Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency

Silvio P. Eberhardt[a] and Lynne E. Bernstein[b]
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Marilyn E. Demorest
Department of Psychology, University of Maryland Baltimore County, Catonsville, Maryland 21228

Moise H. Goldstein, Jr.
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 13 November 1989; accepted for publication 9 May 1990)

The main goal of this study was to investigate the efficacy of four vibrotactile speechreading supplements. Three supplements provided single-channel encodings of fundamental frequency ($F0$). Two encodings involved scaling and shifting glottal pulses to pulse rate ranges suited to tactual sensing capabilities; the third transformed $F0$ to differential amplitude of two fixed-frequency sinewaves. The fourth supplement added to one of the $F0$ encodings a second vibrator indicating high-frequency speech energy. A second goal was to develop improved methods for experimental control. Therefore, a sentence corpus was recorded on videodisc using two talkers whose speech was captured by video, microphone, and electroglottograph. Other experimental control issues included use of visual-alone control subjects, a multiple-baseline, single-subject design replicated for each of 15 normal-hearing subjects, sentence and syllable pre- and post-tests balanced for difficulty, and a speechreading screening test for subject selection. Across 17 h of treatment and 5 h of visual-alone baseline testing, each subject performed open-set sentence identification. Covariance analyses showed that the single-channel supplements provided a small but significant benefit, whereas the two-channel supplement was not effective. All subjects improved in visual-alone speechreading and maintained individual differences across the experiment. Vibrotactile benefit did not depend on speechreading ability.

PACS numbers: 43.71.Ma, 43.71.Es, 43.66.Wv

INTRODUCTION

A. Fundamental frequency as a speechreading supplement

Profoundly deaf individuals depend primarily on vision for communication, either via sign language or speechreading. Sign language requires special shared knowledge between the deaf individual and others and so is only a partial solution to communication problems. Speechreading is frequently inaccurate. Sensory substitutes or supplements such as vibrotactile devices or cochlear implants are being studied as a means to ameliorate the communication problems of the deaf. This report concerns efforts to develop an effective vibrotactile speechreading supplement.

One strategy for developing a supplement is to consider what speech information is not available to the speechreader and to design the vibrotactile supplement to provide the invisible information via the skin. One important but invisible source of linguistic information is conveyed by voice fundamental frequency ($F0$). Experiments in which stimuli derived from $F0$ were presented auditorily to speechreaders have demonstrated the efficacy of $F0$ information (Boothroyd *et al.*, 1988; Breeuwer and Plomp, 1984, 1985; Grant, 1987; Rosen *et al.*, 1981; Risberg and Lubker, 1978). For example, Breeuwer and Plomp (1986) showed that experienced normal-hearing subjects improved their syllable identification scores on an open-set sentence identification task from 33% speechreading alone to 73% with an auditory indication of $F0$.

Fundamental frequency's success as a speechreading supplement is probably due to the several segmental and suprasegmental linguistic functions it serves in English. $F0$ onset, duration, and frequency characteristics are implicated in segmental voicing distinctions (Lisker and Abramson, 1967; Klatt, 1976). $F0$ is a primary means in the language to convey stress and intonation (Lehiste, 1970). On the lexical level, stress can differentiate between words (e.g., "the convert" versus "to convert"). On the sentential level, stress can

[a] Some of the results in this report were reported in an unpublished doctoral dissertation of the first author. Send reprint requests to the first author in care of the second author at: Center for Auditory and Speech Sciences, Gallaudet University, Washington, DC 20002.

[b] Current affiliation: Center for Auditory and Speech Sciences, Gallaudet University, Washington, DC 20002.

be used for emphasis or to diambiguate syntax. Intonation provides syntactic information.

$F0$ is particularly attractive for vibrotactile presentation because: (1) changes in F0 on the sentential and lexical level are relatively slow, generally within the temporal processing limitations of the skin (Rothenberg *et al.*, 1977); (2) the range of $F0$ is close to the range of frequencies best suited to the skin (Rothenberg *et al.*, 1977), although frequency transformations can result in improved performance (Bernstein *et al.*, 1989; Rothenberg and Molitor, 1979); and (3) $F0$ is relatively accessible from the acoustic speech signal.

Noting the close fit between the skin's frequency characteristics and the range of $F0$, Rothenberg *et al.* (1977) raised the possibility that a single vibrotactile channel could be used to encode "at least the stronger variations" in $F0$. Within the range 10–100 pulses per second, subjects could differentiate at least seven frequency steps on the forearm and ten on the fingertip. Rothenberg and Molitor (1979) compared several candidate $F0$-to-vibrotactile single-channel transformations. Subjects were tested in a six-alternative forced-choice procedure in which they identified the stress and intonation of a sentence, "Ron will run," spoken with stress on one of the three words and as either a question or a statement. The transformations that centered pulse frequencies around 50 pulses per second tended to be more successful in conveying intonation and, to a lesser extent, stress. Confirmation that a single vibrotactile channel can convey stress and intonation under vibrotactile-alone conditions was provided by Plant and Risberg (1983) and Bernstein *et al.* (1989). Bernstein *et al.* (1989) additionally showed that a transformation from voice $F0$ to differential amplitude of two fixed-frequency sinusoids presented on a single channel was also effective in conveying stress and intonation.

Few speechreading studies of open-set sentence identification with single-channel $F0$ supplements are available in the literature. Plant (1986) tested two deaf subjects using a single channel that conveyed both $F0$ and the presence of high-frequency speech energy. No tests of significance were reported, but results generally showed the supplement to be of benefit for tasks involving live administration of nonsense syllables, words, sentences, and connected discourse tracking (De Filippo and Scott, 1978). Plant and Spens (1986) studied a deaf adult subject with the same supplement. This subject had developed his own method for obtaining $F0$; viz., he placed his hand on the neck and shoulder of the talker. The subject's method was superior to the artificial device, but both vibrotactile methods were superior to speechreading alone, providing gains in a variety of speech tasks. Boothroyd and Hnath (1986) tested two normal-hearing individuals for 6 h, rotating visual-alone, single- and multichannel $F0$-supplemented conditions. Videorecorded sentences were used and both supplements resulted in significantly higher scores, with neither configuration superior to the other.

Several studies of multi-channel encodings of $F0$ are available. A study by Grant (1980) explored the possibility of using electrocutaneous stimulation. Boothroyd and Hnath (1986) and Hnath-Chisolm and Kishon-Rabin (1988) combined both locus and rate coding. Cowan *et al.* (1988) devised electrocutaneous stimulation that provided $F0$, amplitude, and an indication of second formant frequency. All these studies determined that the supplemental stimulation led to gains in speechreading. For the most part, comparison between single- and multichannel supplements for $F0$ has not demonstrated a clear advantage for either type of coding (Boothroyd and Hnath, 1986; Hanin *et al.*, 1988). Single-channel transformations of $F0$ have clear advantages with regard to device development issues such as power consumption, cosmetic acceptability, and use of additional channels for other speech characteristics, and therefore warrant further investigation in the absence of clearcut evidence favoring multichannel configurations.

In the study reported here, three $F0$-to-vibrotactile transformations were tested in an open-set sentence identification task. Two of the transformations encoded $F0$ as tactile pulse rate. The third transformed $F0$ to differential amplitude of two sinusoids, one at 40 Hz and one at 400 Hz. As $F0$ increased, the amplitude of the 400-Hz sinusoid increased (while the amplitude of the 40-Hz sinusoid decreased) and, as $F0$ decreased, the amplitude of the 40-Hz sinusoid increased (while the amplitude of the 400-Hz sinusoid decreased). A fourth supplement combined one of the rate encodings with a second vibrotactile channel that indicated the presence of high-frequency speech information. The addition of the second channel followed the rationale of the first, viz., that it may be possible to efficiently and effectively code extracted speech cues on independent vibrators.

The literature on vibrotactile supplements provides few other attempts to encode extracted speech cues on independent stimulators. One attempt was made by De Filippo (1984). She reported on a three-channel supplement, with each of the channels assigned to one of three transducers: a microphone signal used for obtaining acoustic envelope information, a nasal accelerometer for nasal coupling, and a throat accelerometer for voicing. After 4 to 10 h of training, benefits were obtained on tasks involving nonsense syllables and open-set sentence identification of 10 to 20 CID Everyday Sentences (Davis and Silverman, 1970). It was reported that benefit was between 0% and 15% for open-set sentence identification, and subjects found it difficult to make use of the vibrotactile information when the stimuli were sentences. Given the minimal training, it might be expected that gains would be small and the task would be difficult.

One other demonstration of encoding extracted speech cues on independent vibrotactile channels is found in work by Martony (1974). Encouraged by a visual speechreading device described by Upton (1968), which presented speech features by illuminating individual elements mounted on an eyeglass frame, Martony compared a vibrotactile indication of voiced, voiceless, and stop features with a visual display. Three vibrators were held against three fingertips of one hand, and three indicator lamps were taped to the video monitor used for stimulus presentation. Results showed that for sentence-level tasks, the tactual supplement was superior to the visual, but for syllable-level tasks, the visual was superior. However, gains in open-set sentence identification were small over the short course of the study. In light of the minimal information available on this strategy of providing speech cues on individual vibrators, the addition of a condition in the present study involving a second channel for high-frequency energy indication was warranted.

B. Methodological considerations

One serious concern that this study addressed was the need for better experimental control such that comparisons of vibrotactile supplements could be more readily achieved across subjects and laboratories. The literature discloses widespread use of live talkers, the use of extremely small sample sizes, frequent inattention to possible effects of visual learning, inattention to initial speechreading ability, inattention to individual differences, and repeated use of small sets of sentence materials. A description of the pitfalls of these practices and the strategies adopted here for controlling or studying each of them follows.

In order to assess speechreading performance, many recent studies have relied on the method of continuous discourse tracking (CDT, De Filippo and Scott, 1978). In CDT, an experimenter reads from a book and the subject is required to repeat verbatim what was said. If necessary, the experimenter repeats sentence fragments several times to obtain verbatim repetition from the subject. The performance measure is words correct per minute (wpm). Although the CDT task does mimic some of the interactive and cognitive/linguistic aspects of natural communication, many uncontrolled variables significantly affect tracking rates, and thus the technique has pitfalls when used for evaluating sensory aids, as has been noted repeatedly in the past (Pickett, 1983; Hochberg *et al.*, 1989; Robbins *et al.*, 1985; Matthies and Carney, 1988; Tye-Murray and Tyler, 1988). Despite its pitfalls, CDT continues to be a primary tool for assessing vibrotactile supplements. An alternative method is to present prerecorded isolated sentences. In this manner, it is possible to control many, if not most, of the factors uncontrolled by CDT. For the current study, a sentence corpus was developed and recorded on video laserdisc for the purpose of achieving the desired stimulus control.

Most studies on vibrotactile supplements have employed only two to four subjects, and comparisons among two or more supplements have typically involved within-subject contrasts. It is not known, however, whether use of several different supplements by one individual results in interference or enhancement of performance. These concerns were accommodated here by testing each of the four supplements with a different group of three subjects.

A well-accepted practice in performance assessment outside the area of vibrotactile supplements is to include no-treatment controls. Here, the inclusion of a visual-only control group that received precisely the same protocol but without the vibrotactile stimulation allowed verification that apparent vibrotactile effects were actually due to the supplement and not to experimental design artifacts.

A question for use of vibrotactile supplements outside the laboratory is whether magnitude of benefits depends on individual differences among potential users. Individual differences are important, because clinical observation and the research literature on hearing aids, cochlear implants, and vibrotactile supplements strongly suggest that there is wide variation among subjects in speechreading ability and in ability to benefit from sensory aids. In order to probe this possibility, it is necessary to assess individual performance. Therefore, in this study, within the group design, each of the subjects was tested in a multiple-baseline single-subject design (Kearns, 1986) wherein the subject was alternately tested in (unaided) baseline and (aided) treatment conditions. This design provided data for testing reliability of group effects on the single-subject level. To insure representation of a range of speechreading abilities and to disqualify subjects with extremely poor skills, a brief screening test was developed and used in assigning subjects to experimental groups. This estimate of initial ability enabled examination of the relationship of benefit to initial ability. Also, to obtain an independent measure of subjects' speechreading improvement across the course of the experiment, pre- and post-tests consisting of sentences and syllables were administered. The two sentence sets, taken from the CID Everyday Sentences (Davis and Silverman, 1970), had been previously balanced for difficulty (Demorest and Bernstein, 1987).

Finally, a procedure was included to investigate the effect of repeated presentation of a small set of sentences. It was of interest to determine whether subjects learn to recognize a small set of sentences that appear at regular intervals amidst a much larger set of randomly sampled sentences. This information could assist in determining the nature of materials required to adequately assess vibrotactile benefit. Therefore, a set of 39 sentences was reserved for presentation at the end of every (unaided) baseline session for comparison with nonrepeated sentences.

I. METHOD

A. Visual stimuli

All stimuli were prerecorded on videotape and transferred to a set of two video laserdiscs (Bernstein and Eberhardt, 1986a,b). Stimuli were recorded by both a male and a female talker. Details concerning videodisc production are given in Bernstein *et al.* (1989) and Eberhardt (1987). Briefly, talkers were seated in front of a dark background. Lighting was provided by two face-level direct 600-W floodlights, at approximately 35° to either side of the midline. Two 600-W fill lights were positioned above the talkers. The talker's face filled most of the screen area. During recording, the signal from an electroglottograph (EGG) (Synchrovoice) was recorded on one of the audio tracks following linear phase high-pass filtering (Glottal Enterprises filter).

The talkers spoke General American English. The female was an actress and the male a singer. Extensive testing has shown that the male is more visually intelligible than the female (Demorest and Bernstein, 1987). Talkers were instructed to produce each sentence as though it had been extracted from a natural conversation. This resulted in natural facial expressions and intonation.

Three sets of materials from the videodisc were utilized in the present experiment: CID Everyday Sentences (Davis and Silverman, 1970), CV nonsense syllables, and a corpus of 1568 sentences. The 100 CID sentences spoken by the male talker were used for screening speechreading (20 sentences) and for pre- and post-testing (40 sentences each).

CV nonsense syllables consisted of two tokens from each talker of the 22 consonants /p, b, m, f, v, θ, ð, w, r, č, j, ʃ, ž, t, d, s, z, k, g, n, l, h/ combined with the vowel /ɑ/. The isolated vowel /ɑ/ was also included for a total of 92 tokens.

Since large corpora of everyday speech appear not to be available, sentences were composed largely by one of the authors (LEB) with the intent of providing a large set of unrelated topics in language representative of casual conversation. Words were selected with reference to the Thorndike and Lorge (1944) word counts so as to omit those with extremely low frequency. The sentence corpus contained 70% statements, 25% questions, and 5% imperatives. Question types included *yes/no*, *wh-*, and tag, as well as those formed only by using a rising terminal intonation contour. Syntax of the statements included simple active declaratives, passives, and negatives. Approximately half of the sentences were recorded by both the male and female talker, and half were unique to one talker. Sentence length varied between 3 and 12 syllables as given by dictionary transcriptions. The distribution of sentence types was the same for each sentence length. However, the majority of sentences were between six and nine syllables. Analysis of the phonemic content of the sentences, based on an automatic (DECtalk) transcription, showed the distribution of phonemes in the corpus to be equivalent to that reported by Denes (1963) for a larger corpus of English.

The sentences and nonsense syllables were presented on a 16-in. color monitor (NEC RGB). The first frame of the stimulus was presented until the subject indicated that he or she was ready. The stimulus was then played in real time, and stopped on the last frame for 1 s before the first frame of the next stimulus was displayed. This procedure allowed subjects to become accustomed to the talker's face and position before utterances began. The effect achieved by frame freezing was natural and continuous, since the start and end video frames had been chosen during videotape editing for a relaxed expression on the talker's face. Also, the use of constant angular velocity (CAV) videodiscs allowed frame-by-frame control of the stimulus with flawless transitions between freeze and play modes.

B. Vibrotactile stimuli

1. Transformations for vibrotactile stimuli

Four vibrotactile supplements were tested, of which three used $F0$-to-vibrotactile transformations from Bernstein *et al.* (1989): *log-scaled F0* (LSF); *difference-scaled F0* (DSF); and *amplitude-modulated dual fixed frequency* (AMFF) transformation. The fourth vibrotactile supplement provided indication of $F0$ on one channel and high-frequency speech energy on another.

For the *log-scaled F0* transformation, the logarithm of the normalized fundamental frequency was shifted down in frequency to a range of 10-100 Hz. This transformation was calculated as

$$P_t = 1/50 + [\log(P_v/M)] \times S,$$

where P_t defines vibrotactile pulse interval, P_v is glottal period, and M is the talker's mean pulse period. The frequency excursion (or expansion) scale factor S was chosen so that at least 95% of the P_t's fell into the range 0.01–0.1 s.

The *difference-scaled F0* transformation was similar, except that an infinite impulse response filter was used to estimate the talker's mean glottal pulse period, as might be required by a wearable device. The above equation was applied except that M was estimated by the recursive filter

$$M_v = \frac{P_v}{2} + \frac{M_{v-1}}{2} = \frac{P_v}{2} + \frac{P_{v-1}}{4} + \frac{P_{v-2}}{8} + \cdots,$$

where v corresponds to the current glottal period. A side effect of this filter is that short-term changes in $F0$ are accentuated, whereas long-term changes are deemphasized.

The *amplitude-modulated dual fixed frequency* transformation applied a second encoding to the log-scaled $F0$ transformation: two fixed frequency sine waves at 40 and 400 Hz, presented by a single stimulator, were differentially amplitude modulated such that when voice $F0$ increased, the 400-Hz signal amplitude increased (and the 40-Hz amplitude decreased); and when voice $F0$ decreased, the 40-Hz signal amplitude increased (and the 400-Hz signal decreased). This transformation was formulated to differentially stimulate two vibrotactile receptor populations in the skin. A more detailed description of all these transformations may be found in Bernstein *et al.* (1989).

The fourth supplement (LSF + E) comprised, in addition to the LSF transformation, a second vibrator that presented high-frequency speech energy. The energy information was coded by 400-Hz fixed-frequency, fixed-amplitude sine-wave bursts with durations equivalent to those of the high-frequency speech energy. This vibrator was activated for most occurrences of the /s/ phoneme, but few instances of other phonemes. The word-final /s/ was often not registered, and occasional strong instances of the phonemes /t, č, ʃ/ were.

2. Generation of vibrotactile signals

Generation of the vibrotactile stimuli was accomplished in software in several processing stages (Eberhardt, 1987). First, the electroglottograph (EGG) sound track on the videodisc was sampled at 15 KHz, and antialias filtered at 5 KHz. The EGG detects glottal pulses and thus gives a more accurate measure of $F0$ than obtainable from the acoustic speech signal. A correlation-based pitch detection algorithm was then used to extract pitch period markers (Eberhardt, 1987). To reduce spurious periods and pitch doubling errors, only periods that were within the range of half to double the talker's mean pitch period were selected, and successive periods of a voiced speech segment were required to be within 10% of each other. These rather conservative rules were applied to filter out the highly variable vocal periods that commonly occur during initiation and termination of voiced speech segments.

Finally, $F0$-to-vibrotactile signal transformations were generated for each sentence. $F0$-to-vibrotactile signal transformations followed glottal pulse detection. Based on the duration of glottal periods, a software program generated a list of times (from sentence initiation) at which a vibrotactile pulse was to be initiated (or the amplitude changed, in the case of the AMFF transformation). The basic strategy was to use the current glottal interval to calculate the new vibrotactile inter-pulse-interval.

Files specifying high-frequency information for the LSF + E transformation were prepared by high-pass filtering the speech signal on the audio track of the videodisc at 5 kHz, and rectifying, smoothing, and digitally sampling the resulting signal. An amplitude threshold was then applied such that most occurrences of /s/ would result in vibrotactile stimulation.

3. Vibrotactile apparatus

The transformed pitch periods were presented by a vibrotactile delivery system modeled after one used by Verillo (1962). A vibrator (AV6, Alpha-M Corporation, Euless, TX) was mounted under a table so that the contactor could protrude through a plate drilled to provide a surround 1 mm greater in radius than the 0.28-cm^2 contactor. An electronic filter network was used to compensate for the resonances and frequency rolloff of the vibrator (see Bernstein *et al.*, 1986, for a more detailed description of the vibrotactile delivery system). The high-frequency energy cue was presented using a small vibrator (V1420, Audiological Engineering Corporation, Somerville, MA) that subjects held in their left hand.

C. Subjects

1. Speechreading screening test

To insure an adequate representation of different levels of speechreading ability and to reject extremely poor speechreaders, a preliminary screening test was devised and administered to all potential subjects. The item pool consisted of 100 CID Everyday Sentences (Davis and Silverman, 1970) spoken by the male talker. A sample of 40 normal-hearing subjects had previously speechread the sentences without feedback, and their performance was scored in terms of keywords correct per sentence (Demorest and Bernstein, 1987). Item analyses were conducted to identify those sentences whose correlation with the total score on all 100 sentences was greatest. This item-selection strategy maximizes internal consistency reliability. The 20 sentences with the highest item-total correlations were selected as the screening test.[1]

2. Selection and assignment

Nineteen normal-hearing naive adults recruited from the Johns Hopkins University student community were recruited and individually administered the speechreading screening test. Each sentence was shown once, and the subjects were instructed to type on a computer terminal what they thought had been said. Two subjects were rejected because their extremely low keyword identification scores suggested that little data would be obtained from them. Of the remaining pool, subjects were evenly divided into categories of low-, average-, and high-speechreading ability, and five subjects per level were randomly assigned to the five experimental groups (four supplements, plus the visual-only control group). This attempt to equalize ability across groups was only partially successful because two subjects dropped out of the experiment within the first month, and subjects available to replace them had relatively high screening scores. The subjects who completed the experiment comprised 4 males and 11 females who ranged in age from 19–36 years. Subjects were paid for their participation.

D. Procedures

1. General test procedures

An IAC soundproof chamber was used for all testing. To ensure that auditory emissions from the vibrator were not heard, during treatment sessions subjects wore earplugs (E.A.R.), as well as headphones presenting pink noise at a comfortable listening level.

a. Sentences. Responses to sentence material were collected using a standard computer video terminal. To receive a stimulus (past the frozen first frame stage), subjects were required to press the ⟨RETURN⟩ key. After viewing the sentence, they were encouraged to type on the terminal any phrases, words, nonsense words, or word fragments that they had perceived. Feedback, given for all baseline and treatment sentences, was achieved by displaying on the terminal the text of the sentence that had just been viewed, and replaying the stimulus following a second depression of ⟨RETURN⟩.

All sentence responses were manually edited in order to correct misspellings, to expand numerals to their alphabetic spelling, and to ensure that contractions were consistently punctuated. Whenever misspellings were ambiguous or not obvious, no changes were made. Scoring was then carried out automatically by a recursive procedure that determined the maximum number of matching words between the stimulus and response strings, while requiring that word order be maintained. Because any nonalphabetic character served as separator, contractions were counted as two words. In the case of the CID sentences, only keywords were included in the stimulus sentence lists.

b. Nonsense syllables. A touch-sensitive terminal (Fluke Infotouch) was used to collect responses for the 23-alternative forced-choice nonsense-syllable identification task. Each consonant was represented on the screen by its two-character ARPABET (Shoup, 1980) identifier, and the isolated vowel was represented by "·" Subjects had available a key sheet that listed words using each consonant. To register a response, subjects merely pressed one of the 23 alternate phoneme choices. This action additionally served to trigger the next presentation. Feedback was not given.

2. Pre- and post-testing

The pre- and post-tests each consisted of 40 CID sentences and 92 CV syllables repeated randomly in ten blocks. The large number of CV trials was to provide the subjects with a simple visual speech perception task that was predicted to help the subject learn to attend to visual stimulation and to obtain a large sample of results that could be used for modeling speechreading performance. Results of those analyses are not presented here. The CV syllables were administered over several sessions, and feedback was not given. The

CID sentences were chosen from the 80 sentences not part of the screening test, such that both sets were of approximately equal difficulty and internal consistency reliability. No feedback was given for identification of the CID sentences.

3. Baseline and treatment testing

The major portion of the experiment consisted of alternate periods of (visual-alone) baseline and (supplemented) treatment testing. There were five baseline periods and four treatment periods. Table I summarizes the sequence of pre-, post-, treatment, and baseline testing. Feedback was given throughout by displaying the text of each sentence on the video terminal, then giving a second presentation of the visual or visual-vibrotactile stimulus. This feedback-repetition sequence was the only form of training provided.

a. Baseline periods. For each baseline period, 39 sentences spoken by the male and 64 sentences spoken by the female were selected without replacement from the corpus. These were presented to all subjects in the same randomized order, generally in two test sessions per baseline period. The distribution of sentence types and lengths was the same as for the larger treatment corpus. Subjects received each baseline sentence in only one trial. The lexically equivalent sentences, spoken by the other talker, were available for random selection during the treatment sessions.

b. Treatment periods. Four 450-sentence treatment sets were generated randomly for each subject from the remaining thousand sentences in the corpus. Sentences were randomly selected without replacement during a session and with replacement across sessions. Thus it was possible for a subject to see the same token more than once during the course of the experiment. However, due to the distribution of sentences on two different sides of the videodiscs, a given sentence could be presented at most every other treatment session. Due to the large number of sentences sampled during each treatment period, the number of sentences from each talker was allowed to vary randomly.

The entire protocol, including the screening, pre-, post-, baseline, and test periods required approximately 36 sessions. Subjects completed three to four sessions per week. Baseline sessions were approximately 30 min in length, test sessions approximately 45–55 min.

TABLE I. Sequence of events during testing, and breakdown of presented material. Numbers in parentheses are the nominal number of sessions required to complete the material. Only during treatment were subjects given tactile stimuli (visual-only control subjects excepted).

Session	Stimulus materials
Screening	20 CID sentences
Pretest	40 CID sentences seen only once (1)
	92 CV tokens each repeated ten times (3)
Baseline 1	32 sentences spoken by the female talker, each presented only once (1/2)
	39 sentences spoken by the male talker, each presented only once (1/2)
	32 sentences spoken by the female talker, each presented only once (1/2)
	39 sentences spoken by the male talker, same set each time (1/2)
Treatment 1	450 sentences chosen at random (5)
Baseline 2	Same protocol as Baseline 1, unique sentences
Treatment 2	Same protocol as Treatment 1
Baseline 3	Same protocol as Baseline 1, unique sentences
Treatment 3	Same protocol as Treatment 1
Baseline 4	Same protocol as Baseline 1, unique sentences
Treatment 4	Same protocol as Treatment 1
Baseline 5	Same protocol as Baseline 1, unique sentences
Post-test	40 CID sentences seen only once (1)
	92 CV tokens each repeated ten times (3)

4. Repeated-baseline sentences

An additional sentence set was incorporated into the protocol in order to determine the extent to which subjects could improve performance when presented with the same sentences repeatedly. Thirty-nine sentences spoken by the male talker were reserved from the corpus for this purpose. One of five randomizations of these sentences were administered after each baseline period. The procedure followed that of the baseline sets.

E. Data analysis

All data analyses were performed using SPSS^x (1986). For the main experiment, a four-factor analysis of covariance was used. The dependent variable was the number of words correct in a single sentence. In order to take sentence length into account, the number of words and the number of syllables in the sentence were included as covariates.

The factors in the experimental design included group (i.e., vibrotactile supplement or visual-alone control), subject within group (i.e., ability), talker, and test period. All four factors were considered fixed. Sentence was nested within cells and was treated as a random factor. *A priori* contrasts were defined for two factors, group and test period. Because the visual-only control group provided a frame of reference for interpretation of effects due to the vibrotactile supplements, each of the four vibrotactile-supplement groups was compared with the visual-only group. Three sets of contrasts were defined for the nine baseline and treatment periods. First, the four treatment periods were contrasted with the five baseline periods to test for the main effect of vibrotactile information (referred to as the *vibrotactile* contrast). Second, within baseline periods, orthogonal contrasts for trend were defined to test for the form of the performance function over time (referred to as the *baseline trend* contrasts). Finally, a second set of trend contrasts was defined for the treatment periods (referred to as the *treatment trend* contrasts).

II. RESULTS

A. Comparisons of test conditions and groups

Two principal questions addressed in this study were whether performance was superior under aided conditions and whether this effect varied as a function of the supplement. The first question was examined initially in terms of the vibrotactile contrast comparing treatment and baseline test periods. This contrast was highly significant [$F(1,34293) = 35.45$, $p < 0.0005$], with adjusted means (with regard to number of syllables and words per sentence) of 1.77 and 1.62 words per sentence for the visual-vibrotactile and visual-alone conditions, respectively. Expressed in terms of the overall percentage of words correctly perceived, the corresponding unadjusted means for aided and unaided conditions were 29.4% and 26.4%, respectively.

1. Differences among groups

Although the vibrotactile contrast (five baselines versus four treatments) revealed that performance was better during treatment sessions, this contrast averaged across all subjects including visual controls. Of greater interest, therefore, was the significant interaction of supplement with the vibrotactile contrast [$F(4,34293) = 4.37, p < 0.0005$], which implies differential effects of treatment across the five groups. Table II(a) illustrates this interaction. The performance of the five groups across baseline periods and across treatment periods is given in terms of the adjusted mean words correct per sentence; Table II(b) expresses the same results in terms of overall percentage of words correct, i.e., the total number of words correctly identified across all sentences divided by the total number of words presented, in the baseline and treatment periods, respectively. Both tables show that the significant interaction of supplement with the vibrotactile contrast arose in the context of fairly small absolute differences among groups and conditions.

Planned comparisons between the visual-alone group and each of the groups receiving supplements were carried out independently in order to determine whether the supplements were effective. A significant interaction between group (aided versus visual control) and the vibrotactile contrast was the basis for concluding that the supplement improved performance during the treatment periods of the experiment. Significant interactions were obtained for three of the four supplements, LSF [$F(1,13652) = 5.87, p = 0.015$], DSF [$F(1,13627) = 8.45, p = 0.004$], and AMFF [$F(1,13649) = 8.70, p = 0.003$]. The interaction for the LSF + E group was not significant [$F(1,13645) = 0.02, p = 0.888$]. Separate analyses of the data for each aid confirmed that the vibrotactile contrast was significant only for the LSF, DSF, and AMFF groups. Across these three supplements, the mean baseline words correct was 26.8%, and the mean treatment words correct was 31.0%, a difference of 4.2%.

Comparisons among the effective supplements were also carried out. No significant differences were found. Thus use of the three single-channel devices resulted in better performance than in the visual-alone conditions, but there was no effect for the two-channel device and, as expected, no effect in the visual control group.

TABLE II. Group means across baseline and treatment sessions. (a) Results given in mean words correct per sentence. (b) Results given as the percentage of words identified over all words presented.

(a)			Group		
Condition	Visual	LSF	DSF	AMFF	LSF + E
Baseline	1.57	1.44	1.73	1.77	1.63
Treatment	1.58	1.62	1.96	2.01	1.66
(b)			Group		
Condition	Visual	LSF	DSF	AMFF	LSF + E
Baseline	25.7	23.4	28.2	28.7	26.1
Treatment	26.4	26.8	32.7	33.4	27.7

2. Differences among subjects

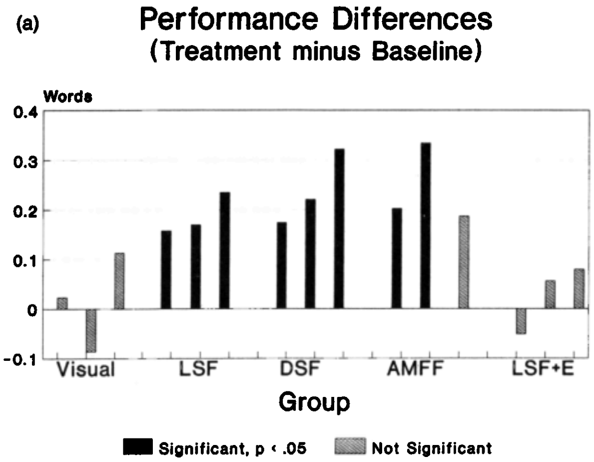
Two questions we may ask are whether subjects varied in performance level across the experiment, and whether these performance levels were related to supplement effects. The main effect of subjects within groups was found to be highly significant, [$F(10,34293) = 86.70, p < 0.0005$], which confirms that there were individual differences in overall level of performance across all sessions of the experiment. Although the screening test had been employed to generate such inter-subject variability, the rank ordering of subjects within groups was not completely consistent with the subjects' screening test scores. In any event, there was no interaction between subjects and the vibrotactile contrast [$F < 1$, *n.s.*]. That is, individual differences did not appear to affect the magnitude of the benefit obtained with the supplements. The difference between adjusted mean words correct per sentence in supplemented and visual-alone conditions was consistent among subjects within each group.

Having utilized a multiple baseline single-subject experimental design, we may also ask whether the vibrotactile contrast was significant for individual subjects. Separate analyses of covariance were performed with test period and talker as the independent variables, and the number of words and syllables in the sentence as the covariates. The vibrotactile contrast was significant for 8 of the 15 subjects. With one exception ($p = 0.075$), all subjects supplemented by a single vibrator performed significantly better in the treatment condition. No significant differences between baseline and treatment were obtained for any control or LSF + E subjects. The upper portion of Fig. 1 summarizes these results by showing the difference, in words correct, between adjusted means for treatment and baseline test periods for each of the 15 subjects. Solid black bars in Fig. 1 indicate subjects for whom the vibrotactile contrast was significant ($p < 0.05$). The lower portion of Fig. 1 presents the results in terms of the percentage difference in words correctly identified during treatment and baseline periods, respectively. These results on individual subjects confirm the reliability of the group analyses. Reliability was an issue of concern in light of the relatively small magnitudes of effects.

3. Trends over time

a. Baseline test periods. In single-subject experiments, alternation of baseline with treatment conditions provides for multiple opportunities to observe the effects of treatment. Ideally, in a multiple-baseline experimental design, baseline measures would reflect only stable levels of speechreading. In fact, trend analyses showed that learning took place during baseline sessions.[2] Across all subjects, adjusted means for the five baseline periods were 1.10, 1.59, 1.89, 1.80, and 1.74 words correct per sentence. The baseline trend analysis permitted examination of this learning in terms of

(a)



(b)

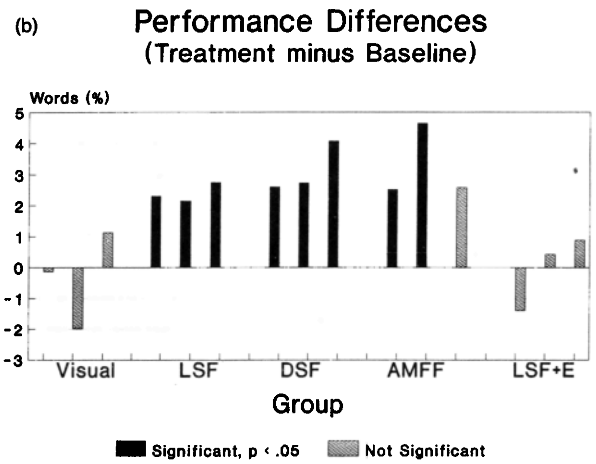


FIG. 1. The difference between (aided) treatment and (unaided) baseline scores for each subject. Solid black bars indicate that the difference was significant at the $p < 0.05$ level, and shaded bars indicate no significant difference: (a) results given in mean words correct per sentence, (b) results given as the percentage of words identified over all words presented.

linear, quadratic, cubic, and quartic trend components. Overall differences among baseline test periods were significant [$F(4,34293) = 43.17, p < 0.0005$], although the exact form of the function varied from subject to subject [$F(40,34293) = 1.52$, $p = 0.019$]. Analysis of baseline trends for individual subjects confirmed significant ($p < 0.05$) linear trends for 12 of the 15 subjects, and significant curvilinear components for 10 subjects.

b. Treatment test periods. Learning occurred across the tactually supplemented treatment periods as well. Trend analyses for treatment periods showed significant change over time [$F(3,34293) = 16.36, p < 0.0005$], varying with group [$F(12,34293) = 3.99$, $p < 0.0005$] and subject [$F(30,34293) = 1.93, p = 0.002$]. Adjusted means for the four treatment periods were 1.41, 1.69, 1.89, and 2.07 words correct. Again, results of individuals were used to confirm group results. All subjects exhibited significant linear trends across treatment periods, and two subjects exhibited significant curvilinear components. These results show that the ability of subjects to recognize words under supplemented conditions climbed steadily during the course of the experiment and, in most cases, subjects did not approach an asymptote in their performance.

4. Talker effects

Built into the experimental design was the opportunity to investigate the role of talker differences under visual-alone and vibrotactile-visual conditions. In previous work (Demorest and Bernstein, 1987), we reported that the female talker on the videodisc was more difficult to speechread than the male. That effect was replicated in the present experiment. The overall covariance analysis showed a significant effect of talker [$F(1,34293) = 381.13, p < 0.0005$]. The magnitude (but not the direction) of the talker effect varied somewhat from group to group [$F(4,34293) = 3.48$, $p = 0.008$], from subject to subject [$F(10,34293) = 9.03$, $p < 0.0005$], and from one test period to the next [$F(8,34293) = 7.60, p < 0.0005$]. Over all conditions of the experiment, the adjusted mean number of words correct was 1.44 for the female talker and 1.94 for the male. The size of the talker effect was larger than any of the differences between aided and unaided conditions (see Table II). This result underscores the need to control talker variables so as to avoid artifactual effects.

Analyses for individual subjects confirmed the talker effect in all 15 subjects, although the magnitude appeared to vary with the subject's overall level of ability. The largest talker effects were found among the subjects with the highest overall levels of performance. For example, the subject with the largest talker effect (1.23 words) had an overall adjusted mean of 2.69 words correct per sentence, whereas the poorest subject (1.02 words correct per sentence) had a talker effect of only 0.26 words.

5. Covariates

The two covariates (sentence length in words and syllables) were significantly related to the dependent variable (number of words correct in the sentence) [$F(2,34293) = 861.00, p < 0.0005$]. The correlations of the covariates with the predicted number of words correct was 0.87 for the number of words in the sentence and 0.48 for the number of syllables.

B. Pre- and post-test effects

One question of interest is whether the baseline and treatment sessions had a general effect on speechreading performance. This question cannot be tested independently of the effects of feedback and repetition given during baselines and treatments but can be evaluated in terms of the visual-only pre- and post-test measures. On CID sentences, there was a significant improvement from pre- to post-tests [$F(1,10) = 35.42$, $p < 0.0005$] but no main effect due to groups and no interactions (both F values < 1). Thus the experience during treatment periods had no differential effect on speechreading-alone learning. The pre-test mean was 24.9% keywords correct vs 35.3% keywords correct for the post-test.

It should be noted that this pre–post difference was approximately equal in magnitude to the gains across baseline sessions. The percentage words correct across subjects in the first baseline (with the male talker) was 21.8% and in the last baseline was 37.5%. These results provide some verification that the difficulty of the sentences generated for the Johns Hopkins Lipreading Corpus (Bernstein and Eberhardt, 1986a,b) is not greatly different from that of the CID sentences.

Identification scores from the CV task were also greater in the post-test. The overall pre-test and post-test CV means were 32.7% and 36.4%, respectively ($F(1,10) = 65.02$, $p < 0.0005$]. There was no overall effect of group but there was an interaction of group with pre/post, attributable to large gains made by some individual subjects. With the exception of one subject, who did not improve, the range of improvement was 0.7% to 8.2%.

Individual differences in speechreading performance were maintained across the interval between the pre- and post-tests. The pre-test/post-test correlation was 0.85 ($p < 0.0005$) for the CID sentences and 0.76 ($p = 0.001$) for the CV syllables. Interesting, however, were the low correlations between CV and CID materials. The correlation between CV pre-test and CID pre-test was 0.26 (n.s.). The correlation between CV post-test and CID post-test was 0.35 (n.s.). This implies that CV recognition performance is not a good estimator of visual word recognition in sentences and vice versa. It should be noted that several other investigators have performed similar correlations between sentence and nonsense syllable scores with differing results (De Filippo, 1984; Walden *et al.*, 1981). In any event, the correlations obtained here should be regarded with some caution due to the relatively small sample size.

C. Repeated-sentence procedure

As noted earlier, 39 sentences were reserved for presentation after each baseline session. The purpose of this procedure was to obtain an estimate of the benefit to performance of repeated viewing of sentences. Results are shown in Fig. 2. Averaged across subjects, the responses to the 39 sentences spoken by the male talker were compared with responses to his unique baseline sentences. Analysis of variance showed a significant interaction between baseline period and sentence group (unique versus repeated) [$F(4,56) = 28.95$, $p < 0.0005$]. Although performance on the unique baseline sentences improved somewhat across the experiment, that improvement was small compared with the increase obtained for the repeated sentences. For the first presentation of the repeated sentence list, subjects recognized 25.4% of the stimulus words, a score similar to that for the first unique baseline sentence list spoken by the male (21.8%). For the second and third baseline sets, scores for the repeated sentences are somewhat higher than the unique sentences, but, by the fifth presentation of the repeated sentences, the mean was much higher (64.4%) than that for the fifth unique sentence set (37.5%).[3]

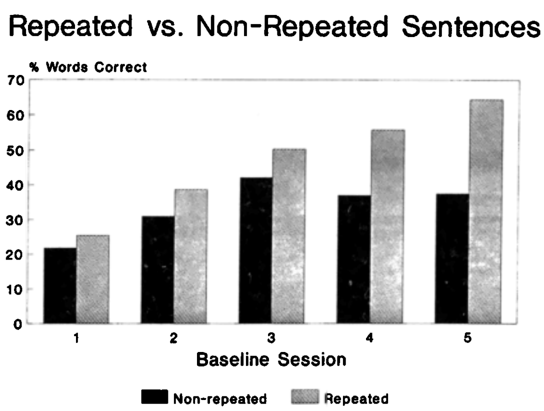


FIG. 2. Comparison of (unaided) baseline scores for the 39 sentences that were seen only once (black bars) and the 39 sentences that were repeated (with different randomizations) each baseline set (shaded bars). Only the sentences spoken by the male talker were included. Results are given as percentage of words identified over total words presented during each of the five sets.

III. DISCUSSION

A. Comparisons among aids

Three of the four supplements (i.e., log-scaled frequency, difference-scaled frequency, and amplitude-modulated fixed-frequency) significantly improved speechreading over vision alone, but there was no difference in performance among the effective supplements. Further, the magnitude of the improvements was small, 4.2% words correctly identified. It is necessary to consider the practical implications of this level of experimental benefit.

The magnitude of the effects must be considered in light of the task with which subjects were confronted and the amount of training they received. The speechreading task in this study was difficult, as evidenced by the mean baseline performance of only 27.71% words correct. Speechreading of isolated sentences given without context is generally very difficult. For example, Breeuwer and Plomp (1986), using hearing subjects and stimulus materials similar to ours, obtained 10.6% mean correct syllables with unsupplemented speechreading. Hanin *et al.* (1988) found gains due to their supplements that were comparable in magnitude to those obtained here. Hearing subjects were tested with either of two vibrotactile supplements that provided $F0$, one via a single channel and one via an array. Mean speechreading-alone performance was 25.38% words correct for isolated sentences whose topic was randomly selected from ten items known to the subjects. The mean difference between visual-alone and vibrotactile-visual performance was 4.65% words, across a total of 576 sentences.

Hanin *et al.* (1988) also report a second study in which three hearing-impaired subjects received videorecorded sentences from a story. The difference between supplemented and visual-alone performance ranged between 11% and 20% words correct. Their results suggest that relatively small absolute benefit observed under difficult conditions such as speechreading of isolated sentences may translate into substantially larger effects under more favorable conditions with more practiced speechreaders. Also, it should be noted that the trend analyses performed for the current study showed that subjects were still improving at the time of the final treatment, suggesting that with further training and experience greater benefit could be demonstrated.

Examination of results for differences among supplements failed to provide evidence that any one of the single-channel transformations was significantly more effective than the others. A similar result was found in Bernstein *et al.* (1989), where significant differences arose only when the visual stimulus was not given. In that report, it was speculated that failure to obtain significant differences in the vibrotactile-visual condition could be attributed to subjects' attending primarily to the visual stimulus. We may ask whether the current study provides indication of a prepotency of the visual stimulus. A comparison between the magnitude of the greatest gain due to a supplement (AMFF), and the overall improvement in speechreading shows that speechreading gains were 2.7 times greater, expressed as words correct per sentence. While this effect may be attributed to visual prepotency, another likely possibility is that vibrotactile learning is slower than visual learning. In any case, had the experiment continued until each subject's performance asymptoted, the benefit of the supplement relative to visual learning might have been greater, and differences among the transformations might have been observed.

An unexpected finding was that the performance of subjects who received the two-channel, LSF + E supplement was indistinguishable from that of the visual control subjects. Considering that the logarithmic scaled-frequency supplement was effective in improving speechreading, it is puzzling that, when a second source of speech information was available, none of the three LSF + E subjects was able to obtain any benefit. While the question of how multiple sources of vibrotactile patterns are processed perceptually is an ongoing area of research, several explanations may be offered for subjects' inability to use the vibrotactile stimuli. One possibility is that the bilateral placement of the stimulators may have resulted in perceptual masking. However, as several studies have demonstrated that masking effects decrease with distance between stimulators (cf. Snyder, 1977; Sherrick, 1964), masking did not likely cause the observed effect.

A second possibility is that integration of information from distant body loci may be more difficult or require a longer course of learning than single or closely located points. A study by Craig (1985) suggested that vibrotactile pattern perception has a fairly long time course that is lengthened when the task is to integrate stimuli from two different fingers. Interestingly, however, classification of patterns using fingers of two different hands is superior to using two fingers of the same hand. Indeed, integration of tactile information from the two hands is not unusual; Gibson (1966, p. 125) noted that subjects tactually exploring unseen objects will generally curl all ten fingers around the object. Results from studies by Geldard and Sherrick (1965) and Gilson (1968) suggest that subjects can attend simultaneously to stimuli presented to as many as ten separate body locations and as many as ten sites on the fingers of both hands, with very few errors associated with two points of stimulation. These findings suggest that bilateral placement *per se* may not be the limiting factor for the two-channel supplement tested here.

Another possibility is that the particular stimuli used for the second channel interfered with use of the first. Phenomenologically, the $F0$ stimulus was intermittent but rhythmically related to the ongoing visual stimulus. The high-frequency indicator was quite different, signaling briefly and unpredictably. The task may simply have been too complex for subjects to learn in the absence of specific individual training in integrating the visual channel with each of the two sources of vibrotactile information. The unexpectedness of our results is a direct and good example of a general problem for developers of vibrotactile devices, viz., that not enough is known about human vibrotactile perception to predict with great accuracy the outcome of new device designs.

B. Methodological considerations

The repeated-sentences experiment, in which the same 39 sentences were presented with different randomizations after each baseline session, was intended to probe whether subjects would learn sentences presented repeatedly in the context of a large set of randomly selected sentences. When compared with the nonrepeated baseline sentences, subjects did indeed exhibit substantially higher scores by the fourth and fifth viewing. The concern that subjects might have learned treatment sentences, some of which were seen four or more times, was not borne out; no differences were found in treatment versus (nonrepeated) baseline sentence scores for any of the visual or two-vibrator, LSF + E subjects. This result implies that limited lists of sentences may be learned if they are presented multiple times, even if different randomizations are used each time and a large amount of other material intervenes between presentations. Sentences that are repeated in an unpredictable fashion amid a much larger stimulus set, on the other hand, are less likely to be learned. These results underscore the importance of using a sizable sentence corpus to minimize effects of learning the specific stimuli.

One goal for studies of the kind reported here is to estimate the possible benefit to deaf individuals of being fitted with a wearable vibrotactile supplement. The results of this experiment suggest that normal-hearing subjects require long periods of laboratory experience to achieve stable speechreading-alone scores and that vibrotactile learning proceeds slowly. The relatively slow rate of learning has been reported almost universally in the literature on vibrotactile supplements. Speechreading improvement was evident from the pre- and post-test results, the difference between first and last baselines, and the changes across treatment sessions. Subjects improved also in their ability to identify CV syllables. It is possible, therefore, that, by employing normally hearing subjects, estimates of benefit are obtained less efficiently and may ultimately be less accurate than might be achieved with proficient deaf speechreaders.

On the other hand, more recent work (in progress) involving deaf speechreaders suggests that visual learning occurs even with individuals who depend primarily on the visual channel for speech communication. Therefore, as long as speechreading continues to improve, and vibrotactile learning progresses slowly, it will remain difficult to estimate maximum benefit with these supplements. Accurate estimates of benefit may require study of long-term users of the tactile supplement. Ironically, development of wearable vibrotactile devices has, for the most part, awaited definitive estimates of benefit from laboratory experiments (cf. Sherrick, 1984).

IV. CONCLUSIONS

(1) Single-channel presentation of $F0$ is a viable method for aiding speechreading. Modest but significant gains were obtained when subjects used the log-scaled frequency, difference-scaled frequency, and amplitude-modulated fixed-frequency transformations.

(2) Presentation of several vibrotactile information channels (possibly at widely separated body loci) may impede the progress of subjects in learning to use this information. None of the three subjects who held a second vibrator presenting high-frequency energy was able to use any of the vibrotactile signals.

(3) When subjects are repeatedly presented with the same group of closed-set sentences, it must be expected that previous viewings will tend to increase recognition, even with hundreds of intervening sentences. Sentences that are unpredictably repeated amid a large corpus of sentences, however, are less likely to be learned.

(4) When hearing subjects are employed in sensory aids research, it is necessary to control for their learning not only the vibrotactile signal, but also the visual speech stimulus.

ACKNOWLEDGMENTS

This research was supported by NINCDS (NS22183 and NS01206) and NIDCD (DC00695). We gratefully thank Dr. Allen Montgomery for useful discussions concerning the design of this experiment and for suggesting the repeated-baseline sentences experiment, and Dr. Andreas Andreou for designing, constructing, and maintaining the vibrotactile presentation system.

[1]CID sentences included in the screening test were: 4, 6, 10, 15, 16, 17, 26, 28, 33, 35, 41, 44, 67, 69, 70, 71, 74, 80, 88, 93.

[2]Since the sentences used for baseline testing were selected by a somewhat different method than that used for sampling treatment sentences, it was important to verify that the distributions of sentences, in terms of length (in words or syllables), were similar during baseline and treatment test periods. Chi square tests confirmed that each subject received, in baseline and treatment conditions, sentences that were similarly distributed in terms of length, measured either in words or syllables.

[3]Since sentences were randomly selected for presentation during treatment sessions, it was of interest to determine the rate at which they were repeated. The number of times that each sentence was presented to each subject was tabulated. Because sentences were selected randomly, each subject never saw, on the average, 142 of the approximately 1000 treatment sentences. Subjects saw fewer than 26% of the sentences more than twice, and fewer than 8% of the sentences four or more times. Of greater importance was the result that the distribution of repeated sentences did not vary among subjects.

Bernstein, L. E., and Eberhardt, S. P. (**1986a**). Johns Hopkins Lipreading Corpus I-II: Disc I (The Johns Hopkins University, Baltimore, MD).

Bernstein, L. E., and Eberhardt, S. P. (**1986b**). Johns Hopkins Lipreading Corpus III-IV: Disc II (The Johns Hopkins University, Baltimore, MD).

Bernstein, L. E., Eberhardt, S. P., and Demorest, M. E. (**1989**). "Single-channel vibrotactile supplements to visual perception of intonation and stress," J. Acoust. Soc. Am. **85**, 397-405.

Bernstein, L. E., Schechter, M. B., and Goldstein, M. H., Jr. (**1986**). "Child and adult vibrotactile thresholds for sinusoidal and pulsatile stimuli," J. Acoust. Soc. Am. **80**, 118–123.

Boothroyd, A., and Hnath, T. (**1986**). "Lipreading with tactile supplements," J. Rehab. Res. Dev. **23**, 139–146.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.

Breeuwer, A., and Plomp, R. (**1984**). "Speechreading supplemented with frequency-selective sound-pressure information," J. Acoust. Soc. Am. **76**, 686–691.

Breeuwer, M., and Plomp, R. (**1985**). "Speechreading supplemented with formant-frequency information from voiced speech," J. Acoust. Soc. Am. **77**, 314–317.

Breeuwer, M., and Plomp, R. (**1986**). "Speechreading supplemented with auditorily presented speech parameters," J. Acoust. Soc. Am. **79**, 481–499.

Cowan, R. S. C., Alcantara, J. I., Blamey, P. J., and Clark, G. M. (**1988**). "Preliminary evaluation of a multichannel electrotactile speech processor," J. Acoust. Soc. Am. **83**, 2328–2338.

Craig, J. C. (**1985**). "Attending to two fingers: Two hands are better than one," Percept. Psychophys. **38**, 496–511.

Davis, H., and Silverman, S. R. (**1970**). *Hearing and Deafness* (Holt, Rinehart, and Winston, New York).

De Filippo, C. L. (**1984**). "Laboratory projects in tactile aids to lipreading," Ear Hear. **5**, 211–227.

De Filippo, C. L., and Scott, B. L. (**1978**). "A method for training and evaluating the reception of ongoing speech," J. Acoust. Soc. Am. **63**, 1186–1192.

Demorest, M. E., and Bernstein, L. E. (**1987**). "Reliability of individual differences in lipreading," J. Acoust. Soc. Am. Suppl. 1 **82**, S24.

Denes, P. B. (**1963**). "On the statistics of spoken English," J. Acoust. Soc. Am. **35**, 892–904.

Eberhardt, S.P. (**1987**). "Vibrotactile presentation of voice fundamental frequency to aid lipreading," unpublished doctoral dissertation, The Johns Hopkins University, Baltimore, MD.

Geldard, F. A., and Sherrick, C. E. (**1965**). "Multiple cutaneous stimulation: The discrimination of vibratory patterns," J. Acoust. Soc. Am. **37**, 797–801.

Gibson, J. J. (**1966**). *The Senses Considered as Perceptual Systems* (Houghton Mifflin, Boston, MA).

Gilson, R. D. (**1968**). "Some factors affecting the spatial discrimination of vibrotactile patterns," Percept. Psychophys. **3**, 131–136.

Grant, K. W. (**1980**). "Investigating a tactile speechreading aid: The transmission of prosodic information in connected discourse and sentences," unpublished Master's thesis, University of Washington, Seattle, WA.

Grant, K. W. (**1987**). "Encoding voice pitch for profoundly hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 423–431.

Hanin, L., Boothroyd, A., and Hnath-Chisolm, T. (**1988**). "Tactile presentation of voice fundamental frequency as an aid to the speechreading of sentences," Ear Hear. **9**, 335–341.

Hnath-Chisolm, T., and Kishon-Rabin, L. (**1988**). "Tactile presentation of voice fundamental frequency as an aid to the perception of speech pattern contrasts," Ear Hear. **9**, 329–334.

Hochberg, I., Rosen, S., and Bell, V. (**1989**). "Effect of text complexity on connected discourse tracking rate," Ear Hear. **10**, 192–199.

Kearns, K. P. (**1986**). "Flexibility of single-subject designs. Part II: Design selection and arrangement of experimental phases," J. Speech Hear. Disord. **51**, 204–214.

Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Lisker, L., and Abramson, A. (**1967**). "Some effects of context on voice onset time in English stops," Lang. Speech **10**, 1–28.

Martony, J. (**1974**). "Some experiments with electronic speechreading aids," Speech Transmission Lab. Q. Prog. Stat. Rep. **2-3**, 34–56.

Matthies, M. L., and Carney, A. E. (**1988**). "A modified speech tracking procedure as a communicative performance measure," J. Speech Hear. Res. **31**, 394–404.

Pickett, J. M. (**1983**). "Theoretical considerations in testing speech perception through electroauditory stimulation," in *Cochlear Prostheses: An International Symposium*, edited by C.W. Parkins and S. W. Anderson; Ann. N.Y. Acad. Sciences **405**, 424–405.

Plant, G. (**1986**). "A single-transducer vibrotactile aid to lipreading," Speech Transmission Lab. Q. Prog. Stat. Rep. **1**, 41–63.

Plant, G., and Risberg, A. (**1983**). "The transmission of fundamental frequency variations via a single channel vibrotactile aid," Speech Transmission Lab. Q. Prog. Stat. Rep. **2-3**, 61–84.

Plant, G., and Spens, K-E. (**1986**). "An experienced user of tactile information as a supplement to lipreading. An evaluative study," Speech Transmission Lab. Q. Prog. Stat. Rep. **1**, 87–110.

Risberg, A., and Lubker, J. (**1978**). "Prosody and Speechreading," Speech Transmission Lab. Q. Prog. Stat. Rep. **4**, 1–16.

Robbins, A. M., Osberger, M. J., Miyamoto, R. T., Kienle, M. L., and Myres, W.A. (**1985**). "Speech-tracking performance in single-channel cochlear implant subjects," J. Speech Hear. Res. **28**, 565–578.

Rosen, S. M., Fourcin, A. J., and Moore, B. C. J. (**1981**). "Voice pitch as an aid to lipreading," Nature **291**, 150–152.

Rothenberg, M., and Molitor, R. D. (**1979**). "Encoding voice fundamental frequency into vibrotactile frequency," J. Acoust. Soc. Am. **66**, 1029–1038.

Rothenberg, M., Verillo, R. T., Zahorian, S. A., Brachman, M. L., and Bolanowski, S. J. (**1977**). "Vibrotactile frequency for encoding a speech parameter," J. Acoust. Soc. Am. **62**, 1003–1012.

Sherrick, C. E., Jr. (**1964**). "Effects of double stimulation of the skin," Am. J. Psych. **77**, 42–53.

Sherrick, C. E. (**1984**). "Basic and applied research on tactile aids for deaf people: Progress and prospects," J. Acoust. Soc. Am. **75**, 1325–1342.

Shoup, J. (**1980**). *Trends in Speech Recognition*, edited by W.A. Lea (Prentice Hall, Englewood Cliffs, NJ).

Snyder, R. E. (**1977**). "Vibrotactile masking: A comparison of psychophysical procedures," Percept. Psychophys. **22**, 471–475.

SPSS[x]. (**1986**). *SPSS[x] User's Guide* (McGraw-Hill, New York).

Thorndike, E. L., and Lorge, I. (**1944**). *The Teacher's Word Book of 30,000 Words* (Teacher's College, Columbia University, New York).

Tye-Murray, N., and Tyler, R. S. (**1988**). "A critique of continuous discourse as a test procedure," J. Speech Hear. Disord. **53**, 226–231.

Upton, H. W. (**1986**). "Wearable eyeglass speechreading aid," Am. Ann. Deaf **113**, 222–229.

Verrillo, R. (**1962**). "Investigation of some parameters of the cutaneous threshold for vibration," J. Acoust. Soc. Am. **34**, 1768–1773.

Walden, B. E., Erdman, S., Montgomery, A. A., Schwartz, D. M., and Prosek, R. A. (**1981**). "Some effects of training on speech recognition by hearing-impaired adults," J. Speech Hear. Res. **24**, 207–216.