JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 42, 144–168 (1986)

Evaluation and Integration of Speech and Pointing Gestures during Referential Understanding

LAURA A. THOMPSON AND DOMINIC W. MASSARO

University of California, Santa Cruz

Two experiments investigated the relative influence of speech and pointing gesture information in the interpretation of referential acts. Children averaging 3 and 5 years of age and adults viewed a videotape containing the independent manipulation of speech and gestural forms of reference. A man instructed the subjects to choose a ball or a doll by vocally labeling the referent and/or pointing to it. A synthetic speech continuum between two alternatives was crossed with the pointing gesture in a factorial design. Based on research in other domains, it was predicted that all age groups would utilize gestural information, although both speech and gestures were predicted to influence children less than adults. The main effects and interactions of speech and gesture in combination with quantitative models of performance showed the following similarities in information processing between preschoolers and adults: (1) referential evaluation of gestures occurs independently of the evaluation of linguistic reference; (2) speech and gesture are continuous, rather than discrete, sources of information; (3) 5-year-olds and adults combine the two types of information in such a way that the least ambiguous source has the most impact on the judgment. Greater discriminability of both speech and gesture information for adults compared to preschoolers indicated small quantitative progressions with development in the ability to extract and utilize referential signals.

When referring to objects that are visually present, people often use gestures in addition to spoken language. For example, an adult might say to a child, *Look at that big doggie!* while pointing in the direction of a dog that is running across the street. The information available to the child during referential acts such as these may be ambiguous. That is, the speech signal itself may be ambiguous so that the child is uncertain

This research was supported in part by a National Institute of Neurological and Communicative Disorders and Stroke Grant 20314, and a National Science Foundation Grant BNS-88-15192. We thank Ray Gibbs, Marjorie Horton, Barry McLaughlin, and Sharon Oviatt for their helpful comments on the manuscript, and Phillip Pillin for his help in data collection. Michael Cohen provided technological support. We are especially grateful to Margaret Ann Krzyzostan and the staff and children at the University Child Care Center, and to the Family Student Housing Center, for their enthusiasm and cooperation. Requests for reprints should be mailed to either author at the Program in Experimental Psychology, Clark Kerr Hall, University of California, Santa Cruz, CA 95064.

0022-0965/86 $3.00

as to which speech sound he or she has heard. The child may also be uncertain as to which of the potentially many objects present, or even which aspect of a single object, is actually being pointed to. Moreover, the child could experience difficulty integrating gestures and speech due to unsophisticated processing strategies. The primary goal of the present experiments was to study the simultaneous processing of spoken information and pointing gestures in the interpretation of an act of reference. We also investigated differences and similarities between preschoolers and adults in the relative reliance on speech and gestures for on-line referential comprehension.

In studying the ontogeny of gestural and spoken language, some investigators have recorded detailed observations of very young children in naturalistic settings, usually interacting with their mothers (e.g., Bates, Camaioni, & Volterra, 1975; Dobrich & Scarborough, 1984; Schnur & Shatz, 1984; Shatz, 1982). It has been shown, for example, that mothers use only a few types of gestures when communicating with their infants, including pointing, holding out, and demonstration (Shatz, 1982). Of course, the fact that mothers provide a manageable repertoire of gestures for their infants in no way demonstrates that the infants know how to interpret them. Indeed, an experimental study reported by Shatz (1984) showed that the presence of experimenter-produced gestures increased the probability of a nonverbal response from the infant, but actually reduced the likelihood of an appropriate verbal response.

By the age of 2 years, children use the pointing gesture frequently (Masur, 1982; Murphy, 1978). Yet, even this evidence does not fully establish that the young child really interprets the referential uses of the pointing gesture as well as adults. In fact, two recent studies revealed a difference between young children and adults in the production of speech and pointing gestures. Dobrich and Scarborough (1984) found that 2-year-old children had not mastered gestural form or function in their productions of pointing gestures. Whereas the mothers' speech and gestures were produced synchronously, children often pointed without speech. Moreover, children used the pointing gesture spontaneously in response to their mother's questions, but rarely in "non-cued" contexts, when referring to objects around them. Deutsch and Pechmann (1982) also found that young children produced more ambiguous gestures and linguistic descriptions compared to older children and adults. Thus, this evidence is suggestive of the general hypothesis that young children's strategies for interpreting pointing gestures may not be fully developed.

How might we characterize exactly what goes on in the mind of the child when both spoken and gestural information are available for interpreting an act of reference? We hypothesized that the processing of gestures and speech proceeds somewhat independently at some initial stage, followed by a stage of processing in which the two sources of

input are integrated into a common representation. Given this conceptualization, several questions arise concerning this hypothetical system for comprehending gestures and speech. How can we best describe the type of perceptual information registered at the earliest stages of comprehension? What are the computations that are performed on these separate representations? Do either of these factors change with development?

Our aim in conducting the present studies was to clarify some of these issues by making explicit, and testing, certain assumptions about information processing during spoken and gestural language comprehension. Unlike previous studies, however, we have begun investigating comprehension using factorial designs to manipulate speech and gesture independently to determine how the child uses these two sources of information and to assess the degree to which they interact. The techniques of information integration (Anderson, 1980) and mathematical model tests afforded us the opportunity to do a fine-grained analysis of information processing during referential comprehension.

We have shown in previous research that the information integration paradigm is quite useful in revealing the underlying cognitive mechanisms used by children in perceiving speech accompanied by visual (facial) information (Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986). In these studies, children and adults watched and listened to a videotaped presentation of a man who was saying either /ba/ or /da/. Their task was to report what the man said. The speech information was actually seven levels of computer-synthesized speech, varying between /ba/ and /da/, which was dubbed onto the tape. Although both sources of information (visual and spoken) influenced identification performance, children were less influenced by the visual source than were the adults. By presenting just the visual source of information, it was observed that the children were poorer lip-readers than adults. In addition, children's ability to read lips was significantly correlated with the contribution of lip information in their final interpretations of auditory–visual speech. This suggested to us that children's knowledge and use of gestures, another source of visual input, might be similarly described.

As in our previous studies, the present experiments were used to test opposing quantitative models of speech identification performance (Massaro, 1984; Massaro & Cohen, 1983; Massaro et al., 1986). These models differ with respect to three important questions concerning the pattern recognition processes used to evaluate and integrate any two sources of information, such as gesture and speech. First, is the perceived information continuous (one of degree) or discrete (categorical)? Second, are the two sources of information evaluated independently of one another, or does the value of one source of information influence the value of the other? In our earlier developmental studies (Massaro, 1984; Massaro et al.,

1986), the assumptions of independent and continuous sources of information in speech perception were supported for all age groups. Finally, what is the nature of the operation used to integrate the two sources of information? The integration algorithm consistent with our previous results is one in which the least ambiguous source has the most influence on perceptual recognition.

We attempted to answer these questions using speech and the pointing gesture in two experiments. In both experiments, all subjects watched a videotape of a man referring to a ball or to a doll by gesturing to and/or vocally labeling the referent. The subjects were to identify whether the man told them to choose the ball or the doll. There were seven levels of speech information, ranging from /ba/ to /da/, and two levels of gestural information, pointing to the ball or to the doll. Three presentation conditions were included: speech information alone, gestural information alone, and speech combined with gestural information. When speech and gesture were combined on a trial, the two were sometimes consistent with one another, sometimes in conflict, and sometimes ambiguous speech appeared with the unambiguous point to the object. Experiment 1 tested 5-year-olds and adults, and Experiment 2 tested 3-year-olds.

Our general predictions were derived from previous investigations of developmental changes in the perception of speech containing both spoken and visual input (Massaro, 1984; Massaro et al., 1986; McGurk & MacDonald, 1976). First, we predicted that all subjects would utilize both forms of reference, although children's utilization should not be as good as adults'. Given that children were found to use the same recognition processes as adults in our previous studies, we predicted the same for the referential task. Thus, speech and gesture should provide continuous and independent sources of information, and gestures should have their greatest contribution when the speech information is ambiguous.

EXPERIMENT 1

Method

Subjects. Two groups of subjects were tested in this experiment. One group was composed of 18 children, 10 males and 8 females, between the ages of 5;0 and 6;3 (mean age = 5;6). Fifteen of the children were from the University Child Care Center, and 3 were from the Family Student Housing Child Care Center. The children received a small toy at the end of each session. Eighteen adults, 9 males and 9 females, aged 19–26 (mean age = 22), were also tested. The adults were university students and received either course credit or payment for their participation in the experiment. One other adult was dropped from the analysis due to a large percentage of missed responses occurring over every trial type.

Design. There were equal numbers of three different trial types presented randomly in each 42-trial block: speech information alone, gestural in-

TABLE 1
EXPERIMENTAL DESIGN FOR EXPERIMENT 1

	Lips moved		Lips did not move	
	Levels	Replications	Levels	Replications
Speech	7	2	7	2
Gesture	2	7	2	7
Speech and gesture	14	1	14	1

Note. The number in the Levels column, when multiplied by the number in the Replications column, yields the number of trials received by each subject for the corresponding condition in each block of 42 trials. There were three blocks in which the lips moved and three blocks in which the lips did not move, giving a total of 252 trials.

formation alone, and the bimodal combination of speech and gestural information (see Table 1). The speech factor contained seven levels of synthetic speech information ranging from an unambiguous /ba/ to an unambiguous /da/. The gestural factor contained two levels, a point to a ball and to a doll. Each block contained two replications of the seven levels of speech information for the speech-alone condition, seven replications of the two levels of gestural information for the gesture-alone condition, and one replication for each of the 14 unique trials created by the combination of speech and gestural information for the bimodal condition. The order of presentation of the trials was randomly determined within each of three blocks of 42 trials. Each block consisted of a different random ordering of the trials, and a Latin square design was used to determine the order of block presentation.

Another factor, whether or not the speaker's lips were moving during the gesture, was included in the design. We included this factor to assess how much of the visual influence could potentially be due to subject's understanding of the articulatory, as opposed to gestural, form of reference. Three blocks contained this source of information, and three did not. When this source of information was present, the speaker's lips moved only on trials containing a gesture. Table 1 shows the breakdown of trials in both types of 42-trial blocks. Half of the subjects from each age group were presented with three blocks of trials in which the speaker's lips moved, followed by three blocks where the speaker's lips did not move. Thus, each subject received a total of 252 experimental trials. For each subject, the order of presentation of the three blocks was repeated in the second half of the experiment, when the lip-moving factor was changed.

Test stimuli. A videotape was recorded to present the experimental stimuli. An adult male (the speaker) was filmed sitting behind a table and in front of a wood panel background. The view included the top of

the table and the speaker from the waist to above the head. A Barbie doll and a small colorful ball of approximately the same size served as the referents. The doll sat on the table on the speaker's right and the ball on his left. The speaker was cued by a computer-controlled monitor in front of him. The cue informed him to provide the appropriate gestural referent by pointing and/or speaking. The pointing gesture was made by extending the right arm out, elbow bent, and pointing with the right forefinger. For the condition in which the speaker's lips moved, the speaker also articulated the word ball or doll during the pointing. For the condition without lip movement, the speaker kept his mouth still during the pointing.

The original audiotrack was replaced by synthetic speech, as in the Massaro and Cohen (1983) study. Due to difficulties in synthesizing the final consonant of the words, /l/, just the initial segments /ba/ and /da/ were presented to the subjects. The synthetic speech was produced by altering the frequency of the second and third formants during the initial 80 ms of the consonant–vowel pairs (CVs). During the first 80 ms, the first formant (F1) went from 250 Hz to 700 Hz following a negatively accelerated path. The F2 followed a negatively accelerated path to 1199 Hz beginning with one of seven values equally spaced between 1125 and 1875 Hz from most /ba/-like to most /da/-like, respectively. The F3 followed a linear transition to 2729 Hz from one of seven values equally spaced between 2325 and 3325 Hz. All other parameter values used to drive a software formant serial resonator speech synthesizer (Klatt, 1980) were constant across the seven 400-ms CVs. The synthetic speech was modeled after the speaker's natural speech, as it was intended to look and sound as though it were actually the speaker's natural speech. The synthesized speech was dubbed in a synchronous fashion onto the original tape. The timing of the events on each trial was as follows: a 250-ms bell (a cue) followed after a silent interval lasting between 1175 and 1375 ms, followed by the referential event(s), and finally, the 2500-ms response interval.

All subjects watched the videotape on a 12-in. NEC 1203 color monitor from a distance of approximately 2.5 ft. The audio was presented at a comfortable listening level (70 dB-A). For the adults, the experiment was controlled and responses collected by a PDP11-34A computer.

Procedure. The children were individually tested in a research van located outside their day-care center. They participated in three 20-min sessions held on different days, with two blocks of trials per day. Adults were tested in separate sound-attenuated rooms for one 1-h session. Both the children and the adults received these instructions:

> You will see a man on the screen sitting behind a table. On the table there is a ball and a doll. You must decide whether he is telling you to choose the ball or

the doll. He tells you in three different ways. Sometimes he says which one you should choose, sometimes he points to the one you should choose, and sometimes he both points and says it. Each time, though, he is telling you to choose one, either the ball or the doll.

Subjects were also told that the word "ba" means "ball" and that "da" means "doll." The children repeated how to play the game and the adults were given the opportunity to ask questions.

The experimenter sat beside the monitor and wrote the childrens' responses onto a sheet of paper hidden from view from the children. In order to reduce the possibility of biasing the children's responses in any way, while maintaining their interest in the task, the experimenter gave the children feedback during the session at predetermined intervals. During the first session, all of the children received positive verbal feedback, regardless of their response. For the other blocks of trials, they received positive feedback on the first and fourth trials, and more often for a few children who seemed unsure of their responses. Responses which were made on trials when the child was not looking at the monitor were not scored. For the most part, the children were very calm and enjoyed participating.

The experimenter played with the children during a short break midway through each of the three sessions. Adults received a 10-min break after three blocks of trials.

Results and Discussion

The proportion of each subject's "doll" responses was computed for each unique stimulus condition. Separate analyses of variance were carried out for each of the three experimental conditions. There were very few responses which were recorded as "missed responses" for both children (3%) and adults (4%).

The analysis of variance for gesture-alone trials showed significant effects for gesture, $F(1, 32) = 1107.05$, $p < .001$, and for age, $F(1, 32) = 4.35$, $p < .043$. In addition, a significant Age $\times$ Gesture interaction was obtained, $F(1, 32) = 10.78$, $p < .003$. Figure 1 shows that adults correctly identified the appropriate referent an average 98% of the time, whereas children identified the appropriate referent only 89% of the time. Thus, we have clear evidence that 5-year-olds do utilize gestural information when it occurs in the absence of speech, although they do not perform as well as adults. The variable for whether the lips were moving or not (L/NL factor) did not produce a significant main effect, nor did it interact with the other variables, $p > .05$.

For the speech-alone condition, there was a main effect for the speech variable, $F(6, 192) = 307.33$, $p < .001$, as well as a significant Age $\times$ Speech interaction, $F(6, 192) = 6.32$, $p < .001$. The right panel of Fig.

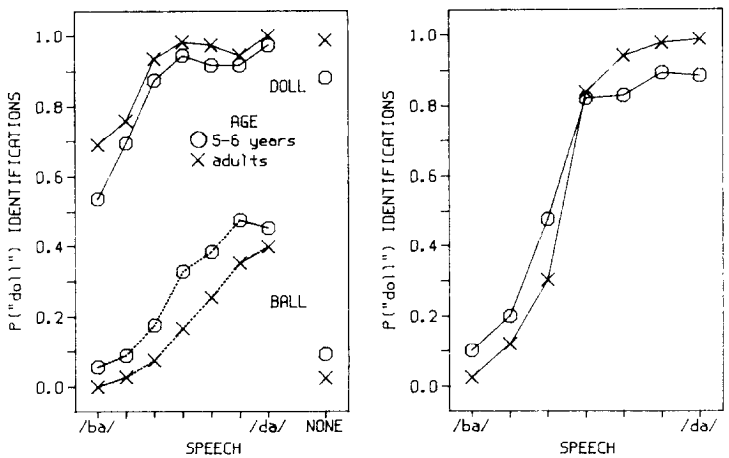

FIG. 1. Observed proportion of "doll" responses as a function of the level of speech and gesture for the adults and 5-year-olds (left panel-conditions containing a "doll" or "ball" gesture; right panel-speech-alone condition).

1 shows slightly better discrimination of the speech dimension for adults in this condition. All other main effects and interactions were nonsignificant, $p > .05$ in all cases.

In the speech-gesture condition analysis, there were significant main effects for the speech factor, $F(6, 192) = 41.04$, $p < .001$, and for the gesture factor, $F(1, 32) = 229.92$, $p < .001$. The left panel of Fig. 1 displays the average proportion of "doll" responses for both age groups in the speech-gesture condition. The significant Speech × Gesture interaction reflects the larger effect of gesture in the middle range of the speech continuum, when the linguistic information is ambiguous, $F(6, 192) = 10.36$, $p < .001$.

In the analysis of the speech-gesture condition, the age group factor did not interact with the speech factor, and the Age × Gesture interaction just missed significance, $F(1, 32) = 3.61$, $p < .063$. When the data from the two age groups were analyzed separately, significant main effects for gesture were obtained, $p < .001$ in both cases. To assess how much of an influence the pointing gesture had on subjects' responses in the speech-gesture condition, the data were first pooled over the five levels of the speech variable. The absolute difference in a given response probability to the "ball" and "doll" gestures was then computed. Specifically, the percentage "ball" responses given a "doll" gesture was subtracted from the percentage "doll" responses when a "doll" gesture was presented. The gesture variable had a larger influence on the adult subjects' identifications (86% effect) as compared to childrens' identifications (78% effect). The results are consistent with the finding of more accurate identifications in the gesture-alone condition for adults than for children.

There was no main effect for whether the lips were moving or not (L/NL), and this variable did not interact with age. However, L/NL did interact with the gesture variable in the speech-gesture analysis, $F(1, 32) = 10.72$, $p < .003$. In addition, a three-way L/NL × Speech × Gesture interaction occurred, $F(6, 192) = 2.16$, $p < .048$. Massaro (1984) and Massaro et al. (1986) found that the influence of articulatory information during bimodal speech was larger for adults relative to children. For this reason, even though age did not interact with the articulatory (L/NL) factor, we chose to analyze the data from this condition separately for each age group. The results were consistent with our previous findings. Separate analyses of variance for each age group again revealed Gesture × L/NL and Speech × Gesture × L/NL interactions but only in the adult group, $F(1, 16) = 9.67$, $p < .007$, and $F(6, 96) = 3.26$, $p < .006$, respectively. Figure 2 displays both of these interactions for the adults. It is evident from the figure that the added articulatory information increased the size of the effect of the gesture variable in the adult group.

To summarize this analysis, the developmental trend appears to be toward greater utilization of the pointing gesture during the interpretation of speech-gesture events. Both 5- and 6-year-olds and adults are very much influenced by pointing gestures in their comprehension of speech-gesture events. The influence of the gesture when combined with speech is somewhat smaller, though, for children. This result substantiates the finding of poorer accuracy of children compared to adults in identifying the referent in the gesture-alone condition. In addition, children show no significant influence of lip movement in the speech-gesture condition,

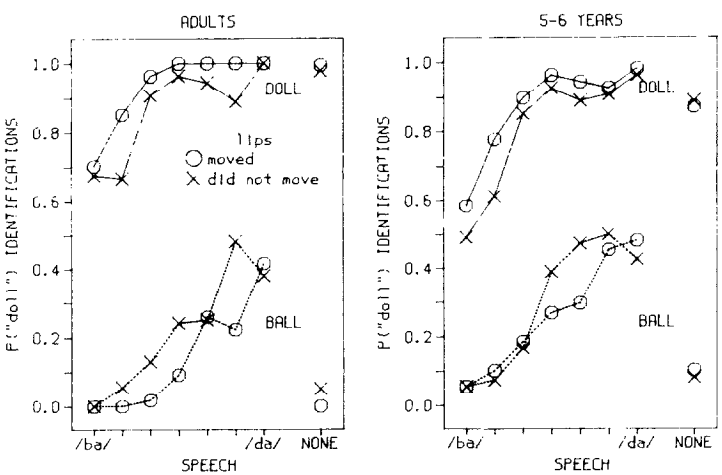

FIG. 2. Adults' and children's observed proportion of "doll" responses in the speech-gesture and gesture-alone conditions for contexts where the speaker's lips did and did not move. The top two curves correspond to the "doll" gesture and the bottom two curves to the "ball" gesture.

in contrast to the significant effect shown by adults. We now apply and evaluate different models of pattern recognition.

Formal models of the pattern recognition processes. As stated in the introduction, we formalized two models of pattern recognition by quantifying three issues pertaining to the evaluation and integration of gestural and linguistic information. These models, the categorical model of perception (CMP) and the Fuzzy Logical Model of Perception (FLMP), are pitted against each other to help determine the nature of information used to interpret an act of reference. Briefly, the categorical model predictions are derived from the assumption that a decision about the likelihood of a particular act of reference is based upon the perception of categorical (discrete) information representing each source of information. On the other hand, the FLMP makes the opposing assumption of continuous information of each source. (Please refer to the appendices for formalized descriptions of both models and their application to the present task.)

The second issue that was tested was whether the two sources of information were evaluated independently of one another. There are limitations in formalizing a workable model of the opposing assumption, nonindependence (see Massaro, 1984). Nonindependence implies a unique effect of each combination of speech and gesture and thus requires as many free parameters as there are independent observations. Rather than testing nonindependence models, the test of independence was evaluated primarily on the basis of the goodness of fit of the FLMP to the results. Finally, the integration issue can also be addressed by comparing the fits of the FLMP and the CMP. As shown by Massaro (1984), the CMP is mathematically equivalent to a weighted averaging model in which the two sources are differentially weighted and then averaged to give the probability of judgment. In contrast, the FLMP uses a multiplication rule to describe the integration of spoken word and gesture.

Two criteria are used to evaluate the goodness of fit of the models to the results, the root mean squared deviation (RMSD) between observed and predicted points, and the number of parameters required by the model to describe the data. Generally speaking, a model provides a good fit to the data to the extent it requires a few parameters and results in a low RMSD.

Both models were fit to the data using the program STEPIT (Chandler, 1969). The data were first pooled over the lips/no-lips variable to increase the number of observations per condition, and because of the relatively small effect of this variable. For each subject, performance in each of the 23 cells representing the three experimental conditions was fit by each model. In addition, the average performance for each cell across all subjects within a group was fitted by each model. These are referred to as "averaged subject fits," and "fit of the average subject," respectively. The average parameter values and RMSD values are given in Table 2.

TABLE 2

ROOT MEAN SQUARE DEVIATIONS (RMSDs) BETWEEN OBSERVED AND PREDICTED VALUES AND BEST-FITTING PARAMETER VALUES FOR CATEGORICAL (CMP) AND FUZZY LOGICAL (FLMP) MODELS IN EXPERIMENT 1

		Speech							Gesture		
Model	RMSD	Ba	2	3	4	5	6	Da	Ball	Doll	Proportion
		Adults									
FLMP											
Fit of the average subject	0.024	0.061	0.093	0.305	0.857	0.924	0.950	0.959	0.028	0.971	
Averaged subject fits	0.048	0.063	0.103	0.329	0.829	0.944	0.969	0.984	0.042	0.963	
CMP											
Fit of the average subject	0.044	0.011	0.116	0.328	0.824	0.931	0.979	0.999	0.020	0.999	.282
Averaged subject fits	0.067	0.056	0.111	0.332	0.830	0.931	0.965	0.997	0.030	0.996	.289
		5- and 6-year-olds									
FLMP											
Fit of the average subject	0.029	0.129	0.223	0.498	0.796	0.822	0.873	0.865	0.119	0.883	
Averaged subjects fits	0.087	0.150	0.225	0.495	0.825	0.830	0.879	0.898	0.128	0.868	
CMP											
Fit of the average subject	0.043	0.071	0.202	0.493	0.819	0.834	0.909	0.913	0.068	0.962	.406
Averaged subject fits	0.098	0.124	0.217	0.497	0.819	0.837	0.900	0.896	0.094	0.926	.416

Note. The parameter values for the FLMP refer to the degree to which "doll" is supported by the speech and gesture sources. The parameter values for the CMP refer to the probability of a "doll" decision given by the speech and gesture sources. Proportion refers to the parameter value for the proportion of trials in which the decision agrees with the decision given to the speech source.

The FLMP yielded lower RMSDs than the CMP for both age groups. The RMSDs for the fit of the FLMP to individual subjects were significantly lower than those for the CMP, $F(1, 34) = 26.14$, $p < .001$. Although the RMSDs came out significantly lower for adults, relative to children, $F(1, 34) = 4.66$, $p < .036$, there was no interaction between age group and model.

Figures 3 and 4 graphically portray the predictions of the FLMP and CMP. It can be seen that the deviations of the observed points from the lines predicted by each model are greater for the CMP. Finally, the requirement of eight parameters for the FLMP versus nine parameters for the CMP is a further basis upon which to favor the FLMP assumption that subjects' interpretations were made using continuous information from the communicative context.

The assumption regarding independence of the two sources of information, as stated previously, is represented by the FLMP. Due to the low RMSDs derived from fits of the FLMP, we have evidence in favor of independent sources of information. Previous tests of this assumption using explicit comparisons of the dependence vs independence models yielded lower RMSDs for the independence model (Massaro, 1984; Massaro & Cohen, 1983). Moreover, in the present study, identical parameter values were used for the single-modality conditions and the bimodal condition. This provides a stronger test of the independence assumption than we have previously reported. For example, the FLMP assumes that the same information that is used for the interpretation of the gesture presented alone is used when it occurs with speech.

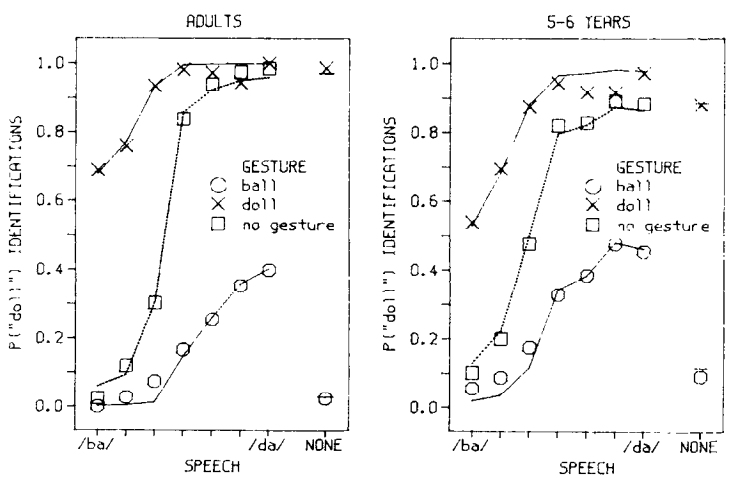

FIG. 3. Observed (points) and predicted (lines) proportion of "doll" responses as a function of the level of speech and gesture for the adults and 5-year-olds. The predictions are for the fuzzy logical model of perception (FLMP).

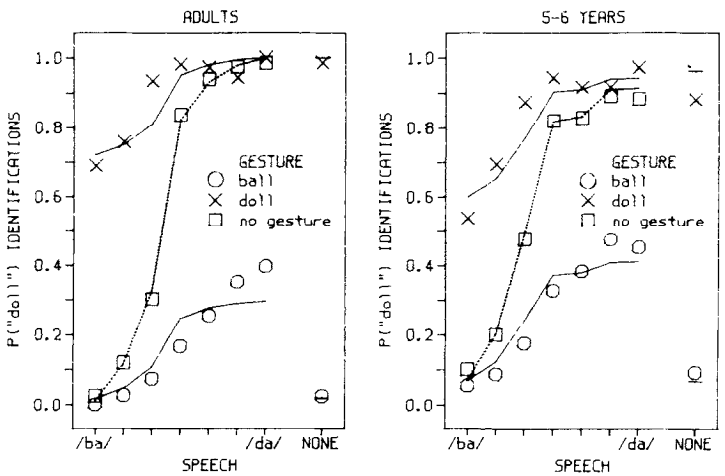

FIG. 4. Observed (points) and predicted (lines) proportion of "doll" responses as a function of the level of speech and gesture for the adults and 5-year-olds. The predictions are for the categorical model of perception (CMP).

Separate analyses of variance were carried out on the parameter values for the speech and gesture parameters determined by the FLMP when fitting the individual subjects. This was done to evaluate age differences in these feature values. The analyses of variances yielded significant main effects for the speech and gesture variables, $p < .001$ in both cases. More interesting are the Age × Speech and Age × Gesture interactions, $F(6, 204) = 5.61$, $p < .001$, and $F(1, 34) = 8.16$, $p < .007$, respectively. These interactions indicate a more compressed range of values for children relative to adults, showing that the informativeness of both sources of information is diminished for children, compared to adults.

The data shown in Fig. 3 also shows support for the argument that gestures and speech presented alone do not convey the full meaning behind the referential act (McNeill, 1985). This figure shows that children's responses to the pointing gestures and to the unambiguous speech tokens are between 84 and 95% correct. However, when the pointing gesture and unambiguous speech converge on the same referent, performance reaches nearly 100% accuracy.

To summarize, the good fit of the FLMP to the results of this experiment supports three conclusions. First, both 5- and 6-year-old children and adults utilize gestural information to interpret acts of reference. Children show less influence of both speech and gestural information than adults, however, suggesting that both sources become more informative with age. Moreover, language users show an understanding of the gesture which can be evaluated independently of their understanding of the linguistic form of reference. Our evidence further illuminates two additional issues

central to investigations concerning the nature of processing involved. Both children and adults base their decisions on continuous, rather than discrete, speech and gestural information.[1] Finally, the integration of these sources is more appropriately described as multiplicative than additive, given that the least ambiguous source has the largest impact on the judgment.

EXPERIMENT 2

Method

Experiment 2 was performed to assess the influence of gestures at an earlier stage of language acquisition. Given the finding that 5- to 6-year-old children were not significantly influenced by articulatory information, we chose to eliminate this variable from the design of Experiment 2. We tested the young children in the conditions containing articulatory information, since this seemed to us a more natural form of communication.

Subjects. Eight children, four males and four females, participated in this experiment. Two others were dropped from the experiment due to too many missed responses. Those that were included ranged between 2-5 and 3-11 years of age (mean age = 3-5). The children were from the University Child Care Center. They received a small toy at the end of each session for their participation.

Design, apparatus and stimuli. These children viewed the experimental tape used in Experiment 1. However, they viewed only the half of the tape containing the three trial blocks in which the speaker's lips moved on gesture trials. Each subject viewed each block of trials once, the order being determined by a Latin square. There were six sessions made by splitting each of the three blocks of trials. Subjects received the first 21 trials of a block during the first session, then the second half of that block for the second session, and likewise for the other four sessions.

Procedure. Certain changes in procedure from Experiment 1 were necessary to enable the young children to understand what was required of them. First, the videotaped toys were given to the children to touch. The children were asked to pick up the ball and the doll, in turn, to ensure that they recognized their labels. In addition, children watched while the experimenter modeled how the "game" was played. A demonstration tape was run, consisting of six trials in this order: the first level of auditory information (a clear /ba/) combined with a point to the ball, the seventh level of auditory information combined with a point to the doll, the first level of auditory information, the seventh level of

[1] Strictly speaking, we do not have direct evidence that gestural information is continuous in the current experiments, since only two levels of gesture were used. In other (unpublished) research from our laboratory, we found evidence for continuous information about gesture in a study with five levels of pointing between a "ball" and a "doll."

auditory information, a point to the ball, and a point to the doll. The experimenter made the appropriate response to these unambiguous stimuli. During the demonstration tape, the child was given informal instructions, and was encouraged to verbalize the experimenter's response. Then, the experimenter instructed the child to

> Watch and listen to the man on the TV. He is going to tell you to choose a toy, either the ball or the doll. When you know which one he told you to choose, you tell me, and I'll write it down.

Prior to the other five sessions, the child was shown the ball and the doll and was asked to pick up each of them. The above instructions were then given to each child.

A final variation in procedure from Experiment 1 involved sometimes pausing the tape after the trial had occurred to prompt the child for his/her response. If the child did not respond immediately after the prompt, the trial was disregarded. As before, a response was not scored if the child made a response without seeing the visual information. Most children responded fairly rapidly and confidently during the entire session. Sessions lasted approximately 6 min after instructions.

Results and Discussion

An average of 10.5% (ranging between 0 and 29%) of each subject's trials had to be excluded from the analysis, either because the subject did not respond, or because the subject was not looking at the screen when the stimuli were presented. Of these "missed" trials, 29% of them were speech-alone trials, 40% were gesture-alone trials, and 31% were speech-gesture trials.

The responses to each unique stimulus condition were converted to the proportion of "doll" responses and submitted to three analyses of variance.

Speech-alone and gesture-alone analyses of variance revealed significant main effects for both variables, $F(6, 42) = 7.46$, $p < .001$, and $F(1, 7) = 11.66$, $p < .011$, respectively. Figure 5 shows that the young children utilized gestural information when it was presented in the absence of speech, with 73% correct identifications of the appropriate referent. Their responses were not biased toward either toy, since there were no differences in performance with the two types of pointing gestures (75% correct for the "ball" point versus 72% for the "doll" point). Even though the young children could discriminate differences along the levels of the auditory continuum, they were biased to identify the speech sounds as "doll." This was not found to be the case for the 5-year-olds in Experiment 1.

Figure 5 also displays the performance in the bimodal condition. There were significant effects for the speech, $F(6, 42) = 7.93$, $p < .001$, and

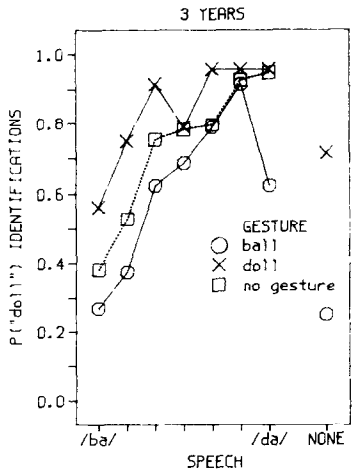

FIG. 5. Observed proportion of "doll" responses as a function of the level of speech and gesture for the 3-year-olds.

gesture, $F(1, 7) = 5.86$, $p < .045$, variables. Consistent with the single modality results, both speech and gestural information influenced identification in the bimodal situation. These young children show smaller effects of both variables relative to older children and adults. A significant interaction between these variables was not obtained for the 3-year-olds, contrary to the results for the 5-year-olds and adults. However, the lack of an interaction for the 3-year olds may simply be due to the smaller effects of both the speech and gestural variables. In addition, even the most /ba/-like speech stimulus produced only 62% /ba/ identifications in the speech-alone condition. This bias to hear the auditory continuum as /da/ for the 3-year-olds may have allowed less opportunity for a speech-gesture interaction to be observed.

In sum, young children do perceive and incorporate gestural information when interpreting acts of reference. However, the extent of the influence of gestures in their interpretations of speech-gesture reference was greatly diminished relative to the 5- and 6-year-olds and adults. This conclusion is made even stronger because the young children also showed less ability to discriminate the auditory continuum, and could have reduced the ambiguity of the act of reference by using the gesture information.

Tests of the models. The results from Experiment 2 were tested against the predictions of the FLMP and CMP, as in Experiment 1. An analysis of variance was used to compare the goodness of fit for the two models. Although the average RMSD for the FLMP was lower than for the CMP (0.1186 as compared to 0.1519), this result was not statistically significant ($p = .110$). However, we believe that the lack of a definitive answer to

TABLE 3

ROOT MEAN SQUARE DEVIATIONS (RMSDs) BETWEEN OBSERVED AND PREDICTED VAULES AND BEST-FITTING PARAMETER VALUES FOR CATEGORICAL (CMP) AND FUZZY LOGICAL (FLMP) MODELS IN EXPERIMENT 2

		Speech						Gesture		
Model	RMSD	Ba	2	3	4	5	6	Da	Ball	Doll	Proportion
FLMP											
Fit of the average subject	0.047	0.390	0.553	0.780	0.793	0.870	0.952	0.841	0.312	0.689	
Averaged subject fits	0.119	0.359	0.528	0.832	0.800	0.847	0.973	0.865	0.264	0.713	
CMP											
Fit of the average subject	0.072	0.377	0.551	0.810	0.802	0.905	0.999	0.920	0.242	0.830	.719
Averaged subject fits	0.152	0.419	0.534	0.799	0.799	0.823	0.939	0.875	0.263	0.766	.744

Note. The parameter values for the FLMP refer to the degree to which "doll" is supported by the speech and gesture sources. The parameter values for the CMP refer to the probability of a "doll" decision given by the speech and gesture sources. Proportion refers to the parameter value for the proportion of trials on which the decision agrees with the decision given to the speech source.

this test between the models is due to variability rather than any inadequacy of the FLMP. The 3-year-olds showed much more variability in their judgments relative to the 5-year-olds and adults. The smaller number of subjects, the fewer number of observations per subject, the larger number of missed responses certainly are partially responsible for the nonsignificant difference between the model fits. In addition, the speech and gestures were less informative for the 3-year-olds, and therefore a larger variance of the judgments should be expected. Given these limitations in the results for the 3-year-olds, it is not surprising that the advantage of the FLMP missed statistical significance even if the young children were basing their interpretations of referential acts on a multiplicative integration of continuous and independent speech and gesture information.

GENERAL DISCUSSION

By 3 years of age, children obviously know that pointing gestures are sometimes used to refer to objects around them. In this sense, children have knowledge concerning the referential use of pointing gestures. However, previous research has not explored how important the pointing gesture is for children in understanding the meaning behind a referential act. Our primary goal was to experimentally compare 3-year-old's, 5-year-old's, and adult's use of gestural information to interpret acts of reference. The results from our previous investigations of children's and adult's comprehension of auditory-visual speech led us to expect that, compared to adults, children's judgments would be less influenced by another source of visual information, the pointing gesture.

To test our hypothesis, we asked subjects to watch a videotape of a person referring to one of two toys in three ways: by vocally labeling it, by pointing to it, and by using both forms of reference. Subjects made decisions as to which object was being referred to on every trial. These decisions were often not straightforward for two reasons. The speech continuum contained ambiguous information and, when presented with gestures, the two referents corresponding to speech and gesture sometimes conflicted.

The results clearly confirmed our hypothesis. In the condition where gestures occurred without speech, subjects' identifications were to the appropriate referent 73, 89, and 98% of the time for 3-year-olds, 5-year-olds, and adults, respectively. Since responses were scored only for trials where the subject was looking at the screen, the fact that children were not 100% accurate is not an intuitive one. Yet, this finding closely parallels our previous result, that children are not as good as adults at reading articulatory (lip) information (Massaro, 1984; Massaro et al., 1986). In the present experiments, the two levels of pointing gestures became more discriminable with development. That is, children have difficulty with environmental signals that adults process without error, due to perceptual,

memory, or decision deficits. Further research is necessary to understand exactly how children learn to capture more information from the pointing gesture.

When gestures and speech occurred together, statistically significant effects of gestures indicated that all age groups used both gestural and speech information to make their decisions. In addition, quantitative estimates of the feature values used in evaluating the gestural input showed developmental differences in the direction of increasing discriminability of gesture and speech feature values with age. A similar result was obtained in our previous investigations of developmental changes in visual-auditory speech perception (Massaro, 1984; Massaro et al., 1986).

Another main purpose of the study was to investigate developmental changes in the strength of the effect for gestures when both sources of referential information occurred together. We predicted, and obtained, a developmental trend toward a greater effect of gestural information in the speech-gesture condition. The percentage "doll" responses to a ball gesture was subtracted from the percentage "doll" responses for the doll gesture, collapsed over all levels of speech, to yield the effect of the gesture in this condition. These values increased from 61% to 78% to 86% for 3-year-olds, 5-year-olds, and adults, respectively. The most plausible explanation for this finding is that when interpreting bimodal referential acts, children were using knowledge of gestural reference that was less informative than it was for adults.

A related goal of the study was to elucidate the nature of the pattern recognition processes involved in referential understanding. Based on our previous work in developmental aspects of speech perception, we expected to find developmental similarities in the evaluation and integration of gesture and speech. We hypothesized that, for all age groups, subjects' responses would be derived from the perception of independent and continuous gesture and speech information. Further, we predicted referential information to be integrated with a rule best described by a multiplicative algorithm. The model tests and statistical results from the 5-year-olds and adults strongly confirmed these hypotheses. Put another way, the pointing gesture influenced subjects' judgments to a greater extent when the speech information was ambiguous. While not as strong, the data from the 3-year-olds indicated support for these predictions. A model assuming independent and continuous referential information also provided the best fit to the data from all age groups.

A plausible interpretation of the lack of a gesture-speech interaction for the youngest age group is due to the particular speech stimuli used in the experiment. Three-year-olds were far worse at discriminating the speech continuum than the 5-year-olds, who did not discriminate as well as adults. Therefore, 3-year-olds were integrating gesture *and* speech

stimuli that were lower in informational content compared to the older subjects. A more discriminable speech continuum may have netted the speech-gesture interaction for the 3-year-olds that we found for the other two age groups.

The poor discrimination of the 3-year-olds is interesting, and may be linked to the type of adult input young children are getting. Garnica (1978) found that adults used a higher-pitched voice when speaking to young children. The male synthesized voice did not share this quality. Perhaps children would have shown better discrimination of the speech continuum if the fundamental and formant frequencies had been higher. Nevertheless, the results from all age groups would seem to suggest that learning to interpret the pointing gesture and to discriminate speech proceeds in tandem. Even by the age of 5, children did not reach the adults' level of performance in our three experimental conditions.

Our results in language understanding can be related to two of McNeill's (1985) recent ideas concerning the role of gestures in language production. McNeill (1985) proposed that the expression of spoken and gestural language stems from a single cognitive structure. The meaning to be expressed undergoes a "computational stage" of processing which is common to gestures and speech. The eventual outcome is exhibited incompletely in gestures or speech alone, so that gestures can be thought to "add to" the meaning of the spoken message. A logical extension of McNeill's proposal would be to reverse the sequence of stages to talk about on-line processing during comprehension of gestures and speech.

The major point of agreement is that gestures and speech share a computational stage of processing in language understanding as they do in language production. More precisely, the computational algorithm that operates to combine gestures and speech into a common representation is best described by a multiplicative rule. This rule ensures that the gesture has a larger impact when the speech information is not clear. Second, since subjects' identifications were more likely to be accurate with more dimensions of input (speech, gesture, and lips), our results and interpretations are consistent with McNeill's claim that speech and gesture "arise from a common cognitive representation, which neither exhibits completely (pg. 353)." Further experimentation using a much wider range of speech and gesture samples could yield valuable support for these claims.

Two empirical studies of the production of speech and gesture also are relevant to the present work. Holender (1980) and Levelt, Richardson, and LaHeij (1985) observed the interaction between speech and gesture in the production of both speech and gesture relative to the separate productions of the two actions. In the Holender study, subjects had to name and/or press one of four keys corresponding to one of four visually presented letters. The reaction time to initiate the naming response was

delayed when the keypress response was also required. Naming the letter did not influence the time to initiate the keypress, unless instructed to give priority to the naming response. Levelt et al. presented one of four lights and subjects were required to point at the light and/or say "this light" or "that light." The authors observed an adaptation of the speech response to the gesture response. The results of both studies are consistent with a two-stage model of the perception-action task. Planning the execution of speech and gesture is interactive, whereas the execution of the two actions become relatively independent of one another once the execution is initiated and completed to some degree.

The nature of the planning and production stages of speech and gesture mirrors our view of their perception. First, the independence of the speech and gesture productions is consistent with the independent evaluations of the two actions, since the feature evaluation stage of the FLMP specifies independent evaluations of speech and gesture. Moreover, the planning and initiation of the two actions originate from a single meaning, and therefore the perceiver must interpret them as specifying a single referential act. In the FLMP, the two acts are integrated in perception with respect to prototypes describing the appropriate actions for different referents.

In summary, our main contribution has been in experimentally mapping out some changes between early childhood and adulthood in the nature of the pattern recognition processes involved in referential understanding. We have shown that the basic architecture for the perceptual processing of referential information is shared by preschoolers and adults. Developmental accomplishments beyond age 3 involve small, but significant, progressions in the ability to extract and utilize information contained in two complementary referential signals, speech and pointing gestures.

APPENDIX A

Formalization of a Categorical Model of Perception (CMP) Assuming Categorical and Independent Speech and Visual (Gestural) Sources of Information

The listener is assumed to have only categorical information representing the speech (S_i) and gestural (G_j) dimensions of the speech event. This model implies that separate categorical (phonetic) decisions are made to the speech and gesture sources (MacDonald & McGurk, 1978). A categorical speech decision would be made in the speech condition and analogously in the gesture condition. Each unique level of the speech and the gesture variables would have a unique probability of producing a "doll" decision. In this case, the probability of a "doll" identification, $P(D)$, in the speech and gesture tasks is equal to

$$P(D|S_i) = s_i \tag{1}$$

$$P(D|G_j) = g_i \tag{2}$$

where $P(D|S_i)$ is the probability of a "doll" identification given the speech level S_i and analogously for $P(D|G_j)$. The values i and j index the levels of the speech and gesture

stimuli, respectively. Given the two-alternative forced-choice task, the probability of a "ball" decision to a given variable would be simply 1 minus the probability of a "doll" decision. No test of this model is possible given only the speech and gesture tasks, since a unique parameter must be assumed for each of the nine (seven speech and two gesture) conditions.

Adding the bimodal task gives 14 additional experimental conditions (7 speech × 2 gesture). Given the assumption of the independence of the speech and gesture sources, the probability of a given decision to a speech level would be identical in the speech and the bimodal conditions. Thus, separate "doll" or "ball" decisions would be made to both the speech and gesture sources in the bimodal condition, and the identification response would be based on these separate decisions. Given categorical decisions, there are only four possible outcomes for a particular combination of speech and gesture information: "doll"–"doll," "doll"–"ball," "ball"–"doll," or "ball"–"ball." If the two decisions to a given speech event agree, the identification response can follow either source. If the two decisions disagree, then the subject is assumed to respond with the decision of the speech source on some proportion p of the trials and respond with the decision of the gesture source on the remainder $(1 - p)$ of the trials. In this conceptualization, the magnitude of p relative to $(1 - p)$ reflects the relative dominance of the speech source of information.

The probability of a "doll" identification response, $P(D)$, given a joint speech/gesture speech event, S_iG_j, would be

$$P(D|S_iG_j) = 1s_ig_j + ps_i(1 - g_j) + (1 - p)(1 - s_i)g_j + 0(1 - s_i)(1 - g_j) \quad (3)$$

$$= ps_i + (1 - p)g_j. \quad (4)$$

As in Eqs. (1) and (2), the s value represents the probability of a "doll" decision given the speech level i and g is the probability of a "doll" decision given the gestural level j. Each of the four terms in Eq. (3) represents the likelihood of one of the four possible outcomes of the separate decisions multiplied by the probability of a "doll" identification response given that outcome. In the present task, there are seven speech levels and two gesture levels. In this model, each unique level of the speech stimulus would require a unique parameter s_i, and analogously for g_j.

Since the parameter p reflects a decision variable, its value would be constant across all of the bimodal conditions. Thus, a total of 10 parameters must be estimated for the 23 independent conditions in the three types of tasks.

It should be noted that the CMP is mathematically equivalent to a weighted averaging model (Massaro, 1984). If the p values in Eq. (3) are interpreted as weights, and the s and g values as continuous information from the speech and gesture sources, then it can be seen that $P(D:S_iG_j)$ is predicted to be a weighted average of the two sources. Thus, rejection of the CMP also implies rejection of an averaging (or additive) integration rule.

APPENDIX B

Formalization of the Fuzzy Logical Model of Perception (FLMP), Assuming Continuous and Independent Sources of Speech and Visual (Gesture) Information

Perceptual categorization is carried out in three operations. The first operation is feature evaluation, during which the stimulus is transduced by the sensory systems and various perceptual features are derived. The features are assumed to be continuous rather than categorical. Thus, the outcome of featural evaluation is not categorical, but is represented by a continuous variable reflecting the degree to which each relevant feature is present.

These continuous values are assumed to be analogous to the truth values in the theory of fuzzy sets (Zadeh, 1965), which explains the first term of the model's name.

The second operation is prototype matching, which involves the integration of the features. The featural information is combined following the rules given by prototype definitions in long-term memory. A prototype defines a perceptual unit of speech in terms of arbitrarily complex fuzzy logical propositions (Massaro & Oden, 1980). The outcome of prototype matching determines to what degree each prototype is realized in the speech event.

The third operation is pattern classification, in which the merit of each potential prototype is evaluated relative to the summed merits of the other potential prototypes (Luce, 1959). This relative merit gives the proportion of times a prototype would be selected as a response. An important property of the model is that one cue has its greatest effect when the second is at its most ambiguous level. The most informative cue has the greatest impact on the judgments.

As in the categorical model, no test of the FLMP is possible given only the gesture-alone and speech-alone tasks. It is assumed that the subject derives the appropriate information and evaluates the degree to which it supports each of the alternatives, "ball" and "doll." Given speech information, the subject evaluates the information conveyed by the F2–F3 transitions to determine how much they are falling and, therefore, support the alternative "doll." A truth value between 0 and 1 is assigned. With just two alternatives, it is reasonable to assume that the truth value supporting the alternative "ball" is 1 minus that for "doll." Given the pattern classification of relative truth values, $P(D)$, the probability of a "doll" decision given a speech event S_i is equal to

$$P(D|S_i) = \frac{s_i}{s_i + (1 - s_i)} = s_i.$$

An exactly analogous situation occurs for the gesture condition G_j,

$$P(D|G_j) = \frac{g_j}{g_j + (1 - g_j)} = g_j.$$

In both cases, the probability of a given decision is predicted to be equal to the truth value of the relevant variable.

Applying the model to the bimodal task with both gestures and speech, both sources are assumed to provide independent evidence for the alternatives "ball" and "doll." Defining the important speech cue as the F2–F3 transitions and the important gesture cue as the direction of pointing, the prototypes are

"doll" : Slightly falling F2–F3 & Pointing at Doll
"ball" : Rising F2–F3 & Pointing at Ball.

Given a prototype's *independent* specifications for the speech and gesture sources, the value of one source cannot change the value of the other source at the prototype matching stage. In addition, the negation of a feature is defined as the additive complement. That is, we can represent Rising F2–F3 as (1-Slightly falling F2–F3) and Pointing at Ball as (1-Pointing at Doll)

"doll" : Slightly falling F2–F3 & Pointing at Doll
"ball" : (1-Slightly falling F2–F3) & (1-Pointing at Doll).

The integration of the features defining each prototype is evaluated according to the product of the feature values. If s_i represents the degree to which the speech stimulus S_i has Slightly falling F2–F3 and G_j represents the degree to which the gesture stimulus g_j is Pointing at Doll, the outcome of prototype matching would be

"doll" : $s_i g_j$
"ball" : $(1 - s_i)(1 - g_j)$.

If these two prototypes are the only valid response alternatives, the pattern classification operation determines their relative merit leading to the prediction that

$$P(D|S_i G_j) = \frac{s_i g_j}{s_i g_j + (1 - s_i)(1 - g_j)}.$$

Given seven levels of S_i and two levels of G_j in the present task, the predictions of the model require nine parameters (seven s_i values and two g_j values), one fewer than the categorical model.

REFERENCES

Anderson, N. H. (1980). Information integration theory in developmental psychology. In F. Wilkening, J. Becker, & T. Trabasso, (Eds.), *Information integration by children.* Hillsdale, NJ: Erlbaum.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill–Palmer Quarterly,* **21,** 205–226.

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

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition,* **11,** 159–184.

Dobrich, W., & Scarborough, H. S. (1984). Form and function in early communication: Language and pointing gestures. *Journal of Experimental Child Psychology,* **38,** 475–490.

Garnica, O. (1978). Non-verbal concomitants of language input to children. In N. Waterson & C. Snow (Eds.), *The development of communication.* New York: Wiley.

Holender, D. (1980). Interference between a vocal and a manual response to the same stimulus. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior.* Amsterdam: North–Holland.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America,* **63,** 905–917.

Levelt, W. J. M., Richardson, G., & LaHeij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language,* **24,** 133–164.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics,* **24,** 253–257.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development,* **51,** 1777–1788.

Massaro, D. W., & Cohen, M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 753–771.

Massaro, D. W., & Oden, G. C. (1980). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice.* New York: Academic Press.

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

Masur, E. F. (1982). Mothers' responses to infants' object-related gestures: Influence on lexical development. *Journal of Child Language,* **9,** 23–30.

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

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

Murphy, C. M. (1978). Pointing in the context of a shared activity. *Child Development,* **49,** 371–380.

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

Shatz, M. (1982). On mechanisms of language acquistion: 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.

Shatz, M. (1984). Contributions of mother and mind to the development of communicative competence: A status report. In M. Perlmutter (Ed.), *Minnesota Symposium on Child Psychology,* **17.** Hillsdale, NJ: Erlbaum.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control,* **8,** 338–353.

RECEIVED: October 17, 1985; REVISED: March 26, 1986.