MEMORY FOR VISUAL-SPOKEN LANGUAGE IN CHILDREN AND ADULTS

Laura A. Thompson, Donna Driscoll, and Lori Markson

ABSTRACT: Two experiments examined how developmental changes in processing speed, reliance on visual articulatory cues, memory retrieval, and the ability to interpret representational gestures influence memory for spoken language presented with a view of the speaker (visual-spoken language). Experiment 1 compared 16 children (M = 9.5 yrs.) and 16 young adults, using an immediate recall procedure. Experiment 2 replicated the methods with new speakers, stimuli, and participants. Results showed that both children's and adults' memory for sentences was aided by the presence of visual articulatory information and gestures. Children's slower processing speeds did not adversely affect their ability to process visual-spoken language. However, children's ability to retrieve the words from memory was poorer than adults'. Children's memory was also more influenced by representational gestures that appeared along with predicate terms than by gestures that co-occurred with nouns.

If given a choice, most people would prefer to watch a person tell a story than to just listen to a person tell a story; choosing the voice-alone option would be like water for chocolate. By watching the speaker, we gain better access to his or her feelings through facial expressions and posture. By noting the pace of eye, head, and arm movements we can discern structural features of communication which help to decide when to take our turn in the conversation. We encode visual articulatory movements of the face which help us to access the correct lexical terms (cf. Massaro, 1987). Moreover, while we are usually not aware of it, the speaker's gestures contribute to our understanding and memory for utterance meaning (Goldin-Meadow, Wein, & Chang, 1992; Thompson, 1995; Thompson &

Laura A. Thompson, Donna Driscoll, and Lori Markson, Department of Psychology, New Mexico State University.

We very much appreciate the research assistance we received from Leslie Beauchamp, Tara Biehl, Melissa Franklin, Jeanne Malmberg and Al Moreno. Thanks also to Steve Helmreich and Ted Dunning, Computing Research Laboratory computational linguists, for helping us code our sentences.

Address correspondence to Laura A. Thompson, Dept. 3452/Psychology, Box 30001, New Mexico State University, Las Cruces, NM 88003; E-mail: thompson@crl.nmsu.edu.

Journal of Nonverbal Behavior 22(3), Fall 1998

Massaro, 1994; although see Krauss, Morrel-Samuels, & Colasante, 1991 for an opposing view).

Since visual-spoken language is informationally complex, like any complex system, it should take years to acquire expert-level understanding of the meanings underlying speakers' visual movements. Moreover, since the acquisition process is made manifest within a developing memory system, the developing mind must impose some limitations upon the amount of visual-spoken language information that can be attended to, encoded, and remembered by the child. The purpose of the present study is to provide an exploratory investigation of the extent to which specific characteristics of visual-spoken language (visual articulatory cues presented by the mouth and face and representational hand gestures) help children and adults understand and remember language, and how visual-spoken language comprehension is affected by age differences in speed of working memory processing and memory retrieval.

Even infants are affected by supportive visual articulatory cues (e.g., Kuhl & Meltzoff, 1984) and gestural cues (e.g., Bates, Thal, Whitesell, Fenson, & Oakes, 1989). Two decades ago, McGurk and MacDonald (1976) showed that three-year-olds are susceptible to the McGurk illusion, whereby an auditory /ba/ paired with a visual /ga/ is identified as the syllable /da/. Yet, there is some evidence that gestures and visual articulatory movements contribute more to understanding as children develop. Thompson and Massaro (1986) showed that the influence of both pointing gestures and visual articulation on syllable identification becomes significantly stronger between the ages of three and five years. In the case of conflicting pointing gesture and speech information, relative to younger children, nine-year-olds are more likely than four-year-olds to report the word for the object that is pointed to, rather than the word that is spoken (Thompson & Massaro, 1994). Thus, the influence of pointing gestures on word comprehension becomes stronger between four and nine years of age.

However, current empirical knowledge of the development of visual-spoken understanding is limited in many respects. First of all, there have been very few reports of developmental aspects of visual-spoken language comprehension in older children. This study attempts to help fill the knowledge gap of the developmental time frame between age nine and adulthood. Secondly, research has not been published on the effect of visual articulation and gestural cues on children's *memory* for language. In addition, the ecological validity of the data collected thus far is limited due to the narrow range of stimuli tested (a small set of syllables and single-syllable words), as well as their brief durations. Thus, one of the goals of the present study is to investigate possible developmental differences in children's and adults' memory for sentences containing visual cues, compared to sentences without visual cues. It is expected that children's and adults' immediate recall for sentences will be aided by the presence of both visual articulatory cues and gestures during language encoding.

Previous research on adult age differences in language comprehension has highlighted the important role of working memory speed in auditory speech processing. Given that the listener cannot control the speaker's rate of speech, much is dependent on the listener's ability to quickly encode and categorize visual-spoken language, access word meanings, and integrate word meanings into propositional and phrase meanings (Marslen-Wilson & Tyler, 1980). A common finding reported in the adult cognitive aging literature is that working memory operations are performed more slowly in old adults relative to young adults (for a review, read Salthouse, 1996). Research has shown that children, likewise, possess slower processing speeds than do adolescents or young adults (e.g., Fray & Hale, 1996; Hale, 1990; Kail, 1986, 1990a; Thompson, & Massaro, 1989). Kail and Salthouse (1994) argue that the increases in speed of processing during childhood and subsequent decreases in processing speed in adulthood are a result of global changes in the architecture of the cognitive system. Yet processing speed factors have not been studied in the context of spoken language comprehension in children. Children's slower processing speeds would be expected to result in diminished speech comprehension, relative to adults, by reducing the effectiveness of the working memory system in connecting propositions of an utterance (e.g., Just & Carpenter, 1992).

If children's slower processing speeds affect their ability to link propositions into coherent meanings, intuitively, one would expect children to be frequently confused by what they are hearing. However, this intuition may not be correct. Cognitive aging research has demonstrated that older adults have strategies at their disposal that compensate for slower working memory speeds, including a greater reliance on prosodic cues (Stine & Wingfield, 1987), visual articulation cues (Thompson, 1995), and syntactic and semantic knowledge (Wingfield, Poon, Lombardi, & Lowe, 1985) as compared to young adults. The present study will attempt to discover whether or not nine- and ten-year-old children have sufficiently slow processing speeds that they perform more poorly than adults under capacity-demanding conditions, and if they show evidence for compensatory strategies.

The method employed to study this was first reported by Wingfield et al. (1985) in a study of language processing in old age. They tested young and older adults' immediate recall for three types of five- and eight-word strings: ordinary meaningful English sentences, sentences deprived of meaning but that were syntactically regular (anomalous sentences), and random word strings deprived of both meaning and syntactic structure. Speech was presented to participants at four speech rates varying between 275 words per minute (wpm) and 425 wpm. Young and old adults alike showed excellent recall for all meaningful sentences, regardless of speech rate. However, in the anomalous sentences, old adults showed far greater declines in recall performance, relative to young adults, as speech rates increased. The researchers argued that grammatical and semantic knowledge could be incorporated by the older adults into their understanding of the meaningful phrases, but that this knowledge was not useful to them in understanding words from the non-meaningful linguistic contexts. Consequently, as speech rates became faster, the low-quality speech encodings could not be "touched up" with previous knowledge of what must have been said. Younger adults presumably do not need to touch up their encodings because, due to their faster processing speeds, they encode enough of the speech that it can be sufficiently understood. Thus, evidence for processing speed deficiencies interfering with performance can be found in an interaction between sentence type and speech rate, where performance declines are greater as speech gets faster for anomalous sentences, compared to meaningful sentences.

Due to the exploratory nature of the study, it was decided to use a broad method of hypothesis testing, one based on Platt's (1964) strong inference technique. That is, many plausible developmental differences in visual-spoken language comprehension were conceived and tested within the context of a single experiment and these same hypotheses were tested in a second experiment to determine their reliability. In the following we describe the background research relevant to each developmental contrast examined in the present study.

The first prediction states that recall should decline with increases in speech rate and also that an interaction is expected to be obtained between sentence type and speech rate. More specifically, young adults' recall of meaningful sentences will not decline as speech rate increases, but should show a decline across speech rate for anomalous sentences. These results would replicate the results obtained in a previous study (Thompson, 1995). If children cannot process spoken language quickly enough to incorporate their semantic knowledge into an understanding of the utterance, then their performance will decline for both meaningful and anomalous sentences as speech rate increases. The second prediction concerns the effect of visual articulatory cues (sometimes called "visible speech") on memory. It is expected that the results of this study for adults will replicate the previous study with regard to the effect of visual articulatory cues on memory. In Thompson (1995), visual articulatory cues improved young adults' recall only in the anomalous sentence condition. Since the influence of visual articulatory cues becomes stronger between the ages of four and nine years (e.g., Thompson & Massaro, 1994) it is predicted that this trend will continue, and that the effect of visual articulatory cues on memory will be greater for adults compared to nine-year-olds.

The third prediction relates to the influence of representational gestures on memory for visual-spoken language. The production of representational hand gestures has been studied very thoroughly by McNeill (1985; 1992) and Kendon (1983). These gestures represent "attributes, actions, or relationships of objects or characters" (McNeill, 1992, p. 377), and they appear in two forms, iconic and metaphoric. Iconic gestures exhibit concrete images of the speaker's thoughts. Iconix can completely parallel the meaning of the spoken words (e.g., crossing the hands over the heart while saying "he loves her"), or they can add meaning (e.g., using a hand to mimic holding a pen while saying "he sat down to write"). The gesture clarifies that the act of writing was not accomplished with a computer, but with a pen or pencil. Metaphoric gestures pictorialize an abstract concept (e.g., using the hands in an offering gesture while saying, "you just gave me an idea"). Children (Kelly & Church, 1997) and young adults (Goldin-Meadow et al., 1992; Thompson & Massaro, 1986) are capable of detecting meanings conveyed by representational gestures. Thompson (1995) also found that the presence of representational gestures at encoding resulted in higher recall for young adults, compared to recall for sentences lacking gestural cues. However, this effect was specific to meaningful sentence recall. Thus, young adults in the present study are predicted to show the same effect. If children cannot interpret representational gestures then their recall should be uninfluenced by the presence of gestures.

Finally, reviews of the memory development literature lead to the prediction that children will not remember as many words as adults in the immediate recall task (Ceci & Bruck, 1993; Kail, 1990b). Importantly, the design of the experiment potentially allows us to distinguish between two reasons for the hypothesized developmental difference in overall level of recall. If children and adults show the same pattern of stability and decline in recall across speech rates, this would indicate that children's processing speeds do not place them at a comparative disadvantage to young adults in encoding the meaning of the sentence. However, if their recall is nevertheless poorer overall, we can conclude that a retrieval deficit, rather than an encoding deficit, is responsible. In contrast, if children show comparatively more decline across speech rates and also a lower level of recall, the source of the age difference in recall could be due either to encoding deficiencies, retrieval deficiencies, or both.

In Experiment 1, nine-year-old children and adults watched and listened to a female speaker who was reciting short sentences. Their task was to recall as many of the words from each sentence as possible, during an immediate recall interval following each sentence. Experiment 2, with some minor variations, replicates the design and procedure of Experiment 1. The aim of both experiments is to identify some of the differences in visual-spoken language processing that exist between older children and young adults.

Experiment 1

Method

Subjects. Sixteen nine-year-old children (M = 9.5 years; range = 9.1–10.1 years; 9 girls and 7 boys) and 16 young adults (M = 19 years; range = 17–24 years; 10 women and 6 men) were recruited through the Introductory Psychology subject pool to participate in the study. Some of the children were friends and relatives of students and faculty at New Mexico State University. The young adults received course credit and the children were paid $10 to take part in the study. All participants reported normal 20/20 vision with or without correction and no participants were aware of uncorrected hearing loss.

Sentence and videotape construction. Thirty-six 16-word sentences were written and divided into four nine-sentence sets. One-half of the sentences were spoken by one adult female and one-half by another. Half of the sentences were ordinary meaningful American English sentences (*meaningful sentences*) containing between 18 and 25 syllables (e.g., "Gripping the blue ax handle tightly in his strong hands, Jim chopped down his first tree.") Within a set of nine meaningful sentences, words were rearranged to form four sets of nine additional 16-word sentences (*anomalous sentences*). These 36 semantically anomalous sentences were consistent with English syntactic usage (e.g., "Even though the appropriate ledge was out of blue money, they blew to brown the hands.") Because the same words were used in constructing the anomalous sentences from the meaningful sentences, the two types of sentences were equated in terms of word length and word frequency.

Each sentence was recorded with MacRecorder. To maintain a high-quality representation of the spoken input, a high sampling rate method (22,000 kHz) was used in recording. The software program SoundEdit was used to delete portions of the original sound files in producing three levels of speech rates: 180, 270, and 360 words per minute (wpm). Deleted speech portions included endings of long vowels, pauses between words or phrases, and mid-sections of fricatives. A 500 msec tone, followed by a 500 msec period of silence, preceded each 2.67, 3.57, or 5.33 sec sentence. The videotapes were recorded by a Panasonic color video camera, connected to a Hitachi Model VT-RM300A stereo high fidelity video cassette recorder.

All conditions contained a view of the speaker. The speakers were filmed sitting at a table in front of a green background, hands folded in front when they were not being used for gesturing. For the condition without articulatory movements, the speaker looked at the camera with a neutral facial expression. This was an "audio-only" condition. For the condition with articulatory movements, the speaker's mouth moved normally, without exaggeration. In the final condition, the speaker used gestures in addition to articulatory movements. The speakers used a wide variety of iconic and metaphoric gestures. An example of an iconic gesture is using the right hand to hold an imaginary spoon while making a full circle in the air parallel to the table, to illustrate stirring ingredients in a pot. An example of a metaphoric gesture is pointing backwards over the shoulder to indicate a moment in the past while saying "long ago." A few of the gestures were not clearly iconic or metaphoric, such as pointing to the temple while saying "she forgot." Gestured sentences contained between three and six gestured words or phrases (the modal number was three).[1] Example sentences and illustrations for the gestures appear in the Appendix to Thompson (1995).

The speakers practiced gesturing and/or speaking in synchrony with the audio recording several times to achieve the illusion that the audio and video components were recorded simultaneously. An unbiased listener judged the video and audio components for each of the final sentences to be perfectly synchronized.[2] On the final videotape, each sentence was preceded by a four-sec visual "Ready" cue, followed by the sentence. After each sentence, a "Recall Sentence" message appeared on the screen for a 20-sec immediate recall period. Participants viewed the tapes on a 19" RCA XL-100 color television. Responses were recorded onto audio tape for later transcription.

Design and procedure. Within the four blocks of nine meaningful sentences and four blocks of nine anomalous sentences, one-third were presented at each of the three speech rates. For every speech rate within a block, there was one repetition of each of the three levels of visual-spoken language: spoken sentences without visual cues, but with a view of the speaker (S), spoken sentences with articulatory (facial) cues (SA), and spoken sentences with articulatory and gestural cues (SAG). Within blocks, the order of the nine different sentence types was random. Half of the participants in each age group received the four blocks recorded by one speaker, the other half by the other speaker. A counterbalanced order of presentation was used for the meaningful and anomalous conditions. Half of the participants who listened to a given speaker began with meaningful, and half with the anomalous, sentences.

During testing, participants received four blocks of trials, for a total of 36 sentences, two sentences at each speech rate and for each visual condition. Twelve practice trials preceded testing. They consisted of meaningful and anomalous sentences spoken at each speech rate, but contained no visual cues. The volume control was adjusted to each individual's preferred listening level, and maintained at this level throughout the experiment. Prior to the experimental phase, all participants demonstrated perfect recall of four 16-word audio-recorded sentences spoken at a normal speaking rate and at a comfortable volume for each individual.

Each participant was told to watch and to listen to the speaker very carefully, and to repeat what they understood word for word during the recall period after each sentence. Participants were tested individually in a single one-hour session.

Results and Discussion

Memory for sentences. Each sentence was scored for the number of words correctly reported, regardless of their placement in the sentence. For a word to be correct, it could not have been a fragment of the word presented (e.g., "hand" for "hands"), although a correct word could contain an additional element such as a prefix or grammatical ending.

These data were analyzed in a 2 (age: children vs. adults) × 2 (sentence type: meaningful vs. anomalous) × 3 (visual language: S vs. SA vs. SAG) × 3 (speech rate: 180 vs. 270 vs. 360 wpm) repeated measures analysis of variance (ANOVA) with age as a between-subjects factor and all other variables as within-subjects factors. Adults recalled significantly more words than children, $F(1,30) = 15.99$, $p < .0001$, and recall was higher in meaningful, compared to anomalous, sentences $F(1,30) = 316.54$, $p < .0001$. Recall also increased as the amount of visual language cues increased, $F(2,30) = 14.15$, $p < .0001$, and declined as speech rate became faster, $F(2,60) = 9.43$, $p < .0001$.

The difference in recall for anomalous versus meaningful sentences was greater for children than it was for adults, and was confirmed by a significant Age × Sentence Type interaction, $F(1,30) = 4.64$, $p < .05$.

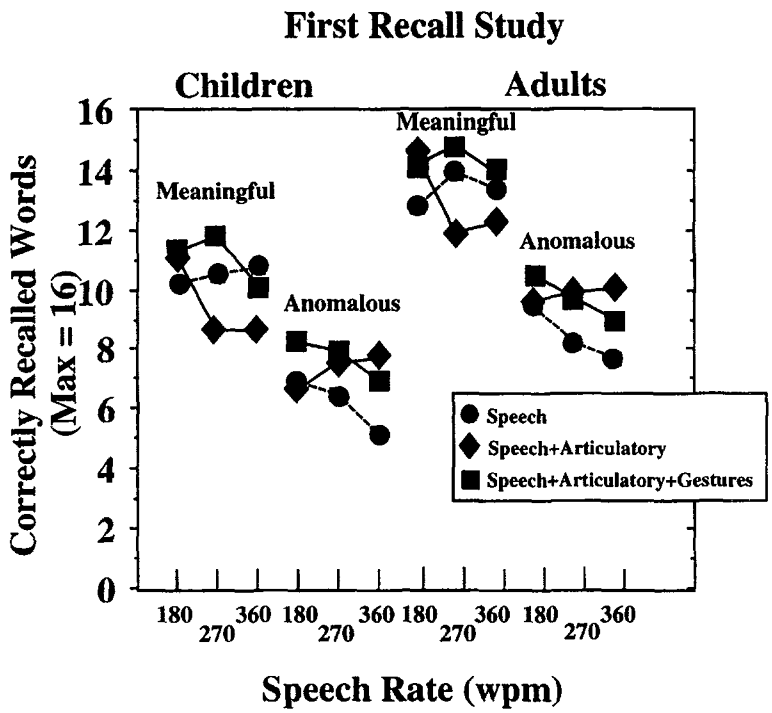

Figure 1. Average number of words correctly recalled by nine-year-old children and adults for all conditions in Experiment 1.

Secondly, sentence type also interacted with visual language condition, $F(2,60) = 18.88$, $p < .0001$. Post-hoc tests of this effect showed that gestures aided recall in the meaningful sentences, $t(31) = 4.4$, $p < .0001$, but in the anomalous sentences articulatory movements, and not gestures, aided recall, $t(31)$, 5.39, $p < .0001$.

This interaction was complicated by a significant three-way Sentence Type × Visual Language Condition × Speech Rate interaction, $F(4,120) = 12.88$, $p < .0001$. It appears from Figure 1 that recall performance was higher in the two visual language conditions compared to the speech-alone condition at the 180- and 270 wpm rates, but only for anomalous sentences. However, since each of the condition means is based on data from two sentences per individual, this level of detail in statistical testing may not be reliable. Figure 1 shows the most salient feature of the data from Experiment 1. Specifically, the pattern of responding across all conditions was virtually identical for children and adults, yet children's recall was lower than adults' in every condition.

Similar performance across speech rates for the two age groups implies that there were no age differences in the speed of processing visual-spoken language. If children had been unable to keep up because of reduced ability to quickly connect propositions, their performance would have shown a more marked decline than adult's performance with increases in speech rate. Children also showed the same effects as did adults for the influence of representational gestures and visual articulatory cues. Specifically, for both age groups, articulatory movements aided recall in the anomalous sentence condition, but additional gestures did not. In contrast, for meaningful sentences, both age groups showed the highest recall performance for the condition that included representational gestures in addition to articulatory movements. Yet the presence of only visual articulatory cues did not improve recall for meaningful sentences. The only age difference thus supported by the data in Experiment 1 is one characterizing poorer word retrieval in children compared to adults.

We were concerned that a primacy/recency effect in the children's data could have occurred for children. To determine whether there were any consistent patterns in the children's data for the position of unreported words in sentences, we analyzed the likelihood of recalling words from the beginning, the middle, and the ends of sentences. No consistent errors were made in this regard, thus, children left out words from all parts of the sentences. The results also imply that the overall age difference in level of recall was not based on the fact that children have lower vocabularies and may not have understood particular words spoken. If this had been the case, the fluctuations in performance across conditions should not have followed the same pattern in children and adults, because children would have known all words in some sentences and they would have not known some words in other sentences. Presumably, adults would have semantic knowledge of all words.

Phrase type and recall. The above analyses suggest that nine-year-old children's ability to interpret representational gestures does not differ from young adults. The following analyses approach the issue in a different way, by focusing on the type of words that were recalled when they were accompanied by gestures. We were interested in the phenomenon that children's first words are mostly nouns (nominal terms), and that they acquire verbs and prepositions (predicate terms) later (Gentner, 1978, 1982; Macnamara, 1972; Nelson, 1973). Gentner (1982) has argued that the nominal term category is conceptually more basic than the category of predicate terms. If nominal terms are conceptually distinct from predicate terms in the lexicon, perhaps gestures which co-occur with nominals and predicates might also be treated differently by children. The data from two of the visual language conditions were compared, specifically, the SAG and SA conditions, since the only visual difference between these two sentence conditions was the presence of gestures.

A prediction concerning the benefit to recall of gestured nominal terms versus gestured predicate terms could reasonably go either direction: (a) Children's recall in the SAG and SA conditions was not at ceiling. Children might focus on gestures accompanying nominal terms to a greater degree than they focus on predicate term gestures. Thus it is possible that recall for gestured nominals in the SAG condition compared to recall of nominals in the SA condition would show a larger difference than the recall difference between gestured predicates in the SAG and predicates in the SA conditions. Alternatively, (b) since nominals are already conceptually easier to understand, perhaps children focus more on the gestured predicate terms in an effort to make up for the difficulty in understanding or remembering the predicate concepts. We expected adults to show a benefit to recall of the presence of gestures from both categories.

Transcriptions of the original data from the SAG and SA conditions were reanalyzed for all participants. Nominal terms and predicate terms were identified in each sentence (in the case of SAG sentences, only nominal and predicate terms that were accompanied by gestures were included).[3] The proportion of correctly recalled words fitting the two categories was then calculated for each participant. The data, collapsed across sentence type and speech rate, appear in Figure 2. Two t-tests for matched pairs were performed on the data from the separate age groups. In the children's age group, the proportion of correctly recalled gestured nominals did not differ from the proportion of correctly recalled nominals in the SA sentences, $t(15) = 1.67$, $p = .12$. Children recalled a significantly greater number of gestured predicate terms from SAG sentences, compared to predicate terms in SA sentences, $t(15) = 4.20$, $p < .001$. Adults' recall for nominals in the two conditions did not differ, but there was a marginally significant difference between their recall of gestured predicate terms in SAG, compared to SA, sentences, $t(15) = 2.04$, $p = .06$. Thus, evidence from both age groups in Experiment 1 suggests that the presence of gestures representing an activity or change-of-state aids children's and adults' memory for spoken predicate terms.

To summarize the results of Experiment 1, memory declined in both age groups as speech rate became faster, recall was poorer for anomalous, compared to meaningful sentences, visual articulatory cues aided recall for

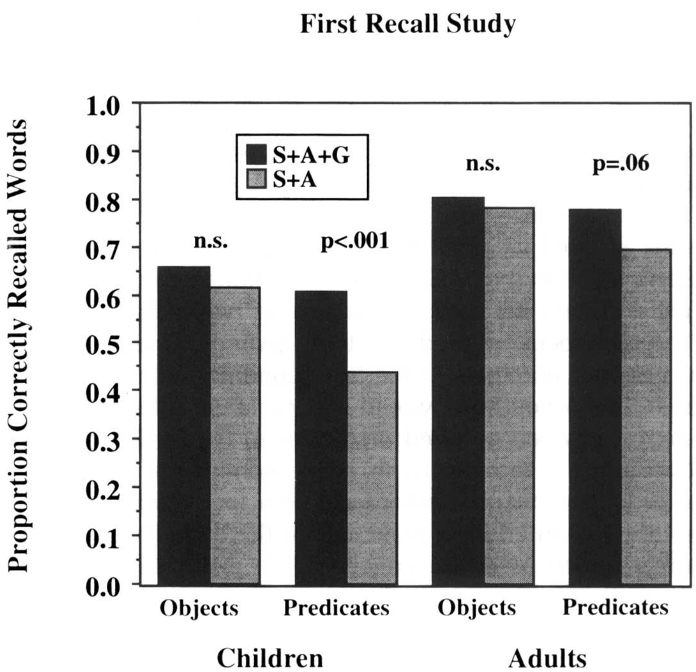

Figure 2. Average proportion of gestured nominal terms, gestured predicate terms, non-gestured nominal terms, and non-gestured predicate terms recalled by children and adults in Experiment 1.(S + A + G = speech plus articulatory cues plus gestures; S + A = speech plus articulatory cues.)

anomalous sentences, and gestures provided benefit to recall in the meaningful, but not the anomalous, sentence condition. Children's memory was overall poorer than adults. Because age did not modify speech rate nor the interaction with sentence type, the data support the view that children's processing speeds do not put them at a comparative disadvantage during the encoding of visual-spoken language. Moreover, both age groups seem to have been focusing more on the gestures which accompanied predicate terms than on the gestures which accompanied nominal terms. This finding will be explored further in Experiment 2.

Experiment 2

The majority of age comparisons in recall performance of visual-spoken language resulted in nonsignificant age effects. The data thus far imply that there is no change between the ages of nine and young adulthood in amount of aid to comprehension which representational gestures and visual articulatory cues can provide. Nor does it appear that the oft-reported increase in processing speed across childhood and adolescence (e.g., Kail & Salthouse, 1994) differentially impacts children's processing of visual-spoken language. However, conclusions based on null effects are problematic, if the results are not replicated. The purpose of Experiment 2 was to test the same hypotheses as were tested in Experiment 1 on a new set of participants, using different sets of sentences and different speakers. If the same pattern of age differences and similarities generalizes across experiments, conclusions regarding these effects can be made with greater confidence.

Method

Subjects. Sixteen ten-year-old children (M = 10.5; range = 10.1–10.9 years; eight males and eight females) and 16 young adults (M = 20.4 years; range = 18–28 years; nine males and seven females) participated in Experiment 2. Both the children and the adult group members were recruited through the Introductory Psychology subject pool (parents of children received course credit for children's participation). The children also received a small toy.

Sentence and videotape construction. The method for making the videotapes in Experiment 2 was exactly the same as in Experiment 1. There were three differences in the videotapes between experiments. First, a new set of thirty-six 16-word sentences was created. Secondly, two different female speakers were audio- and video-recorded. Thirdly, speakers made either three or four gestures in each SAG sentence, most often four gestures, which was less variability than was the case in Experiment 1.

Design and procedure. The only difference in design made to Experiment 2 was that each participant was exposed to both speakers. More specifically, two sets of meaningful sentences were developed for both speakers to recite. As before, the anomalous sentences were written from the same words within each nine-sentence meaningful sentence set. If a participant received a given set of meaningful sentences spoken by one speaker, then the same participant would not receive the anomalous sentences created from that nine-sentence set. Instead, the participant received an anomalous sentence set spoken by the other speaker. This design allowed for less overlap in the words each participant listened to from the meaningful and anomalous sentences.

Results and Discussion

Memory for sentences. Correctly recalled words were coded in the manner described for Experiment 1, and data for all conditions are shown in Figure 3. A 2 (Age) × 2 (Sentence Type) × 3 (Visual Language Condition) × 3 (Speech Rate) ANOVA was conducted on the data. Children recalled significantly fewer words than adults, $F(1,30) = 31.19$, $p < .0001$, meaningful sentence recall was higher than anomalous sentence recall, $F(1,30)$, 349.21, $p < .0001$, recall declined with less visual information, $F(2,60) = 5.49$, $p < .01$, and also declined with increases in speech rate, $F(2,60) = 3.82$, $p < .05$. Several interactions were also significant. As in Experiment 1, the Sentence Type × Visual Language Condition interaction reached significance, $F(2,60) = 9.24$, $p < .0001$. For anomalous sentences, recall was significantly higher in the SA, compared to the S condition, $t(31) = 3.25$, $p < .005$, and gestures did not improve recall compared to the SA condition, replicating Experiment 1. Unlike Experiment 1

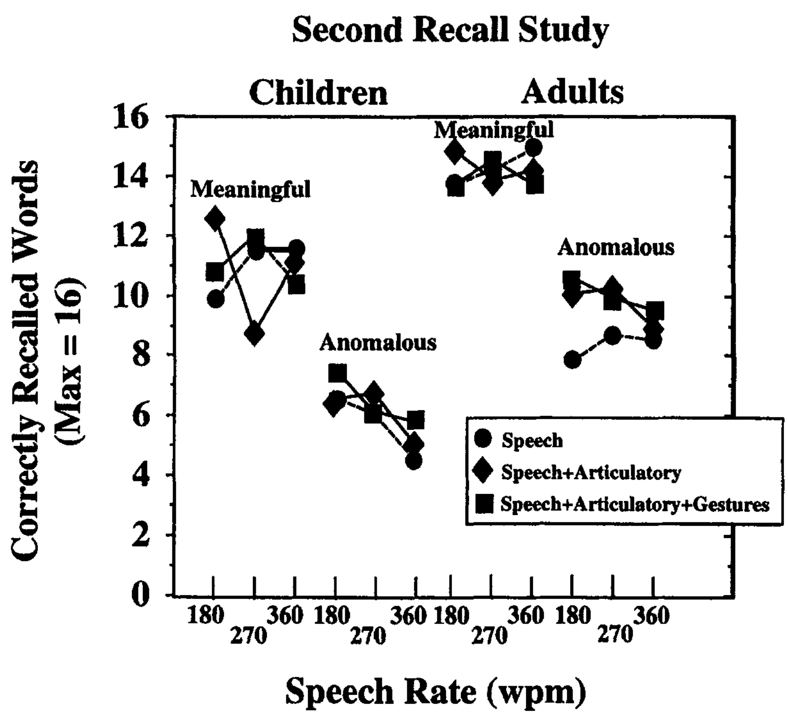

Figure 3. Average number of words correctly recalled by ten-year-old children and adults for all conditions in Experiment 2.

however, the presence of gestures did not result in better memory in the meaningful sentences (all $p > .05$).

The effect of visual language depended on the rate at which it was spoken, as indicated by the Speech Rate × Visual Language interaction, $F(4,120) = 4.72$, $p < .001$. Speech rate also interacted with sentence type, $F(2,60) = 5.56$, $p < .01$. Further, for these three variables a significant Sentence Type × Visual Language Condition × Speech Rate interaction occurred, $F(4,120) = 8.48$, $p < .0001$, and was modified by age group, $F(4,120) = 2.94$, $p < .05$. Looking at Figure 3, it can be seen that the same overall pattern of recall was produced by children and adults, with one exception, children's recall for the medium-speed-SA-meaningful-sentence condition was comparatively lower than their recall for the other conditions. The phenomenon likely represents children not having semantic knowledge of some of the words used in sentences in this condition. The three-way interaction is similar to the one attained in Experiment 1, although for different speech rates. More specifically, in Experiment 1 the visual language effect was greater at the fastest condition (for anomalous sentences) but in Experiment 2 the visual language effect was greatest at the slowest speech rate.

Phrase type and recall. The proportion of each participant's correctly recalled gestured nominal and predicate terms was calculated for the SAG conditions, and for nominal and predicate terms from the SA conditions. The averaged results are presented in Figure 4. In the ten-year-old age group, gestured predicates were recalled at a significantly higher rate than predicates from the SA condition, $t(15) = 5.10$, $p < .0001$, but again, recall for gestured nominals in the SAG condition was not higher than recall for nominals in the SA condition, $t(15) = 1.15$, $p = .27$. In the adult group both comparisons reached statistical significance. Gestured nominals were recalled better than nominals from the SA condition, $t(15) = 3.38$, $p < .005$, and gestured predicates were recalled significantly better than predicates in the SA condition, $t(15) = 2.86$, $p < .01$. Thus, in Experiment 2, the finding of improved recall for gestured predicate terms occurred in both age groups, while improved recall for gestured nominal terms occurred only in the adult age group.

General Discussion

This exploratory study examined the differences and similarities between nine- and ten-year-old children and adults in their ability to use visual

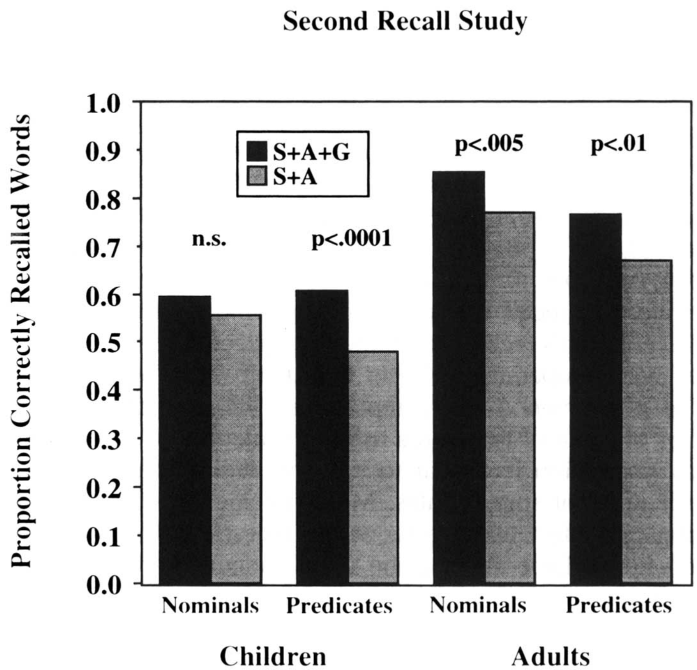

Figure 4. Average proportion of gestured nominal terms, gestured predicate terms, non-gestured nominal terms, and non-gestured predicate terms recalled by children and adults in Experiment 2. (S + A + G = speech plus articulatory cues plus gestures; S + A = speech plus articulatory cues.)

articulatory cues of the face and representational gestures when interpreting and remembering spoken language. Two classes of factors were considered as sources of age differences that could possibly affect memory for visual-spoken language: (a) general factors inherent within children's developing memory systems, including children's comparatively lower level of semantic knowledge, children's slower cognitive processing speeds, and children's difficulty in retrieving information from memory, and (b) factors specifically applicable to the interpretation of visual articulatory cues and gestures, including the type of lexical term-gesture displayed by the speaker. Similarities and dissimilarities between children's and adults' performance were corroborated across Experiments 1 and 2, and each of these findings will be discussed in turn.

The most salient feature of the immediate recall data was the overall age difference in mean level of performance. Adults recalled an average of three words more per sentence than children recalled. Alone, this effect would have left many questions unanswered about the underlying reason(s) for an age difference in recall. However, the multi-factor design utilized in the present study enabled the elimination of some age differences in processing and the support of another. For example, although it is widely accepted that general processing speed is slower for nine-year-old children compared to adults (e.g., Kail, 1990), their performance did not decline more than adults' performance as the speaker's speech rate increased. Thus, nine- and ten-year-old's spoken language processing rates were functionally equivalent to adults' processing rates in the present experiments. This means that children's reduced memory was *not* due to having encoded fewer words while processing the language.

Moreover, children showed the same pattern of decline in performance across speech rate in both types of sentence conditions. Older adults' recall performance drops precipitously as speech rate increases in anomalous sentences, compared to meaningful sentences, which shows that they use a compensatory mechanism of relying to a greater extent on semantic knowledge than young adults under normal conditions of spoken language processing (Stine & Wingfield, 1987; Thompson, 1995). The fact that the children performed in a qualitatively similar manner as adults suggests that they do not use a compensatory mechanism, and also suggests that their level of semantic knowledge did not place them at a comparative disadvantage to adults in these experiments. If the vocabulary chosen for the sentences were more difficult it is likely that children would show greater performance declines across speech rate regardless of sentence condition than would adults. Consequently, the remaining interpretation for children's lower overall performance is that they were poorer than adults at retrieving word meanings from memory.

In terms of the amount of additional visual articulatory information that was incorporated into representations of the visual-spoken utterances, children and adults exhibited no differences. That is, the additional visual articulatory cues aided both groups' ability to process and remember the words from the anomalous sentences, but visual articulation did not improve meaningful sentence recall. Perhaps the listener's attentional focus on visual articulatory cues is stronger when the nature of the spoken language is made more difficult to interpret. Developmental improvement in the reliance on visual articulation cues has been demonstrated in past research testing children up to the age of nine years (e.g., Thompson & Massaro, 1986, 1994). The present results complement this research by revealing no developmental improvement between the ages of nine and adulthood in the benefit to spoken language comprehension given visual articulatory cues.

Moreover, the *amount* of benefit to comprehension provided by representational gestures also did not differ developmentally, since age did not modify the effect of gestures on recall in Experiment 1. However, an important age difference was revealed in the *type* of gestured words which children and adults recalled. Specifically, for children, gestures that appeared along with verbs and prepositions facilitated recall, but gestures that appeared along with nouns did not facilitate their recall. Both nominal term gestures (Experiment 2) and predicate term gestures (Experiments 1 & 2) facilitated adults' recall. These data are very intriguing because they imply that, unlike adults, children differentially focus on predicate term gestures. Perhaps children deliberately pay more attention to predicate term gestures because they are attempting to compensate for the difficulty in learning (Gentner, 1982; Goldin-Meadow, Butcher, Mylander, & Dodge, 1994), and hence memory for, predicate terms. While this interpretation of the data seems somewhat reasonable, there is an alternative explanation.

We suspect that the reason for this finding lies in the nature of the gestures, and not in the language structure itself. Gestures that accompany verbs pictorialize characteristics of actions in a very direct manner, while gestures that accompany nouns often require a more indirect interpretation of an image. Consider the sentence "The library was so quiet that when she *dropped* a stack of *books* it *startled* us." The speaker of this sentence abruptly dropped her hands as she spoke the word "dropped" and moved her hands up to her face in a startled manner when she spoke the word "startled." The gesture she used for books implied a stack of something had been held, but probably was not any more effective in conveying the meaning than the spoken word "books." Perhaps there exists a more direct mapping between the action explicitly communicated by predicate term gestures and their verbs than by nominal term gestures and their parallel words. Children might be especially susceptible to a processing advantage involving fewer steps of interpretation.

Regardless of the source of the advantage in processing these words, one thing is made very clear by the phrase type findings. Namely, representational gestures can be shown to contribute to the meaning of an utterance, because without them, recall of similar phrases is lower. This evidence is consistent with Thompson (1995), who also found better memory for gestured sentences. However, the phrase type data make a stronger case that the meaning of the *gesture* in the sentence is responsible for better recall, since the comparison involved only those parts of the sentences which either appeared (in the SAG sentences) or did not appear (in the SA sentences) with a representational gesture. Some have forcefully argued against the notion that gestures greatly contribute to the meaning of a spoken utterance (e.g., Krauss et al., 1991), citing evidence that people do not remember more information when the information is accompanied by a gesture. In fact, the present Experiment 2 obtained findings consistent with this view. There are two plausible explanations for why gestures sometimes do, and sometimes do not, improve memory.

First, it is entirely possible that the information value of most gestures is not very high in comparison to other dimensions of the message. Given limited processing capacity, it is to our advantage to focus on aspects of visual-spoken language which most directly activate word meanings. Often in normal conversation, the meaning of a gesture is just not interpreted the same way by everyone (Krauss et al., 1991), presumably because we have learned somewhat different meanings for the same gestures. The gestures used in our second experiment may not have been as effective in conveying their intended meanings as those used in the first experiment. A second explanation is that representational gestures are very powerful indicators of meaning, and may actually disrupt processing of the spoken message. Evidence in favor of this explanation can be found in Experiment 2, where three of the four contrasts in the phrase type analysis showed a significant effect on memory of the presence of gestures. Yet, the overall effect of gestures was not significant in the analysis on the entire spoken sentence corpus. A comparison across analyses thus suggests that gestures aided memory for the words they accompanied and actually disrupted processing of words not accompanied by gestures, cancelling out an overall advantage of gestures. If gestures had an additive effect, the effect in the overall analysis would have been maintained.

In face-to-face communication, speakers make available multiple dimensions of cues for the listener to process. By the age of nine, children have logged many thousands of hours decoding this information, and with considerable success, since they show no deficit in their on-line processing capabilities compared to young adults. Yet, when asked to recall the words processed just seconds before, they remember fewer of them, suggesting a memory retrieval deficit for visual-spoken language processing. One practical implication of these results is that both children and adults will learn and remember more when representational gestures and visual articulation cues are included in instructional materials.

Notes

1. Although the amount of gestures produced varies greatly amongst individuals, the spoken sentences admittedly included more gestures than is typical in everyday conversation. We decided to include a higher-than-average number to determine if their presence could

have a noticeable effect on comprehension. Furthermore, the particular gestures we used were created from our own experience. While every attempt was made to create "natural" gestures, some gestures may have seemed opaque and idiosyncratic to some viewers.
2. We were able to achieve full synchronicity of the spoken language and speaker's facial articulatory movements. No participant spontaneously mentioned noticing any asynchronicities. Moreover, those participants who were informed about the dubbing after the experiment were surprised.
3. In the non-gestured (SA) sentences, not every nominal and predicate term was scored for correct recall. This was done to more closely equate SAG and SA conditions in terms of overall numbers of words coded. Regardless, the dependent measure is the *proportion* of words recalled, not the absolute number of words recalled, from each category.

References

Bates, E., Thall, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology, 25,* 1004–1019.
Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439.
Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237–241.
Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49,* 988–998.
Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development, Volume 2: Language, thought and culture* (pgs. 301–334). Hillsdale, NJ: Lawrence Erlbaum.
Goldin-Meadow, S., Butcher, C., Mylander, C., & Dodge, M. (1994). Nouns and verbs in a self-styled gesture system: What's in a name? *Cognitive Psychology, 27,* 259–319.
Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction, 9,* 201–219.
Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61,* 653–663.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.
Kail, R. (1986). Sources of age differences in speed of processing. *Child Development, 57,* 969–987.
Kail, R. (1990a). More evidence for a common, central constraint on speed of processing. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 159–173). North-Holland: Elsevier Science Publishers.
Kail, R. (1990b). *The development of memory in children,* (3rd ed.). New York: W.H. Freeman & Co.
Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86,* 199–225.
Kelly, S. D., & Church, R. B. (1997). Can children detect conceptual information conveyed through other children's nonverbal behaviors? *Cognition and Instruction, 15,* 107–134.
Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, CA: Sage.
Krauss, R., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61,* 743–754.
Kuhl, P., & Meltzoff, A N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development, 7,* 361–381.
Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review, 79,* 1–13.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

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

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

McNeill, D. (1985). Iconic gestures of children and adults. *Semiotica, 62,* 107–128.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago, IL: University of Chicago Press.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38,* 149.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Stine, E. L., & Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychology and Aging, 2,* 272–279.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults., *Psychology and Aging, 10,* 215–228.

Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology 42,* 144–168.

Thompson, L. A., & Massaro, D. (1989). Before you see it, you see its parts: Evidence for feature encoding and integration in preschool children and adults. *Cognitive Psychology, 21,* 334–362.

Thompson, L. A., & Massaro, D. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology, 57,* 327–354.

Wingfield, Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology, 40,* 579–585.