Enhanced Speechreading in Deaf Adults: Can Short-Term Training/Practice Close the Gap for Hearing Adults?

Lynne E. Bernstein
Edward T. Auer, Jr.
Paula E. Tucker
Center for Auditory and Speech Sciences
Gallaudet University
Washington, DC

This study investigated effects of short-term training/practice on group and individual differences in deaf and hearing speechreaders. In two experiments, participants speechread sentences with feedback during training and without feedback during testing, alternating 10 times over six sessions spanning up to 5 weeks. Testing used sentence sets balanced for expected mean performance. In each experiment, participants were adults who reported good speechreading and either normal hearing (n = 8) or severe to profound hearing impairments (n = 8). The experiments were replicates, except that in one participants received vibrotactile speech stimuli in addition to visible speech during training, testing whether vibrotactile speech enhances speechreading learning. Results showed that (a) training/practice did not alter the relative performance among individuals or groups; (b) significant learning occurred when training and testing were conducted with speechreading only (although the magnitude of the effect was small); and (c) there was evidence that the vibrotactile training depressed rather than raised speechreading scores over the training period.

KEY WORDS: speechread, lipread, learning, deaf, hearing, vibrotactile

Several fundamental questions have been posed in the speechreading literature concerning effects of short-term training/practice and long-term perceptual experience on accuracy of speechreading. For example, can lifelong reliance on speechreading associated with early-onset deafness[1] result in enhanced speechreading accuracy? Can short-term training/practice result in significant learning? Does the method of training/practice affect speechreading accuracy? Previous investigators have concluded that the necessity to rely on visible speech as a consequence of deafness *does not* result in any experientially based advantage for visual speech perception among deaf as opposed to hearing individuals (Clouser, 1977; Conrad, 1977; Green, 1998; Lyxell & Rönnberg, 1991a, 1991b; Massaro, 1987; Mogford, 1987; Owens & Blazek, 1985; Rönnberg, 1995; Summerfield, 1991; cf., Pelson & Prather, 1974; Tillberg, Rönnberg, Svärd, & Ahlner, 1996). Furthermore, throughout the twentieth century, the clinical and educational

[1] In this paper, the term *deaf* refers to individuals with bilateral severe to profound hearing impairments. The term is not intended to connote those who consider themselves members of deaf culture. However, the extension of the term *deaf* to include individuals whose hearing impairment is severe is consistent with usage in the cultural deaf community.

experience involving speechreading training in children has been understood to suggest that training is not very effective in reliably producing outstanding speechreaders among young deaf children (e.g., Heider & Heider, 1940; Jeffers & Barley, 1971). An opinion held by many today is that, in both deaf and hearing populations, speechreading varies widely across individuals and that individuals' relative ability among other speechreaders is primarily an inborn trait, not one that can be affected greatly by either explicit training or need to rely on visible speech for communication.

Bernstein, Demorest, and Tucker (2000) conducted a normative study of speechreading in deaf and hearing adults. They found within their sample a large group of deaf adults with early-onset deafness who were consistently superior to all of the hearing adults in the study across a range of speechreading materials and measures. Approximately 25% of the deaf participants outperformed 100% of the hearing participants. Bernstein et al. interpreted their results as evidence that some combination of reliance on speechreading, long-term severe to profound hearing impairment, and natural talent can result in enhanced speechreading in a subset of deaf individuals. However, several questions could be raised by this interpretation. First, was the deaf group's superiority attributable to their being asked to do a task that was for them simply more familiar (as opposed to a true difference in speech perception capability)? Second, could training/practice result in substantial changes in performance in either or both of the two groups? These questions were investigated in the current study, in which short-term training/practice was given to deaf and hearing participants.

In this paper, *training* refers to explicit procedures intended to enhance performance. *Learning* refers to demonstrated enhancements in performance. The combination "training/practice" acknowledges the possibility that when learning is associated with a particular training method, the learning may not necessarily be due to the training method per se, but to something else (unidentified) that might be no more than the opportunity to repeatedly perform (i.e., practice) the speechreading task.

Effects of Short-Term Training/Practice on Speechreading

Under the implicit or explicit hypothesis that training on phoneme identification might enhance speechreading, several studies have given training with nonsense syllables (e.g., Gesi, Massaro, & Cohen, 1992; Massaro, Cohen, & Gesi, 1993; Walden, Erdman, Montgomery, Schwartz, & Prosek, 1981; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). Improvements in phoneme identification have been demonstrated in participants with impaired hearing (Walden et al., 1977; Walden et al., 1981) and normal hearing (Gesi et al., 1992; Massaro et al., 1993). These studies did not compare deaf versus hearing participants.

Whether the training with nonsense syllables generalized to sentences was investigated by Massaro et al. (1993), whose training and testing of hearing adults involved CV syllables, monosyllabic words, and sentences presented seven times across the experiment: once at the beginning, once after each of five courses of training, and once after a retention period of 7.5 weeks. The same stimuli were observed under conditions of audio alone, video alone, and audiovisual presentation. The sentence tests showed approximately 20 percentage point increases between the first and second sessions and little change after that. However, the 96 sentences were presented under all conditions and were almost perfectly intelligible under the audiovisual condition. Improved scores could have been due to remembering items from the high-intelligibility conditions.

Speechreading in Studies of Vibrotactile Speech Aid Training

The focus for the current study was speechreading of sentences. A substantial source of contemporary information about training to speechread sentences is in the literature on training with a vibrotactile aid. (Comprehensive reviews on vibrotactile speech aids can be found in Plant & Spens, 1995, and Summers, 1992.) Evidence that improvements in speechreading are associated with vibrotactile speech aid training was incorporated into the current study.

Because researchers have observed that learning vibrotactile speech stimuli is a slow process, in many of the studies adults were trained across a lengthy series of aided and unaided sessions/conditions. Novel isolated sentence stimuli have frequently been used. Speechreading *without* the vibrotactile aid has been shown to improve for isolated sentence stimuli in participants who received training with the vibrotactile aid (e.g., Bernstein et al., 1991; Boothroyd & Hnath-Chisolm, 1988; Boothroyd, Kishon-Rabin, & Waldstein, 1995; Eberhardt, Bernstein, Demorest, & Goldstein, 1990; Kishon-Rabin, Boothroyd, & Hanin, 1996; also see results for studies that used continuous discourse tracking: De Filippo, 1984; De Filippo & Scott, 1978; Weisenberger & Broadstone, 1989).

Typically, vibrotactile aid studies have not included controls who did not receive the vibrotactile aid training. One exception was Eberhardt et al. (1990), in which four different F0 vibrotactile aids were studied. Groups of three hearing adults were assigned to each of the F0 conditions or to a speechreading-only control condition. Training and testing with the vibrotactile aid involved five alternations (totaling approximately 35 hours) of aided and unaided speechreading. Controls never received vibrotactile stimuli but received the otherwise identical protocol. Statistical tests confirmed learning in the unaided test conditions for the group and, individually, in 12 out of the 15 participants, including participants in the control condition.

In another study that used unaided controls, Bernstein et al. (1991) tested three multichannel vibrotactile aids that provided spectral speech information. Adult participants were prescreened to be average or better speechreaders. Performance improved during testing and training across participants. Two deaf participants gained approximately 20 percentage points in speechreading alone. Gains were smaller in all the other participants.

Literature Summary

The literature shows that training/practice is associated with improvements in performance, and some results suggest differential effects in deaf versus hearing participants. The vibrotactile speech aid literature shows that training with a vibrotactile speech aid is associated with improved unaided speechreading performance; and in studies with visual-only controls, controls also demonstrated improved speechreading. On the time scale of all the vibrotactile aid training studies, which were relatively long (up to approximately one year), improvements in scores on sentences ranged between approximately 10 and 30 percentage points, with the largest gains observed in deaf participants who were in the aided conditions of the studies.

The Current Study

Given our previous study (Bernstein et al., 2000), we were particularly interested in whether short-term training/practice speechreading would demonstrate learning, and whether it would differentially affect deaf versus hearing speechreaders. Given the observations we and others made of gains in speechreading associated with training on a vibrotactile aid, two different training experiments were conducted. In Experiment I, a speechreading-only (SO) experiment, participants in training and testing speechread sentences *without* auditory or vibrotactile stimuli. In Experiment II, participants received vibrotactile speech stimuli during speechreading training (S+V), and *only* during training, and were tested via speechreading only. In each experiment, half the participants were deaf and half were hearing adults. The main questions investigated were (a) Is training/practice differentially effective across deaf versus hearing groups? (b) Does short-term training affect the relative standing of the individual speechreaders? (c) Does short-term training on sentence stimuli result in significant learning? (d) Does training with a vibrotactile speech aid enhance learning?

Method

Participants

College-educated adults with early-onset (before age 3) severe or profound hearing impairments and English as a first language were recruited for comparison with college-educated adults with normal hearing. The participants reported that they were good speechreaders, but they were not tested on their speechreading before the study.[2]

Deaf Participants

Eight deaf participants, a different group of eight for each experiment, were screened to have the following characteristics: (a) be between 18 and 40 years old; (b) be a Gallaudet University student; (c) have better pure tone average threshold for 500, 1000, and 2000 Hz greater than 80 dB HL; (d) self-report no disability other than hearing impairment; (e) self-report use of spoken English as the primary language of the family; (f) self-report English (including signs presented with

[2] The argument might be put forward that the appropriate design for this study would have been to use select participants randomly. In Bernstein et al. (2000), 96 hearing and 72 adults with impaired hearing were randomly selected in terms of speechreading ability, resulting in the performance advantage for deaf adults that motivated the current study. Therefore, statistically, the expected means for performance measures with random selection would be ones that favor deaf over hearing participants. By requiring self-report of good speechreading in both participant groups in the current study, the possibility was that no group difference would be obtained—a more conservative approach than random selection for the current investigation. An alternate design would have been to match groups on initial speechreading screening scores. Based on the expected values, matching would have likely required eliminating volunteers who were excellent deaf speechreaders—the actual focus of interest.

Another possible approach would have been to obtain a different sample of participants based on some other set of screening criteria. We chose the criteria of the current study based on previous results obtained in Bernstein, Demorest, and Tucker (1998), in which correlations were presented between speechreading measures (percent correct identification of nonsense syllables, words in isolation, and words in sentences) and audiological variables for the participants in Bernstein et al. (2000). In that study, the following variables did *not* correlate with speech perception measures: (a) age loss occurred, (b) age loss discovered, (c) medical or surgical treatment for loss (d) age first hearing aid obtained, (e) presently owns a hearing aid, (f) anacusic, and (g) better pure tone average. The variable *years since last used hearing aid* correlated with some measures, only at around .3. The variable *frequency of hearing aid use* correlated at approximately .35 with the various visual speech perception measures (but the sign changed across measures). These latter correlations, accounting for approximately 10% of the variance, should not be regarded as highly informative. Based on these results, it would be difficult to predict a priori the speechreading score of a particular deaf individual. The selection criteria, therefore, followed those of the Bernstein et al. (2000) study that motivated the current one.

English syntax) as the participant's native language; (g) have been educated in a mainstream and/or oral program for 8 or more years; (h) self-report of good speechreading skills; and (i) vision at least 20/30 in each eye. Participants were paid by the hour.

Table 1 gives audiological information about the participants in each of the experiments and shows the number of years hearing aids were used and the age at which they were first used. The table also shows etiology of the hearing impairment. It was discovered after testing in the speechreading-only (SO) experiment that two of the participants had a better pure tone average threshold of slightly less than 80 dB HL. These participants were not replaced because the deviations (78 and 77 dB HL) from the 80 dB HL criterion were not deemed likely to have had a significant effect on the results. Age at onset for participants in the SO experiment was before 20 months. Seven of the participants received hearing aids by age 3, and one participant received a hearing aid at age 6. Only half were still using them at the time of the study. In the S+V experiment, 7 participants listed birth as the age at onset of hearing impairment. One participant listed age of onset as 30 months. Six participants received their hearing aids at age 3 or older, and 5 out of the 8 were still using their aid at the time of the study.

Participants With Normal Hearing

Eight participants with normal hearing were recruited from the Gallaudet University community for each experiment. They met the same requirements as the deaf participants, except for the requirement of normal hearing. Ages of participants in the SO experiment ranged between 23 and 39 years. Ages of participants in the S+V experiment ranged between 23 and 33 years. Participants were paid by the hour.

Materials

Table 2 lists the materials for training and testing in order of their presentation, including which of the lists were counterbalanced. Sentence stimuli (Bernstein & Eberhardt, 1986a, 1986b) were produced by a male and a female who spoke General American English. The B-E (Bernstein-Eberhardt) Sentences (male talker) and CID Everyday Sentences (Davis & Silverman, 1970) were sorted into lists with equal expected means based on previous results (Bernstein et al., 2000). The B-E Sentences (female talker) comprised four lists of 25 sentences, divided between pre- and posttest occasions. Two of the counterbalanced lists comprised sentences for which normative data were not available for computing expected means.

Training sentences were assigned randomly to lists, as there was not a previous database of responses upon which to calculate expected means. Participants never saw the same stimulus twice in the experiment. A 20-sentence practice list presented first to participants was constructed using sentences spoken by a third talker seen only during practice (Bernstein, 1991). Vocabulary

Table 1. Deaf participant information for SO and S+V experiments.

ID	Age (years;months)	Age at onset (months)	Etiology	L Ear PTA dB HL	R Ear PTA dB HL	HA use now?	Age HA first used (months)	Length of HA use (years)
				SO Experiment				
SP01	30;10	birth	unknown	98	108	no	30	11.5
SP03	28;2	birth	unknown	78	102	yes	12	27
SP04	21;4	birth	unknown	110 (1 Freq)	110 (1 Freq)	no	36	13
SP05	23;3	20	unknown	97	120 (1 Freq)	no	20	5
SP06	21;0	birth	unknown	103	108	yes	20	20
SP07	24;0	birth	genetic	77	83	yes	24	22
SP08	22;7	18	spinal meningitis	107	120 (1 Freq)	no	72	12
SP10	22;6	6	maternal rubella	85	88	yes	24	20
				S+V Experiment				
TSD01	27;1	birth	renal tubular acidosis	100	85	no	120	15
TSD02	23;3	30	spinal meningitis	NMH*	NMH	no	36	11
TSD03	18;11	birth	maternal rubella	85	88	yes	12	18
TSD04	23;3	birth	hereditary	98	95	no	36	15
TSD05	27;11	birth	unknown	90	90	yes	72	21
TSD06	25;6	birth	maternal rubella	102	85 (1 Freq)	yes	16	24
TSD07	28;6	birth	maternal rubella	95	83	yes	48	24
TSD08	21;6	birth	unknown	98	85	yes	36	18

Note. NMH = no measurable hearing

Table 2. Number of sentences per training and testing session.

Session	Materials type	B-E Sentences, female	CID Sentences, male	B-E Sentences, male
1	Pretest	50 (2 lists counterbalanced, pre- vs. posttest)	10 (List A) (Counterbalanced with List J, pre- vs. posttest)	5 (List A) (Counterbalanced with List J, pre- vs. posttest)
2	Training #1			66 (List 1)
	Test #2		10 (List B)	5 (List B)
	Training #2			66 (List 2)
	Test #3		10 (List C)	5 (List C)
3	Training #3			66 (List 3)
	Test #4		10 (List D)	5 (List D)
	Training #4			66 (List 4)
	Test #5		10 (List E)	5 (List E)
4	Training #5			66 (List 5)
	Test #6		10 (List F)	5 (List F)
	Training #6			66 (List 6)
	Test #7		10 (List G)	5 (List G)
5	Training #7			66 (List 7)
	Test #8		10 (List H)	5 (List H)
	Training #8			66 (List 8)
	Test #9		10 (List I)	5 (List I)
6	Training #9			66 (List 9)
	Posttest	50 (2 lists counterbalanced pre- vs. posttest)	10 (List J)	5 (List J)

in the CID Sentences and in the B-E Sentences consisted of high-frequency English words.

Vocabulary Test

The Peabody Picture Vocabulary Test–Revised (PPVT-R; Dunn & Dunn, 1981) was administered to all of the participants. Scores on the PPVT were obtained to investigate the possibility that vocabulary knowledge interacts with speechreading learning. The test is a four-alternative, forced-choice test of receptive vocabulary with a standardized method for selecting starting and stopping points in the list of age-normed test words. This test is normed for oral presentation of words. As we have done in other studies (e.g., Auer, Bernstein, & Tucker, 1998), we administered the test by presenting the words on printed cards. The results of presenting the test with this method were shown (Auer et al., 1998) to be highly correlated with reading vocabulary test scores from the Stanford Achievement Test (SAT-8; The Psychological Corporation, 1989).

Vibrotactile Stimuli

The vibrotactile speech stimuli in the S+V experiment were presented via a vibrotactile vocoder, GULin, described in detail in Bernstein et al., 1991. The vocoder presented the envelope of 16 consecutive filters to the underside of the lower forearm of participants via a set of 16 transducers corresponding to the 16 filters. The transducers were small solenoids that vibrated at 100 Hz. The vocoder spanned the center frequencies of 15 pass bands from 260 Hz to 3115 Hz, with the 16th filter being a 3565 Hz high pass filter. The GULin vocoder was shown by Bernstein et al. (1991) to be superior to two other vocoders tested and to provide significant speech information to enhance speechreading. Vibrotactile stimuli were presented only during training, not testing. All participants wore EAR earplugs and headphones through which pink noise was delivered to mask possible audible components of the vibrotactile stimuli.

Procedures

Table 2 shows the plan of training and testing. Each type of pre- versus posttest material was counterbalanced across participant groups within each experiment. Table 2 shows the counterbalanced lists. None of the training lists were counterbalanced, and none of the test lists for Sessions 2 through 5 were counterbalanced. By presenting lists in a fixed order in Sessions 2 through 5, it was possible to observe patterns attributable to list differences (as well as practice). By counterbalancing between pre- and posttest, it was possible to control for list effects in the analysis of pre- to posttest change in scores. It was not intended to analyze the training list scores, and therefore Training 1 versus Training 9 lists were not counterbalanced.

Participants sat in a darkened, sound-attenuating room. A personal computer was used to control stimuli and to record responses. A 15-in color monitor was placed on a small table in front of the participant, who viewed the screen at a 60-cm distance. Frontal images of the talker whose face filled the screen were presented. A computer key-press was used to initiate the first stimulus, and the return key was used after each subsequent stimulus and response. Participants were instructed to type exactly what they thought the talker had said. Partial responses and guesses were encouraged.

During training, the correct sentence was printed on the computer screen following each response. During testing, no feedback was given, and participants gave confidence ratings on each response. It was possible that confidence could change with performance or independently of performance. Participants were given the following instructions for giving confidence ratings: “Rate your confidence in the correctness of your response.” Anchors for the rating scale were 0 = “No confidence—I guessed” and 7 = “Complete confidence—I understood every word.” Numbers between 0 and 7 were used to represent intermediate degrees of subjective performance. As in Demorest and Bernstein (1997), the zero rating included the phrase “I guessed” so as to acknowledge that participants had been encouraged to respond and that they might therefore lower their criteria for responding (Van Tasell & Hawkins, 1981).

Before beginning the sequence of testing and training, participants were given practice performing the task with a 20-sentence practice list. Then pretesting was initiated. Training and testing alternated over 6 sessions, as indicated in Table 2. Deaf participants completed the protocol within 7 to 34 days, and hearing participants completed it within 7 to 21 days.

Sentence Scoring

Responses to all stimuli (in testing and training) were checked for spelling errors. Words correct were scored with a computer program that compared (in order) the stimulus words with the response words for each sentence. The words were counted as correct only if they were exactly correct, including contracted forms. Previous work in our laboratory has shown that scores do not change in a meaningful manner when scoring is less strict. Percent-words-correct was calculated across sentences for each stimulus list (in training and testing) for each participant.

It was possible that word-level scoring was insensitive to significant factors in the study that took place at the phoneme level. For example, participants could have become more accurate for certain phonemes, yet not have been able to achieve better accuracy for word identification. Therefore, responses were also scored at the phoneme level. To accomplish this, responses were first phonemically transcribed using software that looked up response words in the PhLex database of phonemically transcribed words (Seitz, Bernstein, Auer, & MacEachern, 1998). Then the transcribed response was submitted to a software sequence comparison program (Bernstein, Demorest, & Eberhardt, 1994) that aligned the stimulus and response, phoneme by phoneme. Sequence comparison (of which dynamic time warping is a specific case) takes into account differences in symbol strings attributable to substitutions, deletions, and insertions. The sequence comparator used here includes a minimization algorithm (Bernstein et al., 1994; Sankoff & Kruskall, 1983) that seeks the lowest total cost for aligning the phonemes from the stimulus and response. The costs for phoneme-to-phoneme alignments were derived from phoneme confusion data obtained from experiments involving phoneme identification for speechread nonsense syllables (Bernstein et al., 1994). Phoneme-to-phoneme costs were Euclidean distances obtained via multidimensional scaling. Costs for insertions and deletions were selected so that perceptually implausible phoneme-to-phoneme alignments would not occur.

The following is an example of an alignment obtained with the sequence comparator:

Stimulus: p r u f r i d y u r f ɑΔn ə l r ɪ s ʌ l t s

Response: b l u f - ɪ ʃ ɑ - r f ɑ n ɪ - - - - - - - -

The stimulus was “proofread your final results,” and the response was “blue fish are funny.” The example has 5 correct phonemes (/u f r f n/), 6 phoneme substitutions (/p-b/, /r-l/, /i-ɪ/, /d-ʃ/, /y-ɑ/, /ɑΔɑ/, /x-ɪ/), 0 phoneme insertions (“-”), and 10 phoneme deletions (“-”).

A measure of mean proportion phonemes-correct was obtained for each stimulus sentence list for each participant. Mean proportion phonemes-correct was the mean taken across all of the sentences scored individually in terms of total correct response phonemes weighted by the number of phonemes in the respective stimulus.

A measure of mean confidence was obtained for each stimulus list for each participant. This was simply the mean of all confidence ratings across sentences within each stimulus list.

Analyses

Within each experiment, omnibus analyses for each of the measures (percent words-correct, mean proportion phonemes-correct, and mean confidence) on each of the test sets (CID Sentences, B-E Sentences with the male talker, and B-E Sentences with the female talker) were conducted. For CID Sentence Tests and the B-E Sentence Tests (male talker), 2 × 2 × 10 (Hearing Group × Experiment × Test) mixed analyses of variance were performed for each of the measures. The stimulus lists, analyzed in terms of their temporal order, are referred to as the factor *test*. Hearing group (hearing vs. deaf) was the between-subjects factor, and test was the within-subjects factor. A similar analysis was performed for the pre- versus posttest results of the B-E Sentence Tests (female talker). Analyses took into account counterbalancing between pre- and posttest sentence lists. That is, the data were analyzed in terms of the temporal order in which the participants received the stimulus lists, not in terms of list number. Linear trends were examined for the test factor and its interaction with group. Whenever group and/or test was shown to be a significant factor in the omnibus analysis, a reduced ANOVA was performed with pre- versus posttest as the repeated factor and group as the between factor. Because only the pre- and posttest scores were obtained with counterbalanced tests, this ANOVA was a more stringent method for evaluating learning than was the linear trend analysis, which included tests that were not counterbalanced.

Whenever percent words-correct scores were analyzed statistically, they were first submitted to arcsine transformations, which take into account the sampling variance of the score, and those results are reported here. However, whenever percent-correct means are reported, they are untransformed measures, to afford the reader direct access to the results.

Results and Discussion

Experiment I: Speechreading-Only (SO)

Within this section, percent words-correct scores are presented first, followed by mean proportion phoneme and confidence scores for each type of stimulus material. Table 3 is a summary of all of the means and ranges of test scores in terms of percent words-correct and mean proportion phonemes correct as a function of experiment.

CID Sentences, Percent Words Correct

In the omnibus ANOVA, when the CID Sentences were scored in terms of percent words correct, mean performance by the deaf group was not more accurate than that of the hearing group [$F(1, 14) = 3.107$, $p = .100$]. There was a significant effect of test [$F(9, 126) = 7.330$,

Table 3. Means and ranges across materials, groups, and experiments.

	Percent Words Correct				Proportion Phonemes Correct			
	Pretest		Posttest		Pretest		Posttest	
Group	Mean	Range	Mean	Range	Mean	Range	Mean	Range
CID Sentences, Test								
Exp. I								
Deaf	45	(27–69)	49	(30–94)	.55	(.34–.73)	.61	(.41–.97)
Hearing	28	(16–58)	35	(6–63)	.34	(.13–.63)	.47	(.29–.68)
Exp. II								
Deaf	64	(30–86)	60	(38–86)	.68	(.49–.83)	.66	(.36–.89)
Hearing	33	(16–56)	40	(11–67)	.51	(.27–.68)	.36	(.21–.60)
B-E Sentences, Male Talker, Test								
Exp. I								
Deaf	44	(26–72)	49	(34–92)	.53	(.43–.77)	.58	(.44–.96)
Hearing	36	(5–62)	35	(18–52)	.44	(.08–.68)	.42	(.26–.57)
Exp. II								
Deaf	58	(52–69)	55	(34–97)	.64	(.58–.71)	.59	(.31–.98)
Hearing	28	(5–48)	35	(11–74)	.37	(.12–.55)	.45	(.16–.83)
B-E Sentences, Female Talker, Test								
Exp. I								
Deaf	25	(17–45)	26	(10–55)	.34	(.20–.57)	.36	(.16–.64)
Hearing	17	(10–22)	20	(10–31)	.25	(.15–.32)	.28	(.17–.40)
Exp. II								
Deaf	31	(25–47)	32	(21–56)	.40	(.22–.55)	.39	(.21–.63)
Hearing	19	(7–31)	22	(14–36)	.26	(.18–.38)	.31	(.24–.44)

p = .000], but the linear trend for test was not significant (*p* = .080). The reduced analysis for group and test concerned with assessing learning showed no significant effects.

Figure 1 shows the percent words correct means in the SO experiment in terms of group and test. The error bars represent ±1 standard error of the respective mean.[3] The significant main effect of test can be seen in Figure 1 as list-to-list variations across the two participant groups. Tests 2 through 9 were not counterbalanced or randomized, allowing the list-to-list pattern to be clearly evident. Between pre- and posttests, the effect of list-to-list variations was controlled via counterbalancing list order across participants, within groups. Therefore, list-to-list variation would not have been a factor in a reduced analysis, had it been required because of the results of the omnibus analysis. Demorest, Bernstein, and DeHaven (1996) showed with the same sentences and method of scoring that variance of test scores has several components. One of the largest is the error variance, followed by the subject, and the sentence. This suggests that variability is to be expected, even when an attempt has been made to equate across lists, and caution should be exercised in interpreting fluctuations as experimental effects (see Abelson, 1995, particularly Chapter 2). The apparently consistent nature of the list-to-list variation for Tests 2 to 9 (Lists B to I) suggests, however, that error was not the primary source of the variation—rather that the source was stimulus attributes that at this time have not been identified.

CID Sentences, Mean Proportion Phonemes Correct

In the omnibus analysis for CID Sentences scored in terms of phonemes, the deaf group was more accurate [$F(1, 14) = 5.078$, $p = .041$], and tests varied significantly [$F(9, 126) = 9.117$, $p = .000$]. However, test × group and the linear trends were not significant. The group difference remained in the reduced analysis [$F(1, 14) = 6.414$, $p = .024$], as did the effect of test [$F(1, 14) = 11.984$, $p = .004$]; however, their interaction was not significant. The learning effect was due to an increase in scores pre- to posttest of .097.

Figure 2 shows the proportion phonemes correct means for this experiment. The figure shows, as do the analyses, that groups were different and that overall scores showed a learning effect. Generally, scoring using phonemes correct resulted in higher scores than the corresponding percent correct words (see Table 3). This is straightforwardly explained by noting that partially correct words are counted as incorrect when word scoring is used but receive credit when phoneme scoring is used.

[3] The standard error of the mean is a measure of the sampling variance of the mean for samples taken from the same population (SPSS, 1998).

CID Sentences, Mean Confidence

In the omnibus analysis, mean confidence was the same across groups and varied across tests [$F(9, 126) = 5.967$, $p = .000$], including linearly [$F(1, 14) = 5.734$, $p = .008$]. In the reduced analysis, both test and the test × group interaction were significant [respectively, $F(1, 14) = 26.947$, $p = .000$ and $F(1, 14) = 3.990$, $p = .000$]. The interaction was investigated with *t* tests. Deaf participants

Figure 1. Percent words correct means for CID Sentences in the SO experiment in terms of group and test. The error bars represent ±1 *SE* of the particular mean.

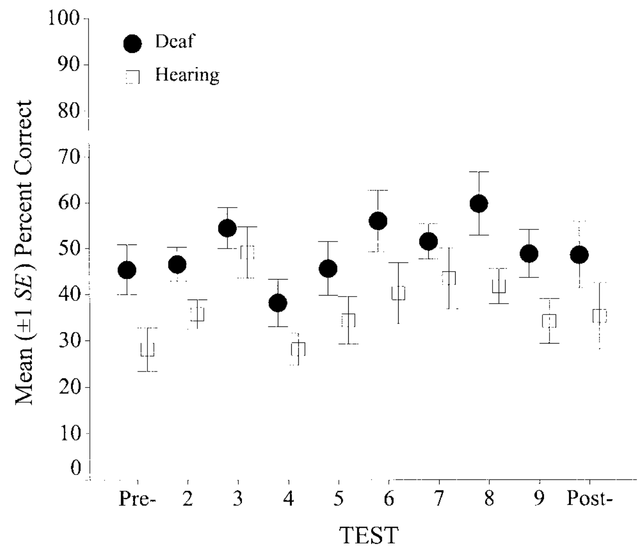
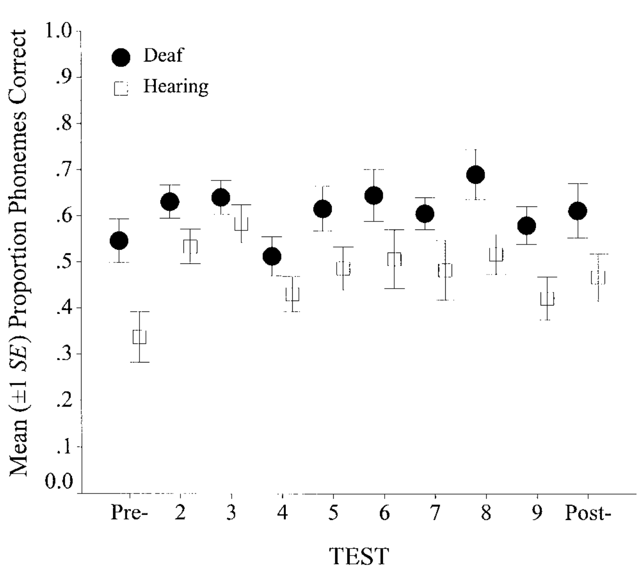


Figure 2. Proportion phonemes correct means for CID Sentences in the SO experiment in terms of group and test. The error bars represent ±1 *SE* of the particular mean.



did not vary in their confidence across pre- and posttest (p = .200), but hearing participants became more confident [$t(7) = -6.804$, p = .000].

B-E Sentence Tests, Male Talker, Percent Words Correct

None of the factors was significant when scores for these materials were based on percent words correct.

B-E Sentence Tests, Male Talker, Mean Proportion Phonemes Correct

When scores for these materials were based on mean proportion phonemes correct, only group was a significant factor in the omnibus analysis [$F(1, 14) = 4.919$, p = .044]. The reduced analysis for group and test resulted in no significant factors.

B-E Sentence Tests, Male Talker, Confidence

In the omnibus analysis, mean confidence was not different across groups (p = .219) but did vary across tests [$F(9, 126) = 2.082$, p = .036]. However, this test effect was not due to a linear trend. In the reduced analysis, the test × group interaction was significant [$F(1, 14) = 11.430$, p = .004]. The interaction was investigated with t tests. Deaf participants maintained their level of confidence across pre- versus posttests. But the hearing participants lost confidence [$t(7) = 2.722$, p = .030].

B-E Sentences, Female Talker, Percent Correct

By pre- and posttesting with a talker seen only on those two occasions, it was possible to investigate whether learning generalized. In the analysis of these sentence tests, within-subject factors were day (pre- vs. posttest) and test (first or second sentence set on each of the days). The between-subjects factor was group.

There was not a mean difference between groups (p = .174). However, both day and test were significant [respectively, $F(1, 14) = 6.615$, p = .022 and $F(1, 14) = 5.785$, p = .031]. The mean score was 21% words correct on pretest and 23% words correct on posttest. On each day, the second test was more difficult than the first. Although the 2 percentage-point difference from pre- to posttest was reliable, its magnitude was extremely small, likely of marginal practical significance, and could be attributed to retesting with the same talker rather than carryover from training. Because no interaction took place between day and group, it can be concluded that there was not a differential effect of training/practice across groups.

B-E Sentences, Female Talker, Mean Proportion Phonemes Correct

Analysis failed to reveal any significant factors when sentences were scored in terms of phonemes.

B-E Sentences, Female Talker Confidence

Analysis failed to reveal any significant factors.

Summary Experiment I

In this experiment, group differences were not obtained when responses were scored in terms of words correct. Group differences were observed with the phoneme scores. The phoneme scores obtained with the CID Sentences resulted in a consistent advantage on the part of the deaf group for the omnibus and reduced analyses. The phoneme scores with the B-E Sentences (male talker) resulted in an advantage only in the omnibus analysis. No group effect was obtained with the B-E Sentences (female talker). A possible reason for the reduction in the group effect across materials was that the sets varied in difficulty. Table 3 shows that the CID Sentences were the easiest, followed by the B-E Sentences (male), and then the B-E Sentences (female). The magnitude of the difference between groups was compressed as the scores became lower.

The only evidence of learning in this experiment was with the CID and the B-E Sentence (female) Sentences when scored in terms of phonemes. None of the interactions between test and group was reliable for the performance measures (words or phonemes). That is, there was no evidence in the performance measures for a differential effect associated with group. However, confidence revealed a somewhat different picture. With the CID Sentences, deaf participants did not vary in their confidence, but hearing participants became more confident. The opposite pattern was observed with the B-E Sentences (male). In that case, hearing participants lost confidence.

The conclusion from this experiment was that training/practice did raise speechreading scores in deaf and hearing groups by small increments that did not favor either group.

Experiment II: Vibrotactile Speech Stimuli During Speechreading Training (S+V)

CID Sentences, Percent Words Correct

In the omnibus analysis, when the CID Sentences were scored in terms of percent words correct, mean performance by the deaf group was more accurate than that of the hearing group [$F(1, 14) = 12.203$, p = .004] (see Figure 3). There was a significant effect of test [$F(9, 126) = 2.552$, $p = .010$], but the linear trend for test was not significant nor were any interactions. Group continued to be a significant factor in the reduced analysis [$F(1, 14) = 12.300$, $p = .003$], but test and test × group were not significant (ps > .474). Groups were reliably different, but there was not a reliable difference pre- to posttest.

CID Sentences, Mean Proportion Phonemes Correct

In the omnibus analysis, when the CID Sentences were scored in terms of phonemes, mean performance by the deaf group was more accurate than by the hearing group [$F(1, 14) = 14.193$, $p = .002$]. There was a significant effect of test [$F(9, 126) = 4.320$, $p = .000$], and the linear trend for test was significant [$F(1, 14) = 28.433$, $p = .000$], as was the test × group interaction [$F(1, 14) = 5.499$, $p = .034$]. In the reduced analysis, only the main effects of group and test were significant [respectively, $F(1, 14) = 15.089$, $p = .002$ and $F(1, 14) = 5.639$, $p = .032$]. However, the change in scores across pre- versus posttest was actually towards significantly lower scores (see Figure 4). Across groups, the decline was .086.

CID Sentences, Mean Confidence

In the omnibus analysis, mean confidence differed across groups [$F(1, 14) = 6.225$, $p = .026$] and across tests [$F(9, 126) = 2.996$, $p = .003$], but not linearly. In the reduced analysis, only the group effect was significant [$F(1, 14) = 6.787$, $p = .021$].

B-E Sentence Tests, Male Talker, Percent Words Correct

When the B-E Sentences were scored in terms of percent words correct, mean performance by the deaf group was more accurate than that of the hearing group [$F(1, 14) = 12.867$, $p = .003$]. There was a significant effect of test [$F(9, 126) = 2.922$, $p = .004$], and the linear trend for test was significant [$F(1, 14) = 5.557$, $p = .033$]. However, the group × test interaction was also significant [$F(9, 126) = 1.975$, $p = .047$], as was the linear trend for group × test [$F(1, 14) = 9.062$, $p = .009$]. However, only the group effect was significant in the reduced analysis [$F(1, 14) = 10.507$, $p = .006$]; therefore, there was no learning effect.

B-E Sentence Tests, Male Talker, Mean Proportion Phonemes Correct

Across tests, the groups were different [$F(1, 14) = 11.819$, $p = .004$]. Also the test and test × group effects were significant [respectively, $F(9, 126) = 3.061$, $p = .002$ and $F(9, 126) = 2.361$, $p = .017$], as was the linear test × group effect [$F(1, 14) = 9.532$, $p = .008$]. In the reduced analysis, only the difference between groups was significant [$F(1, 14) = 10.814$, $p = .005$]; therefore, there was no learning effect.

Figure 3. Percent words correct means for CID Sentences in the S+V experiment in terms of group and test. The error bars represent ±1 *SE* of the particular mean.

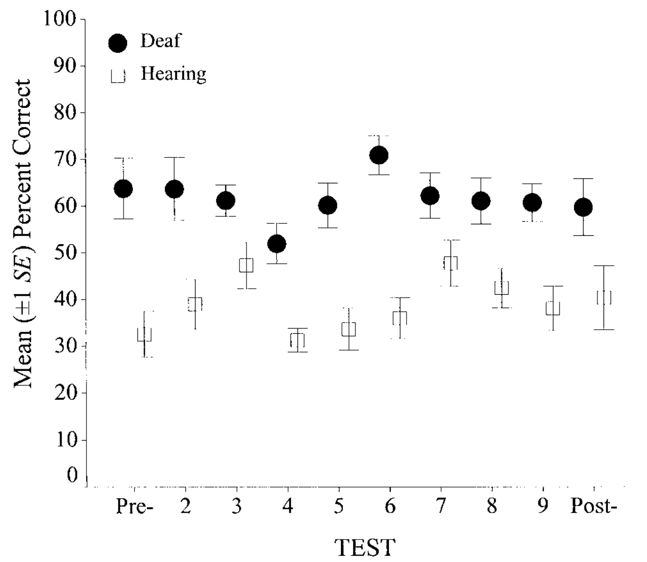
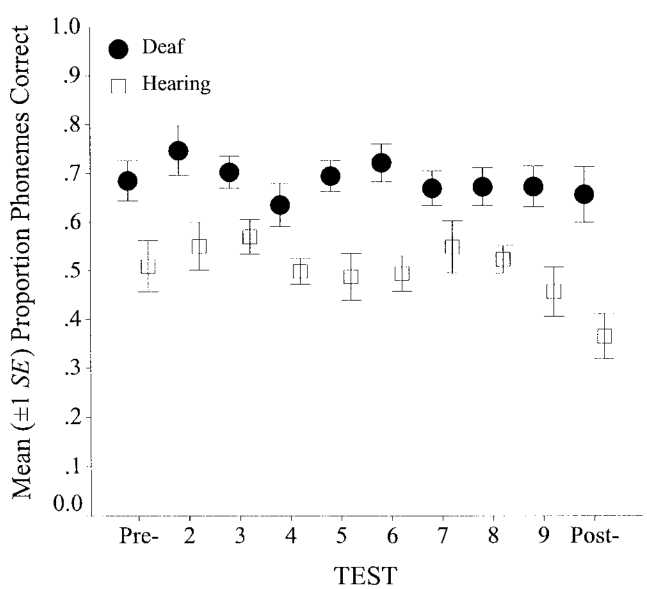


Figure 4. Proportion phonemes correct means for CID Sentences in the S+V experiment in terms of group and test. The error bars represent ±1 *SE* of the particular mean.



B-E Sentence Tests, Male Talker, Confidence

There were no significant effects.

B-E Sentences, Female Talker, Percent Correct

Across pre- and posttest, the groups were different [$F(1, 14) = 7.204$, $p = .018$]. On each day, the second test was more difficult than the first [$F(1, 14) = 10.065$, $p = .007$]. However, there was not a learning (i.e., day) effect. There were no interactions.

B-E Sentences, Female Talker, Mean Proportion Phonemes Correct

The results with phoneme scoring were similar to those with word scoring. Across pre- and posttest, the groups were different [$F(1, 14) = 5.467$, $p = .035$], and the test effect was significant [$F(1, 14) = 12.984$, $p = .003$]; however, the test × group effect was not. There was not a learning effect.

B-E Sentences, Female Talker Confidence

There were no significant effects.

Summary Experiment II

In this experiment, groups were consistently different. Significant interactions of group × test in the omnibus analyses were due to list-to-list effects only, because the group × test interactions were not significant in the reduced analysis, which investigated the pre- versus posttest scores. There was not any indication of learning in this experiment. The only pre- to posttest change that was statistically reliable was in the direction of a decline in performance. The only significant effect associated with confidence was a group difference favoring the deaf group that was obtained with the CID Sentences.

PPVT Scores

The PPVT results were used to probe the differential learning effects across the two experiments. The PPVT standard score equivalent (SSE) for each of the participants was submitted to a univariate analysis of variance with group (deaf vs. hearing) and experiment (SO vs. S+V) as between-subject factors. Participants in Experiment II, S+V, had significantly higher PPVT scores than those in Experiment I, SO (mean 117 vs. 95, respectively) [$F(1, 32) = 7.477$, $p = .011$]. Hearing participants had significantly higher scores than deaf participants (114 vs. 97, respectively) [$F(1, 32) = 4.834$, $p = .036$]. The range of scores in the hearing group was 96 to 160. The range in the deaf group was 40 to 160. Group and experiment did not interact. Given the higher vocabulary knowledge of the participants in the S+V experiment, more learning might have been predicted in the S+V experiment; this did not occur. To probe whether PPVT scores were related to speechreading, the PPVT scores were submitted to Pearson correlation analyses with pre- and posttest scores from the CID and B-E (male talker) Sentences within each experiment. None of those correlations was significant. Thus, there was no evidence that learning was related to PPVT scores, either across experiments or within experiments.

CID Sentence Scores Across Experiments

Results of these experiments can be used to investigate whether short-term training/practice affects the relative standing of individual speechreaders. Although the two experiments differed in their training stimuli, in their evidence for learning, and in their overall scores on the PPVT vocabulary tests, they can be combined to investigate whether the relative standing of the individual participants was substantially affected by their participation in the study.

Figure 5 is a scatterplot of the CID Sentence preversus posttest percent correct words scores for each of the participants. Figure 6 shows the results in terms of mean proportion phonemes correct. The figures show that several hearing participants were the lowest scoring participants at pre- and posttest. Many participants, both deaf and hearing, performed in the midrange, as might be expected. The highest scoring participants at pre- and posttest were deaf. No hearing participant exceeded the performance levels of the most proficient deaf participants (Participants SP03, TSD03, and TSD05; see Table 3). The highest three CID Sentence phoneme scores at posttest for deaf participants were .82, .89, and .97. The highest three phoneme scores for hearing participants were .60, .65, and .68. Participants SP03, TSD03, and TSD05 had severe to profound hearing impairments at birth. Bernstein et al. (2000) obtained similar examples of highly accurate speechreading from a different and larger sample of deaf speechreaders. The consistency of this observation supports the hypothesis that early severe to profound hearing impairment in some yet-to-be understood combination with reliance on visible speech can result in enhanced levels of speechreading accuracy.

Several hearing participants with scores in the vicinity of 25% words correct raised their scores to the vicinity of 60% words correct at posttest. But these large changes did not overcome the advantage of the most accurate deaf participants. Most important, overall participants did not change relative standing from pre- to posttest. This is demonstrated by the R square of .857 (R of .926) for the regression line through the total population of words correct scores in Figure 5, indicating high predictability

Figure 5. Scatterplot of the pre- versus posttest percent correct words scores (CID Sentences) for each of the participants across experiments. The regression line was calculated through the origin for the entire sample population.

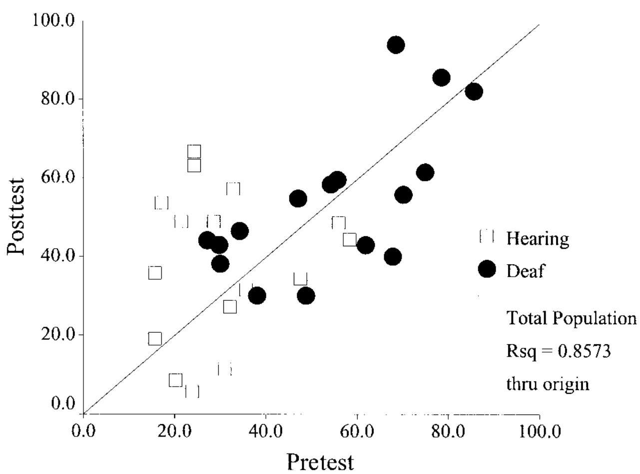
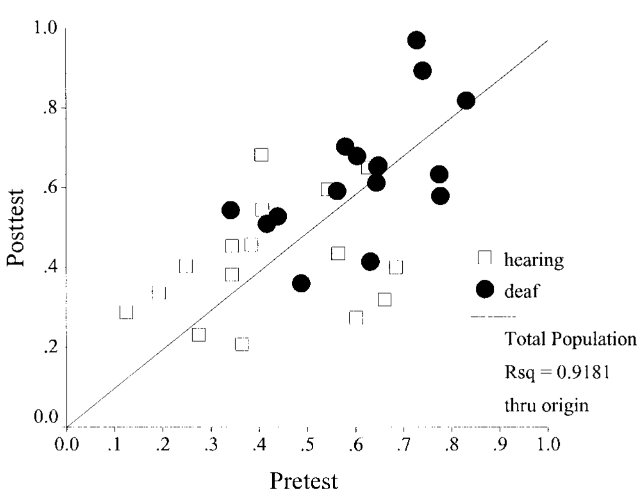


Figure 6. Scatterplot of the pre- versus posttest mean proportion phonemes correct scores (CID Sentences) for each of the participants across experiments. The regression line was calculated through the origin for the entire sample population.



from pre- to posttest performance levels. In Figure 6, the *R* square of the total population regression line when phoneme scores were used was even higher: .918 (*R* of .958).

General Discussion

Several conclusions can be drawn from this study. First, consistent evidence was obtained for a speechreading advantage among the deaf versus hearing participants in this study who were self-selected as good speechreaders. Second, whether or not learning occurred, it did not alter the relative standing of individual speechreaders. Finally, short-term training/practice on sentence stimuli was shown to result in a small but statistically reliable learning effect when training and testing were conducted in the same manner (SO), but not when the training was conducted with vibrotactile stimuli (S+V). Some additional comments about these results follow.

Group and Individual Effects

Training/practice was not found to be differentially effective across deaf versus hearing groups. No evidence for interactions between groups and tests was obtained for speechreading performance measures. Instead, a relatively consistent pattern of advantage for accuracy in the deaf groups was obtained. The current results do not encourage the conclusion that substantial group differences can be erased through a period of training/practice comparable in extent to that in the current study.

The relative standing of individual speechreaders was not affected by training/practice. Training did not bring any of the individual hearing participants up to the level of the most accurate deaf participants. Given the range of individual differences in the deaf and hearing populations (Bernstein et al., 2000), it is clear that groups with similar abilities can be recruited that are likely to remain similar across a period of laboratory testing. Groups that are recruited to be substantially different from each other are likely to maintain their differences across a period of laboratory testing.

Short-Term Training Effects

The results showed that short-term training can result in significant change, but its direction depends on the training conditions and perhaps also the test materials. The tests with CID Sentences (scored in terms of phonemes) showed that mean proportion phonemes correct increased .097 from pre- to posttest in the SO experiment, but decreased .086 in the S+V experiment. This pattern was not observed in the performance measures (phonemes and words) obtained with the B-E Sentences (male talker). However, the latter materials were more difficult (see Table 3), and the tests were possibly less sensitive given their brevity (see Demorest, Bernstein, & DeHaven, 1996). A significant improvement of approximately 2 percentage points was obtained in Experiment I with the B-E Sentences (female talker) tested at pre- versus posttest only. In Experiment II, there was not a pre- versus posttest improvement. This difference across experiments suggests that carry-over might have influenced results with these materials.

Effect of Training With the Vibrotactile Speech Aid

We can only speculate at this time why the vibrotactile stimuli might have resulted in depressed scores. Perhaps previous enhancements to speechreading alone in experiments with vibrotactile stimuli were the result of more extensive training. Another possibility is that the S+V participants construed that their main task was to demonstrate that vibrotactile stimuli enhance speechreading and during testing unconsciously (or consciously) reduced their attention or effort to speechreading. A third possibility is that sensitivity to visual information actually declined as a result of relying on the vibrotactile stimuli (Sherrick, 1984). However, how the two modalities in Experiment II might have enhanced or interfered with each other during perception and in the total experimental context cannot be determined within the parameters of the current study.

Methodological Implications

A methodological issue relevant to this study is whether studies that compare deaf and hearing speechreaders should provide initial training in order to obtain stable performance that is not somehow favorable to one or the other group. Our results support the general finding in the literature that with repeated training/practice, speechreaders can show modest improvements. But our results do not support providing differential training/practice to hearing versus deaf speechreaders. The groups change modestly at the same rate.

Conclusions

These results suggest that speechreading differences between deaf and hearing populations are quite stable. Among deaf speechreaders there are individuals with enhanced speechreading capability whose superiority cannot be matched via short-term training by less accurate speechreaders from either the deaf or hearing population. The existence of stable individual differences sets the stage for future studies that can investigate the underlying sensory/perceptual and psycholinguistic processes responsible for the wide range of individual differences among speechreaders.

Acknowledgments

We thank Brian Chaney for his assistance in processing data and Sven Mattys, PhD, for commenting on the paper. A subset of the results was presented at Auditory-Visual Speech Processing, 1998, Terrigal, New South Wales, Australia. The research was supported by grants from NIH/NIDCD DC01207 and DC00695.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Erlbaum.

Auer, E. T., Jr., Bernstein, L. E., & Tucker, P. E. (2000). Is subjective word familiarity a meter of ambient language? A natural experiment on effects of perceptual experience. *Journal of Memory and Cognition, 28,* 789–797.

Bernstein, L. E. (1991). Lipreading Corpus V-VI: Disc 3 [videodisc]. Washington, DC: Gallaudet University.

Bernstein, L. E., Demorest, M. E., Coulter, D. C., & O'Connell, M. P. (1991). Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America, 90,* 2971–2984.

Bernstein, L. E., Demorest, M. E., & Eberhardt, S. P. (1994). A computational approach to analyzing sentential speech perception: Phoneme-to-phoneme stimulus-response alignment. *Journal of the Acoustical Society of America, 95,* 3617–3622.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (1998). What makes a good speechreader? First you have to find one. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: The psychology of speechreading and auditory-visual speech* (pp. 211–228). East Sussex, U.K.: Psychology Press.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62,* 233–252.

Bernstein, L. E., & Eberhardt, S. P. (1986a). Johns Hopkins Lipreading Corpus I-II: Disc 1 [videodisc]. Baltimore, MD: Johns Hopkins University.

Bernstein, L. E., & Eberhardt, S. P. (1986b). Johns Hopkins Lipreading Corpus III-IV: Disc 2 [videodisc]. Baltimore, MD: Johns Hopkins University.

Boothroyd, A., & Hnath-Chisolm, T. (1988). Spatial, tactile presentation of voice fundamental frequency as a supplement to lipreading: Results of extended training with a single subject. *Journal of Rehabilitation Research and Development, 25,* 51–56.

Boothroyd, A., Kishon-Rabin, L., & Waldstein, R. (1995). Studies of tactile speechreading enhancement in deaf adults. *Seminars in Hearing, 16,* 328–342.

Clouser, R. A. (1977). Relative phoneme visibility and lipreading performance. *Volta Review, 79,* 27–34.

Conrad, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology, 47,* 60–65.

Davis, H., & Silverman, S. R. (Eds.). (1970). *Hearing and deafness* (3rd ed.). New York: Holt, Rinehart & Winston.

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

De Filippo, C. L., & Scott, B. L. (1978). A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America, 63,* 1186–1192.

Demorest, M. E., & Bernstein, L. E. (1997). Relationships between subjective ratings and objective measures of performance in speechreading sentences. *Journal of Speech, Language, and Hearing Research, 40,* 900–911.

Demorest, M. E., Bernstein, L. E., & DeHaven, G. P. (1996). Generalizability of speechreading performance on nonsense syllables, words, and sentences: Subjects with normal hearing. *Journal of Speech and Hearing Research, 39,* 697–713.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised.* Circle Pines, MN: American Guidance Service.

Eberhardt, S. P., Bernstein, L. E., Demorest, M. E., & Goldstein, M. H. (1990). Speechreading sentences with single-channel vibrotactile presentations of voice fundamental frequency. *Journal of the Acoustical Society of America, 88,* 1274–1285.

Gesi, A. T., Massaro, D. W., & Cohen, M. M. (1992). Discovery and expository methods in teaching visual consonant and word identification. *Journal of Speech and Hearing Research, 35,* 1180–1188.

Green, K. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 3–25). East Sussex, U.K.: Psychology Press.

Heider, F., & Heider, G. M. (1940). An experimental investigation of lipreading. *Psychological Monographs, 52,* 124–153.

Jeffers, J., & Barley, M. (1971). *Speechreading (lipreading).* Springfield, IL: Charles C. Thomas.

Kishon-Rabin, L., Boothroyd, A., & Hanin, L. (1996). Speechreading enhancement: A comparison of spatial-tactile display of voice fundamental frequency (F_0) with auditory F_0. *Journal of the Acoustical Society of America, 100,* 593–602.

Lyxell, B., & Rönnberg, J. (1991a).Visual speech processing: Word-decoding and word-discrimination related to sentence-based speechreading and hearing-impairment. *Scandinavian Journal of Psychology, 32,* 9–17.

Lyxell, B., & Rönnberg, J. (1991b). Word discrimination and chronological age related to sentence-based speechreading skill. *British Journal of Audiology, 25,* 3–10.

Massaro, D. W. (1987). *Speech perception by eye and ear.* Hillsdale, NJ: Erlbaum.

Massaro, D. W., Cohen, M. M., & Gesi, A. T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception and Psychophysics, 53,* 549–562.

Mogford, K. (1987). Lip-reading in the prelingually deaf. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 191–211). Hove, U.K.: Erlbaum.

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research, 28,* 381–393.

Pelson, R. O., & Prather, W. F. (1974). Effects of visual message-related cues, age, and hearing impairment on speechreading performance. *Journal of Speech and Hearing Research, 17,* 518–525.

Plant, G., & Spens, K-E. (Eds.). (1995). *Profound deafness and speech communication.* London: Whurr.

Rönnberg, J. (1995). Perceptual compensation in the deaf and blind: Myth or reality? In R. A. Dixon & L. Bäckman, (Eds.), *Compensating for psychological deficits and declines* (pp. 251–274). Mahwah, NJ: Erlbaum.

Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison.* Reading, MA: Addison-Wesley.

Seitz, P. F., Bernstein, L. E., Auer, E. T., Jr., & MacEachern, M. (1998). *PhLex (Phonologically Transformable Lexicon): A 35,000-word computer readable pronouncing American English lexicon on structural principles, with accompanying phonological transformations, and word frequencies.* Los Angeles: House Ear Institute.

Sherrick, C. E. (1984). Basic and applied research on tactile aids for deaf people: Progress and prospects. *Journal of the Acoustical Society of America, 75,* 1325–1342.

SPSS. (1998). SPSS for Windows 9.0 [Computer software]. Chicago: Author.

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 117–137). Hillsdale, NJ: Erlbaum.

Summers, I. R. (Ed.). (1992). *Tactile aids for the hearing impaired.* London: Whurr Limited.

The Psychological Corporation. (1989). *Stanford Achievement Test Series* (8th ed.). San Antonio, TX: Harcourt Brace Jovanovich.

Tillberg, I., Rönnberg, J., Svärd, I., & Ahlner, B. (1996). Audio-visual speechreading in a group of hearing aid users: The effects of onset age, handicap age, and degree of hearing loss. *Scandinavian Audiology, 25,* 267–272.

Van Tasell, D. J., & Hawkins, D. B. (1981). Effects of guessing strategy on speechreading test scores. *American Annals of the Deaf, 126,* 840–844.

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

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., & Jones, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20,* 130–145.

Weisenberger, J. M., & Broadstone, S. M. (1989). Evaluation of two multichannel tactile aids for the hearing-impaired. *Journal of the Acoustical Society of America, 86,* 1764–1775.

Received January 19, 1999

Accepted November 11, 2000

DOI: 10.1044/1092-4388(2001/001)

Contact author: Lynne Bernstein, PhD, Department of Communication Neuroscience, House Ear Institute, 2100 W. Third Street, Los Angeles, CA 90057.
Email: lbernstein@hei.org