Exploring long-term modality effects:
Vocalization leads to best retention

SUSAN E. GATHERCOLE and MARTIN A. CONWAY
MRC Applied Psychology Unit, Cambridge, England

Five experiments employing an incidental learning procedure explored the effects of different input modalities on the long-term retention of word lists. In each of the first four experiments, mixed-mode presentations featuring three different modes of presentation were employed. In the baseline presentation mode in each experiment, subjects were to read words silently. In the two additional presentation modes, subjects were to vocalize or read and hear (Experiment 1), read and hear or mouth (Experiment 2), vocalize or write (Experiment 3), and vocalize or write without seeing the written word (Experiment 4). In Experiment 5, separate groups of subjects were presented with pure-mode lists that they were to read silently, write without seeing, write and see, mouth, hear, read and hear, or vocalize. The principal findings were that auditory presentation procedures led to best memory performance and that, of these, only vocalization was found to consistently enhance retention. These findings are conceptualized within a framework that proposes that both the temporal distinctiveness of auditory information and self-generated cues are employed in the process of retrieval.

Recent research has demonstrated that the long-term retention of words may be determined by modality of input. For instance, Conway and Gathercole (1987) found superior retention of words that were spoken aloud by subjects at input compared with retention of words that were silently read. This vocalization advantage extended throughout the input list, and occurred despite the interpolation of spoken instructions between presentation and test. Moreover, Conway and Gathercole found similar memory advantages for words that were heard at input compared with words silently read. The aim of the present study was to investigate these through-list long-term modality effects in further detail, and in particular to explore the role of acoustic information in elevating memory performance.

The finding of a persistent vocalization advantage in incidental learning provides a major challenge for current conceptualizations of modality and memory, which either do not predict modality differences in long-term memory (e.g., Watkins & Watkins, 1980) or can accommodate differences only in the recency portion of the memory list (e.g., Glenberg, 1984). Conway and Gathercole (1987) speculated that the notion of perceptual distinctiveness may be extended to provide a general framework for these modality differences, but, currently at least, this conceptualization is not sufficiently well specified to yield any strong predictions.

In the absence of any specific model of the beneficial effects of vocalizing over silent reading in incidental learning, it is important to empirically evaluate the critical components of the act of vocalization that give rise to this phenomenon. In this paper we report a series of experiments in which we assessed whether it was the presence of acoustic information, the execution of motor actions, or a combination of the two that were critical to the retention differences.

Some evidence related to these issues was reported by Conway and Gathercole (1987), who found that silently mouthing a printed word failed to give subjects any consistent memory advantage over simply reading a word silently. This finding suggests that the act of articulation itself is not the critical component of the vocalization advantage. On the other hand, there was superior incidental learning of words that were spoken by the experimenter, compared with retention of words that were read silently. On the basis of this combination of results, it might appear that it is simply the presence of acoustic information that is critical to these long-term modality differences, although comparisons across experiments indicated better retention of vocalized than of heard words. Conway and Gathercole (1987) did not, however, directly compare memory for vocalized words with memory for words that were only heard.

Clearly, it is crucial to our understanding of these modality differences that we establish whether the memory advantage is a product of acoustic information *and* an explicit motor action or whether the memory advantage can simply be localized to the presence of acoustic information only. Therefore, in Experiment 1 of the present study, we compared recognition of words that were silently read, silently read and heard, and vocalized by subjects at input. If both an active input response and acoustic infor-

mation enhance memory performance, then memory for vocalized words should be superior to memory for passively heard words. On the other hand, if it is merely the presence of acoustic information that is important, and explicit actions are not critical, there should be no difference in memory for vocalized over heard words.

A finding that vocalization led to superior memory performance in Experiment 1 would not necessarily indicate that passively presented acoustic information has no beneficial effect upon memory whatsoever. In order to examine the effects of passive acoustic input in more detail, we decided to compare memory for words that were heard and read silently with memory for words that were mouthed and read silently (Experiment 2). Conway and Gathercole (1987) found that hearing and reading silently led to better memory performance than reading silently, but found no such advantage for mouthing; therefore, we expected to find an advantage for hearing and reading silently over mouthing and reading silently in Experiment 2, demonstrating that even passive acoustic information may lead to comparatively high levels of retention.

Experiments 3 and 4 focused on the explicit action component of vocalization and investigated whether the execution of a nonspeech motor action that specified the to-be-remembered item might also enhance incidental learning. In these experiments the effects of vocalizing versus writing a word were compared. The question was, Does writing give rise to levels of retention similar to those following vocalization? Previous investigations of long-term modality effects had not compared memory for written and for silently read material, but intuitively, at least, it seemed quite likely that writing would indeed confer a memory advantage, if only because it might guarantee attention to the visually presented word. Thus, direct reference to the speech-based characteristics of the to-be-remembered words might not be critical to the vocalization advantage.

A finding that acoustic presentation procedures did in fact lead to best retention would not necessarily indicate that some property of acoustic information per se determines this memory advantage. Rather, it may be that acoustic presentation procedures have a beneficial effect on retention only in the context of other presentation modalities. Thus, it may be that the acoustic modality effect arises because acoustic information is *relatively* distinct, in comparison with other input modalities (see Conway & Gathercole, 1987, for further discussion of this point). This possibility was explored in a final experiment, in which we compared memory for all the input conditions in Experiments 1–4—silently read, mouthed, vocalized, read and heard, written (seen), and written (unseen)—using pure lists rather than the mixed-list presentation employed in Experiments 1–4. The aim of Experiment 5, then, was to examine whether acoustic presentations conferred a general memory advantage or whether the advantage was restricted to mixed-mode presentations.

A further measure of the effects of input modality was also taken: after making recognition (old/new) judgments, the subjects in Experiments 1–4 were required to judge the input condition of the words that they recognized as old. A similar measure was employed by Conway and Gathercole (1987), who found that their subjects were biased toward judging the false positives—new words incorrectly judged "old"—as having been read silently, and were very unlikely to judge them as having been vocalized. Our interpretation of this phenomenon is that it reflects the application of metamemory knowledge, which associates active input responses, such as vocalization, with good retention, and passive input responses, such as silently reading, with poor retention. If this metamemory heuristic is generally employed when people make judgments about the modality of a previously experienced event, then we would expect to observe a similar pattern of findings in the present study.

EXPERIMENT 1

The purpose of this experiment was to examine whether an active acoustic presentation procedure—having subjects read a word and say it aloud—would give rise to better memory performance than would a passive acoustic presentation procedure, such as having subjects read and hear a word. In the baseline input condition, the subjects read a word silently. If reading and hearing were to lead to poorer memory performance than vocalizing, this would demonstrate that the vocalization action as well as the resulting acoustic information contributed to the vocalization advantage observed by Conway and Gathercole (1987).

Method

Subjects. Separate groups of 18 subjects each participated in each experiment. The subjects were Open University students attending a residential summer school, and their ages ranged from 18 to 45 years. The subjects were randomly allocated to experiments, and each subject was tested individually.

Design. Each subject received the same set of 30 words during the presentation phase of the experiment. Three input conditions were tested—"read silently," "read and vocalize," and "read and listen." Ten words were presented in each of these conditions. Three different pseudorandom sequences of input conditions were employed. In each of these sequences, the 30 words were presented in the same predetermined random order. Within each sequence, the 10 words assigned to each input condition appeared in a different input condition in each of the remaining two orders. Each ordering was presented to a different group of subjects. Thus, across all subjects tested in the experiment, each word was tested equally often in each input condition.

Following the presentation phase of the experiment, the subjects were given a printed questionnaire containing the 30 words presented in the presentation phase and 30 filler words, presented in an unsystematic sequence. The subjects were to judge each of these 60 words as "old" or "new" and to provide a rating of confidence for each "old" judgment. The subjects were also asked to judge whether each "old" word had been read silently, said aloud, or read and heard, and to provide a confidence rating for this "input judgment."

Materials. The 60 four-letter words employed were all highly imageable, with a minimum imagery rating of 4.00 (Paivio, Yuille, & Madigan, 1968) and a mean imagery rating of 5.28 ($SD = 1.52$). All words were frequent, with a minimum frequency of 50 per million and a mean frequency of 78 per million, according to Thorndike and Lorge (1944).

Apparatus. The experimental stimuli in the initial phase of the experiment were presented on a videocassette recording shown to subjects on a black-and-white video monitor. All words were displayed in white against a black background. The acoustic information for the read-and-heard stimuli had been recorded synchronously with the corresponding visual display.

Procedure. At the beginning of the experimental session the subjects were seated in front of the video monitor and were told that they would see a series of words, presented one at a time, each of which would be preceded by an instruction. If instructed to "say aloud," the subjects were to say the word clearly in a normal speaking voice, loud enough for the experimenter to hear. If instructed to "read and listen," the subjects were to read the word silently and listen to the spoken word. If instructed to "read silently," the subjects were to repeat the word silently to themselves without making any mouth movements. The importance of complying with the presentation instructions was emphasized.

In order to minimize intentional learning, the subjects were told a cover story. They were led to believe that they were a normal control group piloting stimuli to be used with brain-damaged patients. No further information about the purpose of the experiment was provided to the subjects at this stage.

On each of the 30 trials of the presentation phase, the presentation instruction was displayed on the screen in lowercase letters for 2 sec, followed by a blank screen for 1.5 sec. The word that was to be said aloud, read silently, or read and heard then appeared on the screen in uppercase letters for 2.5 sec. A blank interval of 2 sec preceded the beginning of every trial. The duration of the presentation phase of the experiment was about 3 min.

The subjects were then provided with written instructions. These instructions informed the subjects that they were to complete a written questionnaire containing the 30 words they had just seen on the monitor plus 30 filler words, and that their main task was to identify the words that they had seen as "old" and the fillers as "new." The subjects were to make the recognition (old/new) judgments for the words in the sequence in which they appeared on the questionnaire. After judging a word as new, the subject was to proceed to the next word. After judging a word as old, the subject was to provide a rating of confidence in his/her judgment by circling 1 (*very confident*), 2 (*mildly confident*), or 3 (*uncertain*) on the questionnaire. Then the subject was to make an input judgment for the old word by circling "read silently," "said aloud," or "read and heard" (guessing if necessary), and to indicate a confidence rating for this judgment by circling 1, 2, or 3, as before. After the subjects had read these instructions, the experimenter verbally summarized them and ascertained that the subjects fully understood the procedure. This phase of the experiment took between 7 and 10 min.

After completing the questionnaire, the subjects were informally asked by the experimenter whether they had expected the memory test. Over the entire series of experiments only 12 subjects—2 in Experiment 1, 1 in Experiment 2, 3 in Experiment 3, 1 in Experiment 4, and 5 in Experiment 5—claimed to have suspected during the presentation phase of the experiment that there would be a memory test. Of these, none claimed to have actively rehearsed the items in the input list, and memory performance for these subjects was well within the range of memory performance for the subjects who had not anticipated the memory test.

Results

The two principal measures of memory performance taken in this experiment were recognition and input con-

dition judgments. The recognition data are considered first. For each subject, the number of words presented in each input condition that were correctly recognized as old was calculated. The results are summarized in Table 1. Recognition was best for vocalized words (86%), intermediate for read-and-heard words (75%), and poorest for read-only words (66%).[1]

The number of new words incorrectly identified as old was also calculated for each subject, and was used to compute d' values for each of the three input conditions. The mean false positive rate was 22%. The mean d' values in each condition are shown in the second column of Table 1. The results corresponded closely to the recognition data, showing greatest sensitivity to the vocalized words ($d' = 1.94$), lower sensitivity for the read-and-heard words ($d' = 1.59$), and lowest sensitivity to the read-only words ($d' = 1.33$). An analysis of variance was performed on the d' values of each subject in each of the three input conditions, and this showed a significant effect of input condition [$F(2,34) = 8.157$, $MSe = .208$, $p < .01$]. In this analysis, as in all other analyses reported in the present paper, planned comparisons of condition means were performed to explore main effects, using the pooled error term (Winer, 1971) for all contrasts; for all significant comparisons, $p < .05$ unless stated otherwise.

The planned comparisons showed that sensitivity was significantly higher to vocalized words than to either read-and-heard words or read-only words. The greater sensitivity to words that were both read and heard than to those that were read silently was significant only at the 10% level.

The percentage of correct input judgments for words correctly recognized as old was calculated for each subject in each of the three input conditions; these accuracy data are summarized in the final column of Table 1. An analysis of variance performed on the percent accuracy of each subject's input judgments in the three input conditions showed that the variations in input judgment accuracy across the different input conditions were not reliable [$F(2,34) < 1$].

Separate analyses were also made of the frequency of the three input judgments in each input condition for each subject, although, because few systematic biases emerged in these data, they are not fully reported here.[2] In all three cases, correct input judgments were significantly more frequent than either of the possible incorrect judgments. Only for the read-and-heard words was there any significant bias in the frequency of input judgments, and for correct recognitions 36% were judged to have been read only, whereas only 12% were judged to have been vocalized.

Table 1
Mean Percentage of Words Correctly Recognized in Each Input Condition, Mean Sensitivity Values (d'), and Mean Percent Accuracy of Input Judgments in Experiment 1

Input Condition	% Recognized	d'	% Accuracy
Vocalized	86	1.94	57
Read and heard	75	1.59	52
Read only	66	1.33	61

When the input judgments for the false positive items are considered, a clear pattern emerges that corresponds closely to that observed by Conway and Gathercole (1987). Of the false positives, the majority (68%) were judged to have been read only, 25% to have been read and heard, and only 7% to have been vocalized. An analysis of variance performed on the number of false positives attributed to each of the three input conditions for each subject showed the effect of judged condition to be significant $[F(2,34) = 26.398, MSe = 3.310, p < .001]$. Judgments of "read only" were significantly more frequent than either "vocalized" or "read and heard" judgments, although the greater number of "read and heard" than "vocalized" judgments for the false positives was only marginally significant ($p = .052$).

It was found that the differences in subjects' confidence ratings simply corresponded to the pattern of differences for recognition and input judgments. (This was also the case in Experiments 2-5; therefore, inasmuch as these data do not add anything to the findings, they are not reported.)

Discussion

Recognition of words that were vocalized by subjects at input was superior to recognition both for words that were read and passively heard and for words that were read only. Thus there was a beneficial effect on incidental learning of active over passive vocalization. These results have important implications for interpretation of the results reported by Conway and Gathercole (1987). In that study, it appeared that it was simply the presence of acoustic information that gave rise to the advantage of vocalization over reading silently. The present results indicate that although passive exposure to acoustic information may enhance memory performance, its beneficial effect is somewhat inconsistent, and certainly not of the same magnitude as the beneficial effects of vocalizing. Therefore, it seems that memory performance in the present task is enhanced by both the presence of acoustic information *and* the explicit action of vocalization.

Interestingly, the assignment of input judgments for false positives—the new words incorrectly judged to be old—corresponded in some respects to the recognition accuracy of the different input conditions. Subjects were most likely to judge the false positives as having been read only, which was the least-well-remembered input condition. Conversely, subjects were least likely to judge the false positives as having been vocalized, the input condition that led to the best recognition performance. Thus it appears that the association between readily remembered items and vocalization may be explicitly represented at a metamemory level.

Experiment 2 provided a further test of the hypothesis that both the act of vocalization and the presence of acoustic information are critical to the vocalization advantage. In Experiment 2, subjects (1) both read and heard, (2) silently mouthed, or (3) silently read words. If the combination of action and acoustic record were the criti-

cal factor in a reliable incidental learning advantage over silent reading, there should be no benefit of reading and hearing over mouthing. On the other hand, if acoustic information alone were sufficient to enhance performance, retention should be better for read-and-heard than for mouthed words. If acoustic information were sufficient to enhance performance, the difference between vocalized and read-and-heard words in Experiment 1 could simply be attributed to differences in the acoustic quality of self-generated and heard speech, the former possibly being more distinctive than the latter.

EXPERIMENT 2

Method

The method was the same as in Experiment 1, with the following exceptions.

Design. The three input conditions used in this experiment were "read and listen," "mouth silently," and "read silently." The same design and materials were used as in Experiment 1, except that in each of the three different input lists, the words that had been spoken by the subjects in Experiment 1 were now silently mouthed. The recognition questionnaire was amended so that the choices for input judgments for old words were now "read and heard," "mouthed silently," and "read silently."

Procedure. In all critical respects, the procedure was the same as that followed in Experiment 1. Before each experimental session began, the subjects were told that when they were instructed to "mouth silently" it was crucial that they mouth the word in as natural a way as possible and that they generate no acoustic information. In order to ensure that these instructions were followed, the experimenter sat in full view of the subjects and monitored their mouthing actions.

Results

The mean percentage of words recognized in each input condition in Experiment 2, the corresponding d' scores, and the mean percent accuracy of input judgments are shown in Table 2. Recognition performance was highest for read-and-heard words (76%), intermediate for mouthed words (69%), and lowest for read-only words (62%). The d' scores revealed the same pattern, demonstrating highest sensitivity for read-and-heard words, and poorest sensitivity for read-only words. The analysis of variance performed on the d' values for each subject in each of the three input conditions showed a significant effect of input condition $[F(2,34) = 5.391, MSe = 0.154, p < .01]$. This reflects the greater sensitivity to read-and-heard than to read-only words. There was no significant difference between read-and-heard and mouthed words, and the greater sensitivity to mouthed than to read-only words was significant only at the 10% level.

Table 2
Mean Percentage of Words Correctly Recognized in Each Input Condition, Mean Sensitivity Values (d'), and Mean Percent Accuracy of Input Judgments in Experiment 2

Input Condition	% Recognized	d'	% Accuracy
Read and heard	76	1.73	47
Mouthed	69	1.55	44
Read only	62	1.30	59

The percent accuracy of the input judgments made for correctly recognized words was calculated for each subject in each of the three input conditions. An analysis of variance performed on these data showed no significant differences in accuracy of input judgments across conditions [$F(2,34) = 2.145$, $MSe = 548.597$, $p > .05$].

Although there did not appear to be any systematic misattributions of input conditions for the correctly recognized old words in this experiment, the input judgments for the false positives (the overall false positive rate was 21.5%) once again demonstrated a clear bias. Of these new words incorrectly judged as old, 63% were judged to have been silently read, 18% to have been read and heard, and 19% to have been mouthed. An analysis of variance performed on the number of judgments of each input condition made for the false positives by each subject showed that this bias was significant [$F(2,34) = 13.229$, $MSe = 3.859$, $p < .001$]. "Read only" judgments were significantly more frequent than either of the other two judgments.

Discussion

Recognition of words that were both read and heard was better than recognition of read-only words, with mouthed items occupying an intermediate position. The beneficial effect of reading and hearing replicated the results of Conway and Gathercole (1987, Experiment 4), although in Experiment 1 of the present study the difference was found to be of marginal significance only (but see Experiment 5 below). On this basis it appears that the presence of acoustic information at input does enhance the incidental learning of words in this task, although the difference is not of the same magnitude as that obtained with subject vocalization.

As in Conway and Gathercole (1987), in Experiment 2 there appeared to be only a weak effect of mouthing—although recognition of mouthed words was not reliably greater than that of words that were read only, retention of mouthed words was not significantly poorer than that of words that were both read and heard. Thus, any benefits of the act of articulation alone in incidental learning are at best weak and, across experiments, unreliable. The better retention of vocalized than of passively heard words does not, therefore, appear to be simply due to the additional articulatory information involved in vocalizing. Instead, it seems that the vocalization advantage may be a product of an explicit action in the context of acoustic information. This is considered further in the General Discussion.

In Experiments 3 and 4, we attempted to determine whether memory is enhanced by the execution of an explicit action that does not involve speech by comparing recognition of written, vocalized, and read-only words. Writing, like vocalization, requires the subject to perform a motor action that specifies the stimulus item, and guarantees attention to the visually presented item. Perhaps, then, writing, like vocalization, should lead to enhanced memory performance.

On the other hand, the printed word bears only an indirect and arbitrary relationship to the speech representation of that word, and as such represents a secondary linguistic input. Some theorists have argued that, at least in short-term memory, advantages to speech-based inputs over written ones are due to the direct mapping of these inputs onto the specialized speech representational system of the hearing population (Shand & Klima, 1981). If this principle applied as well to longer-term retention, written words would not be as well remembered as vocalized words.

EXPERIMENT 3

Method

The method was as in Experiment 1, with the following exceptions.

Design. The three input conditions used in this experiment were "say aloud," "write," and "read silently." Each word in the input list was preceded by one of these three instructions. The same design and materials were used as in Experiment 1, except that in each of the three different input lists, the words that had been read and heard by the subjects in Experiment 1 were now written. The recognition questionnaire was amended so that the choices of input conditions for recognized words were now "said aloud," "written," and "read silently."

Procedure. At the beginning of the experimental session, each subject was seated in front of a TV monitor, with a pen and a blank sheet of paper positioned to the side of the monitor. When the instruction on the videotape was "write," the subjects were to write the word legibly in their normal handwriting; the words were to be written in a column. The intertrial interval was increased from 2 sec to 3.5 sec in order to ensure that all subjects had finished writing before the display of the next instruction. In all other respects, the procedure was the same as that employed in the previous experiments.

Results

Table 3 shows the mean percentage of correctly recognized words in each input condition, the corresponding d' values, and the mean percent accuracy of the input judgments in those conditions. A mean correct rejection rate of 81% was found for new words. Recognition was highest for the vocalized words (79%), intermediate for the written words (70%), and poorest for the silently read words (60%). The mean d' values were 1.82 for the vocalized words, 1.56 for the written words, and 1.23 for the silently read words. An analysis of variance on the d' scores of each subject showed a significant effect of input condition [$F(2,34) = 7.466$, $MSe = 1.58$, $p < .01$]. Sensitivity was higher to the vocalized words than to the words read silently, with the advantage of writing over

Table 3
Mean Percentage of Words Correctly Recognized in Each Input Condition, Mean Sensitivity Values (d'), and Mean Percent Accuracy of Input Judgments in Experiment 3

Input Condition	% Recognized	d'	% Accuracy
Vocalized	79	1.82	51
Written	70	1.56	53
Read only	60	1.23	65

reading silently reaching significance at the 10% level only. There was no reliable difference between vocalizing and writing.

The mean percent accuracy of input judgments in the three conditions is shown in the third column of Table 3. An analysis of variance performed on the percent accuracy of input judgments of each subject in each condition showed no differences in accuracy of judgments across the three input conditions [$F(2,34) = 2.014$, $MSe = 417.819$, $p > .05$].

For all input conditions, correct input judgments were significantly more frequent than either of the two possible incorrect judgments. For the written words, however, there were systematic misattributions of input conditions: 40% of the written words were judged to have been read silently, whereas only 6% were judged to have been vocalized. This difference was reliable by planned comparison.

Judgments of input conditions for the false positives were also analyzed. Once again, there was a large bias toward "read silently" judgments (79% of all judgments of false positives, vs. 6% "said aloud" and 15% "written" judgments). The effect of judged input condition for false positives was significant by analysis of variance [$F(2,34) = 43.189$, $MSe = 2.110$, $p < .001$], reflecting the greater frequency of "read silently" judgments than either of the other two judgments, which did not differ reliably.

Discussion

Writing a word at input failed to significantly enhance incidental learning over simply reading the word silently at input. In contrast, the usual advantage of vocalization over silent reading was established. The fact that recognition was no better for written than for silently read words suggests that the execution of an explicit action that specifies the stimulus will not necessarily lead to improved retention of that stimulus. It appears to be critical that the explicit action map directly onto a speech-based presentation of the input, as is the case for vocalized words. It should be noted, however, that there was no significant difference in Experiment 3 between recognition for vocalized words and that for written words (but see Experiment 5 below). This clearly weakens any argument concerning fundamental differences between vocalizing and writing. It is possible that the written condition may benefit from the availability of the visual script generated. The acoustic information generated by vocalization is transient, whereas the typographic information generated by writing is available to the subject until the next word appears on the screen. Thus, although there may not be any memory advantage conferred by the writing action itself, writing's intermediate position between reading and vocalizing might have been due to the availability of the additional visual record specifying the stimulus.

We investigated this possibility in Experiment 4 by preventing subjects from seeing the words they had written. If writing's intermediate position between reading silently and vocalizing in Experiment 3 were due to a

beneficial effect of the additional typographic information generated by writing, or indeed if the slight nonsignificant advantage of writing over reading in Experiment 3 were unreliable, unseen writing in Experiment 4 should not lead to better recognition than reading silently, and recognition should be significantly poorer for these words than for vocalized words.

EXPERIMENT 4

Method

The method was the same as in Experiment 1, with the following exceptions.

Design. As in Experiment 3, the three input conditions were "say aloud," "write," and "read silently." Subjects were, however, prevented from seeing their handwritten script.

Apparatus. In addition to the apparatus employed in the previous experiments, a box was constructed that enabled the subject to write but obscured from the subject's view what he/she had written. The box had an opening in the front through which the subject placed his/her dominant hand and wrist. The box was lined with paper, so that a record of what was written was retained for the experimenter.

Procedure. At the beginning of the experimental session, each subject was seated at a table in front of a TV monitor, and his/her dominant hand was placed in the obscuring box. Prior to the start of the experiment, the subject wrote his/her name 10 times inside the box. The paper was then removed and shown to the subject, who checked that he/she had not overwritten words and that the handwriting was legible.

Results

Table 4 shows the mean percent recognition of words in each of the three input conditions, the corresponding d' scores, and the mean percent accuracy of input judgments in each case. The mean rate of correct rejections of new words was 77%. Recognition was higher for the vocalized words (81%) than for either the written (66%) or the silently read (63%) words. The corresponding d' values were 1.70, 1.22, and 1.15. The analysis of variance performed on the d' values of each subject for each of the input conditions showed a significant effect of input condition [$F(2,34) = 7.416$, $MSe = 0.219$, $p < .01$]. This reflects the greater sensitivity to vocalized words than to written or read-only words. There was no reliable difference between sensitivity to the written and read-only words.

Analysis of the percent accuracy of input judgments of each subject in the three conditions revealed no significant differences. For both the vocalized and read-only words, there was a significantly greater frequency of correct input judgments than of either incorrect judgment.

Table 4
Mean Percentage of Words Correctly Recognized in Each Input Condition, Mean Sensitivity Values (d'), and Mean Percent Accuracy of Input Judgments in Experiment 4

Input Condition	% Recognized	d'	% Accuracy
Vocalized	81	1.70	60
Written (unseen)	66	1.22	50
Read only	63	1.15	69

In both cases, there were no biases in the frequency of the two incorrect judgments. Once again, however, for the written words there were systematic misattributions of input conditions: whereas 50% of the correctly recognized written words were correctly judged, 36% were judged to have been read only, and only 14% to have been vocalized. Although there were significantly more correct input judgments than vocalized judgments, there was no reliable difference between the frequency of correct "written" and incorrect "read only" judgments.

Finally, an analysis of variance was performed on the distribution of input judgments for the false positive items. The significant effect of judged condition [$F(2,34) = 10.336$, $MSe = 40.96$, $p < .001$] reflected the greater number of "read silently" judgments (57%) than either "said aloud" (12%) or "written" (31%) judgments.

Discussion

A clear pattern of findings emerged from this experiment. Subjects' recognition was better for vocalized words than for words they had written or silently read, and there was no advantage of writing over reading. These results unambiguously establish that the act of writing per se does not have any beneficial effects on incidental learning in this task. It is not completely clear whether the apparent reduction in retention of written words in this experiment—in contrast to Experiment 3, they were significantly more poorly recognized than vocalized words—was due to the unavailability of the handwritten script, which may have provided an additional cue, or whether the intermediate effect of writing in Experiment 3 was not robust. (In fact, to anticipate, the results of Experiment 5 demonstrated that there was no consistent advantage of writing over silently reading, whether or not the writing was seen.)

The pattern of results observed so far shows that input conditions that involve acoustic information—especially subject vocalization—lead to enhanced recognition performance. There are, however, some apparently intermediate effects—in particular, of mouthing and seen writing—whose status is not completely clear. On some occasions, there appears to be an advantage to these input conditions over reading alone, and on others there does not. In order to gain a clearer picture of the effects of these different input conditions, we conducted Experiment 5, in which each group of subjects was tested under only one input condition. It was possible, of course, that the effects of different input conditions established in Experiments 1-4 had been specific to the mixed-mode presentation procedure. Perhaps, for example, vocalized words in a mixed-mode list had been better remembered only because of the relative distinctiveness of this input condition. It was clearly important to establish whether this was the case. If, however, we found that these long-term modality differences generalized to pure-mode lists, we should be able to gain a better idea of the ranking of the different input conditions. Obviously, it would be

predicted that vocalizing would lead to better recognition than reading silently, and the passive acoustic input conditions of hearing alone and reading and hearing might be expected to give rise to recognition performance midway between reading silently and vocalizing. The outcomes of the mouthed and written conditions were more uncertain.

EXPERIMENT 5

Method

Subjects. Seventy-seven undergraduate students from Cambridge University, aged between 18 and 23, participated in this experiment. Each subject was tested individually.

Design. The same input list of 30 words was presented to each of the seven groups of 11 subjects. Each group was tested in one of the following input conditions: "read silently," "write (seen)," "write (unseen)," "mouth," "hear only," "read and hear," and "say aloud." As in the previous experiments, the questionnaire contained the 30 test words plus 30 fillers. For each word, the subjects were required to make a recognition judgment, followed by a rating of confidence for each word judged to be old.

Materials and Apparatus. These were the same as in Experiments 1-4.

Procedure. The general procedure was the same as that employed in Experiments 1-4, except that each word was preceded by a warning signal—a row of asterisks on the center of the screen—rather than an instruction. The subjects were verbally instructed as to input condition. The timing of each trial was the same as the time employed in Experiments 1-4.

In all conditions except heard-only, each word appeared printed on the screen. In the heard-only condition, a row of dollar symbols was substituted for the visually presented word, and, as in the read-and-heard condition, the experimenter was heard speaking the word simultaneously with stimulus onset.

Results

The results of this experiment are summarized in Table 5, which shows the mean percent recognition in each of the seven input conditions and the corresponding mean d' values. Consider first the percent recognition scores for each group. There were no differences in performance for the read-only, written (seen), written (unseen), and mouthed conditions, all of which led to 75% recognition. Recognition scores were higher for the two heard conditions, heard-only and read-and-heard, which had mean values of 82% and 83%, respectively. Recognition was highest for the vocalized words (91%). The same general pattern of results emerged for the d' measure of sensitiv-

Table 5
Mean Percentage of Words Correctly Recognized in Each Input Condition, Mean Sensitivity Values (d'), and False Positive Rates in Experiment 5

Input Condition	% Recognized	d'	% False Positives
Vocalized	91	2.55	15
Read and heard	83	2.00	17
Heard only	82	1.94	18
Mouthed	75	1.74	19
Written (seen)	75	1.42	22
Written (unseen)	75	1.32	25
Read only	75	1.35	24

ity, except that d' values for the mouthed group were greater than for the two written and the read-silently groups.

A one-way analysis of variance was performed on each subject's d' score as a function of input condition. The effect of input conditions was significant [$F(6,70)$ = 5.655, MSe = 0.384, p < .001]. The following contrasts were significant by planned comparisons: Vocalized presentation led to higher scores than all of the other input conditions. The two heard conditions were superior to the read-only, written (seen), and written (unseen) conditions, although not to the mouthed condition. Mouthed presentation, however, did not differ significantly from either the read or the two written conditions.

Discussion

The results of this experiment demonstrate that differences in incidental learning between different input conditions are present in pure-mode as well as mixed-mode presentations. Vocalized presentation led to best recognition performance, with the two passive acoustic presentation conditions—heard and read and heard—occupying an intermediate position. Poorest retention was found for read-only and written words. There was no recognition advantage at all for written words, whether or not subjects saw their handwritten scripts. The only input condition whose status remains slightly ambiguous is mouthed presentation, which did not differ significantly from either the two heard conditions or any of the silent conditions. However, a cautious interpretation—and one that receives support from previous experiments using the mixed-mode technique—is that mouthing does not significantly enhance incidental recognition.

GENERAL DISCUSSION

The principal finding of this study may be summarized as follows: Acoustic presentation procedures led to better long-term recognition memory than nonacoustic presentation procedures, and this was the case whether the acoustic information was part of mixed- or pure-mode presentations. Furthermore, the acoustic memory advantage was present throughout list positions and auditory recency was not observed. Thus, the present findings replicate and extend the long-term through-list modality differences reported by Conway and Gathercole (1987).

Of the acoustic presentation procedures employed in the present study—vocalization, reading and hearing, and hearing only—only vocalization consistently gave rise to significantly enhanced memory performance. Reading and hearing did not lead to significantly better memory performance than reading silently in Experiment 1, and reading and hearing and hearing alone did not lead to recognition rates that differed significantly from the rates for mouthed presentations in Experiment 5. It seems reasonable to conclude on the basis of these findings that although acoustic information may generally enhance memory (especially when compared to reading silently), this enhance-

ment is greatest when the acoustic information is self-generated.

How can these findings be explained? One approach taken by previous investigators has been to assume that auditory information receives some form of privileged encoding. For example, Conway and Gathercole (1987) argued that auditory information may be encoded in a perceptually distinct way—at least in comparison with visually presented orthographic information. Similarly, Shand and Klima (1981) proposed that acoustic inputs may benefit memory because the memory system—or perhaps the cognitive system generally—is specialized for speech representations, and that inputs that map directly onto this primary linguistic representation may therefore be better retained. Thus, sound-based inputs will be better remembered than visual inputs, which first require access to stored orthographic information. Perhaps the most influential of these encoding views of modality effects was proposed by Gardiner (1983), who suggested that audition gives rise to temporally distinct memory representations. In particular, Gardiner proposed that auditory presentations may lead to memory representations in which ordinal information is well preserved, but that in corresponding representations of visual events, temporal features are not well preserved.

These encoding views of modality effects, however, suffer from a number of shortcomings. None of the encoding approaches specify mechanisms that might mediate the privileged encoding of auditory information—in fact, one of the only views to do this was the no longer favored echoic memory approach proposed by Crowder and Morton (1969; see Gardiner, 1983; Greene, 1985; Glenberg, 1984; and Glenberg & Swanson, 1986, for critical appraisals of echoic memory). Thus, the encoding views are, to differing degrees, underspecified. Moreover, the present findings challenge any approach that simply assumes privileged encoding of *all* auditory information, inasmuch as only vocalization was found to enhance memory performance consistently. Also, nonacoustic but speech-based modalities such as mouthing were not found to enhance memory performance in the present studies, contrary to proposals that mouthing may be functionally equivalent to audition (e.g., Campbell & Dodd, 1980; Glenberg & Swanson, 1986; Shand & Klima, 1981; but see Gathercole, 1986, and Turner et al., 1987, for evidence against this view).

An approach that appears to be more compatible with the present findings was recently proposed by Glenberg and Swanson (1986). These authors assume privileged encoding of auditory information in the form of a modified version of Gardiner's (1983) notion of temporal distinctiveness. They propose that auditory information leads to representations that preserve in detail the time of presentation of the stimuli (fine temporal grain), whereas visual input leads to representations that preserve only general characteristics of the time of presentation (wide temporal grain). According to this approach, retrieval involves the construction of "search sets" that specify temporal do-

mains relating to the time of presentation. Narrower search sets are assumed for recent auditory than for visual to-be-remembered items. In addition, a cue-overload assumption is made, according to which "the greater the number of representations included in the search set the less likely any one of them is to be recalled" (Glenberg & Swanson, 1986, p. 5). Thus the modality effect arises because of the greater temporal distinctiveness of final auditory than visual items in their respective search sets.

The Glenberg and Swanson (1986) model can be used to account for the long-term acoustic advantages observed by Conway and Gathercole (1987)—and, to a lesser extent, in the present paper—in the following way. The recognition test cues the subject to particular parts of the list, and the finer grained representation of acoustic information ensures better recognition than the wider grained representation of visual information. More importantly, this model also predicts a through-list auditory modality advantage in the case of a recognition test, but only auditory recency in the case of free recall—which was exactly the finding of Conway and Gathercole (1987, Experiment 3). According to this model, the restriction of the auditory advantage in free recall arises because the initial narrow search sets access temporally distinct auditory representations of the final list items, and recall of earlier list items is impaired as the search set widens and the set is overloaded.

The Glenberg and Swanson model is, however, intended to apply to the nonacoustic speech-based modalities of lipreading and mouthing as well as to all acoustic information (1986, p. 13). Therefore, the model cannot account for the advantage of vocalization over other acoustic presentation procedures, and mouthing, observed in the present experiments. Furthermore, because temporal distinctiveness is a central component of this model, it would surely predict that a temporally distinct presentation procedure such as writing should lead to temporally narrow-grained representations and so to enhanced memory performance. Enhanced memory performance following writing was not, however, observed in the present studies.

Nevertheless, the emphasis on retrieval processes in the Glenberg and Swanson (1986) model is important and represents an advance over the encoding views discussed previously. Indeed, it seems possible that retrieval processes may have played a critical role in the present pattern of findings. Consider the recognition tests administered in the present experiments. Unlike most recognition tests, these did not involve the use of copy cues, because input modality was not reinstated at recognition. This argument also extends to the read-silently words, which were presented originally on a video monitor and at recognition in printed lists.

Following Glenberg and Swanson (1986), we assume that the recognition items cued subjects to temporally relevant parts of the list in all conditions, and that auditory information was more available to the retrieval process because it was more temporally distinct. To account for

the memory differences in the different auditory conditions, however, further assumptions must be made. One possibility is that search sets include self-generated cues that supplement the partial cues available at recognition. These internally generated cues would primarily relate to features of the encoding context—in particular, to the modalities (or modality) employed in that phase of the experiment. We will refer to these cues as content cues. Perhaps it is easier to generate effective content cues relating to one's own voice than to others' voices. Thus, the retrieval process may capitalize on both the temporal distinctiveness of a representation and the familiarity of the represented information.

Internally generating content cues that supplement the retrieval of representations of self-vocalized events may be comparatively effective because preexisting long-term knowledge structures are specialized for one's own voice. It may be that the same or similar knowledge structures facilitate internal cue generation for representations of events voiced by other people. The present findings indicate, however, that internally generated content cues for another's voice are not as effective as cues generated for one's own voice. (See Nairne & Pusen, 1984, for related evidence on the failure of imaged voices of others to disrupt auditory recency.)

In the case of mouthed and written presentations, it is unclear whether lack of temporal distinctiveness and/or content cues led to poorer recognition memory for these conditions. One possibility is that because mouthing constitutes an unfamiliar activity there may be few, if any, preexisting knowledge structures that could facilitate content cue generation. Given that mouthing does not lead to consistently enhanced memory performance (in the present study and in Conway & Gathercole, 1987), it may be that retrieval in this case is mediated by some process of temporal distinctiveness only. Writing, however, constitutes a more difficult case for this approach. We tentatively suggest that although writing may be a temporally distinct activity and may produce stimuli with which subjects are highly familiar, these aspects do not influence the recognition of written events. It is not clear why this should be the case, although it may reflect the fact that as writing leads to externally verifiable events, no particular encoding and retrieval processes are typically required for written stimuli.

A model that emphasizes retrieval processes may account, at least in part, for the pattern of recognition memory performance observed in the present experiments, but can this putative model also account for the pattern of findings relating to judgments of input modality? Rather surprisingly, there were no differences in the accuracy of input judgments for the different input conditions contrasted in the four mixed-mode experiments. Clearly, differences would be expected if access to representations of modality were mediated by self-generated content cues, as outlined above. It may be, however, that judgment accuracy for the read-only items was inflated by guessing. In all experiments subjects were

biased toward judging false positives as having been read only. Therefore, any words that were read only at input and that subjects (correctly) guessed were old would have a high probability of being correctly judged to have been read only.

Perhaps this aspect of the findings reflects the use of a metamemory heuristic. Conway and Gathercole (1987) proposed that such a metamemory belief may operate by associating different input conditions with different degrees of memorability. In the context of the present experiments, subjects may have (correctly) *believed* that any word they read silently would be more poorly remembered, hence the bias toward erroneous "read only" input modality judgments. This bias, then, may have obscured differences in memory performance. A more sensitive technique will clearly have to be employed if this aspect of modality is to be used to further explore the role of self-generated content cues in memory for modality. The metamemory heuristic may itself require further investigation, particularly as this heuristic may lead subjects to make modality judgments dissociable from recognition memory performance (see Conway & Gathercole, 1987), although such dissociations were not observed in the present series of experiments.

In summary, then, the present findings demonstrate that not all auditory presentations lead to similar levels of enhanced long-term retention. This finding is problematic for current models of auditory modality effects in long-term memory (e.g., Glenberg & Swanson, 1986). We propose that models that focus primarily on retrieval processes, emphasizing temporal distinctiveness and the effectiveness of internally generated cues, may be a useful way of conceptualizing these complex long-term modality effects.

REFERENCES

BISHOP, Y. M. M., FIENBERG, S. E., & HOLLAND, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* London: MIT Press.
CAMPBELL, R., & DODD, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, 32, 85-99.
CONWAY, M. A., & GATHERCOLE, S. E. (1987). Modality and long-term memory. *Journal of Memory & Language*, 26, 341-361.
CROWDER, R. G., & MORTON, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 5, 365-373.
GARDINER, J. M. (1983). On recency and echoic memory. *Philosophical Transactions of the Royal Society of London*, 302, 267-282.
GATHERCOLE, S. E. (1986). The modality effect and articulation. *Quarterly Journal of Experimental Psychology*, 38A, 461-474.

GLENBERG, A. M. (1984). A retrieval account of the modality effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 526-541.
GLENBERG, A. M., & SWANSON, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 12, 3-15.
GREENE, R. L. (1985). Constraints on the long-term modality effect. *Journal of Memory & Language*, 24, 526-541.
LUCE, D. R. (1959). *Individual choice behavior: A theoretical analysis.* New York: Wiley.
NAIRNE, J. S., & PUSEN, C. (1984). Serial recall of imagined voices. *Journal of Verbal Learning & Verbal Behavior*, 23, 331-342.
PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, 76(1, Pt. 2).
SHAND, M. A., & KLIMA, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of American Sign Language. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 464-474.
THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words.* New York: Teachers College Press.
TURNER, M. L., LaPOINTE, L. B., CANTOR, J., REEVES, C. H., GRIFFETH, R. H., & ENGLE, R. W. (1987). Recency and suffix effects found with auditory presentation and with mouthed visual presentation: They're not the same thing. *Journal of Memory & Language*, 26, 138-164.
WATKINS, O. C., & WATKINS, M. J. (1980). The modality effect and echoic persistence. *Journal of Experimental Psychology: General*, 109, 251-278.
WINER, B. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

NOTES

1. Recognition performance in each input condition was also scored as a function of the position of the memory item in the list. For this purpose, the mean percent recall in each condition for the initial 10 items, the middle 10 items, and the final 10 items in the list was calculated. No significant effects of input position were found, and this was also the case in Experiments 2, 3, and 4. For this reason, the input position data are not reported for any of the present experiments. Significant serial position effects were found for some of the nonacoustic groups in Experiment 5; these, however, were uninterpretable and are not reported.

2. In this experiment and in Experiments 2, 3, and 4, a further form of analysis was employed. The input judgment data were entered into a log-linear model fitted to an incomplete table (see Bishop, Fienberg, & Holland, 1975, p. 177); correct responses were not employed in this analysis. The model was an extension of Luce's (1959) model and by an iterative process identified the best-fitting incomplete table. The purpose of this analysis was to identify systematic biases in subjects' error modality judgments. This analysis, however, did not reveal any biases other than those observed in the analyses of variance reported for the four experiments and is, therefore, not reported.