

Electromyography and lipreading in the detection of verbal rehearsal

JOHN L. LOCKE

Yale University, New Haven, Connecticut 06512

and

MICKEY GINSBURG

University of Connecticut, Storrs, Connecticut

Adults memorized lists of visually presented consonant letters while monitored simultaneously by a lipreader and an electromyograph (EMG) for rehearsal activity. Lipreading was more sensitive than EMG in detecting who was verbally rehearsing.

Verbal rehearsal is a common and telling strategy in memorizing words or word-like material. What gets rehearsed, and how it gets rehearsed, are important pieces of information; they can permit fairly accurate estimates of what later may be retrieved either from short-term or long-term storage. It also is important, then, to perfect methods of scrutiny which show whether a given subject is a rehearser, and more precisely, when he is verbally rehearsing.

A method which has been used in several experiments recently involves the use of electromyography (EMG). In this work (Locke & Fehr, 1970a, b; 1972; Goldstein, Locke & Fehr, 1972), surface electrodes are attached to speech-muscle areas likely to be active in subjects' rehearsal of certain classes of stimuli. With an electrode situated between the chin and lower lip, adults and older children have consistently produced greater microvolt values when learning a list of words or pictures with labial names (e.g., baby, favor) than with nonlabial names (e.g., ticket, hugging). Though the method has proved useful, there are several disadvantages. One is that it is hard to tell whether a particular subject is rehearsing a particular list, though he may, as the result of receiving multiple lists, provide enough tracings to suggest that he did or did not rehearse *most* of the lists. That is a measurement problem, and there is in our minds no reason to assume that this could not be rectified. Probably the more serious problem with EMG is that many of the people interested in rehearsal (a) do not have an electromyograph or (b) have one, but cannot easily use it with some of the more interesting populations with respect to rehearsal (e.g., the deaf, the

retarded, the speech-disordered), who often are located miles from the laboratory. So it makes sense to ask if one can get similar information in a more practical way.

Flavell and his associates (Appel, Cooper, McCarrell, Sims-Knight, Yussen, & Flavell, 1972; Flavell, Beach, & Chinsky, 1966; Kenney, Cannizzo, & Flavell, 1967) watch the subject's mouth during stimulus presentation and attempt, through lipreading, to determine whether he is articulating stimulus names. Since they have observed stimulus-relevant articulation to increase with age, with the amount recalled and with the instructions, one might conclude that the method works, for their purposes at least.

The main purpose of this study was to determine which of the two methods, lipreading and electromyography, provide for more accurate predictions of who is rehearsing what. It should be understood at the outset that both methods were used in a particular way in this experiment, and that either may give different results depending on how and with whom it is used. Less central to the experiment, but still of interest, was the level of recall and degree of speech coding shown by the subjects who obviously rehearsed

This project was supported, in part, by a National Institutes of Health Fellowship (1 F03 HD-53445-01) and Grant HD-05951, both from the National Institute of Child Health and Human Development. The authors acknowledge the assistance of Ella Futrell in preparing the manuscript for publication. This paper is sponsored by J. C. Campione who takes full editorial responsibility. John L. Locke is now at the Speech and Hearing Laboratory, Children's Research Center, University of Illinois, Champaign, Illinois 61820. Requests for reprints should be addressed to the senior author at that address. Mickey Ginsburg is now at Denver General Hospital, Division of Psychiatry, West Eighth Avenue and Cherokee Street, Denver, Colorado 80204.

relative to those who apparently did not.

METHOD

Subjects

The subjects were 18 undergraduate females who served voluntarily and without pay or other formal reward. They were told the experiment was concerned with the physiological correlates of memory.

Stimuli

Memory stimuli were 24 lists of eight upper-case consonant letters. Each letter list contained two ensembles of four letters. An easily observed (EO) ensemble, PBMF, was considered highly lip-readable when articulated silently. That is not to say the individual letters can be easily discriminated one from the others, but that the set can be discerned from the less observable (LO) set of acoustically similar cognates, TDNS. Ideally, stimulus letters in an experiment designed to elicit rehearsal would not rhyme, since that might discourage subjects from rehearsing. However, the experiment also was interested in the degree to which phonetic recording was related to speech rehearsal, and it was presumed that the presentation rate would be sufficiently fast to minimize alternate strategies. Each of the two four-letter ensembles were arranged into their 24 possible permutations and randomly assigned to a list. A second ordering procedure assigned EO letters either to the first or second half of the eight-item lists, with the LO set occupying the remaining half. A constraint on this second procedure was that LO and EO sets could not initiate a list more than three times in succession. The four carousels of 35-mm slides, with six lists to a carousel, were presented in a different order to each subject. In summary, then, each subject saw 24 eight-letter lists in which the EO and LO four-letter sets occupied the first and second halves of the list exactly half the time and were themselves in 24 different orders.

Procedure

The subjects were tested individually, seated approximately 3 ft from an 18 x 24 in. white screen. The experimenter and the EMG recording equipment were situated about 9 ft to the left and behind the subject. A grass silver chlorided surface electrode, 9-mm in diam, was placed with electrode cream and clear, surgical tape between subject's chin and lower lip; another was laterally displaced about 2 in. to the left. A ground was placed on the left ear lobe and dummy electrodes (used to give the subject the impression that her whole body was being monitored) were situated on the ankle and forearm. EMG recordings were obtained with a Beckman Type R.S. dynagraph, Type 462 amplifier, Type 461B preamplifier and Type 9852A integrator.

The subjects were asked to write the eight letters in ordered,

Table 1
Number of Lists at Each Observer-Assigned Confidence Interval and Percentage of Those Lists in Which Gross Characteristics of the Stimulus Input Could be Identified Correctly from EMG and Observer Recordings

	Observer Confidence Rating				
	Low 1	2	3	4	High 5
Number of Observer Judgments	161	47	92	55	77
Percentage Observer Hits	53	78	91	100	98
Percentage EMG Hits	50	58	60	64	77

left-to-right fashion, beginning immediately following the last letter in a list. Each list was to be recalled on a separate page of the subject's response booklet on the eight spaces provided. Stimulus exposure time was 350 msec with interstimulus intervals of 700 msec. A blank (ninth) slide, arriving in time with the stimuli, signaled the 10-sec written recall period. A STOP WRITING slide signaled the end of the write-down period, followed by a READY slide and the next list.

The slide-projected letters, which measured approximately 3/4 x 1/2 in. on the screen, were perceived perfectly by four pilot subjects who read them aloud as they appeared. Stimuli were projected by a Kodak Extographic Model AF slide projector. A second channel of the dynagraph simultaneously recorded the slide onsets and offsets and write-down period. This channel and the projector were driven by pulses fed by a Sony Model TC-500 tape recorder through a voice-activated relay.

Two female undergraduates were trained as observers and were instructed to determine whether the EO or LO set came first, rating their confidence on a five-point scale. Training for the observers consisted of scoring three pilot subjects with immediate feedback on accuracy for each list. The observers sat in an adjoining room behind a one-way mirror. They had a clear view of the subject, a nearly head-on, 10-20-deg angle, but could see neither the slide screen or subject's response booklet. Audio pulses on the second channel of the cueing tape recorder indicated to the observers the onset of the first and fifth slides and the beginning of the subject's response period. By listening through a headset, then, the observer was perfectly aware of when the subject was viewing the first and second halves of each list. Each observer worked alone, observing nine sessions. None of the 18 subjects, in subsequent questioning, expressed the suspicion that she was being observed, or that the experiment dealt with anything (e.g., rehearsal) but the stated purpose.

RESULTS AND DISCUSSION

Observer Accuracy

Observers registered correct judgments on 78% of the 432 trials (24 lists by 18 subjects). Table 1 shows the number of observer judgments at each of the five confidence intervals and observer accuracy at each interval. An observer "hit" here does not mean, in the strictest sense, correct detection of a rehearsal, because the subjects may not have rehearsed on every list. A correct detection means that the observer, by looking at subject's lips, figured out what must have been on the screen at the time. If the subjects rehearsed on only those trials where observer was correct, it could be said that observer-detection was perfect, i.e., 100% accurate. There is no desire on our part to say this, only to point out that "accuracy" here is a reflection not only of the observation method but of the likelihood that *these subjects* would rehearse *these* stimuli at the prevailing input rates, task requirements, and so forth.

Rank correlation analyses revealed a highly significant relationship ($z = 3.64$; $p < .001$) between the observer accuracy and confidence estimate. If the 161 confidence ratings of 1 are omitted, their accuracy increases from 78% to 93%. However, as we suggested above, this procedure only increases their accuracy in determining what was on the slide screen; ratings of 1 may mean, in

fact, that the subjects did not rehearse many of those 161 lists.

The two observers were individually correct on 72% and 84% of the lists. When rehearsal difference in their separate groups of subjects were controlled, on the basis of EMG records, the two observers were brought into perfect agreement at 84%.

EMG Analysis

EMG tracings from two pilot subjects who were instructed to mouth the EO set and to observe passively the HO set showed that list halves functionally were divided at the end of the interval following the fourth slide. That point is where the observer heard the pulse marking the end of the first half of lists. Test tracings were analyzed by assessing the magnitude (in microvolts) of pen excursions for first and second halves of each list. Since even at rest there is a certain amount of vegetative, nonspeech activity, it was important to count only those deflections which reasonably could be attributed to subvocal rehearsal. A criterion, a 10-mV increase within 200 msec, was formulated from a rough comparison of test and nontest tracings. Application of that criterion simply meant that a conforming peak would be assessed, the value being determined by the microvoltage it reached at its maximum. The figure entered for each list half was the sum of microvolt values for all peaks meeting criterion.

EMG Accuracy

Table 1, in addition to reporting observer data, shows EMG "hits," cases where the sum of the measured peaks was greater for EO (i.e., labial) than for LO (i.e., nonlabial) list halves. The percentage of hits is shown at each of the observer-registered scale points, the data base for each being the number of observer judgments at that confidence interval. For example, where the observers assigned a confidence estimate of 1 on 161 occasions, their accuracy was 53%, the EMG accuracy being 50% for those 161 lists.

EMG tracings, as scored here, produced 60% correct predictions where 123 flat (hence unscorable) tracings were assigned 50% accuracy. With these ambiguous trials excluded, EMG accuracy rises to 64%. It is interest to note that, while the EMG measure was 50% correct, by definition, in the case of the 123 flat patterns, it also was 50% correct for the 161 trials in which the observers assigned a confidence rating of 1. The observer was correct on 56% of the 123 flat EMG tracings and correct on 53% of the 161 trials receiving a confidence rating of 1. When the two observers were not especially confident of their judgments, no meaningful EMG patterns presented: 78% of the 123 flat patterns occurred on trials which received a confidence rating of 1, 15% of "2" trials, 5% on "3" trials, and 1% on "4" and "5"

trials.

It appears that both measures increase in accuracy as observer confidence increases, both methods apparently tapping the same general behavior ($r = .64$; $p < .01$). However, it is clear that when one wants to know whether and on which trials the subject is subvocally rehearsing, the observer's judgment is the more accurate measure given our paradigm, electrode placement, tracings analysis, and so forth. EMG was conservative; it, however, is not insensitive. On the contrary, the primary weakness may be an indiscriminate and highly sensitive detection of all activity in the area of the electrode placement. The observer can reject nonspeech activity, EMG accepts everything, and there are no characteristic nonspeech patterns on an EMG tracing that we easily can recognize and confidently discard.

Whether visual observation would retain its apparent advantage in child work is not clear. It is precisely these subjects who would give the least evidence of rehearsal, and in whom a retention interval is most important. That is to say, children of 4 and 5 years of age are likely to articulate stimulus names as they see them, but not during a subsequent period in which the stimuli are absent (Locke & Fehr, 1970b). Since the retention interval is where adult activity increases (Locke & Fehr, 1970a), rather than decreases, measurement of rehearsal during this period is critical. The question here is whether EMG can detect rehearsal where the subject's lips are motionless, as may be the case under these conditions. In any case, the high accuracy of observers in this study suggests that the term "covert" may be less accurate in describing adults' rehearsal than "intrapersonal," or some other term which carries the implication that the activity is not done for public benefit, though it may be observed by others.

REFERENCES

- Appel, L. F., Cooper, R. G., McCarrell, N., Sims-Knight, J., Yussen, S. R., & Flavell, J. H. The development of the distinction between perceiving and memorizing. *Child Development*, 1972, 43, 1365-1381.
- Conrad, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, 193, 1314-1315.
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. Spontaneous verbal rehearsal in memory task as a function of age. *Child Development*, 1966, 37, 283-299.
- Goldstein, J. I., Locke, J. L., & Fehr, F. S. Children's prerecall phonetic processing of pictures and printed words. *Psychonomic Science*, 1972, 26, 314-316.
- Keeney, T. J., Cannizzo, S. R., & Flavell, J. H. Spontaneous and induced verbal rehearsal in a recall task. *Child Development*, 1967, 38, 953-966.
- Locke, J. L., & Fehr, F. S. Subvocal rehearsal as a form of speech. *Journal of Verbal Learning and Verbal Behavior*, 1970a, 9, 495-498.
- Locke, J. L., & Fehr, F. S. Young children's use of the speech code in a recall task. *Journal of Experimental Child Psychology*, 1970b, 10, 367-373.
- Locke, J. L., & Fehr, F. S. Subvocalization of heard or seen words prior to spoken or written recall. *American Journal of Psychology*, 1972, 85, 63-68.

(Received for publication December 5, 1974.)