Kindergarten, second-grade, fifth-grade, and college subjects were tested in a serial-position recall task under each of four conditions: Visual stimuli/visual recall cue, visual stimuli/auditory recall cue, auditory stimuli/visual recall cue, auditory stimuli/auditory recall cue. Visual stimuli were pictures of common animals and objects; auditory stimuli were the tape recorded names of these animals and objects. Two serial positions were probed on each trial. At all grade levels: (1) recall of visual stimuli was markedly superior to that of auditory stimuli, (2) modality of recall cue had no effect, (3) primacy and recency effects were found for both visual and auditory stimuli. Delayed testing (Probe 2) produced a decrement in auditory stimulus recall but not in visual stimulus recall. It was suggested that overt labeling of pictorial stimuli functionally changes the stimuli from being purely visual to being both visual and auditory. [Author]
A Developmental Study of Visual and Auditory Short-Term Memory

Alexander W. Siegel and Judith P. Allik

Department of Psychology
University of Pittsburgh
Pittsburgh, Pa. 15260

Visual and Auditory Short-Term Memory

Alexander W. Siegel
Department of Psychology
University of Pittsburgh
Pittsburgh, Pa. 15260
A Developmental Study of Visual and Auditory Short-Term Memory

Alexander W. Siegel and Judith P. Allik

The experiment was a developmental study of modality effects in a serial-position recall task. At all grade levels: 1) recall of visual stimuli was markedly superior to that of auditory stimuli; 2) modality of recall cue had no effect; 3) primacy and recency effects were found for both visual and auditory stimuli; 4) delayed testing produced a decrement in auditory, but not in visual, stimulus recall.
Abstract

Kindergarten, second-grade, fifth-grade, and college subjects were tested in a serial-position recall task under each of four conditions: Visual stimuli - visual recall cue, visual stimuli - auditory recall cue, auditory stimuli - visual recall cue, auditory stimuli - auditory recall cue. Visual stimuli were pictures of common animals and objects; auditory stimuli were the tape-recorded names of these animals and objects. Two serial positions were probed on each trial.

At all grade levels: 1) recall of visual stimuli was markedly superior to that of auditory stimuli, 2) modality of recall cue had no effect, 3) primacy and recency effects were found for both visual and auditory stimuli. Delayed testing (Probe 2) produced a decrement in auditory stimulus recall but not in visual stimulus recall. It was suggested that overt labeling of pictorial stimuli functionally changes the stimuli from being purely visual to being both visual and auditory.
A Developmental Study of Visual and Auditory Short-Term Memory

Alexander W. Siegel and Judith P. Allik

University of Pittsburgh

The present study investigates modality effects on children's short-term memory for serially presented items. The questions asked were: 1) Are there modality effects in children's short-term memory; 2) If so, what is their developmental course; 3) To what extent are variations that have been found in the shape of the serial position curves of children in STM studies influenced by the modality of stimulus presentation; 4) Does the modality of the recall cue exert a significant influence on performance? The answers to these questions should provide developmental evidence relevant to current theoretical models of STM and storage systems (Neisser, 1966; Crowder & Morton, 1969).

Procedures frequently used in studies of children's STM are based on the technique introduced by Atkinson, Hansen, and Bernbach (1964). In their procedure, individual children are shown a series of pictures of common objects or animals, one at a time; the cards are then placed face down in a horizontal row. The subjects are then shown a card identical to one of the original stimuli and are asked to turn over the card that matches it. Atkinson et al.'s initial study used four- and five-year-old children. A puzzling result was that, although recency effects were found, there was no evidence of a primacy effect for either group of children. In similar experiments with older children and college students the characteristic bow-shaped function had been found, indicating that the first (primacy) and last (recency) serial positions had a higher probability of being correctly recalled than did the middle positions.

Several subsequent attempts to account for this result have centered around "rehearsal" (Bernbach, 1967), "mediational deficiency" (Reese, 1962) or "production deficiency" (Flavell, Beach, & Chinsky, 1966) hypotheses. Bernbach's
rehearsal hypothesis claims that the absence of a primacy effect is due to
the failure of the young subjects to rehearse. The mediational deficiency
hypothesis implies that the labels are produced but fail to mediate, while the
production deficiency hypothesis states that at certain ages children fail to
produce the verbal labels even though they are capable of doing so.

To test these hypotheses, and to explore the developmental effects of
verbal labeling on children's short-term memory, several studies have used a
procedure based upon the Atkinson et al. (1964) paradigm but have included an
overt label as well as a no-label condition (Bernbach, 1967; Hagen & Kingsley,
1968; Kingsley & Hagen, 1969; Hagen, Meacham, & Mesibov, 1970; McCarver & Ellis,
1972). Each of these studies included a condition in which the subjects were
required to label the stimuli aloud. The results indicated that overt labeling
produced either no effect or decreased performance at the primacy portion of the
serial position curve, while it significantly improved performance on recent
positions.

One purpose of the present study was to test the hypothesis that the in-
creased recency effect that has been found in the labeling conditions was not
caused by the production of verbal labels (Flavell, Beach, & Chinsky, 1966)
or by an increase in the rehearsal of the items (Bernbach, 1967). Rather,
overt labeling of the items radically changes the functional modality of the
stimulus. In no-label conditions, the stimuli are purely visual (i.e., there
is no auditory component); when items are overtly labeled, they become both
visual and verbal-auditory. Thus, the requirement that subjects overtly label
the stimuli changes the Atkinson et al. paradigm from a study of the processing
of spatially presented visual information to a study of both spatially pre-
sentated visual information and temporally presented auditory information. It
was felt that the superior recency of the labeling groups was due to this addi-
tional auditory component of the input which is stored in Precategorical Acoustic
Storage (PAS) (Crowder & Morton, 1969). This store is conceptualized as similar
to a visual precategorical store but having a substantially longer storage time. PAS could account for the superior recency effects found when subjects are required to overtly label the stimuli.

We further hypothesized that the auditory component of the overt labeling condition interferes with rehearsal, under the assumption that the rehearsal is taking place in a verbal-auditory representation system (Brooks, 1968; Atwood, 1971; Paivio, 1971). Brooks has found poorer performance when subjects must process information and report on their processing in the same modality (Brooks, 1968). Corballis (1966) suggested that cumulative rehearsal is more restricted when presentation is auditory than when it is visual. This would account for the decrement in performance that has been found at the primacy portion of the serial position curve (Hagen, Meacham, & Mesibov, 1970; McCarver & Ellis, 1972) in the overt label conditions.

A procedure based on the Atkinson et al. paradigm was developed to test our hypotheses: In an attempt to separate visual and auditory components, stimuli were either visual (line drawings of animals and objects) or auditory (the tape-recorded names of animals or objects) (Murdock, 1969). Since our procedure utilized a fairly slow presentation rate, and since research with adults indicates that slow presentation rates facilitate recall of visual materials (Dornbush, 1968; Corballis, 1966), it was predicted that overall performance would be superior when stimuli were presented visually. In light of Crowder and Morton's (1969) evidence for PAS, it was predicted that a greater recency effect would be found with auditory presentation of stimuli.

To further examine the mediational deficiency (Reese, 1962) and production deficiency (Flavell et al., 1966) hypotheses, the modality of the recall cue was either visual or auditory (within a condition). If younger subjects were either not producing labels or were not using them, their recall of the position of a visually presented item should be significantly poorer when the
recall cue is presented in the auditory modality. While it was not expected that cue modality would affect the accuracy of the performance of older subjects, it was hypothesized that the latencies of correct responses in the cross-modal conditions would be longer than those in the intra-modal conditions for all subjects.

Method

Subjects

Sixteen children at each of three grade levels: kindergarten, second and fifth grades, and 16 college students enrolled in an introductory psychology course participated in the experiment. The mean age of the kindergarteners was 6 years - 0 months (range = 69-78 months), that of the second-graders was 7 years - 9 months (range = 88-99 months), and that of the fifth-graders was 10 years - 10 months (range = 123-136 months). All children were average or above in intelligence and were from middle-class socioeconomic backgrounds. Eight males and eight females were tested at each of the four grade levels.

Stimuli

Stimuli for the visual presentation conditions were 135 black line drawings of common, easily-labeled objects and animals, drawn on 3 X 5-inch white cards. These pictures were presented to the subjects in sets (or series) or seven cards at a rate of approximately one picture per 4 sec.

Stimuli for the auditory presentation conditions were the tape-recorded names of the same objects and animals spoken in a woman's voice (JPA). For each of the two auditory presentation conditions, 16 series of different object and animal names were recorded (two 4-item warmup series, and 14 7-item test series) on Scotch 290 Magnetic Tape. The items in the series were recorded and presented such that it took approximately 22-25 sec to present a series of 7 items. An attempt was made during the recording process to keep a 4-sec constant interval.
between the beginning (i.e., initial pronunciation) of item \( n \) and the beginning of item \( n+1 \), thus equating the amount of time required to present an entire series in both visual and auditory conditions.

No stimulus was used more than once within any one of the four experimental conditions.

**Apparatus**

The same display and response console was used in all experimental conditions. The console consisted of a wooden platform measuring 30 in. (long) X 8 in. (deep) X 2½ in. (high) situated horizontally on a table in front of the subject. The console was divided lengthwise into two areas, one for response buttons, the other for stimulus presentation. The top surface of the console platform was divided into seven equal spaces with \( \frac{1}{4} \)-in. black stripes separating each space. Within each space closest to the subject was a black 1-in. diameter response button. Stimulus loci were directly in back of the response buttons. These loci were set at an angle of 15 degrees from the perpendicular. In all conditions, the subject responded by pushing the response button beneath his stimulus choice. The response buttons were connected to a Lafayette digital timer, allowing recording of the subject's latency to the nearest 1/100 sec.

For the auditory presentation conditions, the output jack of a Sony TC 230 tape recorder was connected to a multiple position rotary switch. The contacts of each of seven positions was connected to one of seven small auxiliary speakers. Each speaker was the same size as the visual stimuli (3 X 5 in.) and was covered with black cloth.

**Procedure**

Each subject was tested individually in all four experimental conditions. Testing of each child was accomplished in two sessions (approximately a week apart) or approximately 40 to 60 minutes in length. Testing of college subjects was accomplished in one session lasting approximately 90 minutes. Order of
testing of the four conditions was counterbalanced such that each condition was presented first, second, third, and last to an equal number of subjects.

For each condition, the subject was seated at a table opposite the experimenter and facing the display-response console. The task was introduced to the children as a game. The experimenter gave the instructions appropriate to the conditions (as detailed below), using the two practice problems of four items each. After the experimenter was sure that the subject understood the instructions, the subject was given the 14 seven-item test series. On all series, stimuli were presented in left-to-right order.

Experimental conditions and instructions

The following prototypic instructions were given to fifth-graders and adults; for younger children the language and timing of the practice trials were modified when necessary to ensure comprehension of the task.

1. Visual Presentation - Visual Recall Cue:
   "I'm going to show you a series of seven pictures and I want you to try to remember where you saw each picture. At the end of the series I'll show you two of the pictures again, one at a time. When I show you the picture, you push the button at the position where you saw the picture. We'll do a few problems for practice so that you can get the idea. (The two 4-item practice problems were given.) Any questions?"

2. Visual Presentation - Auditory Recall Cue:
   "I'm going to show you a series of seven pictures and I want you to try to remember where you saw each picture. At the end of the series, I'll ask you where you saw two of the pictures. When I name the picture, you push the button at the position where you saw the picture. We'll do a few problems for practice so that you can get the idea. (The two 4-item practice problems were given.) Any questions?"
3. Auditory Presentation - Auditory Recall Cue:

"These seven black boxes are speakers and a word is going to come out of each one of them, one at a time, from left to right. I want you to try to remember where you heard each word. At the end of the series, I'll ask you where you heard two of the words. When I say one of the words, you push the button at the position of the speaker where you heard the word. (The two 4-item practice problems were given.) Any questions?"

In the visual conditions, the pictorial stimuli were initially placed, face down, in their predetermined spaces on the response console. They were then turned over and shown to the subject, one at a time. To present the cue (probe), the experimenter either held up a picture identical to one of the stimuli or spoke the "name" of one of the stimuli. Upon presentation of the cue, either visual or auditory, the experimenter manually activated the digital timer.

In the auditory conditions, the seven speakers were in place on the console and the experimenter manually operated the rotary switch to sequentially present the auditory stimuli from the speakers. The probe procedures were identical to that of the visual conditions. Pretesting had indicated that after practice trials, children could readily identify from which speaker a single word was spoken. (The separation between two adjacent speakers was approximately 15°, and it is known that young children can localize auditory stimuli with an average error of the order of 5° [Warren, 1970]).

On each trial (7-item series), two non-adjacent serial positions were probed and the subject's response was recorded. Probes were made non-adjacent to minimize any cueing effect the first probe might have had on the recall of the second probed item. Time between probes was variable, depending on the subject's latency of response to the first probe. Within each condition, each serial position was probed four times; twice as a first probe and twice
and a second probe. Response latencies were recorded for each probe (time between presentation of the recall cue and the subject's button-press response). No feedback as to correctness of response was given. General social reinforcement was given to subjects on all four conditions, when necessary, to maintain an appropriate level of attention and motivation.

Results

Percent correct recall

The percentages of correct responses at each of the seven serial positions (summed over probes) were tabulated and subjected to an arc sin transformation. These transformed scores were subjected to a 4 (Grade) X 2 (Sex) X 8 (Subjects per cell) X 2 (Presentation modality) X 2 (Recall cue modality) X 7 (Serial position) mixed analysis of variance, with repeated measures on the last three factors.

The "between subjects" portion of the analysis yielded a highly significant main effect of grade level, $F (3/56) = 34.04$, $p < .0001$. Scheffé (.05) confidence intervals indicated that a) the mean performance of the adults (76% correct) was significantly greater than the performance of all three groups of children; b) the performance of the kindergarteners (47%) was not significantly less than that of the second-graders (52%), but was significantly less than that of the fifth-graders (61%). The main effect of sex was also significant; the mean performance of the girls (62% correct) was significantly greater than that of the boys (56% correct), $F (1/56) = 7.00$, $p < .05$. The Grade X Sex interaction was not significant, $F < 1$.

In line with our prediction, visually presented stimuli produced a higher percentage of correct responses (64%) than did auditorially presented stimuli (54%), $F (1/56) = 82.81$, $p < .0001$. Neither the main effect of recall cue modality nor the Presentation modality X Recall Cue modality interaction was significant, $F < 1$. Thus, the modality in which the stimuli were presented had a clear effect...
on performance, whereas the effect of the modality of the recall cue was negligible.

The overall effect of serial position was, of course, highly significant, \( F(6, 336) = 48.85, p < .0001 \). The percentage of correct responses was as follows (for serial positions 1 through 7): 73%, 54%, 48%, 48%, 48%, 63%, and 78%. Scheffé (.05) confidence intervals indicated a significant primacy effect and a significant recency effect. The Grade X Serial position interaction was highly significant, \( F(18, 336) = 4.33, p < .005 \), and is presented graphically in Figure 1.

---

It is clear from this figure that performance on the middle portion of the serial position curve (positions 2 through 5) increased regularly with increasing developmental level. Of more central concern are the primacy and recency effects. Scheffé (.05) confidence intervals indicated that for adults, although the shape of the curve indicated primacy and recency effect, performance at the intermediate positions was not significantly lower than performance at positions 1 and 7; significant primacy and recency effects were found with all groups of children.

The significant Presentation modality X Serial position interaction, \( F(6, 336) = 12.42, p < .0001 \), is portrayed graphically in Figure 2. Scheffé (.05) confidence intervals indicated that the difference in performance as a function of presentation modality was due primarily to the superior performance with visual stimuli on serial positions 1 through 5 only; performance on the last two serial positions (summed over probes) was very similar for both presentation modalities.
Effects of first and second probes

Difference scores were computed for each subject in each of the experimental conditions (the number of correct responses on the first probe on a trial minus the number of correct responses on the second probe). A 4 (Grade) X 16 (Subjects) X 2 (Presentation modality) X 2 (Recall cue) mixed factorial analysis of variance with repeated measures on the last two factors was performed on these difference scores. The main effect of grade level was highly significant, \( F(3/60) = 5.33, p < .005 \). Scheffé (.05) confidence intervals indicated that all groups of children made more correct responses under Probe 1 than under Probe 2: The difference score for kindergarteners (0.48) was not significantly different from that of either second- (0.94) or fifth-graders (0.97). Most importantly, the main effect of presentation modality was highly significant, \( F(1/60) = 8.54, p < .01 \); whereas there was practically no difference in performance between Probe 1 and Probe 2 when stimuli were presented visually (0.06), when stimuli were presented auditorially, performance was much better on the first than on the second probe (difference of 0.89). Post hoc analysis (\( t \) tests on within-subject differences) indicated that there were no significant differences in the number of correct responses between Probes 1 and 2 for the visual presentation modality at any grade level, \( t < 1 \). However, for the auditory presentation modality, significantly more correct responses were made under Probe 1 than under Probe 2 by all three groups of children, \( t (31) = 2.73, p < .05 \), but not by adults, \( t < 1 \).

In order to examine more closely the significant performance differences between Probes 1 and 2, separate serial position curves for Probe 1 and Probe 2 for both the visual and auditory presentation modalities were plotted and are presented in Figure 3.
It is clear from Figure 3 that the superiority in performance under Probe 1 in the auditory presentation conditions was due almost entirely to the performance differences at serial positions 6 and 7 relative to no differences at positions 1 through 5, \( t (63) = 6.06, p < .0001 \). There were no other significant differences between Probes 1 and 2 for any other serial position for either visual or auditory presentation conditions, \( t < 1 \). At the first 5 serial positions, recall of visual stimuli was superior to recall of auditory stimuli for both Probes 1 and 2, \( t (63) = 3.75, p < .001 \), with the single exception of performance on Probe 2 at position 4. At position 6, a striking reversal occurred. At both positions 6 and 7, performance under Probe 1 was significantly better under auditory than under visual presentation conditions, \( t (63) = 3.53, p < .001 \); performance under Probe 2, however, was significantly better under visual than under auditory presentation conditions, \( t (63) = 2.56, p < .05 \).

**Analyses of response latencies**

For each subject, the mean latency for correct responses (collapsed across serial position and probe) was computed for each Presentation X Recall Cue condition. These latencies were subjected to a 4 (Grade) X 2 (Sex) X 8 (Subjects) X 2 (Presentation Modality) X 2 (Recall Cue) mixed analysis of variance with repeated measures on the last two factors. The highly significant main effect of grade level indicated that latencies decreased regularly with increasing age, \( F (3/56) = 13.92, p < .0001 \). The main effect of presentation modality was also significant, \( F (1/56) = 5.20, p < .05 \). Mean latency was significantly longer when stimuli were presented auditorily (1.28 sec) than when they were presented visually (1.10).

In line with prediction, the interaction of Presentation modality X Recall Cue modality was highly significant, \( F (1/56) = 18.44, p < .0001 \). Scheffé (.05) confidence intervals indicated that the mean latency in the auditory presentation - visual recall cue conditions (1.46 sec) was significantly greater
than those in the visual presentation - auditory recall cue (1.17), visual presentation - visual recall cue (1.04), and auditory presentation - auditory recall cue (1.10) conditions; these latter three conditions did not differ significantly from each other.

Discussion

As was expected, overall performance on both visual and auditory stimuli improved with age, with the most marked increase coming between second and fifth grades. As predicted, at all grade levels recall of visually presented stimuli was superior to recall of auditorially presented stimuli. It would appear that the relatively slow presentation rate used permitted considerable cumulative rehearsal of the visually presented items, i.e., visual stimuli were immediately labeled (covertly) by the subjects and then rehearsed within the verbal-auditory system. It is assumed that the intake of auditory stimuli and verbal-auditory rehearsal are component processes within the same system, whereas the intake of visual stimuli and their verbal rehearsal are component processes of two different systems. The data of Brooks (1968) and Atwood (1971) suggest that attention can be more easily divided between the visual and verbal-auditory system than within either one separately. Thus, the presentation of auditory stimuli was detrimental to subjects' attempts to cumulatively rehearse the material.

Results indicated that modality of recall cue had no influence on either overall recall performance or on the shape of the serial position curve. Even the youngest children were able to recall the position of visually presented stimuli when an auditory recall cue was used. This would seem to indicate that they, as well as the older subjects, were producing labels for the pictures at the time of presentation and that these labels were serving as mediators. If they had not done so, performance on visually presented stimuli should have been markedly poorer when an auditory, rather than a visual, recall cue was used.
On each trial, two non-adjacent serial positions were probed. The comparisons of recall performance on the first and second probes yielded several theoretically interesting results. When visual stimuli were presented, overall performance on the first and second probes was not significantly different at any age level. When auditory stimuli were presented, however, children's performance was markedly better on the first probe. Thus, the time between probes and/or the retroactive interference of the first probe on the second had no effect on memory for visually presented items, but had a significant debilitating effect on memory for auditorially presented items. Apparently, rehearsal after visual presentation was sufficiently effective to enable all subjects to perform equally well on both probes; rehearsal after auditory presentation is relatively difficult, and thus only the college students were able to maintain their performance over both probes.

The superior overall performance found with Probe 1 relative to Probe 2 in the auditory presentation conditions can be attributed almost entirely to the last two serial positions: At positions 6 and 7, Probe 1 recall was significantly greater than that found with Probe 2. This finding is consistent with the Precategorical Acoustic Storage (PAS) model proposed by Crowder and Morton (1969). The superior immediate retention (Probe 1) of auditory information (PAS) can account for our marked auditory recency effect and can also explain the finding of Bernbach (1967) and Hagen and Kingsley (1968) that overt verbal labeling had a facilitative effect upon the recall (of recent items) of young children. Thus, it is clear that great caution should be exercised when drawing inferences about the facilitative effects of labeling on short-term memory. If the effects of labeling are confounded with those of modality, the relative contribution of either is difficult to assess. It is important to note that at serial position 7, Probe 1 recall was extremely high for all subjects -- the performance of kindergarteners (89%) was not significantly different from that of college students (96%).
The significant Presentation modality X Recall Cue modality interaction found in the analysis performed on mean latency of correct responses was due primarily to the relatively long latency found when stimuli were presented auditorily and the recall cue was visual. This can be accounted for if it is assumed that pictures are labeled and are then rehearsed in a verbal-auditory system. In the other three conditions there was an immediate "match" between the recall cue and the stimulus item. With auditory stimuli and a visual recall cue, however, the pictorial cue must be given a label prior to the search for the corresponding stimulus item.

Although the superiority of the visual presentation conditions in the present experiment may not seem congruent with Murdock's (1967, 1968, 1969) findings that auditory presentation of verbal information results in superior retention, the differences in results may be attributed to the slower presentation rate used in the present study. Murdock used presentation rates ranging from 2 per sec to 1 per 2 sec; in the present study, the presentation rate was 1 per 4 sec. This slower presentation rate should not only permit rehearsal in the visual condition, but should also limit the effect of PAS in the auditory condition. Furthermore, the authors feel that it is not unlikely that the pictures themselves had a facilitative effect on performance because a picture can be stored simultaneously in both a visual and an auditory-verbal system.
References


Footnote

1This research is derived from a project of the Learning Research and Development Center, University of Pittsburgh, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position of policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

The authors wish to thank the principal, Ms. Louise Brennan, and teachers of Colfax Elementary School for their generous cooperation in supplying subjects, Mr. Walter Luniewski for the design and construction of the apparatus, and Ms. Laura Bell for her help in testing subjects.
Figure Captions

Fig. 1. Serial position function for percent correct recall for each of the four age groups.

Fig. 2. Serial position function for percent correct recall of visual and auditory stimuli.

Fig. 3. Serial position function for percent correct recall of visual and auditory stimuli for both the immediate (Probe 1) and delayed (Probe 2) probes.