JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 57, 327–354 (1994)

Children's Integration of Speech and Pointing Gestures in Comprehension

LAURA A. THOMPSON

New Mexico State University

AND

DOMINIC W. MASSARO

University of California at Santa Cruz

We examined 4- and 9-year-old's referential comprehension when given pointing gestures and spoken labels, in two types of contextually ambiguous situations. In one situation, speech/gesture discordance was produced in conditions where labels for one of four objects being referred to sounded either alike, or different from each other. In the other, the contextual set contained the same two objects, and ambiguity was produced by factorially combining speech on a continuum ranging between /bɔl/ and /dɔl/ with a pointing gesture from a continuum ranging between an unambiguous point to a ball and to a doll. Results showed that the speech modality had a far greater influence on word comprehension than gestures. Second, the influence of gestures was greater for the older children. Mathematical models of speech–gesture understanding were tested against the data. Selection models assume that one dimension of information is used on a given trial, and that the selection of a modality depends on the ambiguity of information encoded on the dominant dimension. The Fuzzy Logical Model of Perception (FLMP) assumes that both modalities are evaluated independently of one another and then integrated to achieve comprehension. The results from both age groups were best described by the assumptions of the FLMP. Results are related to general claims about perceptual development during childhood concerning the quality of representations formed and dimensional selectivity of visual–spoken language. © 1994 Academic Press, Inc.

The research reported in this article and the writing of the article were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), a James McKeen Cattell Fellowship, and the graduate division of the University of California, Santa Cruz, to D.W.M. The authors thank Michael M. Cohen for assistance at many stages of the project.

0022-0965/94 $6.00

In understanding messages, children must evaluate and integrate multiple dimensions of input into a representation of the speaker's meaning. One of the issues raised concerning the prelinguistic phase of language development is whether or not words and gestures map onto common meanings (cf., Bates, Thal, Whitesell, Fenson, & Oakes, 1989; McNeill, 1985). Most of the empirical work in this area involves naturalistic observation and testing of infants whose naming skills are just beginning to emerge (for a review see Bates & Snyder, 1987). While it is clear that mothers frequently use gestures in their conversational routines with infants (e.g., Shatz, 1982), there is controversy over the extent to which young infants can integrate gestural and spoken cues into representational symbols for objects. Early in development, infants appear to be influenced by just a single dimension of the input. Allen and Shatz (1983) found no effect of gestural input on the vocal responses of 16- to 18-month-old infants. Schnur and Shatz (1984) found that 16-month-old infants benefitted from the attention-directing aspects of gestures, although they argued that the mapping relationships between gestures and words were probably too difficult for children to use in acquiring grammar at this age. Evidence has been provided that recognitory gestures (e.g., drinking from a cup) and deictic gestures (giving, showing, and pointing) of 13- to 15-month-old infants in the earliest stages of word comprehension are unaffected by supportive spoken cues (Bates et al., 1989). However, these gestural responses of infants whose comprehension skills are slightly more advanced *are* affected by supportive versus contradictory gestures and speech (Bates et al., 1989). They are more accurate when speech and gesture agree and they are confused when they do not. Thus, the combination or integration of speech and gesture emerges sometime before children are 2 years old. By this age, children understand the referential use of both pointing gestures and spoken language, and they can integrate speech and gesture into a common representation of the speaker's meaning.

We call a message multimodal when there are several modalities of input relevant to the memory of the message. To date, there are only a handful of published empirical studies of multimodal language understanding in children beyond the age of two (Massaro, 1984; Massaro, Thompson, Baron, & Laren, 1986; McGurk & MacDonald, 1976; Thompson & Massaro, 1986). Each of these studies has shown that children's comprehension is substantially influenced by all available sources of information. For example, McGurk and MacDonald (1976) found that even 3-year-old children are influenced by both the lip movements and auditory speech in syllable identification. The children identified an auditory /ba/ paired with a visual /ga/ as the syllable /da/, for example. In another study, Thompson and Massaro (1986) showed videotaped sequences of a male speaking syllables that were either clear tokens of /ba/ or /da/ or

in the ambiguous region in between, presented with or without pointing gestures toward a doll or a ball. At all ages, both the speech and the pointing gestures influenced choice of the object identified as the speaker's referent. Three-year-old children's responses were the least affected by the gestural information, while 5-year-old children's responses showed a stronger effect compared to younger children, but they were less influenced by it than were adults. Thus, even young children are influenced by gestures, and their syllable identifications become more influenced by gestures as they get older.

One possible explanation of a smaller gestural effect for young children is that they normally devote less attentional capacity to the visual modality than older children or adults. If so, when their attention is directed to visible speech, their identifications should show a greater effect of the visual modality compared to a neutral attention condition. Massaro (1984) tested this hypothesis by requiring 4-year-old children to watch and listen to a videotape of a talker saying a series of test syllables. The children identified the auditory–visual (bimodal) speech patterns as /ba/ or /da/ in single- and dual-task conditions. The single-task condition was a simple identification of the bimodal syllable. In the dual-task condition children were also required to indicate whether or not the speaker's lips moved during the speech event. Both the auditory and visual speech influenced syllable identification, but the children's identifications were no more influenced by the visual syllable in the dual-task, compared to the single-task, condition. Therefore, encouragement to devote more processing capacity to the visual modality does not result in a greater influence of visible speech in 4-year-old children.

Our approach to studying language comprehension is to develop and test theoretical models that specify the nature of the information sources, how the sources are perceived or evaluated, and how they are integrated during language comprehension (Massaro, 1987). To address these issues, we adopt a method of scientific inquiry called strong inference (Platt, 1964). The experimental task, data analysis, and model testing are devised specifically to decide among several theoretical alternatives. Within this framework, it is important to distinguish between the information processing operations involved in multimodal language understanding, and the nature of the language information being processed. Developmental changes can occur in either avenue of information or information processing, or both. The theoretical distinction between information and information processing has been usefully adopted by other researchers of perceptual development, for example, Aslin and Smith (1988).

In the present experiments, children were presented with speech, a pointing gesture, or both speech and gesture and were asked to choose the intended referential object. Although children's gestural–spoken language understanding could differ from adults in many possible respects,

three specific theoretical alternatives have their origins in related developmental research. First, much of the research on infant speech perception would suggest that young children perceive speech categorically rather than continuously (e.g., Eimas, 1974, 1985). Second, speech and gesture could be processed in a holistic, nonindependent manner, instead of as independent sources of information (e.g., Smith & Kemler, 1977). Third, the integration of speech and gesture could follow an additive rather than a multiplicative combinatory rule (Anderson & Cuneo, 1978). Previous research testing children and adults in bimodal speech perception of lip movements and audible speech has been informative with respect to these three theoretical alternatives (Massaro, 1984; Massaro et al., 1986; Thompson & Massaro, 1986). The results indicated that: (a) visible and audible speech provide the young perceiver information about the degree to which a given speech category occurred (noncategorical perception), (b) visible and audible speech are each independently (not holistically) processed at an initial stage, and (c) are integrated without any information loss in such a way that the least ambiguous source has the most impact on ultimate word comprehension (multiplicative rule).

We found that the most important features of the Fuzzy Logical Model of Perception (FLMP), first proposed by Oden and Massaro (1978) and empirically supported in many subsequent speech perception studies with adults (Massaro, 1987), are also applicable to child populations ranging in age from 3 to 6 years old (see also Thompson & Massaro, 1989, for a developmental comparison in visual perception). Thus, we identified and tested many possible developmental differences in the *processing* of multimodal speech understanding, and found none. The developmental differences that were obtained could be attributed to the *nature of the information that is evaluated* at the earliest stages of processing. That is, visible and audible speech are integrated at all ages; however, during childhood, featural representations become more discriminable. This conclusion was supported by the finding that young children are poorer lipreaders than older children and also by the finding of less difference in the parameter values representing both information sources.

The picture of development is not nearly complete, however, because there are theoretical alternatives which have not been subjected to empirical test. Aslin and Smith (1988) posit a curvilinear trend in perceptual development where, during infancy, attributes (or parts) of objects are dominant in perception. Later, in early childhood, parts are relatively inaccessible because they are becoming bundled into whole representations. Finally, by at least the age of 10, the child develops the ability to analyze the whole into its constituent parts. They claim that older children, compared to younger children, exhibit a greater *tendency* to analyze an integrated representation into the parts from which it was constructed. If one were to apply this notion to visual–spoken language processing, the

interesting possibility is that, compared to younger children, older children's classifications should be more consistently based on one channel of information than the other, especially in situations where the speaker is communicating contradictory referents through speech and gesture. In other words, compared to younger children, older children are predicted to show a greater tendency to selectively attend to one modality when gestural and spoken referents clearly conflict with each other.

From the language learner's perspective, one can imagine natural situations of conflicting gestures and speech, resulting in a degree of ambiguity in the speaker's intended meaning. To take two hypothetical examples, a child may receive speech input that supports the representation "ghost" while her father is pointing toward a field of goats; similarly, she may perceive a point in the direction of a llama standing next to a goat in the barnyard zoo but the speech she hears is "goat." Given these contradictory language cues, which modality has the greater impact in children's interpretations of referential acts? Is the tendency to base judgments of reference on one dimension and not the other stronger for older children? Alternatively, both young and older child perceivers may more naturally integrate the separate modalities contained in the constructed representation, and may not differ in their tendency to analyze multimodal speech on the basis of its modality-specific information. Our goal is to determine if the tendency to analyze separate modalities of speech–gesture messages is stronger for older children in a task which does not specifically request modality-specific information.

SUMMARY OF PREDICTIONS

Preschool and fourth-grade children were tested in two phases of a single experiment. In Experiment 1a, we explored whether there were developmental differences in the salience of gestural and spoken dimensions of language. With conflicting gestural and spoken language cues, which dimension carries more influence? Is dimensional salience modified by the acoustic similarity of the names for the objects, and, if so, is this effect modified by age? Children watched a videotape of a person who both pointed at one of four objects and simultaneously spoke an object name. The spoken names for objects were clearly articulated, and the pointing gesture also showed a clear line of regard toward the intended object. In one condition, the objects were a tea bag, a pea (pod), the letter D, and a bee; in the other condition, the objects were the letter K, a shoe, a bow and a pie (slice). The speech and gesture signals either agreed or conflicted, and the child was asked to indicate which object the "person on the tape wants you to choose." If, as Aslin and Smith (1988) claim, older children have a greater tendency to analyze component dimensions of stimuli, their judgments should be highly consistent with either the speech or the gesture, regardless of the similarity of the words,

whereas younger children's judgments should be less consistently based upon a single modality. If, on the other hand, the separate information sources have equal value and their values do not change developmentally, the two groups of children should be equally influenced by both modalities. As a third possibility, the tendency to selectively attend to separate dimensions of bimodal speech may grow weaker with development. This would be manifested in a stronger effect on older children's responses of the less-dominant modality, especially in the condition where the choices sound similar to each other.

In Experiment 1b, we held the choice of referents constant; however, we varied referential uncertainty by manipulating the ambiguity of the speech and gesture signals. We presented combinations of speech and gesture tokens that came from a continuum between two clear alternatives, "ball" and "doll." Consistent with our previous study (Thompson & Massaro, 1986), we expect a larger contribution of gesture information in older children's responses, compared to younger children's responses. We predict that, for both age groups, children's responses will be described by the assumptions of the FLMP. According to the FLMP, both of these sources of information are integrated on every trial, thus, both will contribute to speech–gesture understanding. The contrasting models, on the other hand, do not assume that integration has occurred. Rather, one source of information (either speech or gesture) is used on a given trial. All three models are more fully described in a later section.

METHOD

The experiment was divided into three conditions. Experiment 1a is composed of the two Similar- and Dissimilar-Sounding-Words conditions, while the Speech–Gesture–Continuum condition constitutes Experiment 1b. A Latin-square design was used to determine the order of presentation of the three conditions.

Experiment 1a: Similar- and Dissimilar-Sounding Words

Subjects

Nine preschoolers, four males and five females, ranging between 4.0 and 4.11 (mean age = 4.6), and nine fourth-graders, five males and four females, ranging between 9.0 and 10.4 (mean age = 9.6) participated in this study. The preschoolers were enrolled at the University Child Care Center on campus, and the fourth-graders came from public school. Children received a toy at the end of each experimental session.

Design

There were four objects and three different trial types used to refer to these objects in the Similar- and Dissimilar-Sounding Words conditions.

One type of trial contained speech in the absence of a pointing gesture, a second contained a pointing gesture without speech, and the third trial type combined speech and the pointing gesture to refer to the objects. A 24-trial block was produced by including one replication of the four speech- and four gesture-alone trial types, and one replication of the 16 trials produced by the factorial combination of speech and gesture. Subjects were tested in six 24-trial blocks for each condition, producing a total of 288 trials per subject.

The preschoolers were tested in two-block test sessions that included a break between blocks. Fourth-graders received three blocks in each session, without a break.

Stimuli and Apparatus

Stimuli were presented on a video tape. A woman's right arm was filmed from the shoulder down to the hand, pointing to one of four objects. When not pointing to an object, her arm rested in between the two center objects. The objects were placed on top, and in front of, a black cloth and were equidistant from each other. The distance from her index finger to each object was approximately 6 in. In the Dissimilar-Sounding Words condition, the objects were (from left to right): a wedge of apple pie on a plate (PIE), a wooden block of the capital letter K (K), a small red ribbon bow (BOW), and a child's tennis shoe (SHOE). The Similar-Sounding Words objects were (from left to right): a tea bag (TEA), a wooden block of the capital letter D (D), a yellow and black bee (BEE), and a large pea pod (PEA). The objects were approximately the same size.

The production of the visual portion of the videotape was aided by a computer, which presented cues to the actress for both the timing of her gesture and object to point to. Spoken words for the objects were then dubbed onto the tape. These words were natural speech tokens, recorded from a female speaker. All eight speech tokens were standardized for duration (497 ms) and amplitude (49.5 db-A) using the software program ILS. The timing of the trial events was as follows: a 250-ms bell (a cue) followed after a silent interval of 1050 ms followed by a 350-ms pause, then the 500-ms stimulus, and finally, the 2850-ms response interval.

The children watched the videotape on a 12-in. NEC 1203 color monitor from a viewing distance of 2.5 feet. The audio was presented at a comfortable listening level (70 db-A).

Procedure

All subjects were tested in a research van located outside of the school (Mayer, 1982). Prior to playing the videotape, children were shown the four objects and were given the names for each of them. The experimenter said, "Sometimes (the woman) just points to the thing you should choose.

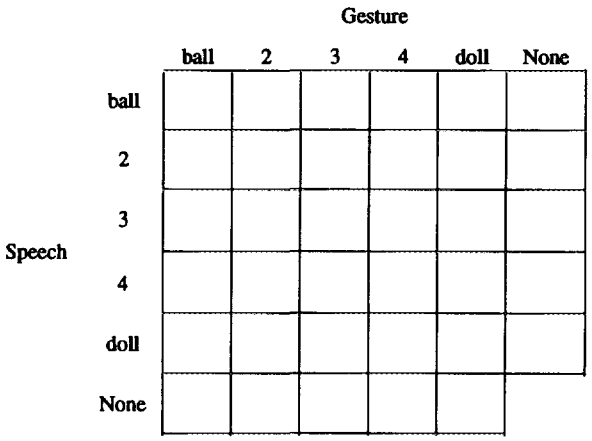

FIG. 1. Expansion of a typical factorial design to include speech and gesture conditions presented alone. The five levels along the speech and gesture continua represent words varying in equal physical steps between ball and doll.

Sometimes she just says it, and sometimes she both points and says which thing she is telling you to choose. Each time though, she wants you to choose one thing, either the (list of objects in view)."

After repeating the instructions, the children began watching the tape. Children reported their responses aloud to the experimenter, who recorded them on paper. If the child looked away from the screen during the trial but gave a response, that trial was recorded as a "miss." Misses were also recorded when the child was paying attention, but made no overt response.

Experiment 1b: Speech–Gesture–Continuum

Subjects

One of the nine preschoolers from Experiment 1a was unavailable for testing in Experiment 1b. The fourth-grade group comprised the same subjects as were in Experiment 1a.

Design

Five levels of the pointing gesture (pointing to the ball, to the doll, or to one of three locations between them), and five levels of speech (ranging from a clear /bɔl/ to a clear /dɔl/), were factorially combined to produce 25 speech with gesture trials. One replication of each of the levels of speech (without gesture) and gesture (without speech) were also included in each 35-trial block. Trial order was randomly determined. Six different blocks were given to each subject, for a total of 210 trials. The preschoolers received two blocks per session, and the fourth-graders received three

blocks per session. The design of this phase of the experiment is shown in Fig. 1. Using an expanded factorial design, we describe how the identification of each speech–gesture event occurs as a function of the identifications of the single modality words that compose it. This design is more powerful than a simple factorial design in differentiating among different models of categorization behavior (Massaro & Friedman, 1990).

Apparatus and Stimuli

A woman's right arm was filmed from the perspective of standing just behind the individual pointing. One of five locations on the viewing screen were referred to: the pointing gesture was aimed directly at one of the two objects, or at one of three positions equally spaced between the two objects. Given the results, it is important to note that the objects were set against a wall. The objects were a Barbie doll and a small colorful ball, approximately 4 in. in diameter. The ball and the doll sat on top of a dark cloth and appeared at the far left and far right corners of the monitor. As in the previous study, a computer controlled a monitor presenting cues for the timing of the gesture during filming.

Five levels of synthetic speech were produced by a software formant serial resonator speech synthesizer (Klatt, 1980) and dubbed onto the tape. By altering the parametric information specifying the first three formants of the consonant-vowel-consonant syllable, a set of five 440 ms syllables covering the range from /bɔl/ to /dɔl/ was created. Although the words ball and doll have different vowels for some speakers, both vowels are pronounced as /ɔ/ in some dialects. We used this vowel in our test words. Figure 2 illustrates how the parameters of the synthetic speech changed for the five stimuli along the continuum.

Procedure

The procedure for Experiment 1b was exactly the same as for Experiment 1a.

RESULTS AND DISCUSSION

Experiment 1a: Similar- and Dissimilar-Sounding Words

There were three types of trials where proportion correct responses could be computed: speech-alone trials, gesture-alone trials, and speech-gesture trials consisting of speech and gesture referring to the same two objects. Overall accuracy on these trials was high, averaging 95%, showing that all children understood the experimental task. Table 1 displays the proportion correct for children in both age groups. Given that performance was so good in the single modality conditions, the accuracy of children's identification responses did not have much room for improvement when both speech and gesture were presented. However, performance was just

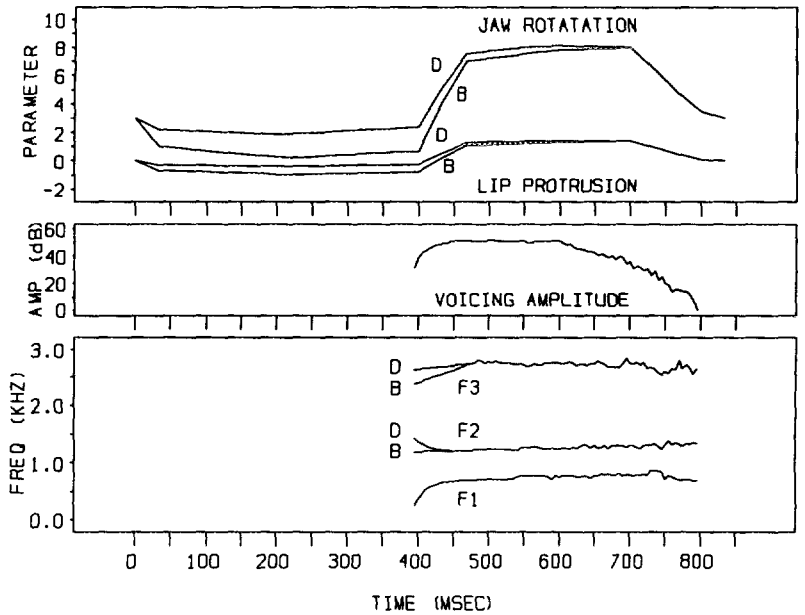

FIG. 2. Schematic illustration of the voicing amplitude of the energy source (top panel) and the values of the first (F1), second (F2), and third (F3) formants for the five speech stimuli during the test word.

poor enough that this expectation was borne out in the Similar-Sounding Words condition for preschoolers. Otherwise, children's accuracy in the speech plus gesture trials did not exceed their accuracy when just one source of information was presented.

The remaining types of trials (when speech and gesture referred to different objects) provide the most pertinent information to determine if the pointing gesture was used differentially, depending on the similarity of the word sounds. An analysis of variance was performed on the data for conflicting speech and gesture trials. In this analysis, age (preschoolers,

TABLE 1
PROPORTION CORRECT RESPONSES FOR CHILDREN IN EXPERIMENT 1a

	Similar-sounding words			Dissimilar-sounding words		
Words	Speech	Gesture	Both[a]	Speech	Gesture	Both
Preschool children	84	88	94	99	93	99
Fourth-grade children	93	97	97	99	96	100

[a] These values are computed from trials in which the gesture and speech referred to the same object.

TABLE 2
PROPORTION OF RESPONSES MADE IN FAVOR OF SPEECH, GESTURE, OR NEITHER, WHEN SPEECH AND GESTURE CONFLICTED

	Similar-sounding words			Dissimilar-sounding words		
	Speech	Gesture	Neither	Speech	Gesture	Neither
Preschool children	74	18	8	98	1	1
Fourth-grade children	71	25	4	79	21	0

fourth-graders) was a between-group factor, while condition (similar, dissimilar), and response type (speech, gesture, neither) were within-group factors. The analysis revealed a significant main effect for response type, $F(2, 32) = 37.83$, $p < .0001$. Children's responses were more strongly influenced by speech when speech and gesture were in conflict. Response type interacted with condition, $F(2, 32) = 14.72$, $p < .0001$, due to a greater degree of influence of the gesture in the Similar-Sounding Words, compared to the Dissimilar-Sounding Words, condition.

However, Table 2 shows that there was also a significant three-way Age $\times$ Condition $\times$ Response Type interaction, $F(2, 32) = 3.65$, $p < .05$. There was a larger difference between the Similar and Dissimilar conditions in the liklihood of a response coinciding with the gesture for preschool than for fourth-grade children. More specifically, when one of the spoken words Pea, D, Bee, and Tea accompanied a gesture to a different object, preschool children were far more likely to be influenced by the gesture compared to the Dissimilar words Shoe, Bow, Pie, and K. However, this was not found to be true for the fourth-graders. For them, the strength of the effect of gesture was relatively high in both the similar- and the dissimilar-sounding word conditions.

Experiment 1b: Speech–Gesture–Continuum

Analyses of Variance

Three analyses of variance (ANOVAs) were carried out on the data for each age group separately, a two-way ANOVA for the speech-alone trials, a two-way ANOVA for the gesture-alone trials, and a three-way ANOVA for the combined speech-gesture trials. Factors in the analyses included the five levels of the two variables in addition to the type of response given. Although there were two specified response alternatives, "ball" and "doll," the children gave a significant number of "wall" responses. Thus, it was decided that the "wall" response should be included as a factor in the analysis.

For the preschoolers, 52% of their responses were "doll," 37% were "ball" and 11% were "wall" in the speech-alone trials. "Wall" responses

occurred more often at the ambiguous levels of the speech continuum. In the analysis of variance for the speech-alone trials, the effect of Response Type was significant, $F(2, 14) = 10.72$, $p < .002$, in addition to the Speech Level × Response Type interaction, $F(8, 56) = 23.34$, $p < .0001$. The children discriminated the levels of the speech continuum very well, as evidenced by the steadily increasing proportion of "doll" responses made to each level of speech as it became more "doll"-like. As a measure of speech discriminability, children's average proportion of "doll" responses to the clearest "ball" sound was subtracted from their average proportion of "doll" responses to the clearest "doll" sound, yielding .89 discriminability.

In comparison, preschool children's visual discriminability measure was .96, for the gesture-alone trials. There were fewer "wall" responses than in the speech-alone condition, 4%, and their "ball" and "doll" responses were divided about equally, 45 and 40%, respectively. In the analysis of variance for the gesture-alone trials, there was a significant main effect for response type, $F(2, 14) = 30.84$, $p < .0001$, in addition to a significant Response Type × Gesture interaction, $F(8, 56) = 24.49$, $p < .0001$. Children's "wall" responses occurred with greatest frequency towards the middle of the gesture continuum, and, not surprisingly, their "doll" and "ball" responses were greatest the closer the pointing gesture was to the appropriate object. Thus, preschoolers' results from the gesture- and speech-alone trials clearly indicate that they could hear and see the rather small differences between stimuli along the levels of the two continua in this experiment.

The preschoolers' results from the combined speech-gesture trials were also submitted to an analysis of variance. Both gesture and speech had a significant impact on preschoolers' responses, $F(8, 56) = 11.03$, $p < .0001$, for gesture, and $F(8, 56) = 19.12$, $p < .0001$, for speech. The size of the effect for speech, calculated by collapsing over all levels of gesture and subtracting the proportion of "doll" responses to the most ball-like speech token from the most doll-like speech token, was .82. This figure was comparable to the discriminability measure when speech was presented alone. However, the size of the gesture effect was only .21, a value far less than the discriminability measure in the gesture-alone condition. Overall, there were 52% "doll," 37% "ball," and 11% "wall" responses, and these percentages were significantly different from each other, $F(2, 14) = 11.65$, $p < .001$. The three-way interaction was nonsignificant.

The fourth-graders' results from the speech- and gesture-alone trials were very similar to the preschoolers' data. In the speech-alone condition, the discriminability measure was .84, with 43% "ball" responses, 49% "doll" responses, and 8% "wall" responses. There were significant main effects for type of response, $F(2, 16) = 28.21$, $p < .0001$, which was

modified by a significant interaction between level of speech and type of response, $F(8, 64) = 27.62$, $p < .0001$. This was primarily a reflection of childrens' tendency to give "wall" responses when the speech was ambiguous. The fourth-graders' gesture-alone discriminability measure was exactly the same as it was for preschoolers', .96. There was a significant main effect for response type, $F(2, 16) = 108.82$, $p < .0001$, in addition to the interaction between response type and gesture level, $F(8, 64) = 89.74$, $p < .0001$. Thus, both the preschoolers and the fourth-graders were about maximally efficient discriminating the differences between gesture and speech in the single modality conditions.

For the combined speech-gesture trials, both gesture and speech had significant effects on the fourth-graders' responses, $F(8, 64) = 6.58$, $p < .0001$, for gesture, and $F(8, 64) = 17.11$, $p < .0001$, for speech. A main effect for response type was obtained, $F(2, 16) = 93.17$, $p < .0001$, indicating a significant difference in the types of responses made in this condition (there were more "ball" and "doll" than "wall" responses). In addition, a three-way Speech × Gesture × Response Type interaction occurred, $F(32, 256) = 1.51$, $p < .05$. For fourth-graders, the type of response was dependent both on the level of the speech and gesture variables.

The size of the effects for speech and gesture were much different for fourth-graders as compared to the preschoolers. More specifically, in the fourth-grade group, the size of the effects for gesture and speech were .39, and .59, respectively. Thus, in the speech-gesture condition, comprehension showed a much stronger influence of gestures in fourth-grade children, compared to preschool children. Moreover, the results were consistent with Experiment 1a both in showing relatively greater influence of speech over pointing gestures in children's comprehension of words in both age groups, and in showing a greater effect of gestures for the older children.

Models of Visual–Spoken Language Comprehension

To test hypotheses concerning developmental differences in the nature of speech-gesture understanding, we compare the FLMP description to two new models. These new models were formulated specifically to address two issues: (a) the selectivity of responding to separate modalities of a representation, and (b) the relative dominance of speech and gestures given ambiguous speech and gesture tokens. The new models are the Speech Selection Model (SSM) and the Gesture Selection Model (GSM). We begin with a description of the FLMP.

The Fuzzy Logical Model of Perception. According to the FLMP, well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns. Three operations assumed by the model are illustrated in Fig. 3. Continuously

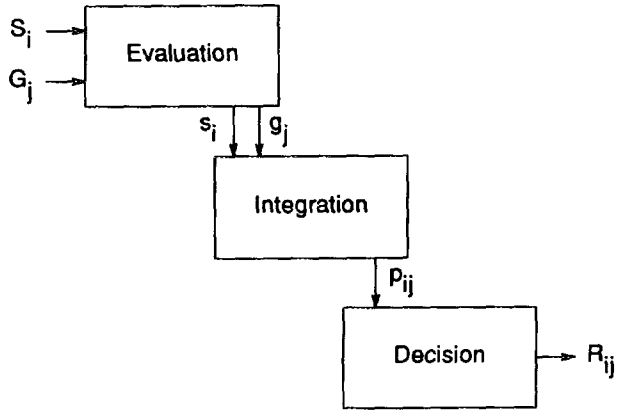

FIG. 3. Schematic representation of the three operations involved in perceptual recognition of a word given speech S_i and gesture G_j information. The evaluation of each source of information produces a truth value indicating the degree of support for each alternative. Integration of the truth values gives an overall goodness of match p_{ij}. The response R_{ij} is equal to the value p_{ij} relative to the goodness of match of all response alternatives. The three operations illustrate their necessarily successive but overlapping processing.

valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions.

Applying the FLMP to this task, both sources are assumed to provide continuous and independent evidence for features along both dimensions of the response alternatives. Defining the speech articulation as the important speech feature and the direction of pointing as the important gestural feature, the distinguishing features for the prototype for "doll" might be something like

"doll" : doll articulation & pointing directly at the doll.

The prototype for "ball" would be defined in an analogous fashion,

"ball" : ball articulation & pointing directly at the ball.

Given a prototype's independent specifications for the speech and gestural sources, the value of one source cannot change the value of the other source. The integration of the features defining each prototype is given by the product of the feature values. If sd_i represents the degree to which the speech stimulus S_i supports the alternative /dɔl/, that is, has a doll articulation, and gd_j represents the degree to which the gesture stimulus G_j supports (is pointing at) the alternative "doll," then the outcome of prototype matching for "doll" would be

"doll" : $sd_i gd_j$

where the subscripts $_i$ and $_j$ index the levels of the speech and gestural

modalities, respectively. Analogously, if gb_i represents the degree to which the speech stimulus S_i has a ball articulation and gb_j represents the degree to which the gesture stimulus G_j is pointing at the ball, the outcome of prototype matching for "ball" would be

"ball" : sb_igb_j.

In the present study, there were three plausible classifications of the stimuli: ball, doll, and wall. The alternative "wall" has a speech sound that is highly similar to the speech sounds of the alternatives "doll" and "ball." In addition, the gesturer was essentially pointing at a wall when the pointing gesture was somewhere between the doll and ball objects. Thus, speech and gesture both support the alternative "wall" to some degree. Thus, if sw_i represents the degree to which the speech stimulus S_i supports a wall articulation and gw_j represents the degree to which the gesture stimulus G_j is pointing at the wall, the outcome of the prototype matching for wall would be

"wall" : sw_igw_j.

The decision operation determines the support for one alternative relative to the sum of the support for each of the possible response alternatives. With only a single source of information, such as speech, the probability of a "doll" response, P("doll"), is predicted to be

$$\text{P("doll")}|S_i) = \frac{sd_i}{sd_i + sb_i + sw_i} = \text{sd}_\text{i}, \tag{1}$$

where the denominator is equal to the sum of the merits of all three response alternatives. Given that the sum of three entries in the denominator are assumed to add to one, P("doll")$|S_i) = sd_i$. An analogous equation holds for the gesture-alone condition so that P("doll")$|G_j) = gd_j$.

With two sources of information S_i and G_j, P("doll") is predicted to be

$$\text{P("doll")}|S_iG_j) = \frac{sd_igd_j}{sd_igd_j + sb_igb_j + sw_igw_j}, \tag{2}$$

where the denominator is equal to the sum of the merit of all three relevant alternatives. Note that the denominator in Eq. 2 is not assumed to add to one.

One important assumption of the FLMP is that the speech source supports each alternative to some degree and analogously for the gestural source. Each alternative is defined by ideal values of the speech and gesture information. Each level of a source supports each alternative to a differing degree represented by feature values. Since we cannot predict the degree to which a particular speech or gesture token supports a response alternative, a free parameter is necessary for each unique speech and gesture token and for each unique response. A speech parameter is

forced to remain invariant across variation in the different gesture conditions and, likewise, for a gestural parameter. For the fit of the FLMP, there are three response alternatives and a free parameter is necessary for each alternative for each level of each source of information. However, we assume that $sd_i = 1 - (sb_i + sw_i)$ and that $gd_i = 1 - (gb_i + gw_i)$. Therefore, 10 levels times 2 = 20 parameters are used in the model to predict 35 conditions times 3 response alternatives = 105 independent data points.

Although perceivers of different ages might process speech and gesture in the same manner, a given level of speech or gesture information will not necessarily have equivalent effects across different ages. In fact, given the inevitable differences in language experience, it is unlikely that a particular speech stimulus will be identified equivalently by two different subjects or by two subjects of different ages (see Massaro, 1992). The hypothesis of no differences in information processing predicts only that the same operations apply across different ages. However, it does not say that the identification results will be exactly the same, because the *quality* of the information processed could differ developmentally. Lower quality would be reflected in less discriminable parameter values representing the encoded feature values for the separate speech and gesture dimensions.

An alternative hypothesis predicts that the FLMP will not give a good description of developmental differences. For example, although a preschool child might identify speech without gesture fairly well, he or she might be overly influenced by gesture when it is paired with speech. In this case the FLMP would fail because identification in bimodal conditions could not be the simple predicted function of identification in unimodal conditions. A second possibility would also falsify the FLMP. Preschoolers might be equivalent to older children in the unimodal condition, but differ significantly in the bimodal condition.

The speech selection model. The selection models make a different assumption about the processing of gestural–spoken messages. These models assume that the bimodal presentation is analyzed into component modality-specific information, and that a judgment is based only on the dominant modality, *unless* the information from the dominant modality is not completely intelligible. In this case, the decision is made in favor of the referent of the less dominant modality. In the SSM, an influence of gesture would occur only when the speech is not completely intelligible.

The SSM was suggested by Sekiyama and Tohkura (1991) in a study of auditory and visual speech perception. The auditory speech corresponded to the auditory syllable and the visual speech corresponded to the face of the speaker articulating the syllable. They tested four labial and six nonlabial consonants in the context of /a/, under auditory and bimodal conditions. The auditory speech was presented either in quiet or in noise. As expected, identification of the speech was very good in quiet

and poor in noise. The influence of visible speech in the bimodal conditions depended on the quality of the auditory speech. There was very little influence with good-quality auditory speech and substantial visible influence with poor-quality auditory speech. In many but not all cases, visible speech had an influence for only those speech segments that were not perfectly identified in the auditory-alone condition.

The speech selection model (SSM) applied to the speech–gesture situation predicts the influence of gesture solely as a function of whether or not the speech is identified correctly. This model is related to a single-channel model (Thompson & Massaro, 1989) in which only a single source of information is used on each trial. According to the SSM, there are two types of trials given a speech–gesture event. On one type, the speech is identified as a particular speech event. On the other type, the speech is not identified as a given alternative. Given these assumptions, the predicted probability of a "doll" response on speech–gesture trials is equal to

$$P(\text{"doll"}) = sd + (Ngd), \tag{3}$$

where *sd* is the probability of identifying the speech source as "doll", *N* is the probability of not identifying the speech source as any specific alternative, and *gd* is the probability of identifying the gesture as a "doll". The value of *N* is 1 minus the sum of the probabilities of identifying a particular speech stimulus as one of the speech categories. Equation 3 predicts that the speech stimulus is either identified or else the subject bases his or her decision on the gesture information. The probability of identifying the speech source as some other known response alternative would be exactly analogous to Eq. 3. In general, the predicted probability of an *r* response is equal to

$$P(r) = sr + (Ngr) \tag{4}$$

where *sr* is the probability of identifying the speech source as *r* and *gr* is the probability of identifying the gesture as *r*. The sum of the *sr* values plus the *N* value are necessarily constrained to add to 1.

On speech-alone trials, the predicted probability of a "doll" response is predicted to be

$$P(\text{"doll"}) = sd + (Nwd)$$

where *sd* is the probability of identifying the speech source as "doll" and *wd* is the bias of identifying the speech source as doll when the speech source has not been identified. In general, the predicted probability of a *r* response is predicted to be

$$P(r) = sr + (Nwr).$$

The *w* values are constrained to add to 1 across the r alternatives.

The probability of identifying the gesture source as alternative r is simply gr, where r indexes a particular response. The gr values are also constrained to add to 1. To apply this model to the results, $(R-1)s+(R-1)g+(R-1)$ free parameters are necessary, where R is the number of response alternatives. With three specific alternatives, $(3-1)5+(3-1)5+(3-1) = 22$ free parameters are necessary.

The gesture selection model. The GSM was formulated in exactly the same way as the SSM, with the exception that speech and gesture values are substituted for each other in the equations. In this case an effect of speech only occurs when the gesture is not completely intelligible. All other aspects of the model are analogous to the SSM. Support for either of the selection models would support the notion that speech–gesture messages are analyzed by subjects selectively attending to the dominant dimension, thus supporting the hypothesis of nonintegration of speech and gesture information.

The FLMP, SSM, and GSM were fit to the individual results of each of six subjects in the two groups. Two of the preschoolers had a large number of "missed" responses, and their data could not be included in the analysis. Two of the fourth graders were randomly deleted from the set for a comparable number of subjects. The quantitative predictions of the model are determined by using the program STEPIT (Chandler, 1969). A model is represented to the program in terms of a set of prediction equations and a set of unknown parameters. By iteratively adjusting the parameter values of the model, the program minimizes the squared deviations between the observed and predicted points. The outcome of the program STEPIT is a set of parameter values which, when put into the model, come closest to predicting the observed results. Thus, STEPIT maximizes the accuracy of the description of a given model. We report the goodness-of-fit of a model by the root mean square deviation (RMSD), the square root of the average squared deviation between the predicted and observed values.

Figures 4 and 5 show that the FLMP provided a reasonable description of the identifications of both the single modality and bimodal conditions for both age groups. The average RMSD was .101 and .092 for the preschoolers and fourth graders, respectively. These averages were computed from the individual fits of the six subjects in each group.

Table 3 presents the average best-fitting parameters of the three models which were fit to the data, including the FLMP. These parameter values index the degree of support for each response alternative by each level of the speech and gesture stimuli. As can be seen in the table, the FLMP parameter values for the gesture variable change in a systematic fashion across the five levels of gesture. However, the first level of the speech continuum from "Ball" to "Doll" proved not to be the clearest token of "Ball." This level was a better match to "Wall." These parameter values

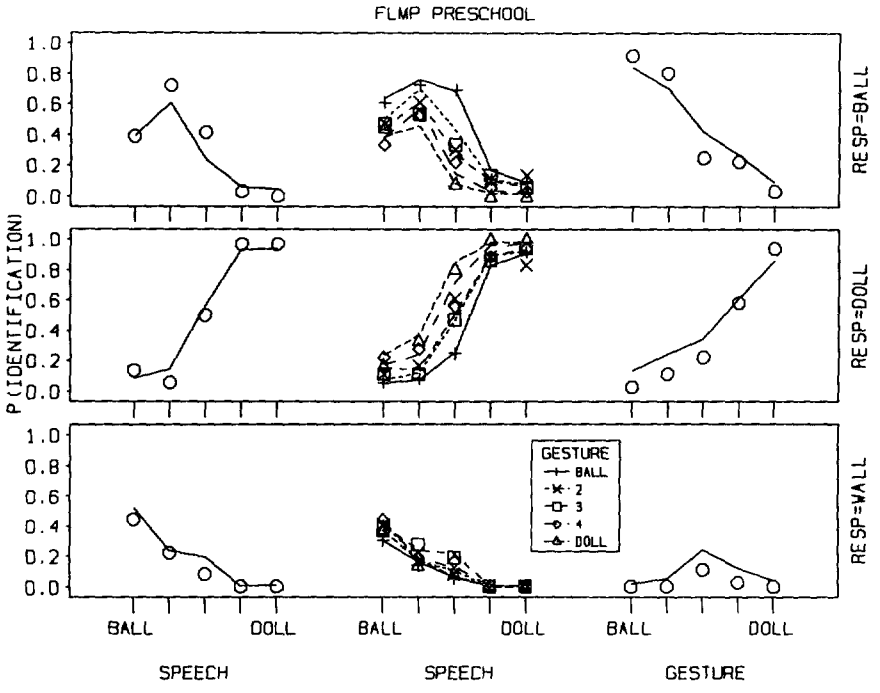

FIG. 4. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for preschool children. The predictions are for the FLMP.

account for the relative contribution of the speech and gesture to the judgments shown in Figs. 4 and 5.

Figures 6 and 7 give the average observed and predicted values for the SSM for the two age groups separately. The average RMSD was .144 and .114 for the preschoolers and fourth graders, respectively. The analysis of variance on the RMSD values showed that the FLMP gave a significantly better description of the results than did the SSM, $F(1, 10) = 5.37$, $p < .041$. This effect did not interact with age.

Figures 8 and 9 give the average observed results and the average predicted results of the GSM. The average RMSD was .137 and .136 for the preschoolers and fourth graders, respectively. An analysis of variance on the RMSD values showed that the FLMP gave a significantly better description of the results than did the GSM, $F(1, 10) = 17.39$, $p < .002$. The interaction between age and model was nonsignificant.

To summarize the model tests, children's judgments were best-predicted by the assumptions embodied by the FLMP. Children's comprehension of words referred to by speech and gestures is the end-product of three

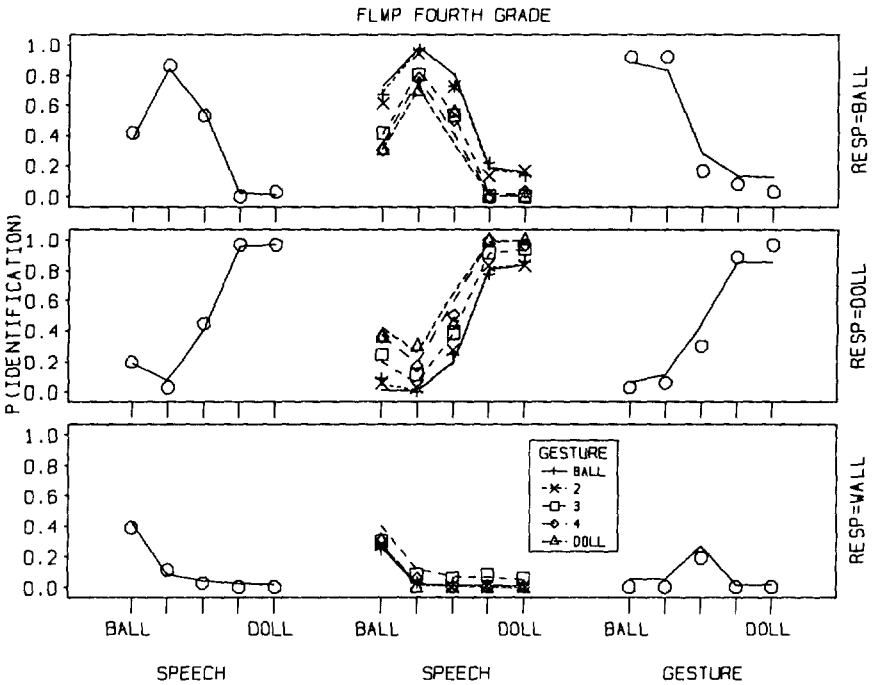

FIG. 5. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for fourth-grade children. The predictions are for the FLMP.

stages of information processing. Gestures and speech are evaluated independently, combined in a multiplicative fashion, and a decision is made based on a relative goodness of match. The SSM and the GSM both assume a process of selective attention to a modality without integration. Both also predict that one source of information dominates comprehension, and that the secondary source will have an influence only when the dominating source is ambiguous. The fact that the FLMP fit the data significantly better than the other two models tested supports the assumption that both sources of information always contribute to comprehension, even though one source may have more influence than the other. Although both sources always contribute, speech has more influence than gesture for both age groups.

GENERAL DISCUSSION

It is not always transparent to children (or adults) what someone is referring to in the observable environment, because aspects of the referential context sometimes ambiguously define the speaker's intended mean-

TABLE 3

ROOT MEAN SQUARE DEVIATIONS (RMSDs) BETWEEN OBSERVED AND PREDICTED VALUES AND BEST-FITTING PARAMETER VALUES FOR THE FLMP, SSM, AND GSM IN EXPERIMENT 1b

			Speech					Gesture				
Model	RMSD	Response	Ball	2	3	4	Doll	Ball	2	3	4	Doll
						Fourth-graders						
FLMP	.101	Ball	.383	.842	.553	.021	.013	.884	.830	.282	.137	.129
		Doll	.120	.073	.266	.947	.960	.129	.147	.488	.697	.720
		Wall	.548	.100	.109	.014	.009	.118	.132	.197	.093	.086
SSM	.114	Ball	.330	.746	.511	.000	.002	.855	.824	.235	.061	.009
		Doll	.067	.000	.293	.836	.867	.025	.046	.400	.798	.865
		Wall	.603	.254	.196	.164	.131	.120	.130	.364	.141	.125
GSM	.136	Ball	.475	.942	.635	.000	.000	.261	.233	.017	.000	.000
		Doll	.129	.000	.338	.990	.100	.000	.000	.135	.294	.344
		Wall	.396	.058	.027	.010	.000	.739	.767	.848	.706	.656
						Preschoolers						
FLMP	.101	Ball	.393	.611	.239	.062	.047	.844	.705	.416	.264	.091
		Doll	.061	.111	.515	.944	.960	.143	.298	.340	.569	.740
		Wall	.593	.339	.188	.017	.013	.148	.194	.257	.169	.084
SSM	.114	Ball	.175	.391	.097	.041	.018	.792	.574	.377	.245	.146
		Doll	.000	.030	.357	.888	.911	.036	.191	.252	.468	.662
		Wall	.825	.580	.547	.071	.071	.172	.235	.371	.286	.192
GSM	.137	Ball	.487	.692	.367	.014	.000	.208	.119	.064	.015	.000
		Doll	.012	.045	.491	.980	1.00	.000	.072	.128	.282	.412
		Wall	.501	.263	.142	.005	.000	.792	.809	.808	.703	.588

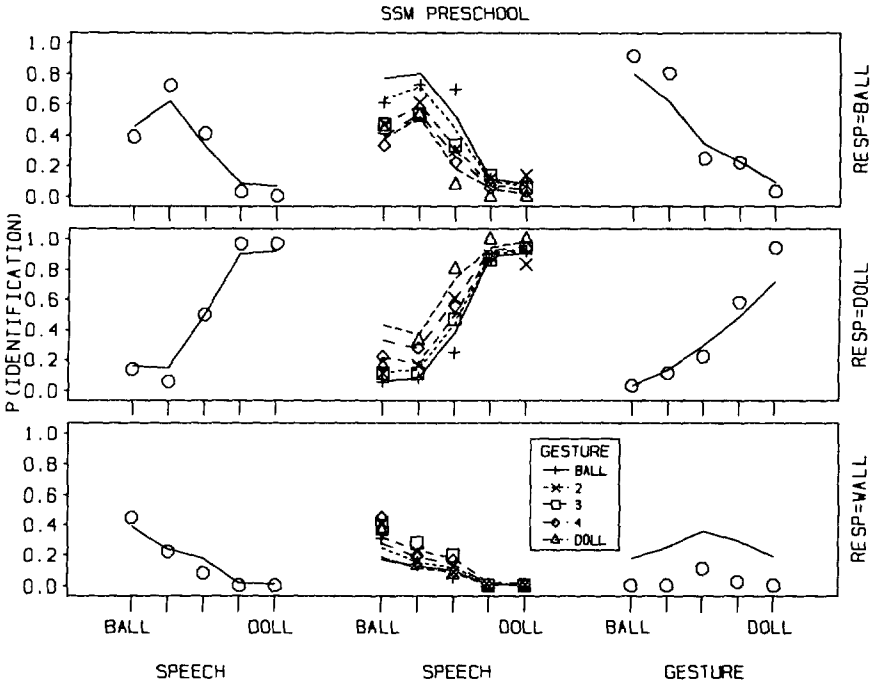

FIG. 6. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for preschool children. The predictions are for the SSM.

ing. Ambiguity in the speech signal itself is of paramount concern for problems of speech recognition because an utterance can differ between speakers or from repetitions by the same speaker. While adults attempt to communicate similar or complementary meanings through gesture and speech (McNeill, 1985), they sometimes may not be understood quite as well as they had intended. That is, an adult, by mumbling, by speaking too softly, or by speaking in a background of noise, could cause the listener to access the meaning of a word which was not spoken, and which perhaps did not even match the meaning communicated by a gesture. A rarer, but real situation, occurs when there is clear gesture/speech discordance. Gesture/speech discordance is sometimes present in the productions of children who are in a phase of transitional knowledge (Church & Goldin-Meadow, 1988). One of our interests was in determining how children assign importance to referential signals which are either in conflict, or, because of the graded nature of the stimuli, simply ambiguous. However, our primary focus was on exploring general issues of dimensional selectivity and integration, which have previously been discussed

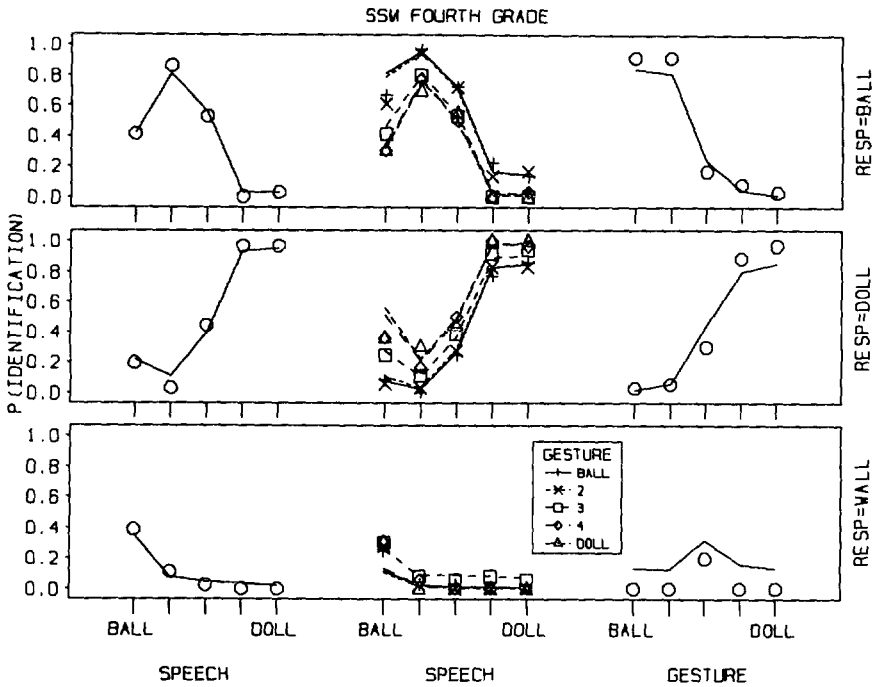

FIG. 7. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for fourth-grade children. The predictions are for the SSM.

solely in terms of separable dimensions of the visual world (e.g., Aslin & Smith, 1988; Cook & Odom, 1992; Smith, 1989).

In visual perceptual classification, Smith (1989) claims that young children become better able to selectively attend to separate dimensions as they get older. Extending this notion to speech–gesture comprehension, compared to the younger children, the older children in our study should have showed a greater tendency for their responses to favor one of the two dimensions. Evidence against a dimensional selectivity account of speech–gesture comprehension can be found in an interesting interaction which occurred in Experiment 1a. In conditions of gesture/speech discordance, the influence of gesture was only slightly greater for fourth graders than preschoolers when the words were similar. However, in the Dissimilar-Sounding word condition, there was no influence of gesture on preschooler's responses, but fourth-grader's responses favored gestures on average 21% of the time. Thus, contrary to expectations derived from Smith's (1989) model of the development of perceptual classification, the

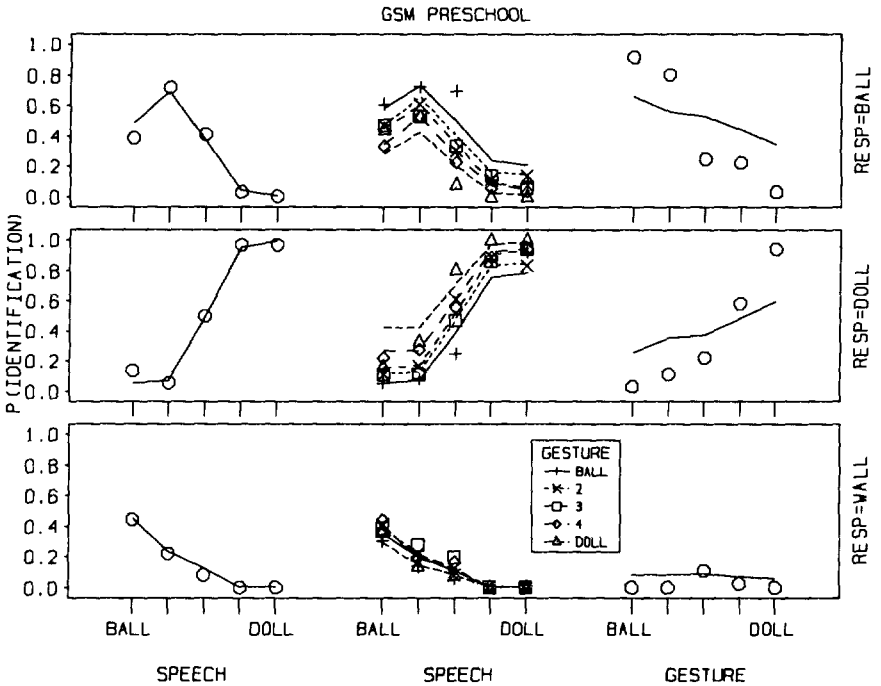

FIG. 8. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for preschool children. The predictions are for the GSM.

younger, not the older, children produced the most classifications reflecting the influence of only one of the dimensions.

Further evidence against a developmental dimensional selectivity account can be found in the model-fitting results of Experiment 1b. While children's judgments of reference were more likely to coincide with speech than gesture, our results are not consistent with the interpretation that they selectively attend to the speech modality. In speech-gesture word comprehension, as in classifying visually perceived objects, a categorical judgment is the end result of several processes, including encoding, integration and comparison operations. The model tests from Experiment 1b allowed us to more precisely define and test the hypothesis that one source of language information might dominate final comprehension due to selective attention to that modality. The selection models make the assumption that information about the intelligibility of the separate speech-gesture dimensions is accessed, and that the listener makes a judgment based partly on this information and according to their dimensional bias. For example, the speech selection model assumes that the effect of

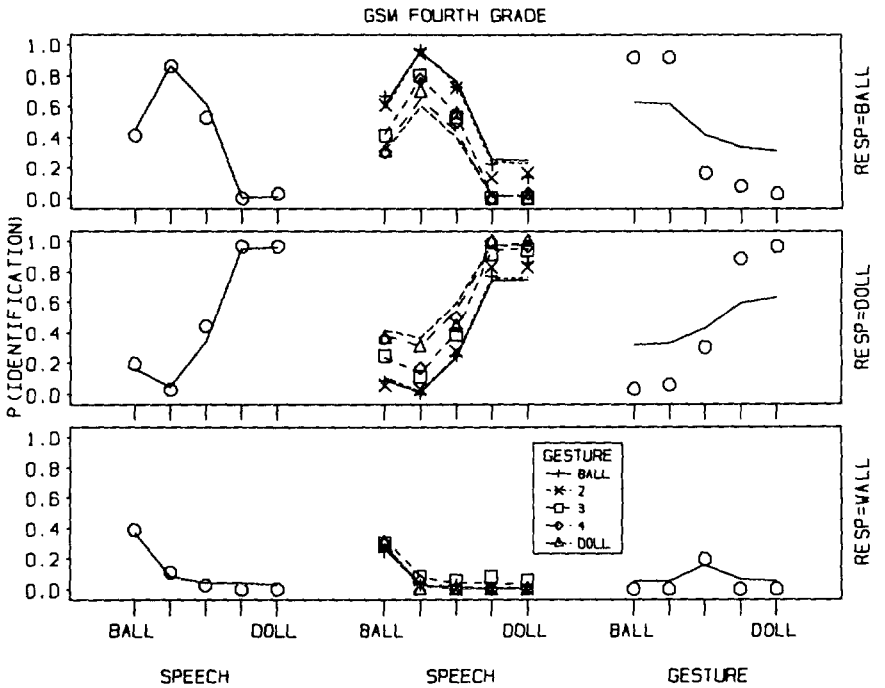

FIG. 9. Observed (points) and predicted (lines) proportion of ball, doll, and wall identifications for speech (left panel), speech–gesture (center panel), and gesture (right panel) trials as a function of the speech and gesture levels of the speech event for fourth-grade children. The predictions are for the GSM.

gesture occurs only when the speech is not completely intelligible. These assumptions of dimensional selectivity which we tested were not supported by the results of these tests. The FLMP assumes a true integration of the separately perceived sources. The significantly better fits of the FLMP over the selection models demonstrates the flexible outcome of human language processing; if the integrated representation has poor quality, or ambiguous, information from either modality, the other dimension will have more influence on final comprehension.

Our previous research showed a developmental trend toward greater influence of gesture information for adults, compared to young children (Thompson & Massaro, 1986). The present study replicates this result using 9-year-old children instead of adults as the age contrast and also extends our previous findings to conditions of ambiguity of gestural reference as well as to conditions of complete gesture/speech discordance. In Experiment 1a, older children's responses were more influenced by the gesture in the Dissimilar words condition compared to preschoolers. In Experiment 1b, with combined speech–gesture stimuli, the size of the

effect of speech was .82 for preschoolers and .59 for fourth-grade condition compared to preschoolers. In Experiment 1b, with combined speech–gesture stimuli, the size of the effect of speech was .82 for preschoolers and .59 for fourth-graders. Our explanation of the effect is that as children get older, the visual modality becomes more informative, consequently, visual language gains more prominence in the integrated representation. On this point, we are in agreement with Aslin & Smith's (1988) claim that the quality of multidimensional representations improves with development.

We distinguished between the quality of information from information processing. Four- and 9-year-old children differ in the quality of the information used in speech–gesture comprehension, not in the evaluation, integration, and decision operations which transmit and transform language information during the process of language understanding. Evidence from the present study showing an age difference in information, not information processing, is consistent with our previous findings on the development of visual-spoken language comprehension (Massaro, 1984; Massaro *et al.*, 1986; Thompson & Massaro, 1986). Moreover, the present study extends our previous work through the derivation and tests of two new mathematical models of visual–spoken language comprehension.

Finally, the present results can be placed in the broader context of developmental changes in the production and representation of speech and gestures. McNeill (1992) has shown that young children's utterances most often are not accompanied by gestures. Children use gestures with greater frequency up until about five when they exhibit adult frequency levels, although, there are major differences in the content of children's and adult's gestures. For example, McNeill (1986) has shown that even 9-year-old children's iconic gestures are very different from adults' iconic gestures. Children's gestures contain indistinct boundaries between them, they have a narrowed temporal locus, they take up more space relative to their smaller body size, and show other spatially different characteristics from the structure of adult iconic gestures. The important parallel to be found between our results and McNeill's work is that many aspects of nonverbal language production are still being acquired in childhood until adolesence (McNeill, 1992). If the comprehension and production of language follow similar developmental courses, many more important changes in speech–gesture comprehension remain to be uncovered in this little-studied period of language growth.

REFERENCES

Allen, R., & Shatz, M. (1983). "What says meow?": The role of context and linguistic experience in very young children's responses to *what*-questions. *Journal of Child Language,* **10,** 321–335.

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgements of quantity. *Journal of Experimental Psychology: General,* **107,** 335–378.

Aslin, R. N., & Smith, L. B. (1988). Perceptual development. *Annual Review of Psychology*, **39**, 435–473.

Bates, E., & Snyder, L. (1987). The cognitive hypothesis in language development. In I. Uzgiris & J. McV. Hunt (Eds.), *Research with scales of psychological development in infancy* (pp. 168–206). Champaign–Urbana: University of Illinois Press.

Bates, E., Thall, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology*, **25**, 1004–1019.

Chandler, J. P. (1969). Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, **14**, 81–82.

Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, **3**, 359–400.

Cook, G. L., & Odom, R. D. (1992). Perception of multidimensional stimuli: A differential-sensitivity account of cognitive processing and development. *Journal of Experimental Child Psychology*, **54**, 213–249.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, **216**, 513–521.

Eimas, P. D. (1985, January). The perception of speech in early infancy. *Scientific American*, **252**, 46–52.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971–995.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, **55**, 1777–1788.

Massaro, D. W. (1987). *Speech perception by ear and by eye*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1992). Broadening the domain of the fuzzy logical model of perception. In H. L. Pick, Jr., P. Van den Broek, & D. C. Knill (eds.) *Cognition, conceptual, and methodological issues*. Washington, DC: American Psychological Association.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, **97**, 225–252.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, D. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, **41**, 93–113.

Mayer, M. J. (1982). A mobile research laboratory. *Behavioral Research Methods and Instrumentation*, **14**, 505–510.

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

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, **92**, 350–371.

McNeill, D. (1986). Iconic gestures of children and adults. *Semiotica*, **62**, 107–128.

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

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172–191.

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

Schnur, E., & Shatz, M. (1984). The role of maternal gesturing in conversations with one-year-olds. *Journal of Child Language*, **11**, 29–41.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, **90**, 1797–1805.

Shatz, M. (1982). On mechanisms of language acquisition: Can features of the communicative environment account for development? In L. Gleitman & E. Wanner (Eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge Univ. Press.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, **96,** 125–144.

Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, **24,** 279–298.

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. W. (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.

RECEIVED: October 2, 1992; REVISED: November 2, 1993