Microprocessor-based hearing aid for the deaf

Yutaka Shimizu, PhD
Industrial Products Research Institute, 1-1-4, Higashi, Tsukuba, Ibaraki 305, Japan

Abstract—A speech reception aid that converts speech into text by use of a microprocessor has been developed for profoundly deaf people. The present system automatically recognizes a spoken word and presents the corresponding text visually and tactually. The system, however, restricts the acceptable number of words and speakers. A new tactile mode applying the psychophysical effect of apparent movement was introduced to reduce the presentation time for the tactile symbols. Laboratory experiments showed that subjects recognized three- to seven-letter words immediately after pronunciation when using the visual display, and after about 5 seconds when using the tactile display. The presentation time of the tactile display was taken up primarily by the activating time of the stimulators. In tactile presentation, subjects received sentences composed of three to four words with more than 90 percent recognition accuracy after about 18 trial sessions.

Key words: microprocessor-based hearing aid, profoundly deaf, psychophysical effect, recognition accuracy, vibrotactile display.

INTRODUCTION

Since the earliest experiments in which the acoustic speech signal was conveyed directly to the skin by tactile means (12), various devices to improve speech communication with deaf individuals have been developed. Various tactile vocoders have been developed for this purpose (3-5,10,11,13,17,24,27,33,34) as well as the visible speech translator (16). Detailed information about this research can also be found in reviews by Elliott (9), Reed et al. (25), Sherrick (28), and Pickett (23).

The choice of display mode for a tactile sensory aid is important in overcoming the perceptual limitations caused by the difference in sensory modalities (14,18,26). Although most tactile speech communication aids have been designed to display the frequency spectrum of speech spatio-temporally, this form of display is not easy to interpret. An alternative to displaying the frequency spectrum is the use of symbolized patterns such as ordinary text, Braille, Morse code, or Blissymbolics (2).

There are two major problems for the designers of speech communication aids by symbolized patterns. One is the technological problem of transforming the acoustic signals into these symbolized patterns. Recent advances in automatic speech recognition by computer promises a possible practical solution. The other problem is perceptual. Psychophysical studies of sensory communication are needed to produce the information needed for the design of effective man-machine interfaces.

This paper describes research on a microprocessor-based system to aid speech reception for profoundly deaf persons. Transformation of a speech signal to text is attempted as a way to realize communication by spatio-temporal patterns. In or-
order to develop a practical device in terms of size and cost, a speech recognition unit that was available commercially (NEC SR-100) was used. In this research, we aimed not at the development of an automatic speech-recognition technique but at the study of visual and tactile symbols as alternative channels for the reception of speech.

**PSYCHOPHYSICAL CONSIDERATIONS FOR SYSTEM DESIGN**

Before assembling the system, the display mode was the subject of psychophysical experiments. A set of Katakana letters (a Japanese system of phonetic symbols) with two special codes was used for this purpose. A Japanese word can be written using these letters, which can be composed mechanically by simple dot elements. Figure 1 illustrates a set of Katakana letters displayed on a 9 × 7-element dot-matrix for tactile communication.

**Visual presentation**

Special attention was not paid to the display mode for this visual presentation, since we already knew from everyday experience that people can read such letters. In the visual mode of operation, the speech recognition unit identifies a word as it is spoken by a speaker and a string of Katakana letters corresponding to the identified word is presented as a visual display.

**Tactile presentation**

Display mode. There are two general types of tactile display. One is a static display in which a stationary pattern is conveyed tactually to the fingers or palm of the hand. The other is a dynamic display in which tactile patterns travel across the skin surface. The dynamic type of display has been used in this investigation. The advantage of a dynamic display is that patterns can be formed in real-time as the speech is recognized. The display is conveyed to a stationary skin surface. It is prefera-
ble not to engage the finger, since it will deprive its use in other occupational activity.

The tactile display developed by Shimizu et al. (32) was used. A matrix of $7 \times 9$ solenoid vibrators spaced 7 mm apart in both rows and columns was used. In order to fit the palm, the surface of the display was made convex, raised to 7 mm at the center. Figure 2 illustrates the vibrotactile display for the palm. The solenoid pins, referred to as tactors, strike the skin with a force of about 3.3 N when activated at 80 pps (pulses per second). These pins have a diameter of 2 mm and a well-defined tactile display can be produced on the palm using this system.

Several modes of tactile display have been studied. In the static mode, all tactors defining a letter are stimulated simultaneously. In the dynamic display, the tactile symbol (e.g., the Katakana letter) is moved across the matrix (6,20). Experiments with well-trained subjects have demonstrated the superiority of the static mode to the moving-letter mode (7,8,21,22). However, accuracy of recognition in reading uppercase letters of the English alphabet has rarely exceeded 80 percent using the static mode. In contrast, the tracing mode (in which the tactors are activated one after another, following the order of strokes used in writing the letter) reveals a high recognition accuracy. In the experiment performed by Shimizu et al. (32), subjects using this mode attained more than 95 percent accuracy, while both the static and the moving-letter modes yielded less than 60 percent when reading Katakana letters. For this reason, the tracing mode was selected for the current investigation.

Time reduction for the tracing mode. Although the tracing mode is effective for recognition accuracy, a shortcoming still exists. It takes time to activate all of the tactors defining a letter, since successive activation is needed to draw each stroke in the tracing mode. A characteristic of this mode is that the tactile stimulus is perceived as a moving point due to the successive activation of pairs of tactors. The effect is known as “apparent movement.” One can perceive a continuously moving-point stimulus by presenting only the initial and final points of the stimulus (19,29). We call the application of this effect an “apparent motion mode.”

In the apparent motion mode only the start, variation, and end-point of a letter-stroke need to be sequentially displayed. Psychophysical experimentation has shown that the optimal sensation of apparent movement appears at the palm when the second pair of tactors are stimulated during activation of the first pair of tactors (31). Therefore, in the present study, the second tactor pair was activated at the mid-point of stimulation of the first tactor pair. This technique reduced the complexity of the hardware. Figure 3 shows a schematic illustration of the apparent motion mode when a one-stroke Katakana letter is defined. To present a well-defined stimulus, a pair of tactors corresponding to each defined position was activated at the same time. Open double circles define the Katakana letter and solid discs identify the tactors that are activated. Three frames are needed to write a stroke in this mode. Note that in the second frame both pairs of activators are stimulated. The perceived effect is as if all the tactors representing the symbol (i.e., tactors represented by open circles) have been stimulated.

Optimal time condition. In the apparent motion mode, the following three parameters are important to the achievement of a high recognition rate. The first is the time duration of stimulators that define a letter-stroke, the second is the time interval between letter-strokes within a single letter, and the third is the time interval between letters that compose a word.

(a) Time interval between letter-strokes—Most Katakana letters are composed of 1 to 4 strokes, as shown in Figure 1. For the tracing and the apparent
motion modes, identification of Katakana becomes much easier when a moderate time interval is used between letter-strokes. Shimizu (30) has investigated the time interval between letter-strokes and has shown that the lower limit of the optimum time interval to be approximately 80 ms. This time interval was used for the apparent motion mode.

(b) Duration of stimulation—An experiment was performed to assess the optimum duration of each tactor in the apparent motion mode. Two female subjects (age 23 and 25) and two male subjects (age 27 and 38) took part in the experiment. A full set of Katakana letters was presented to each subject. Durations of stimulation examined were 25, 50, 100, and 200 ms per letter-stroke (1-4 strokes per letter). An 80 ms inter-stroke interval was inserted when a Katakana letter was composed of more than one stroke. As described above, 80 ms is the briefest interval that produces optimum performance. Total display times for Katakana letters of 1, 2, 3 or 4 strokes were therefore 209, 311, 514, and 921 ms, respectively. A microcomputer system (LSI 11/23) was used to control the tactors in the apparent motion mode. A red LED (light-emitting diode) signal was activated when stimulation was present, while a green LED signal was used for no stimulation. Subjects were asked to respond verbally to each stimulus marker as soon as possible. Response times were recorded between the terminators of the stimulus marker by the first verbal response using a voice sensor attached to the subject's throat. Ten trials using a randomized order of presentation were administered for each time condition.

Figure 4a shows recognition accuracy as a function of mean display time of the Katakana letters, and Figure 4b shows mean response time. Increased recognition accuracy was observed with increased display time, while a decreasing trend in response time was observed. No significant changes in recognition accuracy or speed of response was observed for mean display times over 514 ms. Therefore, it appears that a stimulation duration of 50 ms corresponding to a mean display time of under 514 ms is the lower limit of an optimum stimulation duration. This time condition corresponds to 4 pulses of activation (8 ms of pulse width each) per tactor. As a consequence, the mean display time can be reduced to about one-half of that needed for the tracing mode.

(c) Time interval between letters—Three-to-seven-letter words were presented to assess the effect of varying the time interval between successive letters. The words used were nouns, verbs, adjectives, and adverbs. Six sets of test stimuli composed of 10 words each were used. Inter-letter time intervals examined were 125, 250, 500, and 1000 ms. The same apparatus as described in the previous experiment was used by the same subjects. According to the previous results, the duration of stimulations defining a letter was set to 50 ms and the inter-stroke interval was set to 80 ms. Each subject was able to identify a letter with more than 90 percent of recognition accuracy for this condition. Both recognition accuracy and response times were measured.

Figure 5a shows mean recognition accuracy, and Figure 5b shows mean response time averaged over the four subjects. Recognition accuracy increased and response time decreased with increasing time interval between letters. An analysis of variance showed that neither recognition accuracy nor response time showed a significant change for inter-
letter time intervals greater than 250 ms. Thus, about 300 ms appears to be the lower limit of an optimum inter-letter time interval.

**SYSTEM**

**General**

Figure 6 shows a schematic diagram of the system. A microphone (SHURE SM10A) with FM wireless transmitter (AIWA WT-10) is attached to each speaker. Using this system, it is possible to transmit a voice signal over an area of about 10m × 10m in the laboratory. Two FM systems with carrier frequencies of 77.2 MHz and 78.5 MHz were used to transmit two speakers’ voices separately.

Words identified by the speech recognition unit were coded into eight bits of data (RS-232C) and transferred to the first microprocessor. The first microprocessor is put into a disabled state so as not to accept another speech signal until the signal is processed. (A red LED marker announces the disabled state and a green LED marker the enabled state.)
Figure 6. Block diagram of the system design.

The second microprocessor generates Katakana letters according to the information received from the speech recognition unit. Two display modes can be selected for the tactile display. One is the tracing mode, which is used for the purpose of training in tactile letter perception, and the other is the apparent motion mode, which is intended for actual use. Although optimum times have already been assessed in the previous section, the following time conditions were available for the user's convenience: the durations of tactor stimulation—25, 50, 100, or 200 ms; the inter-stroke intervals—80 or 160 ms; and, the inter-letter intervals—200, 300, or 500 ms.

Two kinds of power sources are available. One is 100 VAC for laboratory use and the other is 12 VDC for outdoor use. The DC power source is rechargeable (NiCd) and lasts about 1.5 hours except when used in a solenoid version of the tactile display.

Figure 7 illustrates the developed system. The main components are included in a case 60 × 38 × 19 cm, weighing about 22 Kg.

Speech recognition

Speech recognition is performed using a commercially available unit (NEC SR-100). This one-board unit can identify one of 120 words pronounced by a known speaker in about 2 seconds. The received speech signal is matched with a standard pattern. Inter-parameter distances between the received signal and a set of standard patterns are computed and the closest match is identified as the recognized word. Recognition time is about 300 ms. The speech recognition unit needs to be trained for each speaker. The first microprocessor assigns a sequence of pronunciations for the speaker in the training mode. It takes about 3 minutes to train the system for 120 words.

The speech recognition unit needs to be trained whenever the system is powered on or a speaker is changed. A mini-floppy disk unit (single-density 48 TPI) is prepared to simplify the training process. Once the system has been trained, the relevant information can be written on or read from the mini-floppy disk under control of the first microprocessor. (It has been found that this information can still be used 3 years after the initial recording onto the mini-floppy disk.)

Two speech recognition boards are used so that the system can accept the speech of two speakers at the same time. Furthermore, to increase the number of words that can be recognized at one time, the system has been trained on three groups of words. The first group is prepared for the purpose of daily greetings and shopping. The second is prepared for education at an elementary school, and the third is for use in demonstrating the system.

Microprocessors

The Z-80 microcomputer handbook (1) was referenced as a source of information about the hardware and software used in the design. Two Z-80A microprocessors with a clock rate of 2 MHz are used. The first has a 2732 EPROM of four kbytes and maintains general control of the system. This processor also controls floppy disk drivers that control the reading, writing, and formatting of a
diskette. The second microprocessor has three 2732 EPROM of 4 kbytes and a 6116 RAM and is the source of the display controls. Programs for the two display modes are written into the first ROM. The three groups of words are written into the second ROM, and the data for the Katakana pattern are written into the third ROM.

**Displays**

Two visual displays are provided. One is a LCD (liquid crystal display) for visual communication and the other is a LED display for tactile perception training. The LCD can represent a 16-letter word at one time. The size of the LCD is $3 \times 8$ mm per letter. The LED display, which is composed of $9 \times 7$ elements, lights on a one-to-one correspondence level with the vibrotactile stimulators.

Two tactile displays are available for tactile communication. One consists of the same solenoid vibrators previously described. This type of display is used in the laboratory and for training on the system. A well-defined stimulus is presented on the palm of the hand by the display. The other consists of PZT (piezo-electric vibrators) for mounting on the wrist in portable occupational use. It is composed of $9 \times 7$ elements with interspacing of 3 mm in both rows and columns. The amplitude of the vibrators is about 230 micrometers when activated at 200 Hz and the weight of the display is about 125 grams. Figure 8 illustrates the portable tactile display.

**Software and operation**

A software program that drives the system was written in the Z-80 assembly language using a microcomputer system (SHARP MZ-80K). After careful checking, the main function of the software was implanted in each ROM without the use of an operating system. Memory requirements are 4 kbytes of ROM for general control of the system,
and 12 kbytes of ROM and 2 kbytes of RAM for control of the displays.

A simplified operational diagram is shown in Figure 9. When a power switch is turned on, the system is automatically initialized. After that, a speaker selects a command function for speech recognition, as described previously, and a deaf user selects a time switch for tactile output. Timing between speech (speaker) and its reception (deaf user) is matched by the LED markers of the system, which announce the enable/disable state.

LABORATORY ASSESSMENT

Psychophysical experiments have been performed to evaluate the system. Subjects taking part in the experiments were a female (age 24), profoundly deaf since birth, a female (age 26) with unimpaired hearing, and a male (age 41). All three are native Japanese and familiar with Katakana.

Subjects were asked to respond verbally to each stimulation as quickly as possible. Response times were recorded between the falls of the output marker by the first response of a voice sensor attached to the subject's throat. The deaf subject was also asked to write each recognized word, since it was difficult for the experimenter to hear her verbal response. A set of earplugs were attached to the ears of each of the other subjects to shield these sound cues.

Three- to seven-letter words were presented. Enough words were prepared to avoid repetitions of the same word. The solenoid type of display was used for tactile presentation, and time conditions of this display were the same as described in the section on psychophysical considerations for system design.
SHIMIZU: Microprocessor-Based Hearing Aid

Figure 9.
Simplified operational diagram of the system. The loop depicted by parallel lines indicates speech transmission from speaker to hearing-impaired subject.

Visual presentation

Twenty words were presented to each subject. Recognition accuracy in from one to five sessions was 100 percent for each subject. Mean recognition times were less than 2 seconds at each session (no statistically significant differences among the sessions were observed). Therefore, no critical training period for the visual recognition of speech was shown by the use of this system.

Tactile presentation

Training for identification of tactile letters. The first subject was trained to perceive tactually presented letters in the tracing mode. The same microcomputer-controlled apparatus as described previously was used for the training. In the first session, the LED visual display was also used to confirm the activated location of stimulators. The apparent motion mode was introduced to each subject when recognition accuracy exceeded 95 percent in the tracing mode.

Each letter was presented three times during a session; thus, it took about 1 hour to complete a session of training. Subjects exceeded 95 percent recognition accuracy in less than five sessions using the tracing mode, and in less than three additional sessions using the apparent motion mode.

Transmission of a word. Thirty different words were presented to each subject. Cued sessions were tried in the first stage. It was announced preliminarily to the subjects what kind of word (e.g., noun, verb) would be spoken. Non-cued sessions were performed when recognition accuracy in the cued sessions exceeded 90 percent for the three consecutive sessions.

Figure 10 shows results of these two experiments. Figure 10a and Figure 10b show data for the deaf subject, Figure 10c and Figure 10d for the normal-hearing female subject, Figure 10e and Figure 10f for the normal-hearing male subject. Figures 10a,c,e show recognition accuracy. Figures 10b,d,f show mean response times. Filled circles show the percentage of correct responses. The open circles show recognition accuracy on the second attempt when the first response was in error.

For each subject, recognition accuracy increased with increased number of sessions. At the same time, mean response times decreased with increased number of sessions. The subjects recognized the presented words very well in the cued experiment. All subjects achieved more than 90 percent of recognition accuracy in fewer than nine sessions. Furthermore, recognition accuracy increased when the same word was spoken twice.

Transmission of sentences. Three- to four-word sentences were presented to the subjects. An experimenter spoke with about a 1-second pause between each word. Thus, each word was presented after about 1.5 seconds of inter-word interval, including the processing time of the system. These sentences were displayed in less than 18 seconds.

The results of this experiment are also shown in Figure 10. The same trends found in the previous two experiments are observed with an increasing number of sessions. Speech reception via the tactile sense of each subject improved with training and exceeded 90 percent of recognition accuracy in fewer than 18 sessions of training.
CONCLUSION

The present experiments have shown that speech reception can be realized using letter symbols via visual and tactual channels. The speech recognition unit used has limitations in that the system must be trained for each speaker and the number of words that can be recognized at one time is limited. However, future devices will overcome these limitations and will be compact and available commercially at a low price.

It was shown that the apparent motion mode was slow but useful for speech reception. Further experiments are necessary to evaluate the minimum transmission time in this mode. Heller (15) performed experiments in which tactile letters were transmitted at several sites on the body at the same time. This knowledge suggests the possibility for reducing the transmission time of the apparent motion mode. Furthermore, it has been proposed to use a symbolized pattern, such as Blissymbolics.

However, it will be necessary to determine how many symbols a person can master at one time and also which mode is most promising in perceiving these symbols tactually.

Tactile communication has the important advantage in that it allows the deaf user to receive normal visual information. Although not assessed in the present study, a wearable tactile display mounted on the wrist will be useful for everyday use. Furthermore, the system can be used for communication in languages other than Japanese since the microprocessor can be programmed to recognize words in any language.

ACKNOWLEDGMENT

The author wishes to thank Professor Barrie J. Frost for his reading of the manuscript and helpful comments.
Figures 10c and 10d show results for the normal-hearing female subject.

Figures 10e and 10f show results for the normal-hearing male subject.
REFERENCES