Tactile speech communication aids for the deaf: A comparison

Spens, K-E.

journal: STL-QPSR
volume: 21
number: 4
year: 1980
pages: 023-039

http://www.speech.kth.se/qpsr
III SPEECH DEFECTS AND AIDS

A. TACTILE SPEECH COMMUNICATION AIDS FOR THE DEAF: A COMPARISON*

K.E. Spens

Abstract

This investigation is a comparison between some of the different speech-conveying tactile systems for the deaf described in recent literature. The correctly conveyed tactile information was measured using the same method, vocabulary and subject for all the systems tested. The results are possible to rank order by two factors. The most important factor seems to be the amount of processing of the speech signal in order to match the tactile stimulation to the characteristics of the skin. The second factor is the place of stimulation and, not surprisingly, loci with a higher receptor density (fingers) are superior to other places of the body.

Many tactile systems for conveying speech information to the deaf have been reported on. The different results obtained are, however, difficult to compare because of different testing procedures, vocabularies, subjects, etc. This investigation is a comparative study of some of the tactile systems described in recent literature. Only the tactile information capacity on a phonetic level was compared, which means that the tests were not made in conjunction with speech reading even if all the aids are meant to facilitate speech reading. Other features like price, weight, period of training needed and acceptability by the user are not discussed. The results must be regarded as indications only since for practical reasons the vocabulary, the number of subjects, and, the training time had to be limited.

DESCRIPTION OF THE SYSTEMS TESTED

Seven different tactile systems with nine different places of applications of the stimuli were tested, three systems stimulating one or more fingers (Schulte, 1972; Spens, 1977; Traunmyller, 1977) and five systems stimulating other parts of the body (Engelmann, 1975; Saunders, 1976; Scott; Schulte, 1972; Sparks, 1978), Fig. III-A-1.

* This paper is a more complete version of a paper presented at the International Congress on Education of the Deaf, Aug. 4-8, 1980, Hamburg, BRD.
Fig. III-A-1. Different tactile speech communication aids with their respective places of stimulation and originators.
Single vibrator systems applied on the fingers

The Phonator (Schulte, 1972) and the Sentiphone (Traunmyller, 1977) are both of the single vibrator type.

The Phonator consists of a power amplifier driving a high quality vibrator which covers the frequency range of the skin down to 30 Hz. The microphone signal is amplified which means that there is no processing of the speech signal at all. This corresponds to the intentions of the originator and is denoted Schulte (b). The test denoted Schulte (a) is a little different in that the high frequencies are purposely overloaded in order to create low frequency distortion products for speech sounds like (s). This kind of distortion can be looked upon as a crude processing of the speech signal in order to match high frequency speech sounds to the low frequency characteristics of the skin.

The Sentiphone (Traunmyller 1977) maps the center of gravity of the speech spectrum of the speech signal on the frequency range 30–350 Hz. The stimulating signal is sinusoidal. The intensity of the vibration corresponds to that of the speech signal compressed by a factor of two in order to match the dynamic range of the speech to the range of the skin. The vibrator is mounted in a cylinder to be held in the hand and the vibration is tangential. The Sentiphone is commercially available.

Single vibrator system stimulating other bodyloci

SRA-10 (Scott) is a further development of the hybrid vibro- and electrotactile aid stimulating the hand which was described by Scott and de Filippo (1977). In order to make the device more convenient to use, the electrotactile part was made vibrotactile and the place of stimulation was changed from the hand to the abdomen.

The system used in this test was a laboratory prototype of the commercially available aid from Scott Instruments. It consists of three vibrators but could still be considered a single vibrator system because the place of sensation is not moving as in a conventional tactile vocoder. The two outer vibrators are connected in parallel, and therefore they do not move the center of the sensation, but make it more diffuse when activated. High frequency speech sounds like (s) activate the middle vibrator in a nonperiodic manner. Middle range speech frequencies activate all three vibrators in the same manner resulting in a more spread
percept. Low frequency speech sounds (voiced speech) activates
all the vibrators with a square wave signal which varies in
frequency between 20 and 200 Hz. The frequency is controlled
mainly by the first formant in such a way that a high formant
frequency results in a low vibration frequency and a low formant
frequency results in a high frequency of vibration. The intensity
of the vibration is controlled by the sound level.

In the test denoted Schulte (c) the high frequencies were
overloaded as in Schulte (b) but in this case the vibrator was
placed on the wrist.

Stimulator arrays (one spatial dimension)

Two array systems were tested, both using the channel voco-
der concept (Dudley 1939), one electrotactile system with 32
electrodes stimulating the abdomen (Saunders, 1976) and one vib-
rotactile with 24 vibrators stimulating the thighs (Engelmann,
1975), Fig. III-A-1. The two systems are very similar function-
ally, even if the technical solutions are different. Each stimula-
tor is activated by a constant frequency which is intensity
modulated by the rectified and low-pass filtered signal from a
corresponding band-pass filter covering a certain part of the
speech spectrum. In this way the spectrum is mapped on the stimu-
lator array.

Matrix applied on the finger (two spatial dimensions)

The vibrator matrix is that of an Optacon (originally a
reading aid for the blind) (Bliss 1974), which is converted to
present tactile information about the frequency and intensity of
the first and second formants (see Fig. III-A-4 from Spens 1976).
The frequency was mapped along the 24 vibrator columns but only
16 of the 24 vibrator rows were used. The vibrators were acti-
vated in an on/off mode, and the intensity was controlled by
activating more or less neighbouring vibrators above and below
the one corresponding to the actual formant frequency. The six
columns of the matrix corresponded to consecutive samples of the
speech signal, each 12 ms long. The total length of the time
window, therefore, was 72 ms.
Matrix system stimulating the abdomen (two spatial dimensions)

This system, MESA (Multipoint Electrotactile Speech Aid, Sparks, 1978) is electrotactile and 36 columns are used to map the frequency scale. In each column there are eight electrodes which are used to convey the intensity of its corresponding frequency band.

More specific data about the tactile systems are available from the references.

TEST MATERIAL

The vocabulary of the test material was inherited from experiments with automatic speech recognition performed at the Dept. of Speech Communication. It consisted of the Swedish numerals: "1" [et:], "2" [tvo:a], "3" [tre:a], "4" [fry:ra], "5" [fem:a] "6" [sek:sa], "7" [ju:], "8" [ot:a], "9" [ni:a], "0" [nol:a].

Typical spectrograms from the stimuli are shown in Fig. III-A-2. Eight of the ten stimuli are two syllable words and hence of about equal length. To make a high correct score the subject had to make use of not only temporal information but also spectral information, see discussion.

Thirteen test and training lists were prepared. Each one consisted of fifty randomized stimuli, five of each type. All the stimuli were independently recorded from one male speaker, MB. This meant that stimuli had inherent small variations in timing, level, pronunciation etc.

The stimuli were read with an interstimulus interval of 5 seconds and recorded on channel 1 of a two-channel tape recorder. On channel 2, the same stimuli were recorded with a 3-second delay compared to channel 1.

SUBJECTS

For practical reasons there was only one subject who tested all the different systems (the author). The subject was unfamiliar with all the different types of tactile systems, except for the Optacon (Spens, 1976). From that system he had experience from a test with a similar vocabulary performed two years earlier. In order to find out if the subject's earlier experience could have any effect on the results with the Optacon system, another subject (J) was tested. He was a blind well-trained Optacon reader, which means he was familiar with the Optacon but had no experience with this kind of experiment.
Fig. III-A-2. Typical spectrograms from stimuli, consisting of the Swedish numerals from 0–9.
Results from training with different speech communication aids using the same stimuli and subject. Each point corresponds to the average result from 250 stimuli. Standard deviation ± 5%.

<table>
<thead>
<tr>
<th>Originator</th>
<th>Schulte (a)</th>
<th>Schulte (b)</th>
<th>Spans</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIG 111-A-3</td>
<td>T1</td>
<td>T8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>T9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>T10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**STIMULATION OF FINGERS**

**STIMULATION OF OTHER BODY LOCUS**

**STIMULATOR ARRAY**

**STIMULATOR MATRIX**
TEST PROCEDURE

The experiments were done during the author's visit to the various laboratories.

The test material was presented to the subject with the aid of the tactile system tested. All systems were used according to the intentions of respective originators. The exceptions were the Schulte (a) and (c) tests in which high frequencies were purposely overloaded in order to create some low-frequency distortion products, see discussion.

The subject immediately wrote down his response. In order to achieve testing and training simultaneously, the correct response was presented auditorily 3 seconds later at a level well above the masking level of the noise. Noise masking was used in order to prevent the stimulators from giving any acoustic cues. Only the tactile information convey ing capacity of the aids was tested, which meant that no visual information was presented.

RESULTS

The learning curves obtained are shown in Fig. III-A-3. In the figure the systems are organized after place of stimulation and degree of speech "processing". This figure also indicates the test order. Implications of the test order will be discussed later. Fig. III-A-6 shows the confusion matrixes from the 500 last stimulations of each test. In Fig. III-A-5 a comparison between the two subjects, the author and subject (J), tested with the Optacon system (Spens, 1977) is shown.

DISCUSSION

The intention of this investigation was not to make an absolute evaluation of the tactile aids tested, but rather to make a relative comparison of the tactile information-conveying capacity of the aids. There is no doubt that all the tested systems convey tactile information on a phonetic level about the spoken words. The range of the results at the end of the training period is between 50% and 75% correct responses. The absolute magnitude of the results are of minor importance, but it indicates that the vocabulary was not too easy to interpret tactually and hence, there are no end effects affecting the more interesting relative' results.

In order to make an absolute evaluation another method should
Fig. III-A-4. An example of the spectrogram, the extracted frequency and intensity parameters and vibrator patterns on the Optacon vibrator matrix displayed from the word "fema" [fema].
Fig. III-A-5. Results from two subjects training with the Optacon system (Spens, 1976). With different experience with tactile speech stimuli, both have reached the same asymptotic level of performance.
be used, and the most promising method currently used is the tracking technique described by de Filippo and Scott (1978), which employs connected speech.

The rank order of the results obtained gives some interesting aspects on the different families of aids tested, which is illustrated in Fig. III-A-3.

**Place of stimulation**

With tactile systems with about the same degree of processing it seems very obvious that stimulating the fingers gives better performance than stimulating other parts of the body. After all, it is not very surprising that employing more receptors, which is the result of stimulating the fingers, even if the stimulated area is smaller, gives a better result. It seems also very likely that this is possible to generalize to connected speech situations.

Stimulating of the fingers, however, is done at the cost of a hand which then cannot be used for other purposes. Therefore, to trade off some efficiency for a more convenient place of stimulation, which has been done by Scott, might be worthwhile.

**Processing**

It is also clearly indicated that the more "processing" of the speech signal up to a certain level the better the results, Fig III-A-3. The concept "processing" includes in this case not only the electronic processing of the speech signal in order to match it to the skin characteristics but also the employment of several stimulators.

Overloading the high frequencies in test Schulte (a) and (c) can be regarded as a crude processing in order to make high frequency speech sounds more adapted to the low frequency characteristics of the skin. A comparison of the results, Schulte (a) and (b) in Fig. III-A-3 and Fig. III-A-6, reveals that the overloading really improved discrimination between stimuli 8 and 6.

Increasing the complexity of the display system beyond that of an array does not increase the performance significantly in this test. It may be the case in a connected speech situation but the literature does not supply any clear answers to that question (Sparks, 1979; Spens, 1977 and 1976; Ifukube, 1978).

The potential capacity of a matrix display may not be possible to benefit from unless one is prepared to take the trouble of a considerably extended training period compared to what usually
Fig. III-A-6. Confusion matrices from the last 500 stimuli from each of the tactile speech communication aids tested.
has been used for evaluation of tactile aids.

In addition to this, it may well be that more optimal strategies, both in extracting the information from the speech signal and in presenting the information to the skin, must be found before it is even possible to benefit from the potentially higher capacity of a matrix display.

Less information conveyed, but information especially matched to support lipreading of connected speech, may be the optimal way to go. This could be the case with Scott's and Traunmyller's aids which are especially developed to support a lipreader.

The results from all the tested aids, however, do nicely fit into a rank order indicating that the more processing the better.

**Transfer**

For practical reasons only, one subject (the author) was used to test all the aids. The subject had experience with these kinds of tests performed with the Optacon two years earlier. With all the other aids he had no previous experience, however. There might be some transfer because of earlier experience with the Optacon. Some transfer between the different tests is also possible. The author is, however, of the opinion that the final results from the tests neither were affected by the author's previous experience, nor by transfer between the tests.

In order to find out if the subject's earlier experience of similar vocabularies could have any effect on the results with the Optacon system, another subject (J) was tested. He was a blind well trained Optacon reader, which means that he was familiar with the Optacon but he had no previous experience with tactile speech stimuli. The results are shown in Fig. III-A-5, and they support the author's subjective impression that earlier experience of a similar vocabulary presented by the Optacon had very little effect on the asymptotic level of performance. Nor did it have any significant effect on the number of stimuli required to reach the asymptotic level. About 1000-1500 stimuli seem to be sufficient.

Due to circumstances, the training with the Sauner's system was interrupted by the training with Engelmann's system, Fig. III-A-3. The prolonged training with Sauner's system supports the author's opinion that transfer between the tests does affect the initial results, but not the asymptotic level of performance.
Vibrotactile versus electrotactile stimulation

The author's subjective impression of electrotactile stimulation with appropriate current levels is that it resembles vibrotactile stimulation very much. The strategies used for extracting data from the acoustic signal and also for presenting the information to the skin are very similar in the two systems. Therefore, it is not surprising to find the results from the systems nicely fit into one smooth learning curve.

In Fig. III-A-6, which shows confusion matrices from the last 500 stimuli of each test, it is indicated that there are three main types of information or spectral changes with different efficiency in being conveyed. They are:

1. Spectral changes which involve large intensity changes as in the initial and the final part of a word, and also in stops.
   Most of the conveyed information is of this type. All confusion matrices exhibit more or less clearly three submatrices containing stimuli with similar intensity patterns, i.e. (6 and 8), (1) and (2, 3, 4, 5, 7, 9 and 0).

2. Large spectral changes with minor intensity changes such as changes from fricatives to vowels.
   In order to discriminate 6 and 8, the high frequencies in the (s) must be taken care of by the processing. (Compare Schulte (a) and Schulte (b).) If so, this kind of information is possible to convey tactually with some variations in effectiveness in the different aids, Fig. III-A-6.

3. Small spectral changes like those from vowels to vowels or vowel like consonants.
   This kind of information seems to be the most difficult to convey. There are differences in performance between different aids which indicate that both the place of stimulation and the processing are important factors. The level of performance seems, however, to be rather low for all systems, Fig. III-A-6. As discussed above, an extended period of training may have some importance.

Categories one and two correspond roughly to the ability of the aids to convey information, about the features sound/silence and voiced/voiceless, respectively. Discrimination of large tem-
poral and frequency differences seem to be perceivable almost instantaneously and the subject's ability to do this is reflected by the initial level of the learning curves. The actual learning is mostly attributed to the information carried by small spectral and temporal changes.

Some conclusions

Important factors affecting the information conveying capacity of tactile systems are:

1. the choice of stimulation area, the fingers being the best.
2. the amount of processing of the speech signal, including the number of spatial dimensions in the tactile display.
3. the information carried by large spectral changes being easier to learn.

There is no significant difference in performance which can be attributed to vibrotactile or electrotactile stimulation.

In order to make training possible during a long period time, the aids should be made wearable.

Acknowledgements

This work was supported by grant no STU-79-5148a, from the Swedish Board for Technical Development.

The author wishes to express his gratitude to the Foundation "Tysta Skolan", which made the travelling economically possible, and also to all persons at the Central Institute for the Deaf, Smithkettlewell Institute for Visual Sciences, Oregon Research Institute and Dept. of Speech and Hearing Sciences, University of Washington, who kindly and willingly collaborated in the preparation for the training and experimenting.
References


Dudley, (1939), "The Vocoder", Bell Laboratories Record 17, 122-126.


Scott, B. L.: "Vibrotactile speechreading aid SRA-10", sold by Scott Instruments, 815 N. Elm, Denton, Texas 76201.


Sparks, D. W. et al. (1979): Investigating the MESA (Multipoint


Spens, K. E., (1977): "Is there an optimal time window for tactually conveyed spectral patterns derived from the speech signal?" paper presented at the Research Conf. on Speech-Processing Aids for the Deaf, Gallaudet College, May.