.

Several generations of tactile aids have been developed at Queen's University. The first two systems had 10 and 14 filter channels that activated a series of solenoids that touched the skin surface. Artificially deafened subjects learned to discriminate upswept and downswept tones chosen to approximate some of the critical formant transitions in speech as described by Martin in 1972 (17). Using the second-generation system, which extended the high frequency range, subjects learned to identify 20 environmental sounds and a 10-word vocabulary as described by Kofmel

*Dr. Brooks and Dr. Frost are members of the Department of Psychology; Dr. Mason and Mr. Gibson are members of the Department of Electrical Engineering.

Address for correspondence: Dr. B. J. Frost, Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6.

in 1979 (18).

Results from the latter study led to modification and design of a third-generation system, which was described in detail by Brooks and Frost in 1983 (19). Extensive evaluation of that system was undertaken by Scilley in 1980 (20): acquisition of a tactual vocabulary was tested using two artificially deafened subjects who obtained information solely through this tactile vocoder. With 80.5 and 40.5 hours of experience, respectively, two subjects learned 250 and 70 words (20, 19, 21). Word pairs that had been initially confused became easier for the subjects to identify as training time increased. Analyses showed that most consonants were discriminated; the few exceptions were confusion of phonemes within the categories of voiced stops, approximants, and nasals. Introduction of novel readers did not interfere markedly with the subjects' performance, indicating that the subjects were extracting general characteristics of the speech signal.

In addition to this evaluation with artifically deafened subjects, a series of experiments was carried out on a postlingually profoundly deaf 13-year-old boy (20). In 12 hours this subject enthusiastically learned to recognize 50 environmental sounds. Results also indicated that the device could be useful in eliminating many confusions inherent in lipreading by supplying the subject with information that helped him discriminate within mouth-movement groups. Lipreading scores within mouth-movement groups (e.g. 'pet-bet-met') averaged 33 percent, whereas identification within these same groups was 88 percent when the subject used only information from the tactile vocoder. Finally, a brief test showed that the intelligibility of the subject's own speech improved 104 percent, after a few hours of training, when feedback from the tactile vocoder was provided.

Fourth-generation design — Analysis of confusions led to a design modification that pre-emphasized high frequencies. A filter with 6 dB/octave gain starting at 3 kHz was inserted after the microphone stage. The system with this addition was referred to as the fourth-generation tactile vocoder; it was evaluated using **open set words** as stimuli in the experiment, described here. ("Open set" refers to the fact that the subject was not previously shown a list of possible alternatives on which to base the response.) Results on identification of words using the tactile vocoder alone, lipreading alone, and lipreading plus the tactile vocoder are reported.

Performance of an experienced subject on identification of open set sentences, and tracking rates using the "tracking procedure" of DeFilippo and Scott, 1978 (22) are described in the paper* following this one. Prelingually profoundly deaf teenagers have also been trained to use the tactile vocoder. Word acquisition rates, placement of CV's in linguistic categories, and results on standardized tests such as the C.I.D. Sentence Test, are reported in Brooks, 1984 (23) and Brooks, Frost, Mason and Gibson (24).

METHOD

Subject

One female graduate student who had normal hearing participated in this study. Prior to this study, the subject had spent 80.5 hours acquiring a tactual vocabulary of 250 words, 8 hours testing on the total 250-word vocabulary, 20 hours of generalization and discrimination training, and 87.5 hours testing various design modifications — a total of 196 hours of experience.

Specialized lipreading training, and practice using lipreading and the tactile vocoder in conjunction, were not given prior to this study; this subject possessed only the lipreading skills that normal-hearing individuals develop through everyday speech comprehension.

Apparatus

Sound information entered the portable fourth-generation tactile vocoder through two Sony ECM-16 electret microphones, one held by the reader and the other held by the subject. A filter with 6 dB/octave gain starting at 3 kHz was inserted after the microphone. The sound spectrum was then analyzed by 18 one-third-octave bandpass filters (center frequencies: 160, 200, 250, 320, 400, 500, 640, 800, 1000, 1260, 1600, 2000, 2560, 3200, 4000, 5000, 6400, and 8000 Hz). Outputs from the 160-Hz and 200-Hz filters, and from the 250-Hz and 320-Hz filters, were combined to form two channels with increased bandwidths.

Each filter channel was followed by an envelope detector and a logarithmic amplifier. The output from a given filter channel was used to amplitude-modulate a 100-Hz square-wave carrier that drove the solenoids.

The system was housed in a cabinet 16x23x28 cm (weight 3.6 kg), which was attached to a 12-V power supply. Power consumption of the system was 7.0 W. Further details on the signal processing

*Brooks PL, Frost BJ, Mason JL, Gibson DM: Continuing Evaluation of the Queen's University Tactile Vocoder — II: Identification of Open Set Sentences and Tracking Narrative.

can be found in the user manual by Gibson, 1983 (25). Five of these portable systems have been constructed at the time of this writing.

The 16 solenoids were spaced 3 cm apart along a supporting piece of Plexiglas which was attached to the subject's ventromedial forearm. Output from the channel of the lowest frequency activated the solenoid closest to the wrist, with increasing frequencies represented along the length of the array.

Procedure

The subject and reader sat in separate adjoining sound-attenuated booths separated by four panes of glass with air gaps between the panes. Earplugs and headphones delivering white noise were worn by the subject, to mask noise from the vibrating solenoids.

The 3000 English words used as stimuli were drawn from Kucera and Francis (26) and O'Rouke (27), and are shown in the appendices of Brooks (23). Test lists of approximately 40 words were created by the reader.

Word identification was tested daily under three different conditions: tactile vocoder alone (TV); lipreading alone (L); and lipreading plus the tactile vocoder (L+TV). In a half-hour session the subject would receive one test in the L condition and another test in the L+TV condition. In a different half-hour session a test was given under the TV condition. The order of the sessions, and conditions within sessions, were randomly determined. Approximately 7 hours of testing occurred for each of the L and L+TV conditions and 14 hours of testing occurred for the TV condition. The number of presentations given in the TV condition, described below, accounts for the difference in testing time between conditions.

In the L and L+TV conditions, visual contact was allowed between the two sound-attenuated booths. The subject wore the vocoder in both conditions and set the ON/OFF switch in accordance with the experimental condition. The reader was not informed of the condition order. During the session, the reader established eye contact with the subject and then said a stimulus word in a clear, normally paced voice. The subject's response was heard through an intercom and was also recorded. A single presentation of each test word was given. Immediate feedback was provided following the subject's response.

In the TV condition, visual contact between the subject and reader was blocked. The reader said a stimulus word and the subject responded. If the response was not correct the stimulus word was presented again — with a maximum of three presentations allowed. No information was furnished that enabled the subject to determine how close she was to identifying the word until after the three presentations.

Training in open set word identification was not given to the subject prior to these tests. It should be noted that, concurrent with this testing, the subject was also receiving a half-hour per day of testing both on sentence identification and on tracking. A minimum of a half-hour break separated the four daily sessions, and the order of the four sessions in a given day was random.

RESULTS

Before any data are presented, it should be noted again that the experiments described were not designed to enable direct comparison between the TV condition and the two lipreading conditions. In the L and L+TV conditions, words were presented a single time, whereas in the TV condition stimuli could be presented up to a maximum of three times. However, since the data for the TV condition underwent the same analyses as the two lipreading conditions, the results are presented together.

Identification of Open Set Words

Initially, the data were analyzed by calculating the number of words that were correctly identified from the 1000 different words presented in each condition. A correct identification was scored when the phonemes in the response word were identical to the stimulus word (e.g. the response 'friend' to the stimulus word 'friends' would be incorrect).

In the L condition 39.4 percent of responses were correct, whereas in the L+TV condition 68.7 percent of response words matched stimulus words; an improvement of 29.3 percent in performance. In the TV condition a total of 88 (8.8%) words, shown in Table 1, were correctly identified. Of the random 1000 stimulus words presented in the TV condition, 64 were also found in the 250-word acquired vocabulary. Forty of these 64 (62.5%) were correctly identified. It is noteworthy that these words were not confused with linguistically similar words when they were included in the larger set.

The errors that occurred in the tactile-vocoder-alone condition demonstrated that substantial correct information was being obtained through the device, even though the response word was not phoneme-for-phoneme identical to the stimulus word. In many cases, only one phoneme would be different in the stimulus and response pairs, as

TABLE 1
Open Set Words Identified Correctly: Tactile Vocoder

About	Mister
Above	Money
Across	Music
Afternoon	My
Age	Need
Allowed	Number
Attack	Open
Babies	Outside
Baby	Park
Bird	Past
Bit	Picture
Breakfast	Please
Car	Poor
Care	Radio
Cars	Relax
Century	Salt
Difficult	Seed
Discussion	Seven
Doesn't	Shall
End	She
Enough	Ship
Feel	Shop
Finished	Show
Fixed	Speak
Forgive	So
Friends	Son
Function	Something
Habit	Standing
Half	Strange
Happen	Strong
Heart	Struck
Hot	Student
Instead	Students
Interested	Stuff
Itch	Success
Keep	Test
Letter	Think
List	Today
Lost	Together
Lunch	Tried
Many	Trust
Married	Understand
Matter	Until
Maybe	Week

TABLE 2
Near Misses Obtained in the TV Condition.

Stimulus	Response	Stimulus	Response
act	at	history	sisterly
afford	afraid	keen	been
annual	annually	loss	moss
applied	applaud	left	lift
at	bat	lines	lens
attention	detention	lives	lines
base	bus	machine	mission
brat	rat	merely	yearly
breath	proof	morning	meaning
brother	rather	muttered	watered
case	curse	off	if
can't	cat	poisonous	business
cat	cot	ring	bring
central	centre	screen	skeen
check	chick	scurry	scary
council	cancel	seen	seed
creepy	teepee	she	shoe
dad	had	shirt	short
date	deet	snow	so
dear	ear	spoke	speak
door	for	staff	stuff
doubt	out	step	stop
eight	at	strive	stove
explain	exclaim	task	pask
facts	acts	taste	cost
far	for	through	three
figure	finger	thus	this
frame	from	tradition	condition
fun	sun	truth	true
further	father	word	world
Germany	chimney	worst	west
glass	gas		
hate	hat		

some examples in Table 2 indicate. In numerous cases the stimulus and response words differed by more than one phoneme, but the similarity between the words indicated that significant information had been acquired through the tactile vocoder. Samples of some of the more interesting substitutions, where two or more phonemes were incorrect, are also shown in Table 2.

Visual analysis of incorrect TV responses

Visual analysis of word pairs in the TV condition revealed that the subject's incorrect responses were not random but followed specific patterns.

1. The subject's response decision was influenced more by the consonants in the stimulus words than by the vowels. In many cases, such as the ones

shown below, the response consonants were all correct but the response vowels were not.

Stimulus	Response	Stimulus	Response
off	if	applied	applaud
hate	hat	step	stop
frame	from	spoke	speak
tape	top	staff	stuff

2. The phonemic substitutions that occurred were often from within the same linguistic category. For example, the response to 'past' was 'cask'. Both the stimulus and response words began and ended with unvoiced stop consonants, although the specific unvoiced stops were confused. Other stimulus-response pairs show substitutions among voiced stops ('bad-dog', 'agreed-abroad'), nasals ('seen-sing', 'can-calm'), approximants ('round-wind', 'clay-tray'), and confusion between nasals and approximants ('watch-much', 'merely-yearly'). When phonemes were not identified correctly, often a phoneme within the same linguistic category was substituted.

3. Stimulus-response pairs such as 'worst-west', 'step-stop', 'staff-stuff', 'spoke-speak', 'task-cask', 'scurry-scary', and 'left-lift' indicate that the subject was successfully identifying many unvoiced-fricative/ unvoiced-stop consonant combinations.

4. The syllable / n/ (e.g. tension) was presented 24 times in the novel words and 11 responses to these words also contained this syllable. An additional 4 responses contained the voiced syllable/ n/ (e.g. decision). Thus the subject reliably identified this syllable, even though the 250-word acquired vocabulary only contained the voiced syllable / n/ once (television).

5. The voiced stop consonants /b/ and /d/ were frequently omitted when they occurred at the beginning of a word, as the stimulus-response pairs 'dear-ear' and 'bring-ring' demonstrate.

6. When the approximants /r/ and /l/ occurred in a consonant cluster they were not always detected, as the following stimulus-response pairs demonstrate: sleep-soap, glass-gas, green-been, creepy-teepee.

The patterns of errors that occurred in the lipreading condition will not be discussed in detail here. A substantial amount of research exists that reports the lipreading confusions that are typically found. See: Woodward and Barber, 1960 (28); Fisher, 1968 (29); Jeffers and Barley, 1971 (30); Franks and Kimble, 1972 (31); Binnie, Jackson, and Montgomery, 1976 (32). The present findings mirror those results. Briefly, confusions occurred between phonemes that have the same mouth-movements.

Examinations of stimulus and response pairs in the L+TV condition indicated that many of the errors could be categorized as follows:

1. Incorrect identification of vowels: e.g. 'done-den', 'press-price'.

2. Voicing errors: e.g. 'bump-pump', 'surface-service' and 'view-few'.

3. The consonant /h/ was omitted or added in several cases such as the following ones: 'older-holder', 'arm-harm' and 'he's-ease'.

Phonemic analysis of stimulus-response pairs

Response patterns in all three conditions reveal that simple calculation of the number of words correctly identified does not accurately reflect the total amount of information that is available through the tactile vocoder. In order to obtain additional information about the capability of the systems, a phonemic analysis of the data was undertaken. Data analysis in the first stage involved phonemic transcription of the 6000 stimulus and response words. A graduate in linguistics, familiar with the Southern Ontario accent, performed the transcription. The 14 vowel and 25 consonant phonemes employed in the transcription are shown in Table 3.

Stimulus and response words were then matched, phoneme-for-phoneme. Two extra symbols denoted that a phoneme existed in the response word that did not exist in the stimulus word and vice versa. For example, if the response 'face' was uttered for the stimulus word 'faces', a symbol indicated that when /z/ was presented no response had occurred. If the stimulus and response words differed markedly, the match was performed in serial order. The phonemic transcription was undertaken on the 3000 stimulus-response pairs in the L, TV, and L+TV conditions.

A reliability check was undertaken on the transcription and matching procedures. Five percent of the stimulus/response pairs were randomly chosen, and transcription and matching carried out by a second linguist was compared to that of the first. Agreement between the two linguists was 97 percent.

A computer program constructed a matrix of the stimulus-response phoneme pairs that occurred in each of the three conditions, and calculated percentage-correct scores.

It should be noted that in all instances the effects of context should not be ignored. Although a phonemic analysis has been undertaken on the data, the original data concerned words. The subject was not asked to identify individual phonemes,

or CV's, from a fixed set, but instead attempted to identify English words. By responding with words, contextual cues were brought into play. Consequently, some phonemes were undoubtedly added or missed so that a real word was given as a response.

Results from the phonemic analysis are displayed in Table 3 for the TV, L, and L+TV conditions. When response words were incorrect, the responses often contained many correct phonemes. Although in the TV condition only 8.8 percent of the words were correctly identified, if the responses were analyzed at the phonemic level the actual number of phonemes correctly identified was 36.7 percent. Similarly, in the L and L+TV conditions, the percentages of words correctly identified were 39.4 percent and 68.7 percent, respectively, whereas at the phonemic level, response words contained 64.9 percent and 85.5 percent correct information.

TABLE 3
Identification of Phonemes within Words

Vowels		TV	L	L+TV
i	beat	33.2%	72.5%	87.3%
I	bit	36.7	60.5	84.2
e	bait	21.9	66.4	80.4
ɛ	bet	31.7	65.0	82.5
æ	bat	39.7	67.1	85.7
ɑ	bought	22.9	82.2	95.5
o	boat	18.5	85.8	96.6
ʊ	put	0	44.0	68.2
u	boot	22.0	82.8	95.8
ʌ	but	37.9	61.8	81.2
ai	bite	7.1	69.6	68.7
aʊ	bout	60.0	88.9	100.0
ɔɪ	boy	0	44.4	100.0
ɜ̊	bird	39.0	72.2	90.2
Consonants		**TV**	**L**	**L+TV**
p		17.5%	74.2%	91.4%
t		66.4	52.1	83.2
k		46.5	40.6	79.3
b		26.0	59.8	82.5
d		41.1	47.7	73.6
g		11.7	53.3	67.3
m		16.3	74.1	95.6
n		43.0	53.0	70.7
ŋ	(sing)	39.0	49.8	89.8
f		45.5	90.6	96.2
θ	(thin)	23.8	96.0	96.7
ʃ	(shin)	48.3	43.7	87.0
tʃ	(chin)	51.7	53.6	86.8
s		76.2	57.2	85.5
h		39.0	74.4	83.3
hw		0	66.6	100.0
v		22.5	77.9	94.1
ð	(the)	27.8	92.8	100.0
z		53.2	39.3	84.7
dʒ	(jaw)	22.8	64.1	88.9
ʒ	(vision)	25.0	37.5	100.0
w		13.1	78.2	92.0
r		21.7	85.2	91.9
j	(yes)	32.2	37.5	70.3
l		15.4	84.5	91.7
Total Correct Phonemes		36.7%	64.9%	85.5%

Once again, it should be noted that the subject received three trials for each word in the TV condition — and one trial for words in the other conditions. Thus the TV condition should not be compared directly with the L or L+TV conditions, or the comparison must be made acknowledging the procedural differences.

An average of 4882 phonemes were presented in each condition, and the frequency of occurrence for each phoneme followed the frequency with which they occur in the English language. In the L+TV condition, 32 of the 39 phonemes were correctly identified more than 80 percent of the time, whereas in the L condition 9 of the 39 identification scores were greater than 80 percent. In every case, performance with the tactile vocoder and lipreading was superior to lipreading performance alone. When the total number of correct phoneme identifications was taken as a percentage of the total number of phonemes presented, it was found that 64.9 percent and 85.5 percent of vowels were correctly identified in the L and L+TV conditions, respectively, an improvement of 20.6 percentage points.

In the TV condition, 36.7 percent of phonemes were correctly identified. The six phonemes with the highest percent correct scores were /t/, /k/, /s/, /ʃ/, /tʃ/ and /f/. One characteristic shared by these sounds is that they contain primarily high frequency energy. Identification of /t/, /k/, /ʃ/, /s/, and /z/ were higher with the tactile vocoder alone than with lipreading alone.

Identification of phonemes within linguistic categories

From visual analysis of Table 2 it is apparent that the majority of errors were not random. Phonemic substitutions in the response words were often from the same linguistic category as the stimulus phoneme. To examine this more thoroughly, frequencies from the stimulus-response matrix of phonemes were summed to form the categories of unvoiced stops, voiced stops, unvoiced fricatives,

voiced fricatives, and approximants. Detailed results were described in Brooks in 1984 (23). Briefly, identification using only the tactile vocoder was 74.7 percent for unvoiced fricatives and 73.4 percent for unvoiced stops. For the L+TV condition, identification of unvoiced stops, unvoiced fricatives, voiced fricatives, and approximants hovered around 90 percent correct, indicating that the features defining the categories were reliably represented through the combination of the tactile vocoder and lipreading.

Identification of number of syllables and stress pattern

One of the striking observations that can be made from Table 2 is that stimulus and response words in the TV condition follow the same rhythmic pattern, in most instances. Even when the response is incorrect, as in 'sisterly-history', the number, duration, and stress pattern of the syllables were identical.

To examine this relationship more closely, the syllable patterns of the words in the TV, L, and L+TV conditions were scored according to the dictionary entry for each word. Of the 3000 stimulus words, 92.0 percent of them fell within the five following categories:

1. Monosyllabic words: one-syllable words
2. Trochees: two-syllable words, first syllable stressed, second unstressed (e.g. father)
3. Iambs: two-syllable words, first syllable unstressed, second stressed (e.g. hotel)
4. Dactyles: three-syllable words, first syllable stressed, second syllable unstressed, third syllable unstressed (e.g. history)
5. Ambibrachs: three-syllable words, first syllable unstressed, second syllable stressed, third syllable unstressed (e.g. together)

To ensure reliable probabilities, only these five categories of stress pattern were examined. Table 4 displays the percentage correct within each category for the three reception conditions. The total number of cases where the stress patterns of the stimulus and response words matched were also taken as a percentage of total stimuli presented. The percentage of response words that had stress patterns identical to those of the stimulus words was 88.0, 86.6 and 96.3 for the TV, L, and L+TV conditions, respectively. The L+TV condition contained the highest performance score for each of the five categories.

The most general error observed was that underestimation of the number of syllables occurred more frequently than overestimation, across conditions. In the TV condition when dactyles were presented, 42.5 percent of the time the response word was a trochee, a very striking confusion. In most of these cases a nasal, approximant, or vowel existed between two syllables in the stimulus word, and the response word given indicated that this syllable was missed.

TABLE 4
Identification of Syllabic Stress: within each of five named categories of stress pattern in one-, two-, and three-syllable words, the percentages indicate percent of words for which response word showed number, duration, and stress pattern of syllables identical to stimulus word, under each of the three reception conditions.

Syllabic Stress	Reception Condition		
	TV	L	L+TV
Monosyllable	95.9%	92.2%	97.8%
Trochee	92.7	83.8	97.6
Iamb	79.2	82.9	92.5
Dactyl	50.0	71.4	89.5
Ambibrach	67.9	82.9	91.1
Total	88.0	86.6	96.3

One vital point should be emphasized. This syllable analysis was undertaken because of patterns that had been observed in the data. But in the experiment, a stimulus word was given and the subject's task was to respond with a word. If the subject had been asked simply to tap out the rhythmic pattern of the word, the results might have been quite different.

DISCUSSION

Two particularly striking results arose from this study. First, after limited training that involved the acquisition of a 250-word vocabulary through the tactile vocoder, a subject was able to identify 8.8 percent of the random English words presented through the system. Words for which the subject had never received prior tactual experience were recognized correctly. In addition, the incorrect responses were not random, but instead often contained many phonemes that were identical to those found in the stimulus words. Further, when the phonemes in the stimulus and response words were not identical, the substitution phonemes often belonged to the same linguistic category as the stimulus phonemes. Thus, important information about speech is available through the tactile vocoder.

The second striking result is the impressive performance obtained when the tactile vocoder and

lipreading were used in conjunction. Response words were identical to the stimulus words 68.7 percent of the time, and at the phonemic level 85.5 percent of the response information was correct. Thirty-two of the 39 phonemes were identified correctly more than 80 percent of the time.

The fact that words were used as stimuli in these experiments complicated the work of differentiating between the effects of the particular reception condition and those of language competence. Language skill is acknowledged to be a crucial factor in speech perception, and the difficulty of the word identification task was reduced because of linguistic predictability within words. The experiment could have been designed to test performance employing nonsense syllables or CV's as stimuli, thereby removing within-word contextual effects. However, the present philosophy of the researchers is that the device should be tested primarily in situations that approximate the normal environment, rather than by concentrating attention on analytic questions or by requiring identification of elements.

It is also apparent that the tactile vocoder and lipreading provided different types of information. If the information available through the tactile vocoder duplicated that supplied by lipreading, one would expect the L and L+TV scores to be similar. However, 20.6 percent more phonemes were correctly identified in the L+TV condition than in the L condition, and for every phoneme, identification was higher in the L+TV condition.

The types of information available through lipreading and the tactile vocoder appear to be complementary, as well as supplementary. Specifically, four features that were detectable through the tactile vocoder (but were difficult to lipread) were voicing, frication, nasality, and plosion. The two types of information used in conjunction may provide sufficient information to allow knowledge of contextual constraints to limit the alternatives even more.

One final point to note about the L+TV condition is the apparent ease with which the two sources of information were integrated. The very first time the subject used both the tactile vocoder and lipreading together, 68.0 percent of the words were correctly identified. It appears that speech features can be extracted by independent modalities and combined by higher-order processors, and that both the extraction and integration can occur with ease.

Number of syllables and stress pattern of words were identified well in the TV, L and L+TV conditions. Performance in the L+TV condition, where 96.3 percent of the response words had stress patterns identical to those of the stimulus words, approached perfect performance. Remember that this value was calculated by taking the total number of cases where the stimulus and response words had identical stress patterns compared to the total number of stimuli presented. Thus, the frequently occurring stress patterns contributed more heavily to the measure than did the infrequent ones. This ensures that the percentage has more meaning for generalizing to real-world situations than it would have if all categories had been presented equally.

Of particular interest was the high correspondence between the stress patterns of stimulus words and response words in the TV condition. Some researchers (33, 34, 35, 36) have expressed concern that since spectral displays divide the acoustic waveform into many channels, there is an absence of a single coherent amplitude envelope. The importance of the envelope, as an integrator for the more rapid spectral changes, has been stressed by Cole and Scott in 1974 (37). Kirman in 1973 (35) has suggested that failure of spectral displays may be due to the lack of an available amplitude envelope. Scott et al. in 1977 (36) stated that the impressive performance of Gaults's subjects, who in 1924 were able to identify any one of 120 sentences, may be due to the fact that Gault's single vibrator provided the rhythmic patterns of stress (1). The present results, where 88 percent of the response words in the TV condition had patterns of stress identical to those of the stimulus words, contradicts these ideas.

Results from the present study also do not agree with more recent studies. Bourgeois and Goldstein in 1983 (38) reported tests of visual (L), tactual (TV), and visual-plus-tactual (L + TV) perception of number of syllables in sentences that contained 1, 3, 5, and 7 syllables. The device used in the tactual conditions was the MESA: see Sparks et al., 1978 (11). MESA is a two-dimensional frequency/amplitude display. Five normal, temporarily deafened subjects were familiarized with the conditions for three one-hour sessions before testing began. The percentage of correct syllable-tapping responses for monosyllabic stimuli was 62.7, 94.0, and 90.0 percent for the TV, L, and L + TV conditions, respectively. The lipreading results were very similar to those obtained in the present study (92.2 percent), so the comparison between studies is probably valid. However, performance in the TV condition differed dramatically. In the TV condition in the present experiment monosyllabic words were correctly identified 95.5 percent of the time, which is 33.2 percent higher then the results obtained by Bourgeois and Goldstein's

subjects. The trend was similar for three-syllable stimuli: identification was 42.3 percent in the Bourgeois and Goldstein study and 60.7 percent in the present study. It is possible that differences between subjects could account for this discrepancy, since the present study uses a single subject, whereas Bourgeois and Goldstein report an average over five subjects. However, the most obvious explanation for the difference in results is the amount of training the subjects in the two studies received using the tactile vocoder. Approximately 200 hours were received by the subject previous to this study, whereas only 3 hours of training were given in the Bourgeois and Goldstein study (38).

In determining syllable number, an inexperienced user generally counts the number of "bounces" felt through the tactile vocoder. Since the frequency content of the incoming auditory signal is divided into many channels, this is not an accurate indication of number of syllables. However, an experienced subject learns to judge syllable stress and number accurately by means other than "bounce counting". The subjects in Bourgeois and Goldstein's study were never provided with feedback, so they were unable to adopt a more reliable strategy for identifying stress.

Discouraging results were also reported by Beachler and Carney in 1981 (39); they examined syllabic stress patterns as well as doing syllabic counting. Normal hearing subjects were tested on their ability to detect syllabic stress through two different vibrotactile instruments.

The first device was the Fonator, described by Schulte in 1970 (40). It delivers an unfiltered vibratory signal to the subject's fingers or palms.

The second device was the 24-channel tactile vocoder designed by Engelmann and Rosov, who published in 1975 (7). Their device more closely resembles the device used in this experiment. The subjects were asked to identify the number of syllables present in 70 words of 1, 2, and 3 syllables, and the syllabic stress of 10 examples each of spondees, trochees, iambs, dactyls, ambibrachs, and anapests.

For identification of syllable number, both trained and untrained subjects scored around 66 percent when tested on the Fonator. Performance scores on their tactile vocoder after six training sessions (58%) were 20 percent higher than pre-training scores, although in general performance was poorer than that found on the Fonator.

In the present study, the data were reanalyzed to examine identification of number of syllables in one-to-three-syllable words through the tactile vocoder alone. It was found that 90.7 percent of the stimulus words had response words with the same number of syllables. Performance of the experienced subject in this study (90.7%) was substantially better than that obtained by Beachler and Carney's subjects (58%) on a similar system.

Results for the identification of syllable stress showed much the same trends. Regardless of training, all subjects tested on the Fonator in Beachler and Carney's 1981 study scored in the 47-49 percent range. Subjects trained on their vocoder scored an average of 44 percent correct, whereas untrained subjects scored 38 percent. Data for the present study was analyzed to examine syllabic stress identification of the six categories of two- and three-syllable words that were used in the Beachler and Carney study. For the 491 words that fell in these six categories, 391 (79.6%) had identical syllabic stress as the stimulus words. This is much superior to the average of 44 percent scored by Beachler and Carney's subjects (39).

Beachler and Carney concluded that caution should be exercised in using vibrotactile devices, particularly multichannel ones, to teach syllable number and stress. However, the results from the present study strongly suggest that these conclusions were premature. Six sessions of training may not be sufficient to judge the capabilities of any tactile vocoder. In the tactile vocoder condition in the present study, 88.0 percent of response words had identical syllabic stress as stimulus words, which indicates that very important stress information is being transferred through the device.

The most noteworthy problem in detection of syllabic stress that existed in the TV condition was the frequent lack of detection of syllables when /l,w,r,j,m,n/ occurred at syllabic boundaries. This is a problem that had been observed previously with our tactile vocoder by Scilley (20). Erber (34) has attempted to show this optically by displaying the envelope patterns of words that have been bandpass-filtered between 200 and 600 Hz. Most two-syllable words produce a pattern that resembles what we hear. However, many words do not produce this typical pattern, and the anomalies occur whenever syllables are joined by phonemes such as nasals or aproximants. For example, the word "lemon" appears as a single energy burst. Thus the lack of detection of syllable boundaries that sometimes occurs when /l, w, r, j, m/ or /n/ are at the boundaries may be a result of the acoustic properties of the waveform, and of our lack of an accurate definition and understanding of the word "syllable."

In conclusion, the results of this paper indicate

that a very large amount of information is available through the tactile vocoder. Open set words were identified correctly with information obtained **only** through the tactile vocoder. When this information was integrated with information available through lipreading, 85.5 percent of the phonemes within words were correctly identified.

The following paper examines the performance of the same experienced subject, using open set sentences plus the tracking procedure. These stimuli provide syntactic and semantic contextual information that is not present in isolated words. Using continuous speech also allows examination of the problem of parsing continuous speech into words ■

REFERENCES

1. Gault RH: Progress in experiments on tactile interpretation of oral speech. J Abn Soc Psych 19:155-159, 1924.
2. Wiener N, Wiesner J: Felix (Sensory replacement project). Q Prog Rep Res Lab Electron MIT, 1949-1951.
3. Guelke RW, Huyssen RMJ: Development of apparatus for the analysis of sound by the sense of touch. J Acoust Soc Am 31:799-809, 1959.
4. Pickett JM, Pickett PH: Communication of speech sounds by a tactual vocoder. J Speech Hear Disord 6:207-222, 1963.
5. Ifukube T, Yoshimoto C: A sonotactile deaf-aid made of piezoelectric vibrator array. J Acoust Soc Jpn 30(8): 461-462, 1974.
6. Kirman JH: Tactile perception of computer-derived formant patterns from voiced speech. J Acoust Soc Am 55:163-169, 1974.
7. Engelmann S, Rosov R: Tactual learning experiment with deaf and hearing subjects. Except Child 41:243-253, 1975.
8. Goldstein MH, Stark RE: Modification of vocalizations of preschool deaf children by vibrotactile and visual displays. J Acoust Soc Am 59(6): 1477-1481, 1976.
9. Saunders FA, Hill WA, Simpson CA: Speech perception via the tactile mode: Progress report. Proc IEEE Int Conf Acoust Speech Sig Processing, Philadelphia. pp 594-597, 1976.
10. Yeni-Komshian GH, Goldstein MH: Identification of speech sounds displayed on a vibrotactile vocoder. J Acoust Soc Am 62:194-198, 1977.
11. Sparks DW, Kuhl PK, Edmonds AE, Gray GP: Investigating the MESA (multipoint electrotactile speech aid): The transmission of segmented features of speech. J Acoust Soc Am 63(1):246-257, 1978.
12. Sparks DW, Ardell LA, Bourgeois MR, Wiedmer B, Kuhl PK: Investigating the MESA (multipoint electrotactile speech aid): The transmission of connected discourse. J Acoust Soc Am 65(3):810-815, 1979.
13. Oller KD, Payne SL, Gavin WJ: Tactual speech perception by minimally trained deaf subjects. J Speech Hear Res 23(4):769-778, 1980.
14. Clements MA, Durlach NI, Braida LD: Tactile communication of speech: Comparison of two spectral displays in a vowel discrimination task. J Acoust Soc Am 72(4):1131-1135, 1982.
15. Reed CM, Durlach NI, Braida LD: Research on tactile communication of speech: A review. ASHA Monographs 20, 1982.
16. Sherrick CE: Research on tactile aids for deaf people: Progress and prospects. J Acoust Soc Am 75(5):1325-1342, 1984.
17. Martin BD: Some aspects of visual and auditory information transfer by cutaneous stimulation. Unpubl M. Sc. thesis. Queen's University, Kingston, Canada, 1972.
18. Kofmel L: Perception of environmental and speech sounds through a vibrotactile device. Unpubl B.A. thesis. Queen's University, Kingston, Canada, 1979.
19. Brooks PL, Frost BJ: Evaluation of a tactile vocoder for word recognition. J Acoust Soc Am 74:34-39, 1983.
20. Scilley PL: Evaluation of a vibrotactile auditory prosthetic device for the profoundly deaf. Unpubl M.A. Thesis. Queen's University, Kingston, Canada, 1980.
21. Brooks PL, Frost BJ, Mason JL, Chung K: Identification of words through the tactile vocoder. J Acoust Soc Am (in press.)
22. DeFilippo CL, Scott BL: A method for training and evaluating the reception of ongoing speech. J Acoust Soc Am 63(4):1186-1191, 1978.
23. Brooks PL: Comprehension of speech by normal-hearing and profoundly deaf subjects using the Queen's University tactile vocoder. Unpubl doctoral dissertation. Queen's University, Kingston, Canada, 1984.
24. Brooks PL, Frost BJ, Mason JL, Gibson DM: Identification of words and manner of articulation features through a tactile vocoder by profoundly deaf subjects. J Speech Hear Res (in press).
25. Gibson DM: Tactile Vocoder User Manual. Queen's University, Kingston, Canada: Dept Electrical Engineering, 1983.
26. Kucera H, Francis WN: Computational analysis of present-day American English. Rhode Island, Brown Univ Press. pp 5-11, 1967.
27. O'Rourke TJ: A basic course in manual communication. Silver Spring, MD: Nat Assn Deaf. pp 107-119, 150-161, 1973.
28. Woodward MF, Barber CG: Phoneme perception in lipreading. J Speech Hear Res 3(3):212-222, 1960.
29. Fisher CG: Confusions among visually perceived consonants. J Speech Hear Res 11:796-804, 1968.
30. Jeffers J, Barley M: Speechreading. Springfield, IL, Chas. C Thomas. pp 41-78, 1971 (Fifth printing, 1977).
31. Franks JR, Kimble J: The confusion of English consonant clusters in lipreading. J Speech Hear Res 15:474-482, 1972.
32. Binnie CA, Jackson PL, Montgomery AA: Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. J Speech Hear Disord 41: 530-539, 1976.
33. Erber NP: Speech-envelope cues as an aid to lipreading for profoundly deaf children. J Acoust Soc Am 51: 1224-1227, 1972.
34. Erber NP: Vibratory perception by deaf children. Int J Rehab Res 1:27-37, 1978.
35. Kirman JH: Tactile communication of speech: A review and an analysis. Psychol Bull 80(1):54-74, 1973.
36. Scott BL, DeFilippo C, Sachs, R, Miller J: Evaluating with spoken text a hybrid vibrotactile-electrotactile aid to lipreading. **In** Papers from the Research Conference on Speech-Processing Aids for the Deaf. Washington, D.C.: Gallaudet College. pp 106-118, 1977.
37. Cole RA, Scott B: Toward a theory of speech perception. Psychol Rev 81: 348-374, 1974.
38. Bourgeois MS, Goldstein H: Visual and tactual perception of syllable number in sentence stimuli. J Acoust Soc Am 73(1):367-371, 1983.
39. Beachler CA, Carney AE: Vibrotactile perception of suprasegmentals: A comparison of single-channel and multi-channel aids. J Acoust Soc Am 69(S1):S123 (abstract), 1981.
40. Schulte K: Fonator System, **In** International Symposium on Speech Communication Ability with Profound Deafness. Washington, D.C.: Alex Graham Bell Assn Deaf:351-353, 1970.