Speech coding in aids for the deaf: An overview of research from 1924 to 1982

Risberg, A.

journal: STL-QPSR
volume: 23
number: 4
year: 1982
pages: 065-098

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

A. SPEECH CODING IN AIDS FOR THE DEAF: AN OVERVIEW OF RESEARCH FROM 1924 TO 1982*

Arne Risberg

Abstract

The aim of speech coding in aids for the deaf is to adapt the acoustic speech code to the sensory capabilities of a limited residual hearing, a direct stimulation of the auditory nerve, or to a transmission through the visual or tactile sensory system. In the 1920’s, Gault (1926) started his work on transmission of speech through the tactile sense. In 1925 Perwitzschky suggested that the speech perception ability of a hearing impaired person with hearing only in low frequencies could be improved by frequency lowering. During the sixty years from 1922 to 1982, many attempts have been made to develop speech coding aids for the deaf. A selection of the more important articles published up to 1979 has been compiled by Levitt, Pickett, & Houde (1980). In spite of great efforts, very few aids based on speech recording techniques are today used by deaf persons. The only exception seems to be some visual aids used for speech training, a few persons that use tactile aids, and 200 to 300 persons that use cochlear implants of different types. There are several reasons for this lack of success. In this article I have tried to give an overview of this area of research and development. Work on amplitude compression systems has not been included.

Is the auditory system special?

In a paper presented at the Washington Conference on Speech-Analyzing Aids for the Deaf in 1967, Liberman & al (1968) suggested that the explanation for the difficulties experienced in learning to read visual transforms of the acoustic speech signal, the sonagram, was that speech is a special code and that a special decoder in the brain of the listener is needed. This decoder only works for the auditory input. A tactually or visually presented speech signal does not have access to this decoder and this cannot be changed by training.

The support that Liberman & al presented for their hypotheses was the limited success achieved by people that have tried to learn to read sonagrams, especially the results of the experiment made by Potter, Kopp, & Green at Bell Telephone Laboratories (1947). In a recent paper, Cole & al (1980) question the hypotheses of Liberman & al. In spite of the rather primitive technical device used by Potter & al, a 12-channel spectrum analyzer, that presented the energy in the frequency range from 300–3000 Hz on a moving phosphorous belt, the most motivated subject could read a vocabulary of 800 words after about 200 hours of training. Compared to the amount of training given a child in learning

to decode the acoustic signal, this result seemed to disapprove the hypotheses of Liberman & al rather than to prove it. The success of the subject V2 in reading sonagrams is also discussed in the article by Cole & al (1980). His results also seem to show that it really is possible to learn to read visible speech if enough training is given.

Other indications that nonauditory transformations of speech can be decoded after training is the use of lipreading by the deaf and the use of the tactile Tadoma-method by the deaf-blind. As an example of speech perception ability obtained by means of a nonauditory code, the results obtained by Nicholls (1979) in an evaluation of "cued speech" can be given. "Cued speech" is a visual code, developed in 1967 by Cornett, consisting of hand shapes and hand movements made in synchrony with the articulatory movements. The code is intended to be used as a supplement to lipreading but in the evaluation study, speech perception ability was also measured when only cues were presented. The subjects were a group of 18 profoundly deaf children that had used cued speech for four years or more. The test material consisted of sentences and the subjects’ ability to identify a keyword in each sentence was measured. In the auditory situation only 1.6% of the keywords could be identified, with lipreading 29.3% and with only cues 46.8%.

Kiron (1974) has reviewed the work on tactile coding of speech and he also discusses the possibility of the existence of a decoder that only works for an auditory speech signal. He came to the conclusion that the most likely reason for the lack of success in the earlier studies of tactual speech transmission was the selected coding technique.

Auditory substitute or lipreading aid?

The auditory sense is used for several different purposes:
1. To monitor activities in the surrounding
2. To perceive different warning signals
3. To perceive speech from others
4. To control our own speech

There is no principle difference between the acoustic signals in these situations. The exception might be some signals with very low frequencies or very high frequencies, that are important in the first two situations but are less important in speech. The aim of a coding system is to make it possible for the deaf subject to perceive the important signals by means of a new sensory modality or in another frequency range than that of the original signal. Speech is the most complex signal and it is, therefore, likely that a recoding technique
that works well for speech also will work reasonably well for other signals. As the result of the combined efforts in many laboratories our knowledge of the acoustic speech signal, the relation between articulation and acoustic output, and the importance of different elements in the acoustic signal for speech perception is good (Pickett, 1980).

The acoustic speech signal can be seen as a combined effect of a source and a filter. The source is either the puffs of air generated by the periodic modulation of the airstream from the lungs by the vocal folds or the friction or transient energy generated by a constriction in the mouth cavities. The filter is the tube from the glottis to the lips plus the nasal cavities. The properties of the filter are changed by movements of the tongue, lips, velum, etc. The acoustic signal that is the result of the different variations in the sources and the filter can be studied by means of the sonagraph. In Fig. 1 the sonagram of the words "speech coding" spoken by a male speaker is seen. The signal in the sonagram has been divided into a number of acoustically different segments. These are described with respect to the source features: voice, noise, transient, and silence. The variation in the fundamental frequency is shown on a separate curve.

Speech perception can, in the same way, be seen as a parallel extraction of information of source and filter features in the acoustic signal. In a technical aid that recodes the speech signal to another modality with other sensory capability than the auditory sense, a code must be selected that transmits as much of the information in the speech signal as possible. A good time resolution is required to transmit source features. The presence of time segments with durations in the order of 10 ms must be perceived. To transmit filter features it requires a good resolution in the frequency domain.

Much of the development work on speech recoding aids for deaf has from the beginning been defined as the development of an auditory substitute, that is, an aid that could replace the lost hearing in all situations. Gradually, it has been realized that the sensory capability of the tactile or visual sense, or the auditory sense in the case of a severe hearing loss, is too limited to enable the perception of speech without support of lipreading. The technical aids have then been defined as lipreading aids and a code has been selected that gives good support during lipreading. In the development of some auditory recoding aids, and in most work on cochlear implants, the goal is still often defined as the development of an auditory substitute. There are reasons to
Fig. 1. Sonoagram of the word "speech coding." In the bottom part of the figure, the fundamental frequency variations and source features are shown.
question this approach. Even if it might be possible to achieve some open speech perception without the support of lipreading with these aids, the noise in most communication situations will make lipreading necessary. It is, therefore, likely that the development work will result in the selection of a better code if the aid primarily is seen as a lipreading aid.

The limitation of lipreading

Many studies have been made on different aspects of speech perception by lipreading (Farwell, 1976). In Table 1 the results are shown of the classical study made by Woodward & Barber (1960) of the visual discriminability of phonemes. The lipreader can discriminate between phonemes belonging to the different groups but cannot discriminate between phonemes within the same group.

From Table 1 it is apparent that the lipreader can identify gross differences in the place of articulation but has difficulties with the manner of articulation. A statistical study of spoken English, made by Denes (1963), shows that the information load in speech is much higher on the manner of articulation than on the place of articulation. The fourth lipreading group, dental, alveolar, and velar consonant phonemes, covers about 75% of all consonants but the manner of articulation is about equally distributed. Vowels are somewhat easier to perceive during lipreading than consonants. Studies by Jackson & al (1976) show that vowel features, such as extended-rounded, lip separation, and overall area, are used by the lipreader. He often cannot see the position of the tongue.

The ability to identify suprasegmental (prosodic) features, such as rhythm, stress-pattern, and intonation during lipreading, has been studied in a few experiments only. In Table 2 some results of an experiment made in our laboratory are shown (Risberg & Lubker, 1978). The stimuli were word pairs with difference between the words in the pair with respect to only one prosodic feature. In the table results are shown from a group of normal hearing subjects (N=7), and a group of hard of hearing subjects (N=12) for the contrasts vowel length, stress-pattern in two-syllable words, juncture (e.g., the difference between "I scream" and "ice cream"), and word emphasis in four-word sentences. Obtained mean value and total range are shown. As can be seen, both the normal hearing group and the hard of hearing group had difficulties in identifying the prosodic contrasts vowel length, juncture, and emphasis.
Table 1. Visually contrastive phoneme groups (Woodward & Barber, 1960).

1. pbm
2. wr
3. fv
4. tdn1 ò szt jdj3 jk gh

Table 2. Results from a study of the ability to identify prosodic contrasts during lipreading (Risberg & Lubker, 1978).

<table>
<thead>
<tr>
<th>Test material</th>
<th>Guessing %</th>
<th>Range %</th>
<th>Mean %</th>
<th>SD %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel length</td>
<td>50</td>
<td>65-80</td>
<td>72.0</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55-80</td>
<td>70.4</td>
<td>7.8</td>
</tr>
<tr>
<td>Syllable stress</td>
<td>50</td>
<td>75-100</td>
<td>85.7</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80-100</td>
<td>90.8</td>
<td>5.9</td>
</tr>
<tr>
<td>Juncture</td>
<td>50</td>
<td>60-75</td>
<td>67.9</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45-85</td>
<td>65.0</td>
<td>14.8</td>
</tr>
<tr>
<td>Emphasis</td>
<td>50</td>
<td>6.3-75.0</td>
<td>42.8</td>
<td>22.0</td>
</tr>
</tbody>
</table>
Stress-pattern in two-syllable words could in most cases be identified.

The selection of code in a lipreading aid

In describing the sonagram in Fig. 1, it was said that the acoustic speech signal could be seen as the combined effect of a source and a filter. The same model can be used in speech perception. The perception is then based on the parallel processing of source and filter information. This model is very suitable to the work on lipreading aids. During lipreading, the filter information, that is, the place of articulation of different vowel and consonant phonemes, can with reasonable success be seen, whereas source features are less visible. A lipreading aid should then primarily be designed to transmit information on source features, voice, noise, transient, and silence, plus changes in the fundamental frequency. These features will also relatively well describe the manner of articulation of consonants. An exception might be nasalization and lateralization.

The result of an experiment that supports this is shown in Fig. 2 (Risberg & Lubker, 1968). In this study, older hearing impaired persons were tested in three situations: only listening to lowpass filtered speech with a cutoff frequency of 180 Hz, only lipreading, and lipreading plus listening to lowpass filtered speech. The test material was unknown sentences read by a female speaker. The great improvement in the audio-visual situation supports the above conclusion. During this severe filtering, mainly source features and prosodic features could be perceived, as shown by Miller & Nicely (1955). Some of the good results obtained by the subjects in this experiment were based on the use of the intonation pattern (fundamental frequency variations) in the sentences. That this information is vital to the lipreader has been shown by the group at The University College in London, that works on extracochlear implants (Rosen & al, 1981; Fourcin & al, 1982). In the experiment the task of the lipreader was to track word by word what a speaker read from a short story in two situations, only lipreading and lipreading supported by acoustically transmitted information of fundamental frequency. The number of correctly reported words per minute was measured. With lipreading only, the subjects could track with a speed of 10 to 20 words per minute. When lipreading was supplemented by information of fundamental frequency variations they tracked with a speed of 25 to 40 words per minute. The tracking procedure has been developed at the Central Institute for the Deaf in USA (De Filippo & Scott, 1979). The method gives a
Fig. 2. Older hearing impaired persons' ability to identify keywords in sentences in three situations: listening to lowpass filtered speech (cutoff frequency 180 Hz), lipreading, listening to lowpass filtered speech plus lipreading (Risberg & Lubker, 1978).
good indication of the effectivity of a communication aid. The results shown in Fig. 2 indicate that there is good hope that speech recording aids can be designed that give substantial help during lipreading. The amount of acoustic speech information transmitted through the lowpass filter in the first experiment is very limited. It ought to be possible to transmit this information in coded form through another sense modality, a small residual hearing, or by a direct stimulation of the auditory nerve.

Recoding aids for the deaf. Some examples

It is beyond the scope of this overview to give a detailed description of all experiments that have been made with different coding approaches in technical aids for the deaf. In the following, some examples are selected that illustrate the problems encountered with the different coding strategies and the different sense modalities used. In the discussion on auditory recording, only frequency lowering systems are included. A review of work both on amplitude compression and frequency lowering has been published by Braida & al (1979).

Auditory recording

Work on auditory recording to help the deaf was started by Perwitschky, who in 1925 suggested that frequency lowering could improve hearing impaired subjects' ability to perceive speech. This suggestion was based on the knowledge of the relative importance of speech information in the frequency spectrum (Fletcher, 1953) and the observation that many hearing impaired persons have better hearing for low frequencies than for high. In the 1920's, frequency lowering systems could not be technically realized. During the last 30 years, however, many different experiments have been made with frequency lowering. The tested systems can roughly be grouped into three categories: linear frequency lowering, nonlinear frequency lowering, and selective frequency lowering, see Fig. 3. A review of published experiments has been compiled by Braida & al (1979).

In linear frequency lowering, the whole frequency range of speech is lowered, or compressed, to fit into the frequency range of the residual hearing. Lowering ratios of up to four have been used. This has been achieved either by means of tape recorders with a rotating head (Oeken, 1963), channel vocoders (Denes, 1967; Pimonow, 1962), or systems based on the FFT-algorithm (Block & Boerger, 1980; Hicks & al, 1981).
Fig. 3. Different technical systems for frequency lowering.
Linear frequency lowering might result in increased masking of the high frequencies by the low frequencies. To overcome this masking, nonlinear frequency lowering systems have been tried (Block & Boerger, 1980; Hicks, 1981). In these experiments, a technique based on the FFT-transform was also used. In selective frequency lowering, only high frequency fricative sounds are changed to low frequency sounds (Risberg, 1969). Different technical solutions have been used, e.g., a channel vocoder technique (Ling & Druz, 1967), modulation with a carrier around 4000 Hz Johansson, 1959; Velmans, 1972), and shift registers with read-out at lower speed than read-in (Knorr, 1976). In some systems a voiced/unvoiced detector is included to avoid lowering of voiced sounds as this might introduce undesired distortions (Risberg, 1969; Knorr, 1976). One hearing aid with frequency lowering has been commercially manufactured by Oticon (Oticon TP72) based on the lowering system developed by Johansson (1959). In an evaluation study some positive results were obtained (Poust & Gengel, 1973).

The evaluation studies of different frequency lowering systems have in many cases been made with normal hearing subjects. A hearing loss has then been simulated by means of a lowpass filter, usually with a cutoff frequency around 1000 Hz. In all experiments it has been shown that the normal hearing subjects can improve their ability to identify speech by means of the lowered speech signal. When experiments have been made with hearing impaired subjects, it has also been shown that they, with training, can improve their ability to understand speech over the system, but when an equal amount of training has been given with the original speech signal no difference has, as a rule, been obtained between direct transmission and lowering (Oeken, 1963; Block & Boerger, 1980). In the experiment by Block & Boerger (1980), however, better results were obtained on CV-syllables with lowering but about the same results on words and sentences.

The results of evaluation experiments with frequency lowering have been disappointing. Theoretically, it seems reasonable to expect that some improvement in speech perception ability could be obtained with this technique. There are probably several reasons for this lack of success. In some systems inappropriate coding techniques that have introduced disturbances have been used. The training time has in all studies been too short. In several studies it has also been shown that the frequency selectivity or frequency discrimination ability has deteriorated in sensorineural hearing loss (Zwicker & Schorn, 1978; Gengel,
1973). It is, therefore, possible that subjects that could not perceive the differences between the lowered speech sounds have been used. This is a very likely explanation when severely hard of hearing subjects have been used (Ling & Druz, 1967; Foust & Gengel, 1973). Even with a moderate hearing loss, frequency discrimination ability can be affected. Fig. 4 shows the audiogram and frequency discrimination ability of a subject with a moderate loss. In spite of a better hearing in the right ear, frequency discrimination ability is poor in this ear but almost normal in the left ear. The speech signal is severely distorted in this subject's right ear.

One of the problems with many of the lowering systems tested has been that they only existed as laboratory equipment or as a computer program. The amount of training that could be given was, therefore, limited. It is very likely that very long training times are needed before full use can be made of a frequency lowering signal. The results from the experiment by Block & Boerger (1980) showed that better results were obtained on syllables with the lowering system than with the direct signal. This might indicate that syllable identification could be improved after the short training period but that word and sentence intelligibility does not improve so quickly.

A possible strategy for future work on frequency lowering might be to concentrate the work on systems that can easily be made wearable. More work is needed on the selection of suitable subjects and in all experiments some measurements of the auditory capacity of the subjects must be included. Fig. 5 shows the audiograms of five subjects with a sensorineural hearing loss. In the bottom part the results obtained by these subjects on tests with monosyllabic phonetically balanced word lists (PB) and on identifying keywords in sentences (S) are shown. In both cases the speech material was presented in a silent room. Subject A had no difficulties to perceive speech in this situation. This is a typical audiogram for an older patient with an acquired hearing loss, often diagnosed as presbyacusis. Typically for these subjects is that they complain about their difficulties to perceive speech in the presence of noise but have no difficulties in silence. A recoding system for these subjects must then be designed for this situation. The high age of these subjects might mean that they have difficulties in learning to use a new acoustic code. Subject B has a more flat audiogram. In this case the dynamic range is also limited due to recruitment. This might, therefore, be a case where amplitude compression can be of some help.
Fig. 4. Puretone audiogram and result of frequency discrimination measurements on a subject with large differences in frequency discrimination ability in the two ears.

Fig. 5. Puretone audiograms and result obtained on speech perception tests with five hearing impaired persons. PB = results on a list with monosyllabic words, S = results on a test with sentences.
Subject C is a case of low frequency hearing with very low results on the speech tests. This is the type of subject that needs help from a frequency lowering system. An ordinary hearing aid is in these cases of very little help. Speech perception is only possible with simultaneous lipreading and even in this situation speech perception is often difficult. Fig. 6 shows results that, at least theoretically, indicate, that selective frequency lowering can be of help in these cases (Risberg, 1965). The experiment was made with the fricative lowering system of Johansson (1959). Normal hearing persons were used and hearing impairment was simulated by means of lowpass filtering. The test material consisted of CV-syllables and monosyllabic words. With frequency lowering, an increase in the per cent correct identified words was obtained. This was coupled with an increase in per cent correctly perceived manner of consonant articulation. As subjects with the kind of hearing loss that was simulated must rely on lipreading, this increased the ability to identify the manner of consonant articulation is important. A technical aid with selective frequency lowering might then be a good lipreading aid. A small study that also seems to show this has been made by Spens & Martony (1972).

Subject D has poor hearing in the whole frequency range. In this case, frequency discrimination ability was also shown to be very abnormal. It is, therefore, not likely that this subject can be helped by means of a frequency lowering system. A system that improves the ability to perceive some source features and fundamental frequency variations might be used. Audiogram E follows closely the threshold of vibration in the ear (Norber, 1967). This is an audiogram from a congenitally totally deaf child. Auditory stimulation in the ear can in this case probably not give more than tactile stimulation on a finger, for example.

Recoding to a visual signal

As an information transmission channel in an aid for speech perception, the visual channel has the limitation that it makes simultaneous lipreading difficult. In 1968, Upton showed a solution to this problem. Speech information was presented on the lipreader's eyeglasses by means of a number of small lamps. In a later design this system was changed to a system based on light-emitting diodes that projected signals on a mirror on the eyeglasses. The lipreader then saw the speaker's face surrounded by a changing light pattern.
Fig. 6. Per cent correct identified monosyllabic words and per cent correct identified manner and place of consonant articulation. Comparison between results obtained with lowpass filtering only and with frequency lowering plus lowpass filter (Risberg, 1965).
Based on the studies made by, among others, Woodward & Barber (1960), Upton chose to transmit information on the presence of friction, voiced, and unvoiced plosive sounds, and some vowel information on the display. These features were extracted from the acoustic signal by a small wearable analyzer. Evaluation studies made by Upton as subject showed that he could get an increase in the number of identified keywords in sentences by 10 to 20% (Gengel, 1976). The device was later tried on a severely hard of hearing girl (Gengel, 1976). Based on her demands, the device was rebuilt to also show speech rhythm on one of the light-emitting diodes. This information Upton could get through his hearing aid. When the girl had used the device for six months, evaluation studies were made that showed an increase by about 20% on sentence material when the aid was used together with lipreading. She also got help from the device to control her own voice and to be aware of different environment sounds.

In spite of the rather promising results with Upton’s lipreading aid, the experiments have apparently not been continued. The way speech information is transmitted to a deaf person in the Upton aid is convenient from a wearing point of view. More experiments with this type of presentation ought, therefore, to be made. As Upton himself had good residual hearing, the experiments were concentrated on the presentation of the manner of consonant articulation. If the aid is to be used by totally deaf persons, the experiments ought to be concentrated on the presentation of prosodic features and the presence of friction.

The visual sense is ideal for use in speech training devices. Many equipments of this type have been built both systems that only show a single speech element, such as nasalization or fundamental frequency (Risberg, 1969), and those that show a frequency spectrum or a time-amplitude-frequency display (sonagram) (Maki & al, 1981). In recent years, microprocessors have been used in systems of this type (King & al, 1982). This makes it possible to select different speech elements for display, vary the display pattern, introduce plays to increase motivation, etc. The research and development problems around visual aids for speech training are not only technically but also pedagogically. Technically it is possible to design a wide variety of aids for speech training. To be of any help, training methods must be developed to suit different speech errors, different age groups, different degrees of hearing loss, etc. This has not yet been done.
Tactile aids and cochlear implants

Tactile aids and cochlear implants can for two reasons be grouped together. The first is that in many cases the speech coding problem has been approached in very much the same way. The other reason is that it often has been questioned if a single channel cochlear implant can give more speech information than what can be obtained by means of a tactile aid.

Coding strategies

The coding strategies used in many tactile aids and cochlear implants can roughly be grouped into three categories: direct transmission, spectral coding, and feature coding. Sometimes approaches have been used that can be described as a combination of these categories.

In the first approach, direct transmission, the acoustic signal is directly transmitted to the selected sensory modality by means of a vibrator or an electrode inserted into the cochlea or placed outside the cochlea. Sometimes the speech signal is adapted to the limited dynamic properties or limited frequency range of the modality. In the second type of code, spectral coding, the main part of the information in speech is assumed to be transmitted by the time-varying distribution of energy in the frequency domain. This approach results in a technical system consisting of a number of filter channels. The energy in the different channels is transmitted to the subject by means of a suitable code. Based on experiments with channel vocoders it seems that six to ten channels are a minimum number required. Fig. 7 shows the results of experiments made by Hill & al (1968) and Zollner (1979). In these experiments a channel vocoder was used where the synthesis filters were replaced by sinusoids with the same frequency as the center frequency of the synthesis channels. The figure shows that with only six channels about 70% of the vowels and the consonants are correctly identified and 90% of the monosyllabic words. These results have been interpreted to indicate that in a speech coding aid a maximum of ten channels are needed to get a good ability to perceive speech over the system. No transmission of the source features, voice and noise, or changes in fundamental frequency have been considered to be necessary. The results shown in Fig. 7 have, however, been obtained with normal hearing listeners. This means that a sensory system with a good time and frequency resolution was used. It is not likely that this good resolution can be obtained when, e.g., the tactile sense is used or in direct stimulation of the auditory nerve.
Fig. 7. Speech perception obtained in a channel vocoder with different manner of channels. As synthesis signals, sinusoids with the same center frequency as the synthesis channel were used (Hill & al, 1968; Zollner, 1979).

Fig. 8. Tactile speech-coding based on a channel vocoder.
In the future, speech perception coding systems are seen as the recognition of a number of acoustic features such as: periodic, noise, fundamental frequency variations, the frequency of the formants, formant transitions, time relation between different acoustic elements, etc. Some of these features are more important than others. Some of them can be identified during lipreading and some cannot. Based on an opinion of the relative importance of the features, a selection is made and the selected features are transmitted in a suitable code. In some systems different phonetic classes, such as voiced plosive, unvoiced fricative, nasalized, etc. are automatically recognized and coded. This automatic recognition increases the possibility to select a suitable code.

Tactile aids

Gault started his work on tactile recoding in the 1920’s (1926). From the beginning he used single channel systems but he gradually realized that the small usable frequency range, only up to 500 Hz, and the poor frequency discrimination ability of the tactile sense made it necessary to use a system where frequency was substituted for place of stimulation. Gault could show that deaf and normal hearing subjects could learn to identify some words with single channel tactile aids and that they could get help during lipreading (1926).

During the 1950’s the interest in tactile recoding increased as a result of the invention of the channel vocoder. Tactile systems were built of the basic type, shown in Fig. 8, with the number of channels ranging from 10 to 24. Evaluation studies showed that subjects could learn to detect syllabic structures and gross spectral properties and that the systems gave support during lipreading (Pickett, 1963). The effect was, however, small. Liberman & al (1968), in the paper quoted earlier, suggested that this was due to the existence of a special decoder in the brain of the listener and that this decoder only worked for auditory signals. Kirman (1974) made a review of published studies on tactile speech communication and came to the conclusion that a more probable explanation was that the tactile code used was less suitable to the transmission of speech signals. He suggested that a matrix of vibrators instead should be used. Based on his suggestion several systems of this type have been tested. In some of them a frequency-intensity pattern is transmitted as in the earlier systems (Sparks & al, 1978) and in some of them a time-frequency-amplitude pattern is transmitted Spens,
1976). This last coding technique results in a vibrating pattern that moves across, for example, a finger. This technique might make it possible to overcome some of the poor time resolutions of the tactile sense (Gescheider, 1970). That this poor time resolution is a real problem in tactile speech transmission has been shown by Keidel (1973). A "tactile cochlea substitute" was arranged that stimulated the forearm of the subject by means of vibrators. When speech was played through the system with the speed reduced by a factor of four the subjects quickly learned to identify some words.

Spens (1980) has compared several existing tactile systems. As speech material the Swedish numerals one to nine were used, and he used himself as the only subject. Fig. 9 shows the placement of the vibrators and the type of system tested. The system of Schulte is a direct transmission system with no processing of the speech signal. The system of Traumnroller and Scott uses some coding of high frequency sounds, the systems of Saunders and Engelmann are traditional filterbank systems but in the system of Sauders electrotactile stimulation is used. The systems of Spens (1976) and of Sparks (1979) are both matrix systems. In Sparks' system a frequency amplitude code is used but in the system of Spens, a time-frequency-amplitude code.

In Fig. 10 the results are shown as learning curves. The dotted line across the figure divides the systems into two groups: systems that stimulate a finger and systems that stimulate other parts of the body. Spens draws the conclusions from his experiment that it is better to stimulate a finger than to stimulate any other part of the body: the more the speech signal is adapted to the limitations of the tactile sense, the better results will be obtained.

The present work on tactile aids seems to be concentrated on aids that use a feature coding. The aids are also directly designed as lipreading aids. Rothenberg & al (1977) have suggested that fundamental frequency could be coded as frequency variations in the range of 20 to 100 Hz. Traumnroller (1980) has coded the gravity of the speech spectrum into frequency variations in the range 30 to 300 Hz. Preliminary results show a good support during speechreading but no studies have been made on deaf persons. At MIT, studies are in progress on the Tadoma-tactile-method used by the deaf-blind (Snyder & al, 1982). Comparison between the Tadoma-method and tactile spectral coding show that the first one is superior (Snyder & al, 1982). An important question is, of course, how the signals used by the Tadoma-reader should be automatically extracted.
Fig. 9. Type and placement of the vibrators in the different tactile speech aids compared by Spens (1980).
Results obtained in the comparison of some tactile speech aids (Spens, 1980).
from the speech signal and how to transmit them in a wearable technical aid.

Another approach in the present studies of tactile aids is to make the aids comfortably wearable. One objection to most of the earlier work on tactile speech transmission is that the systems are so bulky that they can only be used in the laboratory which will result in very short training periods. An alternative approach is then to start from the requirement that the aid must be easily wearable and then try to incorporate as much of the wanted recoding power as possible. This approach has in our laboratory resulted in a very simple aid where an optimum recoding is made only for the time-amplitude information. The electronics is built into a case of a body-worn hearing aid. Preliminary studies with normal hearing subjects show that the aid gives some support during lipreading. Fig. 11 shows the result of an evaluation study. Two groups of normal hearing subjects tried to lipread sentence material. Half of the group started with only lipreading and half started with lipreading plus the tactile aid. After five to six lists they changed the situation. As can be seen, both groups obtained better results when the vibrator is used. The aid has also been tried by several persons with an acquired total deafness. They find it very useful, especially to get information about environment sounds. The tactile signal also makes lipreading less tiring.

Cochlear implants

Work on direct stimulation of the auditory nerve, cochlear implants, has been going on at several laboratories since the end of the 1960's. Most of the work on cochlear implants has been based on the assumption that the only manner in which it could be possible to obtain some ability to perceive speech in such systems was to use a multichannel approach, that copies the frequency-analyzing ability of the peripheral auditory system. Single channel systems were believed only to be able to transmit information about speech rhythm and some information about environmental sounds (Zwicker & Zollner, 1980). The results reported by House & al (1979) and also the results of the evaluation of 13 patients equipped with single channel systems in the study by Bilger (1977) supported this view.

Psychophysical studies during electric stimulation by means of an electrode inserted in the cochlea give pitch sensations that depend on the site of stimulation. This pitch sensation changes, however, with the
frequency of stimulation. The usable dynamic range is very small, often only 6 to 10 dB. Based on these results it seems that a possible code, then, could be to use a multichannel system with constant stimulation frequency and to adapt the amplitude range to the usable range during electric stimulation. This type of coding is used by Chouard (1977). Based on the same studies that were used to design tactile spectrum systems, the necessary number of channels was estimated to be ten (Hill & al, 1968; Zollner, 1979). The selected code very much resembles tactile spectrum coding. In both situations the difference between voiced and unvoiced sounds is not coded efficiently and intonation is not coded at all. Strong energy in one part of the frequency range also masks energy in other parts.

The cochlear implant group in Australia has more directly applied knowledge of the importance of different acoustic speech elements to design their implant (Tong & al, 1981). The selected code can be described as a feature code. The frequency of the midfrequency range (mainly the frequency of the second formant) is coded as the position of stimulation on ten electrodes in the cochlea. Stimulation frequency is proportional to the fundamental frequency and for unvoiced sounds the stimulation frequency is lowered to 70 Hz, which gives a rough percept. From a speech coding point of view this code seem rather promising and the code is also well matched to simultaneous speech-reading. Evaluation studies report that one subject could correctly identify 65% of spondee words in a known set of 16. He could also identify 16% of unknown simple sentences (Tong & al, 1981).

The multichannel systems result in a bulky electronic unit that must be implanted and a bulky, power consuming, external stimulator. Even if electronic circuitry gradually can be miniaturized it is likely that it will be difficult to reduce the size and weight of these systems so that they can be easily wearable. An important question is then if these multichannel systems give much better results than single channel systems. Recently, results have been published that makes it necessary to answer this question (Hochmair-Desoyer & al, 1981; Michelson & Schindler, 1981).

In a single channel approach to electric stimulation of the auditory nerve no attempt is made to simulate the frequency analyzing ability of the peripheral auditory system. Instead the time domain properties of the system are used (Keidel, 1980). Experiments have shown that with electric stimulation in the cochlea subjects can scale pitch up to 400
Fig. 11. Results from evaluation of simple single-vibrator tactile lipreading aids.

Fig. 12. Comparison between frequency difference ability obtained with auditory stimulation (Wier & al., 1977), tactile stimulation (Rothenberg & al, 1977; Goff, 1967), and direct electrical stimulation of the auditory nerve (Bilger, 1977; Fischer, 1981).
Hz and sometimes up to 1000 Hz (Bilger, 1977; Fischer, 1981). With extracochlear stimulation pitch scaling is possible up to 400 Hz (Pourcin & al, 1982). If frequency analyzing ability is expressed as frequency discrimination ability, a direct comparison can be made between measurements on normal hearing, electric stimulation, and tactile stimulation. In Fig. 12 some measurements of this type are shown. The values for tactile stimulation are taken from Goff (1967) (crosses) and from Rothenberg & al (1977) (circles). The values for normal hearing are taken from Wier & al (1977). The values for electric stimulation are taken from Fischer in Vienna (1981) (crosses and circles) and Bilger (1977) (filled marks).

Comparing the values for frequency discrimination ability for tactile stimulation and electric stimulation in the cochlea, given in Fig. 12, it seems that better results ought to be obtained with a single channel cochlear implant than with simple tactile aids. Some frequency discrimination ability seems to be possible below 500 Hz, maybe even up to 1000 Hz. Difference in signal type, periodic or aperiodic, can be detected in the frequency range above 1000 Hz (Bilger, 1977). Studies also show that time-resolution is better with electric stimulation than with tactile (Gescheider, 1970; Fischer, 1981). The very limited dynamic range with electric stimulation, however, requires that a good amplitude coding technique must be used.

The group working on cochlear implants in Vienna has published results from evaluation studies made on four subjects equipped with a single channel stimulator Hockmair-Desoyer, & al, 1981). They report that one subject got 100 % correct on a test with twelve known twosyllabic nouns and that the other three got 50 to 65 % correct on this test. On a test with unknown sentences one subject got about 80 % correct and the three others between 15 and 35 % correct. Results are also reported from confusion tests with vowels and consonants. In Fig. 13 a confusion matrix for the vowel test is shown. The vowels have been ordered with respect to the frequency of the first formant. The results indicate that the subjects have some possibility to extract this information. In Fig. 14 the results for two subjects on the consonant test are shown. The total per cent correct is only about 30. Voiced/unvoiced information is, however, identified 89 % correct, friction-nonfriction 89 %, and transient-nontransient 81 % correct.

The results reported by the group in Vienna with a single channel system are extremely good. Future studies will show if these are results
Fig. 13. Vowel confusions reported by Hochmair-Desoyer & al (1981). The vowels are arranged in groups with about the same first formant frequency.
Fig. 14. Consonant confusions reported in a study of Hochmair-Desoyer & al (1981).

Fig. 15. Results obtained by hearing impaired listeners with hearing aids on a test consisting of 12 known spondees words. In the figure are results obtained on a similar test by patients equipped with a single channel cochlear implant indicated (CK, FW, LP, SS), Hochmair-Desoyer & at (1981) and a multi-channel implant (MCI), Tong & al (1981).
that can be obtained with all subjects or if these four subjects represent an optimum. Similar results have been reported by Michelson & Schindler (1981).

Stimulation outside the cochlea, extracochlear stimulation, is attractive as no irreversible surgical operation has to be made. This technique is studied in several laboratories. Fourcin & al (1982) in London have concentrated their work on the transmission of prosodic features as this will result in an aid that will give good help during speechreading.

Cochlear implants and hearing aids

The very good results that have been reported from experiments with cochlear implants will with certainty result in a rapid increase of experiments in this area. It is then of interest to compare the results that have been reported from these studies with results that are obtained with hearing aids on severely hard of hearing subjects. Fig. 15 shows results from measurements of speech perception ability by means of a test that consisted of twelve known spondee words. The results are plotted against the mean hearing loss for the frequencies 500, 1000, and 2000 Hz.

In the figure results on similar tests reported from cochlear implant studies are shown. CK, FW, LP and SS are from the Vienna group (Hochmair-Desoyer & al, 1981) and MC1 from the Australian group (Tong & al, 1981). As can be seen, the reported results from cochlear implants are comparable to a hearing loss in the order of 90 to 95 dB. The good results from subjects with hearing aids with a mean loss of more than 100 dB are from persons with a combined conductive and sensorineural hearing loss. The results shown in the figure indicate that with a cochlear implant of the type used today, a totally deaf person might be changed into a person with a severe hearing loss.

Some conclusions

The work on speech coding in aids for the deaf has, as was pointed out in the introduction, very rarely resulted in technical aids that are used by the deaf. In this overview some reasons for this lack of success have been discussed. The rapid development around large scale integration of electronic circuitry makes it possible to realize even very complex speech processing algorithms in compact hardware. The problem is, however, to specify what the electronic circuit should do. To solve
this problem basic research is needed around speech processing and the importance of different speech elements for the perception of speech both with and without the support of lipreading. More knowledge is also needed about the capability of the sensory system selected for signal transmission.

The relative success recently achieved in the work on cochlear implants will result in an increased activity in this area. In many cases it is difficult to explain published results from evaluation studies based on our present knowledge about speech perception and the neurophysiology of the auditory system. Basic research is needed around different aspects of speech signal transmission by means of electrical stimulation of the auditory nerve. This can be achieved by means of a closer cooperation between groups working on implants and speech research groups. This coupling can also result in better designed studies for the evaluation of the speech perception ability achieved with different types of implants and different speech coding systems.

References


dimensions underlying vowel lipreading performance", J. Speech Hear.
Res. 19, pp. 796-812.

Johansson, B. (1959): "A new coding amplifier system for the severely

Keidel, W.D. (1973): "The cochlear model in skin stimulation", pp. 27-
32 in Geldard (Ed): Cutaneous Communication Systems and Devices, The
Psychonomic Society.

cochlear prostheses", Audiology 19, pp. 105-127.

computer for use in schools for the deaf", pp. 755-758 in Proc. IEEE
ICASSP 82.

analysis", Psychol. Bull. 80, pp. 54-74; reprinted in Levitt, Pickett, &

Knorr, S.G. (1976): "A hearing aid for subjects with extreme high-


Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M.
(1968): "Why are speech spectograms hard to read?" Am. Ann. Deaf 113,

Ling, D. & Druz, W.S. (1967): "Transposition of high frequency sounds
by partly voicing of the speech spectrum: its use by deaf children", J.
Aud. Res. 7, pp. 133-144.

Maki, J.E., Conklin, J.M., Gustafson, M.S., & Humphrey-Whitehead, B.K.
(1981): "The speech spectrographic display: Interpretation of visual
pattern by hearing impaired adults", J. Speech Hear. Dis. 46, pp. 379-
387.


sions among some English consonants", J. Acoust. Soc. Am. 27, pp. 338-
352.

Nicholls, G.H. (1979): "Cued speech and the reception of spoken
language". Report from School of Human Communication Disorders, McGill
University, Montreal.

intensity sound", The Laryngoscope 78, pp. 2128-2146.

Oeken, F.W. (1963): "Frequenztransposition zur Hörverbesserung bei
Heilk. 181, pp. 418-425.


