Journal of Speech and Hearing Research, Volume 35, 1180–1188, October 1992

Discovery and Expository Methods in Teaching Visual Consonant and Word Identification

Antoinette T. Gesi
Dominic W. Massaro
Michael M. Cohen
University of California
Santa Cruz

An experiment was conducted to examine the processes involved in lipreading as well as to investigate an optimal approach to teaching lipreading skill. We compared discovery and expository methods of learning to lip-read. Twenty-six college students with normal hearing were trained over 3 days to lip-read consonant-vowel (CV) syllables. The training material consisted of a prerecorded videotape of four different talkers. The task was a forced-choice procedure with feedback. Subjects learned with training, but there was no difference between the two learning methods. As a retention measure, subjects returned 4 weeks later and repeated the training. There were significant savings of the original learning. Three weeks after the retention phase, subjects were tested with a 10-item forced-choice monosyllabic word task. Those subjects who had extensive training on CV syllables did no better on identifying the monosyllabic words than did a control group of 9 subjects with no training. Nevertheless, performance for all three groups (discovery, expository, and no training) improved during training in the word identification task.

KEY WORDS: speech, learning, perception

The debate between discovery learning (which is self-taught) and expository learning (which is explicitly taught by another source) persists. Bruner (1961) claimed that one benefit of discovery learning is what he called "conservation of memory." He stated that "the key to retrieval is organization or, in even simpler terms, knowing where to find information and how to get there" (p. 51). Through discovery learning a person is said to organize information in a personal way that is consistent with his or her own cognitive structures; therefore, the information should be more accessible in memory (i.e., retention is better). In addition, discovery learning has been claimed to have a positive influence on transfer to related tasks (e.g., Bruner, 1961; Solter & Mayer, 1978). This ability to transfer supposedly results from the discovery learner's construction of several different routes to the correct solution. Expository learning, on the other hand, supposedly constrains the learner down a single path to the correct solution because the process to the solution is given.

Although the research comparing discovery learning and expository learning has been extensive, the findings are not conclusive. There is evidence for the superiority of discovery learning (Bruner, 1961; Solter & Mayer, 1978), evidence for the superiority of expository learning (McLeod & Adams, 1981; Sweller & Cooper, 1985), and evidence for no difference between the two methods of instruction (Hunt, 1975; Norton, 1977). Egan and Greeno (1973) attempted to account for these conflicting reports by suggesting that either method is better depending on the testing procedure; the test that is most similar to the instruction will be associated with the best performance. This is apparent in the findings of Alcantara, Cowan, Blamey, and Clark (1990) who used electrotactile speech to examine the effects of an analytical approach compared with a synthetic approach. Two groups of subjects with normal hearing participated in 70 hours of training over a 6-month period. One group focused

  0022-4685/92/3505-1180$01.00/0

vowels. The consonants were the same as those used in the Walden et al. (1977) study (i.e., /ð/, /v/, /ʒ/, /z/, /b/, /d/, /w/, /r/, /l/); these were combined with the vowels /i/, /a/, /u/. On a given trial, a syllable was spoken once, and then again for feedback. Each of the 4 talkers recorded 9 randomized lists of syllables; thus, it is unlikely that subjects were able to notice any peculiarities to bias their response. In a given block all 108 utterances (27 CVs × 4 speakers) occurred.

The transfer stimuli were 120 monosyllabic words. The words were chosen from a 437-word set composed of words from the Modified Rhyme Test and other words (Bernstein & Eberhardt, 1986). Twelve blocks, each built from a 10-word set, were constructed with the constraint that each word differed from one other word in the block by one viseme. These visemic minimal pairs consisted of two words that look identical except for one movement/position of the articulators (e.g., /pin/ – /tin/ is a minimal pair; /pin/ – /bin/ is not a minimal pair because they cannot be distinguished from each other based on lip movements). Each block used the 10 words twice sampled randomly without replacement four times. That is, in every 20 trials each of the 10 words was presented twice; this occurred four times. Thus, 80 trials were presented per block. Subjects received four blocks per day over a 3-day period.

Apparatus and Design

The four talkers were seated in a semicircle at approximately the same height against a blank light-blue wall. The talkers were asked to articulate CV syllables in front of a camcorder. They were instructed to enunciate as clearly as possible because the tape would be used to teach people how to lip-read. In order to stress the importance of their pronunciation, we further told the talkers that the tape would be shown to other students who would be asked to indicate what syllable each talker was saying by watching their lip movements. For each trial, the syllable and the talker were indicated by a cue presented on a TVI950 monitor. The cue consisted of the CV syllable and a number from 1 to 4. The monitor was located next to a Panasonic AG170 camcorder, and both were controlled by a DEC PDP-11/34a computer. The camcorder was panned by an operator to the appropriate talker for every trial. The operator pressed a computer key that changed the camcorder from pause to record mode. Following a 1000-msec delay, a ready tone (a 400-msec beep) was presented to the talker over the TVI950 terminal; then, after a 500-msec pause, the screen reversed (black to white) for 500 msec, cuing the talker to say the syllable. The microphone on the camcorder recorded the audio input. After a 5000-msec pause, the screen color reversed again, cuing the talker to repeat the syllable in the same manner as the first presentation. The camera rolled for another 350 msec and then was put in pause mode so the operator could pan to the next talker.

An experimental tape was made from the master tape and was played on a Panasonic NV-8200 VHS VCR. The audio monitor output of the VCR was connected to a schmitt trigger of the KW-11K clock in the PDP-11/34a computer. This was used to detect the beep that marked the start of each trial. It was also used to record subjects' responses and reaction time for each response in msec. During the dubbing process, the audio of the initial CV was gated out by an ICONIX 037 audio gate with a rise and fall time of 2.5 msec under computer control.

Subjects were tested individually in sound-attenuated rooms with overhead lighting. Each room contained a chair facing a table holding two displays: a terminal with keyboard and a NEC C12 202A color monitor. Subjects sat approximately 33 in. away from the monitor, and the face of the speaker filled the greater part of the monitor's screen. The audio feedback portion of the experimental tape was presented to the subjects over the built-in speakers of the monitors at a comfortable listening level of about 67 dB-A measured at the approximate position of the observers head using a B&K 2123 sound level meter with a 4134 microphone.

Subjects made their responses on the keyboard, labeling each CV syllable in the following manner: THEE, VEE, ZHEE, ZEE, BEE, DEE, WEE, REE, LEE along the top row of keys; THAH, VAH, ZHAH, ZAH, BAH, DAH, WAH, RAH, LAH along the middle row; and THUU, VUU, ZHUU, ZUU, BUU, DUU, WUU, RUU, LUU along the bottom row.

During the transfer task, the DEC PDP-11/34A computer controlled an Amiga 1000 computer and a SONY Laser Disk Player (LDP-1500). The stimuli that were displayed on the color monitor were from a laser disk recording created at The Johns Hopkins University of a male talker, approximately 35 years of age, saying monosyllabic words. The Amiga was used to overlay the response possibilities, which consisted of 10 numbered words on the lower right side of the monitor in an area beyond the talker's face. On each trial, one of the 10 possible words was presented. The video frame was displayed for 400 msec just before the word; the frame following the word was left on the screen during the open-ended response interval. No sound was presented. After all responses were recorded, a search for a start point on the video disk was executed, followed by a second presentation of the same stimulus—this time with the sound present to provide feedback to the subjects. After the feedback, there was a blank 500-msec intertrial interval.

Procedure

Training. During the training phase, four groups of subjects, with 3 or 4 subjects per group, were randomly assigned to two different instruction conditions. In the discovery condition, subjects were given just enough information to perform the task. In the expository condition, subjects were also given an explicit lesson on how each of the 27 syllables is produced and the visual characteristics of each. In addition, a summary of helpful phonetic and visual characteristics for each consonant and vowel were displayed on a monitor (see Appendix). The list of descriptors was provided to further assist or cue memory; however, the subjects reported rarely ever looking at the list. The lesson and display occurred on each of the 3 days of training.

The instructions (see Appendix) were read to each group of subjects. For expository groups the experimenter demon-

strated the lip movements of each consonant and each vowel as well as the coarticulatory effects of each CV syllable. Explicit verbal instructions accompanied the visual demonstration of the lip movements in order to emphasize the important characteristics. Each subject was then assigned to one of the soundproofed rooms for testing.

The experimental session consisted of 3 blocks with a short break between blocks. All subjects participated in 3 sessions (1 per day) of about an hour each, completing 9 blocks of 108 CV syllables. Performance was scored as proportion correct.

Retention testing. Approximately 4 weeks following the third training session, subjects returned for another 3-day training period. This retention phase entailed the same procedure as the training session except that subjects were tested individually or in groups of 2 to 4, with each subject in a different testing room. Because the retention phase was designed to provide additional training, we kept the task exactly the same. We refer to it as a retention phase only because of the delay between sessions. We were interested to see if what was learned would be retained and if additional learning could be obtained; thus, we gave feedback during the retention phase. The appropriate instructions were repeated to subjects in the expository and discovery groups. Performance was scored as proportion correct.

Transfer testing and control group. The transfer phase required the subjects in the two training groups to return approximately 3 weeks after the retention phase. Nine subjects without any previous training were tested as a control group. Subjects were tested individually or in groups of up to 4, with each subject in a different soundproofed room. The experimenter read the instructions to the subjects. They were told that they would see but not hear a male talker say 1 of 10 monosyllabic English words that would appear in print on the terminal screen. Their task was to watch the talker and determine what word he said by reading his lips. Subjects responded by pushing one of 10 numbered keys that corresponded to the words. Subjects were asked to respond as soon as they arrived at a decision. After each subject responded, visual and auditory feedback were given. That is, the subjects saw *and heard* the talker repeat the word. The subjects were instructed to watch the lip movements as well as to listen to the sound of the correct response in order to get as much feedback as possible. The subjects were informed that after approximately 10 min a new set of 10 words would appear on the screen and the talker would say words from that set. Before each block, subjects were given 25 sec to view the 10 new words on the screen in order to become familiar with them. Subjects participated for 3 days, during which they were given 4 blocks per day with 10 new test words in each block. Proportion correct and response time were recorded.

Results

All statistical tests were conducted by analyses of variance (ANOVAs). Table 1 gives a description of the different analyses for all of the significant findings and for the important but statistically insignificant findings. No other main effects or interactions reached significance.

Training Phase

For each subject, the proportion correct on each consonant (9), each vowel (3), and each syllable (27) was determined. For example, all syllables with /b/ yielded a score for /b/ as a function of /i/, another as a function of /a/, and a third as a function of /u/, regardless of the correctness of the vowel portion of the response to each /b/ + V syllable. A series of ANOVAs was carried out on the results of the original 10 subjects in each group using the consonant, vowel, and syllable scores as the dependent variable. Consonant, vowel, training block, and instructions were factors in the design. The effects of training and instructions were the same for all three dependent variables. The results indicated in Table 1 for CV accuracy were replicated for both C accuracy and V accuracy (although not included in the table); therefore, only the results for proportion correct on each syllable are examined here to evaluate overall training results. Consonant accuracy and vowel accuracy during training are discussed later in this section. Significance was determined by the .01 criterion level (two-tailed). The left panel of Figure 1 contains the mean proportion of correct identifications of CV syllables for the training phase of the experiment. An ANOVA using a 2 × 9 × 3 × 9 mixed factorial design was performed. Instruction condition, with 2 levels, varied between subjects, whereas the number of blocks (9 levels), number of vowels (3 levels), and number of consonants (9 levels) varied within subjects. For the training results, the left panel of Figure 1 shows that identification accuracy increased across training blocks, $F(8, 136) = 17.542$, $MSe = 0.068$, $p < .001$.

A second ANOVA performed on block 1 ($M = .55$) compared with block 2 ($M = .65$), using the same factors as those just mentioned, found the 10% increase in accuracy significant, $F(1, 18) = 30.801$, $MSe = 0.081$, $p < .001$. This finding suggests that lipreading performance improved after just one block of training. Even when block 1 was eliminated from the analysis, using the same factors as those mentioned above with the exception of number of blocks having 8 levels, the main effect for block was still significant, $F(7, 126) = 7.966$, $MSe = 0.047$, $p < .001$. Thus, significant improvement continued to occur after the initial block of training. Table 2 gives the identification accuracy across blocks for each of the subjects in the experiment. As expected, there were substantial differences among the subjects in both training groups.

There was no significant effect of the two types of instruction, $F(1, 17) = 1.762$, $MSe = 0.8511$, $p = .200$. Because we failed to find a significant difference between the two training groups, we decided to test three additional subjects in each training method in order to increase the power of the experiment. However, this effort did not yield the desired significant differences between the two training methods, $F(1, 24) = 1.144$, $MSe = 1.102$, $p = .296$. Since we failed to find any significant differences, these additional subjects were not asked to return for the retention or transfer phase of the experiment.

TABLE 1. A summary description of all significant effects and important nonsignificant effects obtained from each analysis of variance.

Analysis	Dependent variable	Independent variables	N^a	*df*	*F*-value	*p*
Training	CV accuracy	Block (1–9)	10	8, 136	17.542	.001
		Block (1 vs. 2)	10	1, 18	30.801	.001
		Block (2–9)	10	7, 126	7.966	.001
		Instruction	10	1, 17	1.762	.200
		Instruction (w/added Ss)	13	1, 24	1.144	.296
		Talker	10	3, 54	14.783	.001
	C accuracy	Consonant	10	8, 144	113.118	.001
	V accuracy	Vowel	10	2, 36	15.386	.001
Retention						
	CV accuracy	Block (1–9)	9	8, 128	4.458	.001
		Instruction	9	1, 16	2.211	.154
Training vs. retention						
	CV accuracy	Block (7–9 vs. 10–12)	9	1, 16	0.348	.570
Transfer						
	Word accuracy	Instruction		2, 22	2.470	.106
		Discovery	7			
		Expository	8			
		No training	9			
		Presentation		3, 66	81.547	.001
		Instr X Pres		6, 66	0.836	.547
	Reaction time					
		Instruction		2, 22	0.599	.563
		Presentation		3, 66	87.409	.001
		Instr X Pres		6, 66	0.539	.778

[a]N = number of subjects in the independent groups.

An ANOVA was conducted to determine talker variations. All the factors mentioned in the previous analyses—instruction condition (2), number of blocks (9), vowel (3), consonant (9)—were also used in this analysis. Another within-subjects variable, talker (4 levels), was added. The proportion of overall identification accuracy for the four talkers was .661, .683, .682, and .711, yielding a significant main effect of talker, $F(3, 54) = 14.783$, $MSe = 0.1388$, $p < .001$. Subjects were able to identify the CV syllables at a higher accuracy with the two talkers who were male and native English speakers. No significant difference was obtained between training conditions (discovery, expository) as a function of talker, $F(3, 54) = 0.506$, $p = .684$. In addition, the lack of a significant Talker × Training interac-

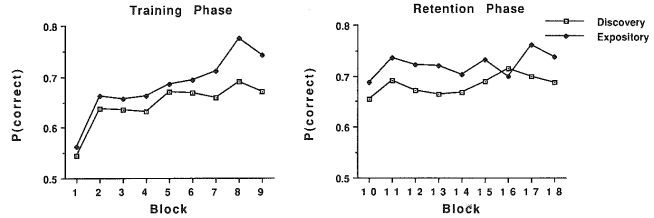

FIGURE 1. Mean proportion of correct identification of CV syllables as a function of instructional method (discovery, expository) for the training phase and the retention phase (9 blocks per phase).

TABLE 2. Proportion of correct responses obtained in two conditions of training and retention (retraining) CV identification.

	Block																	
	Training									Retention (retraining)								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Discovery condition:																		
S#																		
01	.51	.57	.64	.65	.65	.70	.69	.67	.63	.66	.75	.67	.63	.65	.63	.64	.64	.69
02	.47	.57	.62	.56	.67	.81	.61	.56	.58	.57	.57	.62	.60	.63	.62	.57	.63	.67
03	.54	.69	.71	.75	.72	.70	.68	.72	.75	.72	.78	.74	.70	.73	.72	.78	.86	.86
04	.46	.53	.52	.56	.51	.56	.51	.61	.54	.52	.54	.53	.59	.51	.61	.59	.64	.56
05	.68	.80	.69	.68	.74	.74	.64	.71	.75	.63	.70	.72	.74	.67	.72	.73	.77	.69
06	.50	.62	.62	.63	.60	.61	.62	.67	.59	.68	.63	.64	.58	.62	.64	.69	.69	.62
07	.66	.71	.69	.76	.82	.78	.72	.78	.73	did not return								
08	.61	.65	.66	.64	.76	.75	.77	.78	.77	.74	.76	.72	.70	.74	.71	.78	.75	.75
09	.48	.65	.64	.59	.61	.64	.64	.65	.65	.59	.68	.63	.69	.65	.61	.67	.62	.64
10	.64	.66	.61	.62	.77	.69	.76	.77	.77	.76	.82	.78	.71	.79	.75	.78	.83	.78
11	.66	.74	.75	.81	.80	.76	.82	.80	.83	no data collected[a]								
12	.42	.48	.49	.59	.49	.61	.57	.59	.60	no data collected[a]								
13	.41	.61	.59	.57	.65	.61	.62	.57	.62	no data collected[a]								
Expository condition:																		
S#																		
01	.58	.67	.65	.69	.69	.68	.69	.64	.71	.73	.71	.70	.76	.69	.73	.68	.80	.73
02	.39	.63	.68	.66	.74	.77	.77	.77	.82	.78	.79	.76	.80	.78	.76	.72	.79	.74
03	.24	.54	.47	.63	.58	.63	.67	.76	.70	.60	.67	.62	.66	.59	.71	.59	.69	.69
04	.53	.56	.59	.59	.54	.62	.59	.69	.70	.64	.64	.68	.68	.61	.68	.56	.69	.67
05	.62	.66	.64	.64	.72	.69	.67	.76	.76	.53	.70	.75	.67	.72	.69	.72	.81	.80
06	.67	.82	.78	.80	.80	.73	.76	.82	.78	.69	.78	.78	.83	.79	.83	.75	.85	.80
07	.62	.70	.68	.70	.61	.69	.71	.71	.75	.69	.83	.80	.78	.72	.77	.75	.74	.77
08	.68	.69	.73	.61	.68	.71	.69	.75	.70	.77	.76	.69	.65	.73	.69	.77	.76	.68
09	.67	.68	.67	.69	.70	.69	.78	.74	.74	.76	.76	.73	.67	.70	.73	.76	.72	.76
10	.62	.70	.69	.62	.80	.74	.78	.79	.75	did not return—computer malfunction								
11	.47	.54	.55	.67	.68	.68	.74	.63	.59	no data collected[a]								
12	.51	.66	.61	.70	.74	.68	.70	.67	.60	no data collected[a]								
13	.60	.67	.66	.73	.79	.66	.76	.78	.69	no data collected[a]								

[a]Subjects 11–13 were added in the Training phase only, in order to test for a lack of power.

tion indicated that the improvement occurred for the syllables from all of the talkers.

Figure 2 shows the identification accuracy for each of the 9 consonants and 3 vowels averaged over all subjects and blocks. Two ANOVAs, described earlier, were performed with instruction condition (2 levels), number of blocks (9 levels), number of vowels (3 levels), and number of consonants (9 levels) as factors. The dependent variable was, first, proportion correct on each consonant and, then, proportion correct on each vowel. As the visibility of the lip movements decreased (e.g., place of articulation moving from front to back), accuracy tended to decrease for the consonants, $F(8, 144) = 113.118$, $MSe = 0.0065$, $p < .001$. Performance on the three vowels was close to perfect, but the small differences were significant, $F(2, 36) = 15.386$, $MSe = 0.673$, $p < .001$.

Table 3 presents the consonant confusions during the 1st and 18th training blocks. The proportions along the negative diagonal represent correct responses. For most items, the accuracy on block 18 was higher than on block 1, demonstrating a decrease in consonant confusions as training progressed.

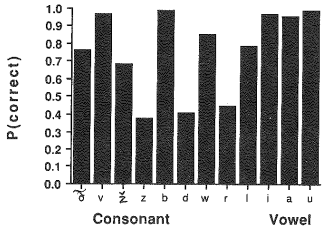

FIGURE 2. The mean proportion correct on each of the nine consonants (when responses are scored for consonant only) and each of the three vowels (scoring for vowel only), averaged across training method and blocks.

TABLE 3. Distribution of responses (in proportion) to each stimulus in the first and last block of training/retraining sessions (Blocks 1 and 18).

		Stimulus								
Response	Block	ð	v	ž	z	b	d	w	r	l
ð	1	.79	.03	.02	.01	.01	.03	.01	.01	.11
	18	.75	.00	.00	.01	.00	.04	.00	.00	.20
v	1	.01	.92	.01	.00	.01	.01	.01	.01	.01
	18	.00	.99	.00	.00	.01	.00	.00	.00	.00
ž	1	.11	.11	.46	.12	.01	.16	.00	.01	.02
	18	.00	.07	.62	.21	.00	.06	.00	.00	.04
z	1	.10	.14	.20	.23	.01	.25	.01	.03	.03
	18	.01	.05	.13	.49	.01	.26	.00	.00	.06
b	1	.00	.01	.00	.00	.98	.00	.01	.01	.00
	18	.00	.00	.00	.00	1.0	.00	.00	.00	.00
d	1	.15	.02	.13	.15	.03	.25	.03	.05	.19
	18	.09	.02	.06	.19	.00	.38	.01	.00	.25
w	1	.00	.01	.01	.00	.02	.00	.88	.08	.01
	18	.00	.00	.00	.00	.02	.00	.79	.19	.00
r	1	.01	.08	.01	.00	.07	.00	.57	.27	.00
	18	.00	.00	.00	.00	.03	.01	.50	.45	.00
l	1	.13	.02	.06	.05	.02	.08	.09	.06	.48
	18	.05	.00	.00	.01	.00	.06	.00	.00	.89

Retention Phase

Nine out of the 10 subjects from each original training condition returned for the second phase of the experiment. The right panel of Figure 1 shows the mean proportion correct for each block in the retention phase. Instruction condition (2), number of blocks (9), vowel (3), and consonant (9) were factors in the ANOVA. The main effect of training block indicated that subjects continued to increase their accuracy across the additional 9 blocks of the retention phase, $F(8, 128) = 4.458$, $MSe = 0.042$, $p < .001$. As in the learning phase, training instructions did not have an effect on performance, $F(1, 16) = 2.211$, $MSe = 0.940$, $p = .154$. To assess whether there was any performance loss over the 4-week interval between phases, performance averaged across the last 3 blocks of training was compared with average performance on the first 3 blocks of retention. Performance was .6995 and .6946 for these two conditions, respectively. As can be seen in Table 1, the ANOVA on these scores indicated no significant differences, $F(1, 16) < 1.0$, $MSe = 0.0165$, $p = .570$.

Transfer Phase

Figure 3 shows performance accuracy on the monosyllabic word test as a function of instruction condition. An analysis of variance was conducted on both proportion correct and reaction times with instruction condition (3 levels), word group (12 levels), and number of presentations (4 levels—the two presentations in every 20 trials were averaged, giving 4 means) of each word as factors (see Table 1). The discovery, expository, and no-training conditions did not differ from one another, $F(2, 22) = 2.470$, $MSe = 0.5740$, $p = .106$.

Figure 3 shows that performance improved across the eight presentations of the test words, $F(3, 66) = 81.547$, $MSe = 0.007$, $p < .001$. Instruction condition did not interact with presentation, $F(6, 66) < 1.0$. Although reaction time was not stressed to the subjects as being important, reaction time has been shown to be informative even if subjects are not instructed to respond as quickly as possible (Massaro, 1987, Chapter 4). Reaction times did not differ significantly as a function of condition, $F(2, 22) < 1.0$ (see Figure 3). Figure 3 also shows that reaction time decreased on subsequent presentations of a test word, $F(3, 66) = 87.409$, $p < .001$. Instruction condition did not interact with presentation, $F(6, 66) < 1.0$. Thus, all subjects learned across the repeated presentations of the test words.

Discussion

The present study contrasted two methods for teaching lipreading. Subjects trained to identify consonant-vowel syllables with the expository method were told where to look and what to look for, whereas the discovery method allowed subjects to discover the information on their own. Both groups of subjects learned over the course of training and retraining. There was no difference between the two types of training methods. It is possible that learning to lip-read is unaffected by such methods when feedback is given on each trial. This feedback may be sufficient for subjects to learn the important cues regardless of the method that is used. Thus,

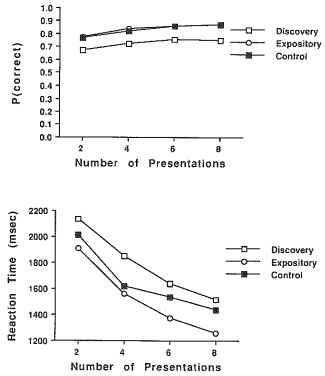

FIGURE 3. Top panel: The mean proportion of correct identifications of monosyllabic words as a function of instruction condition (discovery, expository, no training) and presentation (eight presentations, 2 presentations of each word per 20 trials over 80 trials [in a given block]). Bottom panel: The mean reaction time to the monosyllabic words as a function of instruction condition (discovery, expository, no training) and presentation (eight presentations, 2 presentations of each word per 20 trials [in a given block]).

it is important to determine the role that feedback plays in learning to lip-read.

Alternatively, if one hypothesizes that the learner benefits more from an explicit instructional method, our inconclusive findings could be attributed to subjects in the expository condition not being instructed in sufficient detail or to an insufficient number of training trials. The expository instruction lasted approximately 10 min on the first day of training; thereafter it was repeated only at a subject's request. It is possible that a more elaborate guided training and/or mirror practice would be more effective. Alternatively, it might be the case that time on task is essential for lipreading improvement, and that there is no ideal method for teaching lipreading. One simply cannot bypass the need for direct experience with identifying the items and obtaining feedback on performance. In the same way that we cannot easily instruct an infant or young child to perceive auditory speech, we might not be able to instruct adults how to perceive visible speech. Direct experience with the speech signal is required in both cases. Future research should continue to explore whether there are methods of teaching lipreading that can speed up the learning process.

Significant learning was revealed in overall accuracy as well as through a reduction in consonant confusions. For both instruction conditions, confusion errors (i.e., confusing one consonant with another—for example, /d/ with /l/; Walden et al., 1977) were high at the beginning of training; but as the training progressed, confusion errors decreased. Furthermore, confusions continued to occur among items with similar visual characteristics. One question of interest was whether a subject's performance could be any higher than the level obtained in this study. It is possible that some of the CV syllables are visually indistinguishable, and accuracy is as high as can be expected. The individual differences, however, indicate that there was room for additional improvement for at least some of the subjects.

The results also showed that the improvement in lipreading could be retained for 4 weeks. Both instruction groups retained their attained skill and demonstrated further learning during the retention phase.

We also tested whether the learning would transfer to the lipreading of larger units. Even after extensive training on CV syllables, subjects were no better at lipreading monosyllabic words than the group that had no training on CV syllables. We might conclude that brief CV training does not transfer to monosyllabic words or that feedback during testing is adequate training in itself. Training on vowel-consonant (VC) syllables as well as CV syllables might be necessary to obtain positive transfer because words contain both of these syllable types.

Several other reasons might explain the lack of transfer from CV syllables to monosyllabic words. Most importantly, only eight of the CVs used in the training were contained in the 120 test words. Further, the testing method differed in the transfer phase, again putting the experienced and naive subjects on more equal ground. A third possibility is the significant difference between the talkers during training and transfer test. The talkers used in the training task were college students with no articulation training, whereas the talker used in the transfer task was a trained actor.

Walden et al. (1977) found that most of the learning in their training study took place within the first few hours of training. The present experiment replicated this finding. It appears that most of the learning took place from the first block of trials to the second for both training groups. After these blocks, learning improved only gradually. This result seems to indicate that optimal lipreading skill will require a longer training period beyond the 3 to 6 hours provided here because improvement is so gradual after the initial training session.

In conclusion, the present study examined various factors in the instruction of lipreading for adults with normal hearing. Although no differences were found between expository and discovery types of training, lipreading of CV syllables improved across training and was retained for 4 weeks. Therefore, we conclude that certain aspects of lipreading performance can be improved with experience.

Acknowledgment

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), a James McKeen Cattell Fellowship, and the graduate division of the University of California, Santa Cruz. We thank Carol Lee De Filippo, Al Montgomery, and an anonymous reviewer for comments on the article.

References

Alcantara, J. I., Cowan, R. S., Blamey, P. J., & Clark, G. M. (1990). A comparison of two training strategies for speech recognition with an electronic speech processor. *Journal of Speech and Hearing Research, 33,* 195–204.

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

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31,* 21–32.

Egan, D. E., & Greeno, J. C. (1973). Acquiring cognitive structure by discovery and rule learning. *Journal of Educational Psychology, 64,* 85–97.

Greene, B. G., Pisoni, D. B., & Carrell, T. D. (1984). Recognition of speech spectrograms. *Journal of the Acoustical Society of America, 76,* 32–43.

Hunt, A. P. (1975). A comparison of discovery and reception learning in children at varying operational learning stages. *Dissertation Abstracts International, 36,* 3509A–3510A.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874–886.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

McLeod, D. B., & Adams, V. M. (1981). Locus of control and mathematics instruction: Three exploratory studies. *Journal of Experimental Education, 49,* 94–99.

Norton, R. F. (1977, April). *The effects of practice item variety, practice strategy, and training mode on performance in a rule-finding task.* Paper presented at the meeting of the American Educational Research Association, New York.

Solter, A., & Mayer, R. E. (1978). Broader transfer produced by guided discovery of number concepts with preschool children. *Journal of Educational Psychology, 70,* 363–371.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59–89.

Walden, B. E., Prosek, R., Montgomery, 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.

Received April 1, 1991
Accepted March 17, 1992

Contact author: Dominic W. Masaro, PhD, Program in Experimental Psychology, University of California, Santa Cruz, CA 95064.

Appendix: Instructions and Related Information Given to the Subjects

Discovery Condition. This experiment is designed to teach you how to lip-read. One of four speakers will appear on the screen in front of you, and you will be asked to read the speaker's lips. Each trial will start with a warning tone; then a speaker will say a syllable silently. You are to watch the speaker closely and try to lip-read which syllable the speaker pronounced. The syllable will be one of 27 syllables that have been labeled on this terminal keyboard (demonstrate).* Please be as accurate as possible. Try to respond as soon as you arrive at a decision, avoiding any excess delay; a few seconds should suffice. On some of the trials you may be unsure of what the speaker has said; simply make your best guess. It is important that you respond on every trial; you will have 5 sec to respond. After you have made your response, feedback will be given. The feedback will consist of the speaker repeating the syllable with the sound on. You should continue to pay close attention to the speaker's lips as well as listening to the correct answer. After the feedback has been presented, the next trial will begin. Do you have any questions?

Expository Condition. The preceding paragraph up to the * was read to this group. The examiner then proceeded with the following: Watch my lips as I pronounce each syllable. First concentrating on the 9 different consonants, I will use one vowel with each consonant to demonstrate the lip movements. Notice when I say /ð/ my tongue is between my teeth; when I say /vi/ my teeth grab my lower lip; when I say /ži/ my mouth opens wider than when I say /zi/; when I say /bi/ my lips are pursed changing to open mouth; when I say /di/ the mouth opens rounded; when I say /wi/ the mouth protrudes and changes to open; /ri/ is like /wi/ but it doesn't require as much protrusion; when I say /li/ it looks similar to /di/ but with wider mouth. Now I will demonstrate this with each of the other two vowels (i.e., /a/ and /u/). (After demonstrating all 27 syllables, continue with the following instructions.) You may have noticed that when I used the vowel /i/ my lips were spread and my teeth were close together; for example /ði/, /vi/, and /ži/. The vowel /a/, on the other hand, causes the mouth to drop; for example, /za/, /ba/, and /da/. With the vowel /u/ the lips are protruded and rounded; for example, /wu/, /ru/, and /lu/. Further, when a consonant is followed by a vowel this will alter the visual characteristics of the consonant as well; for example, the rounded vowel /u/ makes all of the consonants protrude to some extent. Was this demonstration clear? Do you have any questions? Now if I can have you direct your attention to the screen, you will notice we have outlined the distinguishing characteristics for each consonant and vowel. You are to use this as a guide when making your choices for the syllable to be lip-read. Again, do you have any questions? [Continue with instructions as for Discovery Condition after *.]

Subjects were given additional cues in addition to the above written instructions. For example, when /l/ was demonstrated the experimenter also brought to the subjects' attention that the tongue position is like the tongue flapping against the front top of the ridge of the mouth and is often visible (or at least more so than for /d/); whereas when /d/ is produced the tongue is rolled back a bit to have the tip of the tongue hitting the hard (middle) part of the alveolar ridge. We also characterized the word *pursed* as being closed. For example, when we described /b/ we specifically added that the lips come together. Also, when we used the word *rounded* we explained we were referring to both vertical and horizontal opening. We described the term *protruded* as being similar to a puckering movement with the lips. Although it was not a systematic or requested task, subjects often produced the CV syllables along with the experimenter in order to feel the lip and tongue movement.

A short description of the phonetic and visual characteristics for each consonant and vowel was displayed on a terminal screen for each subject in the expository condition to view at any time during the experiment. The descriptions were written in the following manner: th: tongue between teeth; v: teeth over lower lip; zh: like Z, mouth opens wider; z: mouth opens like D but slower; B: lips pursed changing to open; D: mouth opens rounded; W: mouth protruding changing to open; R: like W but less protruding; L: like D but wider mouth; EE: teeth close together, lips spread; AH: lips apart, medium wide; UU: lips protruding, rounding.