Veterans Administration

Journal of Rehabilitation Research and Development Vol. 25 No. 3
Pages 51–56

Spatial, tactile presentation of voice fundamental frequency as a supplement to lipreading: Results of extended training with a single subject

ARTHUR BOOTHROYD, PhD, and THERESA HNATH-CHISOLM, PhD
City University of New York, New York, NY 10036

Abstract—An adult with a severe, postlingually-acquired, sensorineural, hearing loss was given 2 hours a week of training in the perception of connected discourse by lipreading, supplemented by voice fundamental frequency (F_o), encoded as locus of vibratory stimulation of the forearm. Three, 3-week blocks of training in the supplemented condition were interspersed with 1-week periods of training by lipreading alone. Training was conducted via a computer-controlled, interactive video system, using a semi-automated connected discourse tracking procedure. Performance was measured as the percentage of words correctly recognized on the first presentation of new sentence material. Scores by lipreading alone averaged approximately 65 percent and remained essentially constant over the 13 weeks of training. Scores under the supplemented condition rose from 65 percent at the beginning of the study to 85 percent at the end. The final supplemented score represented roughly a 50 percent reduction of error rate, when compared with lipreading alone. Performance with the tactile supplement was not as good as with auditorily presented F_o, but was better than has previously been reported in the literature. These data provide evidence to support the notion that subjects can learn to integrate novel tactile codes with the visual stimulus during the lipreading of connected speech.

Key words: *computer-controlled interactive video system, hearing loss, lipreading, tactile codes/display, visual stimulus, voice fundamental frequency, word recognition*

This research was funded by National Institutes of Health Program Project Grant #17764.

Address all correspondence to: Dr. Arthur Boothroyd, Graduate School, City University of New York, 33 W 42nd St., New York, NY 10036.

INTRODUCTION

It has been observed that voice fundamental frequency (F_o), presented auditorily to normally-hearing subjects, is a highly effective supplement to lipreading (11). In our own studies, for example, we have found that typical normally-hearing subjects recognize, by lipreading alone, only about 30 percent of the words in sentences of a known topic. But when they are allowed to simultaneously lipread and hear the fundamental frequency of the voice, their scores rise to approximately 80 percent, even without training (8).

This observation leads to the prediction that the presentation of F_o, suitably encoded for another modality, could form the basis for a highly effective lipreading aid for deaf subjects. The study to be described is one of a series concerned with the development of such an aid. The device in question encodes F_o as the locus of pitch-synchronous vibratory stimulation of the skin (2,4). With this device, the intonation patterns of speech are felt as patterns of movement on the forearm. Psychophysical experiments have confirmed that subjects wearing this aid can perceive F_o changes of one-third octave using an 8-channel transducer array, and one-sixth octave using a 16-channel array (1). In forced-choice tests, this resolution has been shown to be sufficient for the perception of phonologically significant contrasts of intonation contours. The tactile stimulus also provides improved perception of durational cues and consonant voicing cues (9). In a prelimi-

Journal of Rehabilitation Research and Development Vol. 25 No. 3 Summer 1988

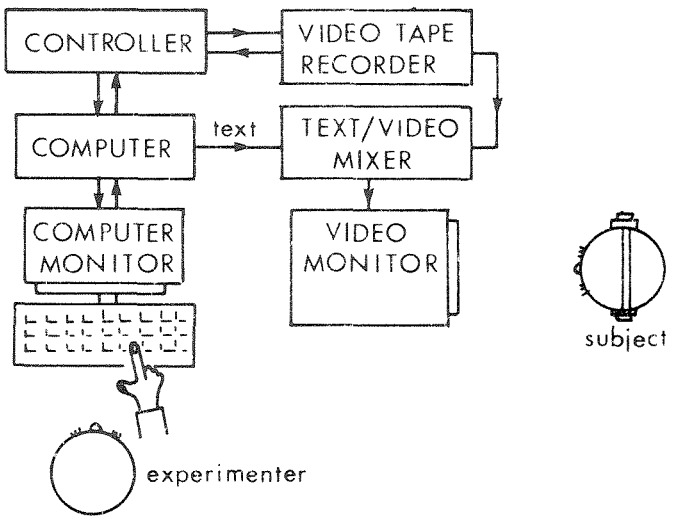

Figure 1.
Block diagram of the computer-controlled, interactive video system used for training.

nary experiment with connected speech, 2 normally-hearing subjects were found to obtain small but significant improvements in lipreading, with the help of the display, after only a small amount of training (4).

Data on the performance of a spatial electrotactile display of F_o have also been reported by Grant *et al.* (6). Again, a small but significant improvement in the perception of connected speech was found when the tactually encoded F_o patterns were combined with the information available from lipreading.

Although the connected speech experiments have confirmed that spatially-coded tactile displays of F_o can supplement lipreading, the magnitude of the effect has been disappointingly small, and much smaller than that found with auditory presentation of F_o. One possible reason is that subjects have so far been given insufficient practice in the perception of the novel sensory code and/or its integration with the visual patterns of lipreading. The present study was designed to explore this issue by providing extended training to a single, highly-motivated, hearing-impaired subject.

SUBJECT

The subject of this study was a hearing-impaired adult male student in the Doctoral Program in Speech and Hearing Sciences at the City University of New York. He has a severe, sensorineural hearing loss with a better ear, 3-frequency average, pure tone threshold of 65 dB. The hearing loss was acquired postlingually, during early childhood. At the time of this experiment, the subject was 32 years old. This subject was chosen because of his long-term experience with the lipreading process.

EQUIPMENT

A desk version of the tactile display was used for this experiment. The spatial patterns were presented to the medial surface of the left forearm, using a 16-channel array of solenoids. The system was adjusted to present a total range of 2.5 octaves, logarithmically mapped onto the transducer array. The channel separation was roughly one-sixth octaves. Only 1 solenoid was activated at any given time, and the stimulus consisted of constant amplitude square waves of frequency $F_o/2$. Input to the display was by direct electrical connection, using a signal derived from a custom-built pitch extractor.

An interactive, computer-controlled video playback system was used in training. The system is illustrated in **Figure 1**. The computer is an IBM-XT with a hard disk and 2 floppy disk drives. The software is stored on the hard disk, while the text of the training material is stored on floppy disks. Video-recordings are played on a half-inch player (Panasonic AG6300). The video image is shown on a 19-inch color monitor (Sony 1901). A PC-Keyer, from Video Associates Laboratories, mixes computer-generated text with the video signal, for display on the subject's video monitor. A BCD Associates VCR controller allows for the control of transport functions, and the numbering, reading, and location of video frames (3).

TRAINING

Material

A video-recorded story was used for training. The story was *Little House in the Big Woods* by Laura Ingalls Wilder. This was chosen because the sentences are of relatively simple syntactic structure, there is limited use of dialogue, there are few proper nouns, and interest level is fairly high.

The story was recorded in color by a female talker. The view was from the clavicle to the top of

the hair. The talker was instructed to avoid facial movements not directly required for speech production, to pause briefly after each sentence; and in the case of long (>13 word) sentences, to pause also at a suitable phrase boundary.

One audio channel was recorded. It contained the acoustic output of an Electrovoice microphone positioned approximately 6 inches below the talker's lips. During the preparation of test tapes, the audio signal was fed to a custom-built pitch extractor. The output of the pitch extractor was recorded onto the audio channel of the test tape.

Method

Training involved a semi-automated version of the connected discourse tracking procedure first described by DeFilippo and Scott (5). In the original procedure, subjects are required to lipread, and repeat verbatim, text read by a speaker. In the case of an error, various repair strategies are used until the subject repeats the phrase or sentence correctly. Performance is measured as the number of correctly repeated words per minute.

In the procedure used in the present study, each phrase, or sentence, is presented a maximum of 4 times. After the first presentation, the subject attempts to repeat the stimulus. He is then shown, on the monitor, the text of the words repeated correctly, together with asterisks to indicate the location of words not yet recognized. After the second presentation, he again sees the text of words recognized, but the computer now adds 50 percent of the words still unrecognized. After the third presentation, the subject is shown the complete text, and the stimulus is presented one more time. If at any time the subject repeats the whole sentence correctly, the computer advances the tape to the next phrase or sentence and the process is repeated. Information about the accuracy of the subject's repetitions is provided to the computer by the experimenter, who sees the text of the current stimulus on the computer monitor, with a number under each word. She simply hits the numbers on the computer keyboard corresponding with the correct words.

In the present study, performance was measured as the percentage of words correctly repeated after the first presentation of a new phrase or sentence. A single score was based on roughly one-half hour of training, which typically involved the presentation of about 30 phrases or sentences.

Procedure

The subject was seen twice a week for 13 weeks. Each session lasted approximately 1 hour and yielded 2 datapoints, that is, there were 4 datapoints for each week of training. Hearing aids were not worn during training, and, even though he was severely deaf, the subject was presented with white noise masking under circumaural earphones, to make absolutely sure he was not hearing the acoustic output of the transducer array.

Training with the tactile supplement took place during 3 blocks of 3 weeks each. These blocks were preceded, separated, and followed by 1 week during which the subject was trained by lipreading alone.

Table 1.
Percentage of words recognized at the first presentation of sentences in story material, during one-half hour training sessions. Scores are shown as a function of modality. There were four training sessions per week.

	Modality																						
Training Block	Lipreading alone (pre)				Lipreading plus F_0												Lipreading alone (post)				Mean	SD	SE
1	70	60	69	63													53	60	68	66	63.6	5.8	2.0
					56	61	61	69	65	65	50	62	75	61	69	80					64.5	8.1	2.3
2	(53	60	68	66)*													61	63	69	74	64.3	6.5	2.3
					68	68	66	65	77	66	75	77	79	68	86	84					73.3	7.4	2.1
3	(61	63	69	74)*													79	79	72	64	70.1	7.1	2.5
					82	83	90	75	85	86	85	76	84	88	84	92					84.2	5.0	1.4

*The "pre" scores for blocks 2 and 3 are also the "post" scores for blocks 1 and 2, respectively.

Journal of Rehabilitation Research and Development Vol. 25 No. 3 Summer 1988

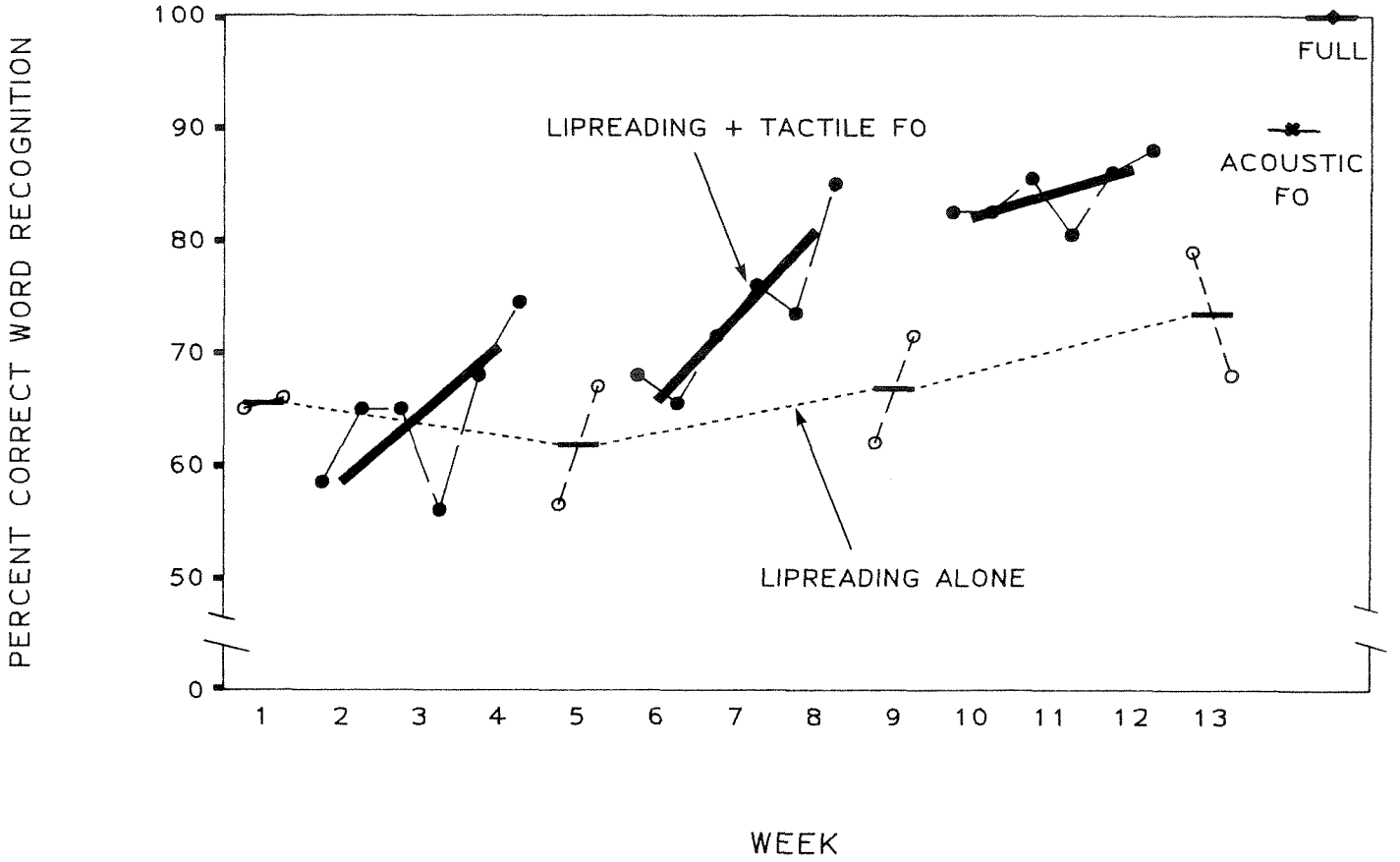

Figure 2.
Percent words recognized at the first presentation of sentences in story material. Data are shown as a function of training time and sensory modality. Each datapoint is the average of 2 scores obtained in a single 1-hour session. On the extreme right are 2 scores obtained by lipreading, supplemented by *auditory* presentation of F_o and of the full speech signal.

RESULTS

The results are presented in **Table 1**. They are also illustrated in **Figure 2**, which shows the average score for each day of training. It will be seen that the percentage of words correctly repeated under the lipreading-alone condition rose only slightly, from around 65 percent to around 70 percent, during the 13 weeks of training. Under the supplemented condition, the initial scores were similar to the lipreading-alone scores, but they rose to around 85 percent by the end of training.

Statistical analyses were performed to answer several questions about these data:

1. *Were there systematic changes in the lipreading-alone scores?* To answer this question, the arcsine transforms ($x_{tr} = \sin^{-1}(x^{(1/2)})$) of the 4 sets of lipreading-alone scores were subjected to a simple one-way analysis, in which the variance between sets was compared with the variance within sets. The results did not support the conclusion that the average scores changed significantly from set to set ($F(3,12) = 2.58$, $p > 0.05$). It must, therefore, be concluded that the lipreading-alone scores did not change during the 13-week training period.

2. *Were the supplemented scores greater than the lipreading-alone scores obtained immediately before and after each block of supplemented training?* To answer this question, a simple t test was performed on each of the 3 blocks of data. For each training block, there were 12 datapoints for the supplemented condition, and 8 datapoints for the lipreading-alone condition (4 "before" and 4 "after"). The difference was not significantly different from zero for the first block ($t(18) = 0.87$, $p > 0.05$), but it was for the second and third ($t(18) = 2.89$, $p < 0.05$ and

$t(18) = 5.22$, $p < 0.01$, respectively). We, therefore, conclude that the supplemented scores were better than the lipreading-alone scores during the second and third blocks of training.

3. *Was there evidence of learning during training with the tactile supplement?* To answer this question, we examined the correlations between supplemented score and session number during each of the 3 training blocks. The results showed a significant correlation during the first and second blocks ($r(10) = .53$, $p(1 \times \text{tail}) < 0.05$ and $r(10) = .74$, $p(1 \times \text{tail}) < 0.01$), but not during the final block ($r(10) = .31$, $p(1 \times \text{tail}) > 0.05$). We, therefore, conclude that the subject evidenced learning under the supplemented condition, but that a plateau was reached by the time he reached the third block of training (i.e., after a total of about 13 hours of training).

FOLLOW-UP STUDY

To aid in the interpretation of results, we obtained 2 additional scores at the end of training, both with auditory supplements to lipreading. For the first, we allowed the subject to listen to the pitch-extracted signal. For the second, we allowed him to hear the complete speech signal. In both cases, he used his personal hearing aids for amplification. The resulting scores, based in each case on approximately 60 sentences, were 90 percent with the auditory F_0 supplement, and 100 percent with the full speech signal. These scores have been included in **Figure 2**. The score of 90 percent, obtained with auditorily-presented F_0 is significantly greater than the mean supplemented score obtained during the last block of training with the tactile supplement (difference = 5.8 percentage points, $t(11) = 4.07$, $p < 0.01$).

DISCUSSION

These results provide striking evidence that a competent lipreader, given sufficient training, can learn to integrate the visual patterns of lipreading with unfamiliar spatial, tactile, patterns representing F_0. It could be argued that the design was biased in favor of the tactile supplement, since the subject received over twice as much training with it than by lipreading alone. But if the supplement were providing no benefit, then all of the training would, effectively, have been by lipreading alone and there would have been no difference between the 2 kinds of score. The fact that there was a difference confirms that the tactile input was providing useful information.

Further support for the notion that these data are showing a real effect comes from the subjective reports of the subject. At the beginning of training, he found the tactile input distracting and felt that he must attend either to the visual or to the tactile channel, but could not do both. By the end of training, he was not aware of the source of sensory input, but was simply focusing on the message. He further reported that he appeared to "hear" the intonation contours of the speech, even though he knew this to be impossible.

At first sight, the magnitude of the improvement afforded by the tactile supplement seems quite small, amounting to only 15 percentage points. This is significantly larger than the improvements found in previous studies, however, and in this particular subject, who is an extremely competent lipreader, a 15 percentage point change represents a 50 percent reduction of error.

It will be noted that this subject obtained more lipreading supplementation from listening to F_0 than from feeling it. Under the auditory condition, the reduction of error was close to 70 percent, this being similar to the results we have obtained with highly competent, normally-hearing, lipreaders (10). It is, of course, possible that with more extended training, performance with the 2 modalities would converge. This seems unlikely, however, since the tactile display is subject both to pitch extraction errors, and to quantization errors arising from the representation of smooth intonation contours with an array of discrete transducers. (Note that, if the subject were perceiving the frequency of vibration of the solenoids directly, there would be no quantization error. But if, as we assume, it is the spatial transform that is providing the salient cue, then quantization is imposed by the display itself.) In other research, we have found that these errors can significantly reduce the effectiveness of auditory supplements to lipreading (7).

These data support the concept of a wearable tactile aid that provides a spatial transform of F_0. It should be noted, however, that in a wearable unit,

the problems of signal level and signal-to-noise ratio impose severe demands on the design of an effective pitch extraction system. Nevertheless, the results of this study do offer existence proof of the feasibility of developing effective tactile aids to lipreading.

CONCLUSIONS

1. With extended training in the perception of connected discourse, the subject of this study was able to integrate a spatial tactile display of voice fundamental frequency with the visual patterns of lipreading, thereby enhancing his lipreading performance.

2. The effect of the supplement began to show after about 5 or 6 hours of training, and reached a plateau after about 13 or 14 hours.

3. By the end of training, the supplemented score was approximately 85 percent, this being approximately 15 percentage points higher than the lipreading-alone score. This improvement represents a 50 percent reduction of error.

ACKNOWLEDGMENTS

The pitch extraction system was built by Eddy Yeung, and the transducer array by Cecil Redmond. Laurie Hanin served as talker for the video tapes. Data were collected by Susan Langtry, Nancy Plant, and Liat Kishon-Rabin. We are particularly grateful to Larry Medwetsky, who served as subject. This research was supported by NIH Program Project Grant #17764.

REFERENCES

1. **Boothroyd A:** Experiments with a wearable tactile pitch display. Paper presented at *Proceedings of the 2nd International Conference on Rehabilitation Engineering*, Ottawa, 1984.
2. **Boothroyd A:** A wearable tactile intonation display for the deaf. *IEEE Trans Acous Speech Sig Proc* 33:111-117, 1985.
3. **Boothroyd A:** CASPER: A computer-assisted system for speech-perception testing and training. *Proceedings of the 10th Annual Conference on Rehabilitation Technology*, 734-736, Washington, DC: Association for the Advancement of Rehabilitation Technology, 1987.
4. **Boothroyd A, Hnath T:** Lipreading with tactile supplements. *J Rehabil Res Dev*, 23(1):139-146, 1986.
5. **DeFilippo CL, Scott BL:** A method for training and evaluating the reception of ongoing speech. *J Acoust Soc Am* 63:1186-1192, 1978.
6. **Grant KW, Ardell LA, Kuhl PK, Sparks D:** The transmission of prosodic information via an electrotactile speechreading aid. *Ear Hear*, 7:328-335, 1986.
7. **Hnath-Chisolm T:** The effects of quantization on the utility of voice fundamental frequency as an acoustic speechreading supplement. PhD diss., City University of New York, 1987.
8. **Hnath-Chisolm T, Hanin L, Boothroyd A:** Sentence perception with and without auditory F_0 supplementation. Report #TSL15, City University of New York, 1986.
9. **Kishon-Rabin L, Boothroyd A, Hnath-Chisolm T:** Lipreading with and without tactually transmitted fundamental frequency: Speech contrast perception. Paper presented at convention, American-Speech-Language-Hearing Association, New Orleans, LA, 1987.
10. **Langtry S, Boothroyd A, Hnath-Chisolm T:** Supplementing lipreading with fundamental frequency: Talker and viewer differences. Paper presented at convention, American Speech-Language-Hearing Association, New Orleans, LA, 1987.
11. **Rosen SM, Fourcin AJ, Moore BCF:** Voice pitch as an aid to lipreading. *Nature*, 291:150-152, 1981.