JOURNAL OF MEMORY AND LANGUAGE 24, 133–164 (1985)

Pointing and Voicing in Deictic Expressions

WILLEM J. M. LEVELT, GRAHAM RICHARDSON, AND WIDO LA HEIJ

Max-Planck-Institut für Psycholinguistik, The Netherlands

The present paper studies how, in deictic expressions, the temporal interdependency of speech and gesture is realized in the course of motor planning and execution. Two theoretical positions were compared. On the "interactive" view the temporal parameters of speech and gesture are claimed to be the result of feedback between the two systems throughout the phases of motor planning and execution. The alternative "ballistic" view, however, predicts that the two systems are independent during the phase of motor execution, the temporal parameters having been preestablished in the planning phase. In four experiments subjects were requested to indicate which of an array of referent lights was momentarily illuminated. This was done by pointing to the light and/or by using a deictic expression (*this/that light*). The temporal and spatial course of the pointing movement was automatically registered by means of a Selspot opto-electronic system. By analyzing the moments of gesture initiation and apex, and relating them to the moments of speech onset, it was possible to show that, for deictic expressions, the ballistic view is very nearly correct.

The general issue addressed in this article concerns the synchronization of speech and gesture. More specifically the aim is to investigate how the frequently noted interdependence of speech and gesture is realized in the course of motor planning and execution. Do the two systems operate interactively, in the sense that mutual adaptation takes place during the phase of motor execution, or do they rather operate in a ballistic or independent fashion in so far as their coordination is established entirely during the planning phase, that is, before motor execution takes place? Since the variety of gestures which can accompany speech is very large, it was necessary to limit the investigation to a subclass of coordinated speech/gesture activities, and it was therefore important to select a subclass for which the synchronization is particularly marked.

Gestures accompanying speech may be classified in a number of different ways. Most classifications reported in the literature acknowledge an element of *directness* in the relationship between speech and certain categories of gesture. An early classification, on which later ones have been based, is that of Efron (1972), who identified a broad category of gestures having what he called an "objective" meaning. This category includes, on the one hand, *deictic* and *iconographic/kinetographic* gestures, which generally exhibit a direct relationship with the content of speech and, on the other, *emblematic* gestures which function as complete utterances in themselves, independent of speech. The latter two subcategories appear to have served as

The authors wish to gratefully acknowledge the contributions to this study of the following persons. Design and installation of the facilities was carried out by Peter Wittenburg, head of the Max-Planck-Institute's Technical Group, assisted by Gerd Klaas, Johan Weustink, and Rob van der Male. Programming advice and assistance in the use of computing facilities was provided by Franz Maurer and John Nagengast. Herman Woltring (now at Philips) was responsible for the development of a software package to implement his photogrammetric calibration method SMAC2. Inge Tarim prepared the figures for this paper. Wolfgang Klein, David McNeill, and two anonymous reviewers did a very thorough and helpful job in analyzing an earlier version of this paper; we are much indebted to them. Wido La Heij is now at the Unit of Experimental Psychology, Department of Psychology, Leyden University.

0749-596X/85 $3.00

models for Ekman and Friesen's (1969) *illustrators* and *emblems,* respectively. These authors introduced a further category, which they termed *self-adaptors* and which involve hand-to-hand and hand-to-body contact. Such gestures also bear no direct relation to speech; it has been suggested that their occurrence is related to either motivational state or the attentional demands of the speech production process. Freedman's (1972) *object-focused* and *body-focused* gestures are analogous to Ekman and Friesen's illustrators and self-adaptors, respectively; as such, object-focused gestures exhibit a direct relation to the conceptual content of the message. McNeill's (1981) "iconic" gestures are like the just-mentioned object-focused ones in that they are concrete depictions of the meanings expressed in the concurrent speech. This paper is concerned with the synchronization between speech and a particular class of gestures directly related to speech, namely deictic gestures.

Deictic gestures are of a special kind in that they can be obligatory in deictic utterances. Deictic terms, such as *here, there, I, you, this, that,* derive their interpretation in part from the speaker/listener situation in which the utterance is made. Among these terms only *here, I,* and in some cases *you* are directly referential; given the situation, their reference is unambiguous. The other deictic terms, however, require the speaker to make some form of pointing gesture, for example, by nodding the head, visibly directing the gaze, turning the body, or moving arm and hand in the appropriate direction. Without such a paralinguistic gesture, the utterance is incomplete in an essential respect. The crucial role of the gesture is evident when one considers that an utterance of this sort could not function unambiguously over the telephone. A pointing gesture which exhibits this essential relation to a deictic utterance will be called a *deictic gesture.* (This situation must be carefully distinguished from one involving the *anaphoric* use of *that* or *there.* An utterance containing such a word can be complete provided that the referent has already been linguistically introduced, for example, "I went to Amsterdam. I saw an accident *there.*")

Deictic hand gestures make particularly good candidates when it comes to studying the synchronization between gesture and speech. On the one hand, their obligatory nature makes them strictly dependent on the message being expressed, while on the other, they have a temporally very marked "apex," insofar as the hand comes to rest, if only momentarily, when the extreme indicating position is reached. The deictic terms which accompany them are also clearly marked, mostly stressed, and of short duration.

Though it is well known that gesture and speech are synchronized in subtle ways (see especially Condon & Ogston, 1971; Kendon 1980; McNeill, 1979, 1981), little is known about the *process* of synchronization. The way in which coordination is achieved is open to a number of possible theoretical interpretations. At one extreme is the view that speech and gesture function as "modular" systems, each generating its output in a fully autonomous fashion. Fodor (1983) has argued that modular organization is characteristic of *input* systems. An important feature of such modularity is what Fodor calls "informational encapsulation," by which he means that the system's operation is insensitive to feedback from other systems. According to this view, visual processing is largely independent of whether the perceiver believes or likes what he sees, or whether it corresponds to what is simultaneously heard. The integration of these different sources of information is a matter of *central* processing, which follows the autonomous perceptual activities. This notion of informational autonomy can be extended to the description of output systems, such as speech and gesture. What this would entail is that the relationship between speech and gesture is established during the planning

phase by virtue of the two systems having access to the *same* central source of information, the conceptual structure or message to which they are both related. But as soon as the systems develop and execute their motor programs there can be no feedback from one system or "module" to the other. Mutually interactive adaptation is precluded in both the planning phase and during motor execution; the systems are entirely independent in their operation. On this account, the observed synchrony between pointing and voicing is considered to be the result of central premotor decisions. There is no possibility for on-line interaction between motor systems; they are autonomous processing modules.

The alternative view is that gesture and speech are at no stage informationally encapsulated, allowing for the possibility that the two systems may achieve a degree of mutual adjustment by means of interaction during both the planning and execution phases. More specifically, it is envisaged that the presence of continual feedback from the gesture system would enable the delivery of the deictic expression to be triggered at the appropriate moment in the execution of the pointing gesture. Also, the apex of pointing may be accelerated or delayed, depending on the moment a particular expression is uttered. This interactive theory will have to specify the nature of the information exchange by establishing what modality (visual, kinesthetic) is employed as a feedback channel, the latency of information transmission in this channel, and the degree to which adaptation is thereby achieved.

Between the extremes of full modularity and full interdependence there is a range of intermediate possibilities. The present study addresses the tenability of one particular theoretical stand, namely, that the motor systems for gesture and speech are interactive during the planning phase, but modular during motor execution. This is close to Arbib's (1981) theoretical analysis of Jeannerod and Biguer's (1981) results on the coordination of reaching and grasping responses. It will be called the *ballistic* view, since motor execution will fly blind on whatever it was set out to do in the planning phase, or at least without concern for the other motor system involved. The theoretical alternative will be that there is no informational encapsulation during gesture execution; parameters of speech and gesture can be mutually adapted during that phase if the speaker wishes to do so. This will be called the *on-line interactive* or short *interactive* view. Both alternatives allow for interdependence in the planning phase, and the character of this interdependence is also addressed in the present study.

The four experiments reported are designed to investigate these processing issues. The first experiment explores the character of synchronization when the hand performs deictic movements to near and far referents in both the ipsilateral or contralateral visual fields. The second experiment evaluates the two theories by comparing "speech-only" and "gesture-only" conditions to the normal "dual" condition, where speech and gesture accompany one another. In the third experiment the information processing load on the two systems is independently manipulated by varying the number of alternatives to be indicated or mentioned. Finally, the fourth experiment provides the most direct test of the ballistic and on-line interactive viewpoints by unexpectedly hampering the pointing gesture during its execution and determining the effect on voicing latencies.

Before reporting these experiments, however, we will describe the equipment used, one of the principal components of which was an opto-electronic system of movement monitoring (Selspot—*Sel*ective *Spot* Recognition). Essentially the same equipment was used throughout the four experiments.

Apparatus

The experimental apparatus was de-

signed to allow the subject's pointing movement and voice onset to be recorded in a situation where his or her task was to respond "this light" or "that light" while indicating which of a series of lights was momentarily illuminated. The subject was seated at a table as shown in Figure 1. The light sources to be indicated consisted of up to four red light-emitting diodes (LEDs) which were mounted on a track, 5 centimeters above the table, in such a way that they could be adjusted horizontally in a frontoparallel plane over a range from about 0 to ±100 centimeters from the centerline. A push-button switch on the centerline of the apparatus served to define the rest position of the hand and was also used by the subject to actuate a sequence of LED operations. The program controlling the operation of the LEDs generated sequences in which the order of operation varied randomly from one trial to the next, subject to the condition that over the complete set of trials each LED operated the same number of times as the others.

The movements of the subject's hand were recorded using a single-diode infrared-emitting assembly, which was attached by means of a clip to the index finger. The two Selspot infrared cameras were positioned about 3 meters above the table at which the subject sat and about 2 meters on either side of the centerline. The diode's x and y image coordinates were recorded every 3.2 milliseconds by each camera and transmitted via a First In/First Out (FIFO) buffer and a Direct Memory Access (DMA) interface to the PDP 11/55 memory, and then to disk or tape.

The program controlling the operation of the LEDs also activated the running of the data acquisition program, which was started up at the instant the LED was turned on, and continued for a predetermined interval of 2 seconds. This period of time sufficed in the present experiments to capture both the outward and return phases of the gesture.

Before using the Selspot system to record experimental data, it was calibrated by means of the procedure known as Simultaneous Multiframe Analytical Calibration, or SMAC, originally developed as a single camera procedure by Brown (1969), and extended by Woltring (1980) to a situation in which two or more cameras with converging axes are used, thereby effecting a considerable improvement in determinacy along those axes. Another feature of this method is that it dispenses with the need for a three-dimensional distribution of landmarks and allows for calibration by a two-dimensional calibration grid, which is tilted at various angles. A set of condition equa-

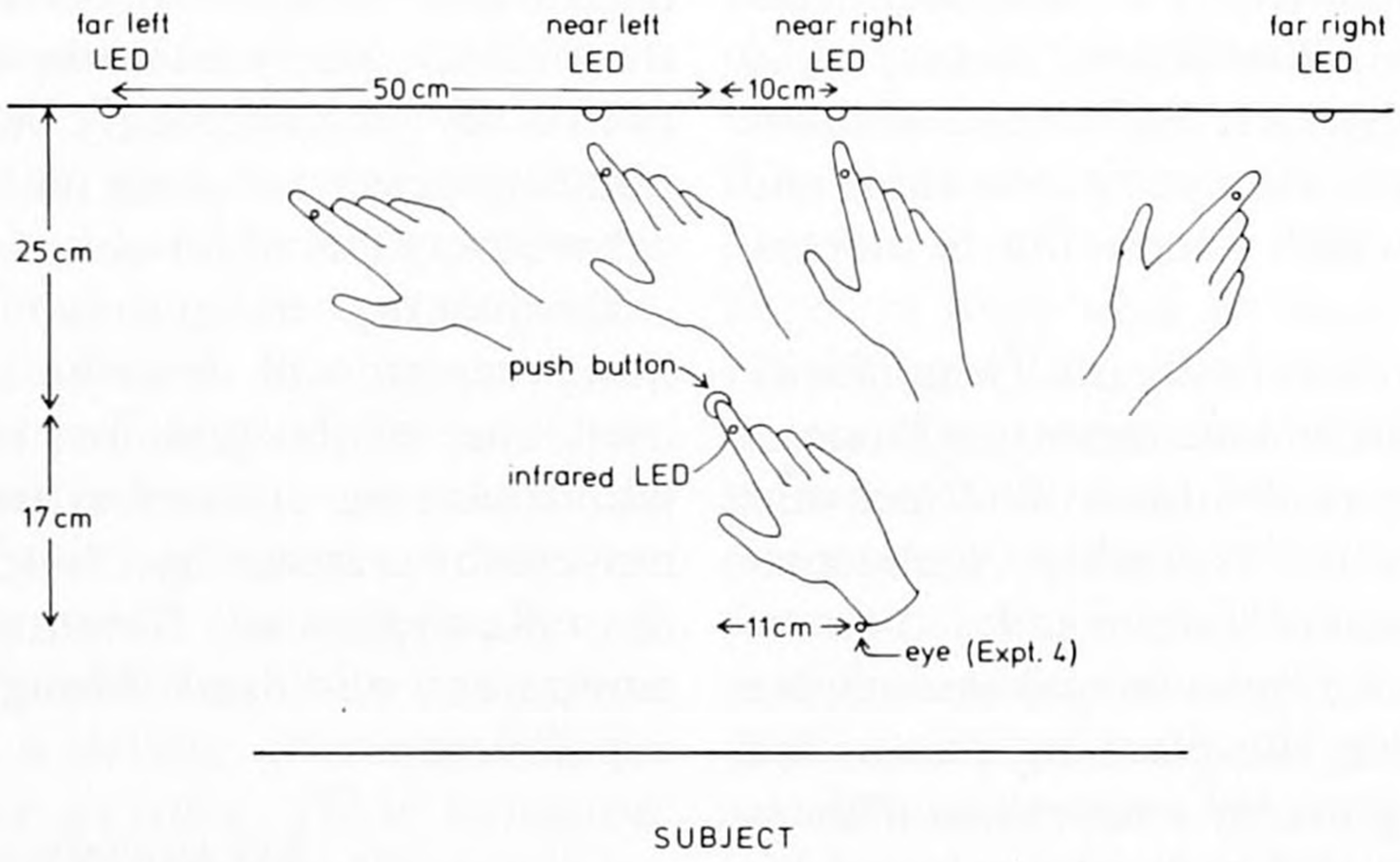

FIG. 1. Apparatus: The spatial arrangement of the four referent LEDs and the push-button, and the attitude of the subject's hand in indicating each of the LEDs.

tions relating points in image space to points in control space is generated and solved for the system parameters by means of a linearized, least-squares, iterative adjustment procedure. The parameters derived in this way are used in subsequent reconstructions of the observed targets.

In order to perform the experiments to be described, it was necessary to measure three variables:

(1) T_I, the elapsed time between the turn-on of a LED and the *initiation* of the pointing movement. It was computed off-line by a movement display and analysis program which accepts a file of x-, y-, z-coordinates generated for each movement by the 3-D reconstruction program. For each new sample point j this program determines the "incremental distance" INCD_j from the previous data point j − 1, according to the following formula:

$$\mathrm{INCD}_j = ((x_j - x_{j-1}))^2 + (y_j - y_{j-1}))^2 + ((z_j - z_{j-1}))^2)^{0.5}$$

where j ranges over the sample numbers, and $j = 1$ corresponds to the instant the LED is turned on. Hence, INCD is proportional to the finger's velocity. The initiation time was defined as that corresponding to the first data point for which, within the next six points, there were at least three exceeding a predetermined INCD value. This value was generally set at either 2 or 3 millimeters, depending on the noise level in the subject's data.

(2) T_A, the time between the turn-on of a LED and the instant at which the pointing movement reaches its extremum or *apex*. In order for this point in the movement to be consistently defined, even in cases where the hand dwelt at the apex, the apex time was defined as the instant at which the movement reached 99% of its maximum extent. The movement analysis program obtained this apex position by first determining the point in space at which the INCD function reached a minimum; the distance d of this point to the finger's starting position was computed. Next the time index j was decremented until a position was found on the gesture trace for which the distance to the finger's starting position was 99% of d. This defined the apex position. The corresponding time value, measured from LED onset, was taken to be apex time T_A.

(3) T_V, the time between turn-on of a LED and the onset of the *verbal response*, which in the present experiments was one of the two Dutch expressions "dit lampje" or "dat lampje" ("this light" or "that light"). This time interval was measured by recording the subject's voice on one track of an audiotape, and pulses generated at the turn-on of each LED on another track. To analyze the speech data, the tape recorder was interfaced with the PDP-11/55, the speech channel being connected via a voice key which produced a pulse at voice onset, and the tape is played back under the control of a program which computed the time interval between the LED pulse and the voice onset pulse.

Experiment 1

The main objective of this experiment was to study the degree to which voice–gesture synchrony is maintained as the distance of the referent, that is, the illuminated LED, is varied. It should be noted that both the ballistic and the interactive theories allow for synchronization to be achieved. From the standpoint of the interaction theory the synchronization is, at least in part, established during the execution of gesture and speech, whereas in terms of the ballistic theory the synchrony results from the preprogrammed instructions governing the activation of the two systems. But how much can be preprogrammed? What can be available to the speech system in terms of the temporal parameters of the gesture before that gesture is executed? If full synchronization is found, the ballistic theory can only be maintained on the assumption that the time pattern of the gesture is (i) completely predetermined, and (ii) accessible to the speech system. For the inter-

active view such a result would be less problematic; there is on-line synchronization of speech and gesture.

The empirical issue to be investigated, then, is whether, and if so to what degree, voicing time T_V covaries with apex time T_A when gestural movements are made to referents at different distances. (It seemed reasonable to assume that T_A would vary with LED distance, even though the subject was not required to reach or touch the referent light.)

The second objective was to determine whether the synchronization between gesture and voice is affected by requiring a fast response to be made. More specifically, two conditions were compared, one (the "on-line" condition) in which the subject is asked to react immediately to the onset of the LED, and the other (the "off-line" condition) in which the subject observed the LED onset, but only responded on hearing the subsequent question of the experimenter: "Which light?" It was thought that, in the latter more relaxed condition, voice timing would stand a better chance of adapting to the duration of the movement.

Method

Subjects. There were 20 subjects, 13 male and 7 female, all of whom were right handed. In this and the other experiments, subjects were paid for their services.

Procedure. Referring to Figure 1, four LEDs, two in each field, were positioned at 10 and 50 centimeters from the midline and about 52 centimeters from the front edge of the table at which the subject was seated. The push-button was on the midline, 25 centimeters in front of the array of LEDs. Pressing the push-button actuated one of the four lamps within an interval which varied randomly from one trial to the next about a mean of 1 second, with a standard deviation of 0.15 second. The LED remained on for 0.5 second. A "ready" light integral with the push-button was turned on at the end of the data acquisition interval, as a signal to the subject that the next trial could commence. The subject was instructed not to lift his finger from the push-button until one of the lamps came on, and to make expansive gestures.

There were four experimental series, each consisting of 40 test trials (i.e., 10 operations of each LED, in random order). In order to acquaint the subject with the new situation, each series was preceded by four practice trials. The first two series were presented in the "off-line" condition, in which the experimenter, seated across the table, cued the subject's response by means of a question which followed the operation of the LED at an interval of 2 or 3 seconds. After each trial, the experimenter "noted down" the subject's response. The second two series were presented in the "on-line" condition in which the subject was required to respond as soon as the LEDs came on. Again, the experimenter wrote down each response. If the subject made an error on a trial, it was immediately repeated as the next trial. The error rate, however, was negligible; most subjects never made any errors, and if an error was made it consisted in lifting the finger before LED onset or giving the wrong verbal response. Within each of the two conditions, half the subjects performed the first series with the right hand, and the second with the left, and half in the reverse order. The room in which the experiment took place was only dimly lighted in order to minimize the level of noise in the Selspot data, but the whole array and the experimenter were visible to the subject.

Results

The values of T_I, T_A, and T_V obtained were the subject of a series of analyses of variance. Considering first the results for the "on-line" condition, Figure 2 shows the mean values of the three variables for each of the four LEDs as target referent, and for the left and right hand separately. Figure 3 gives the corresponding values of the (linear) distances traveled by the finger from push-button to apex position. (It also

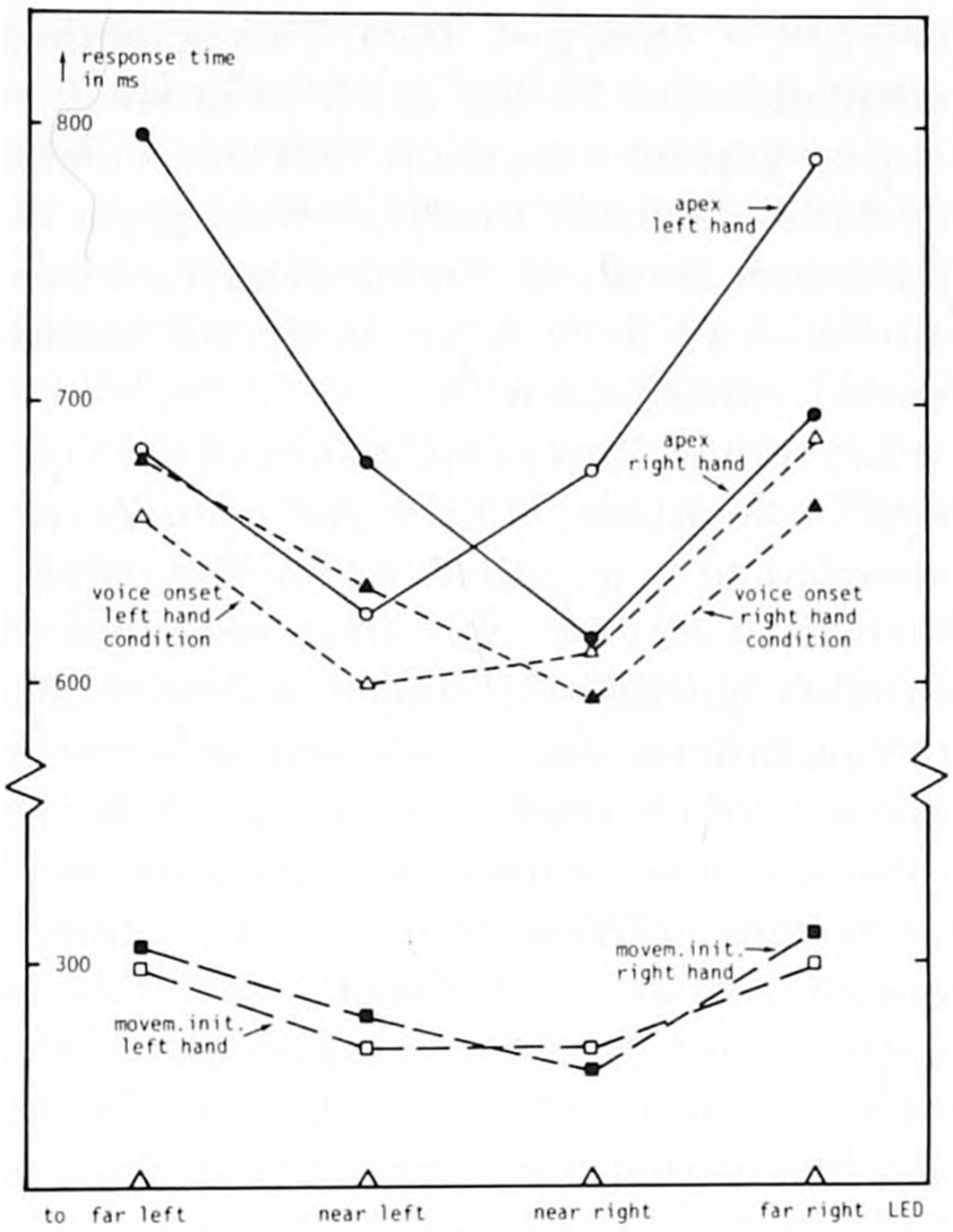

FIG. 2. Experiment 1: Response times for movement initiation, apex and voice onset in referring to four LEDs in the on-line condition. Right-hand and left-hand data.

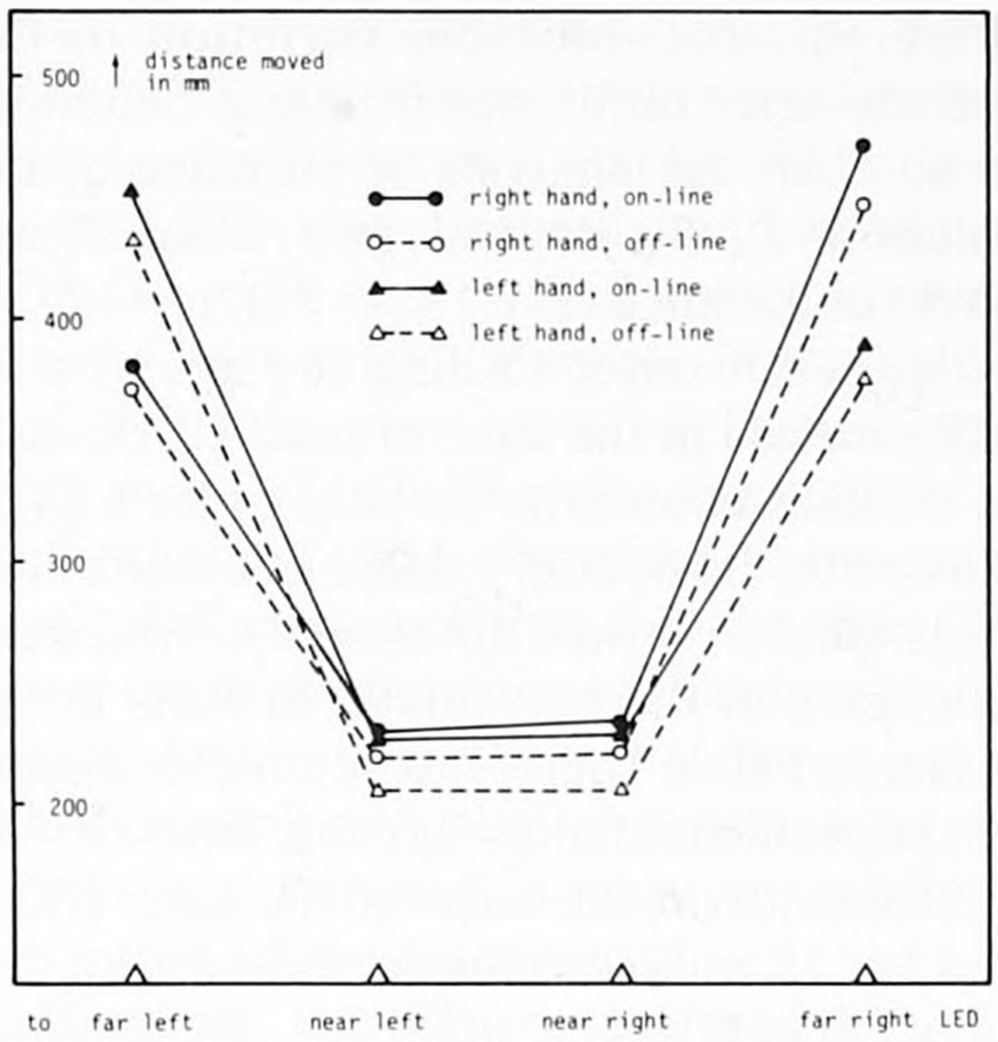

FIG. 3. Experiment 1: Distances moved by right and left hands from movement initiation to apex of gesture. On-line and off-line conditions.

includes the values for the off-line condition discussed below.)

It is clear from Figure 3 that the experimental manipulation was effective in that the extent of movement was greater for the far LEDs as referent than for the near ones, and this, in turn, was reflected in the corresponding values of T_A shown in Figure 2. An analysis of variance of the T_A values showed a main effect of distance ($F(1,19) = 185.4$, $p < .0001$), the mean time to reach apex being 652 milliseconds for near LEDs, and 742 milliseconds for far LEDs. However, these times depended on which field the target referent was situated in relative to the pointing hand, movements to contralateral LEDs requiring more time (mean = 738 milliseconds) to complete than those to ipsilateral ones (mean = 656 milliseconds). This Hand × Field interaction was significant ($F(1,19) = 183.9$, $p < .0001$). The three-way interaction between hand, field, and distance was also significant ($F(1,19) = 16.3$, $p < .001$), the effect of distance being more pronounced in the contralateral field than in the ipsilateral.

The main object of this experiment was to determine whether voice onset time, T_V, would also be affected by these experimental manipulations. Figure 2 shows that it is. There is, clearly, a degree of synchronization of speech and gesture; the T_V curve is not flat, but covaries with gesture apex time. An analysis of variance showed a significant effect of distance ($F(1,19) = 44.5$, $p < .0001$), the mean voice onset latency for the near LEDs being 611 milliseconds and for far LEDs 676 milliseconds. The two-way interaction between hand and field was also significant ($F(1,19) = 11.9$, $p = .0027$), with speech being produced faster when referring to LEDs in the ipsilateral field (mean = 630 milliseconds) than when indicating LEDs in the contralateral field (mean = 657 milliseconds). So, clearly, there is synchronization; the two motor systems show some interdependence. The next question, then, is to what degree do speech and gesture align themselves? Or, in other words, to what extent does the difference between T_A and T_V vary over the experimental conditions?

The data show that, overall, speech onset led the apex by 53 milliseconds, though

there was considerable variation in the speech–apex difference from one condition to another. An analysis of variance of the values of T_A–T_V showed, first, a significant effect of distance ($F(1,19) = 8.0, p = .011$), with speech onset leading the apex by 41 milliseconds in the case of near LEDs, and by 66 milliseconds in the case of far LEDs. However, the overall pattern of results conceals quite a marked difference between the situations in the two fields. In fact, in the ipsilateral field there was almost complete adaptation insofar as speech leads by 23 milliseconds in the case of the near LED, and by 28 milliseconds in the case of the far LED ($F(1,19) = 0.2, p = .6$), whereas in the contralateral field, the respective intervals were 60 and 103 milliseconds ($F(1,19) = 24.5, p = .0001$). On the whole, one does not find the full synchronization which would have made the ballistic theory less likely.

It will be recalled that the apex time T_A consists of two components, the latency to movement initiation T_I, and the gesture execution time (T_A–T_I), which will be referred to as T_E. One might suppose that voice onset simply adapts to the former component, without taking into account the longer execution times required for far LEDs. This would be an attractive result for the ballistic theory; the speech system should only be informed about the moment of gesture initiation. However, the data show that this is not so. The mean value of T_I when pointing to near LEDs was 270 milliseconds, and to far LEDs 303 millisecond; the difference of 34 milliseconds was significant ($F(1,19) = 48.8, p < .0001$). Yet, as was noted above, speech onset times to near (611 milliseconds) and far (676 milliseconds) LEDs differed by 65 milliseconds so that there was an additional 31 milliseconds of voicing adaptation, which cannot be explained in terms of adaptation to the timing of initiation alone. An analysis of variance of T_V–T_I values, that is, of latencies from gesture initiation to voice onset, shows that this value of 31 milliseconds is significant ($F(1,19) = 10.6, p < .005$). This additional adaptation can be the result of interaction during gesture execution, but one cannot exclude a ballistic interpretation; some information about the timing of gesture execution may have been available before movement initiation.

Let us now turn to the results of the "off-line" condition, where the subject responded to a question put by the experimenter. It may be that, when not required to react quickly, the subject is better able to synchronize voice onset and apex as the distance of the referent LED varied. In the off-line condition the time from LED onset to movement initiation, T_I, is of no consequence, since the subject's response is cued not by the LED coming on, but by the experimenter's question. In the off-line condition, therefore, we measured voice onset time, as well as apex time, with respect to movement initiation denoting them by T_V', and T_E (execution time), respectively. The top half of Table 1 shows T_V' and T_E values under the different experimental conditions. For the purposes of comparison, the corresponding values obtained in the on-line condition are presented in the bottom half of the table.

An analysis of variance of the off-line values of T_E showed a significant effect of distance, movements to far LEDs taking 86 milliseconds longer than those to near LEDs ($F(1,19) = 245.2, p < .0001$). The mean distance traveled by the finger in pointing to far LEDs was 407 millimeters, and to near LEDs 212 millimeters. It was found that for each of the four LEDs, the movement was significantly less expansive ($p < .05$) in the off-line condition than in the on-line one. With T_E as the dependent variable, there was also a Hand × Field interaction ($F(1,19) = 187.7, p < .0001$) insofar as movements in the ipsilateral field were executed 110 milliseconds faster than those in the contralateral. The three-way Hand × Field × Distance interaction was also significant ($F(1,19) = 73.0, p < .0001$), on account of the fact that the difference

between near and far T_E values was greater in the contralateral field than in the ipsilateral. Nevertheless, a separate analysis of variance showed that for the ipsilateral field alone, the difference between execution times (T_E) for the near and far LEDs (49 milliseconds) was still significant ($F(1,19) = 86.8$, $p < .0001$). Thus, the general pattern of results is very similar to that found in the on-line condition.

The same was true for the pattern of voice onset times T_V'. Analysis of variance showed a significant effect of distance ($F(1,19) = 55.4$, $p < .0001$) and, moreover, pointing movements in the ipsilateral field were associated with shorter voice onset times than contralateral movements ($F(1,19) = 61.5$, $p < .0001$). The voice onset times for pointing to near and far LEDs differed slightly less in the ipsilateral than in the contralateral field ($F(1,19) = 4.8$, $p = .041$), but the difference in the former case was nevertheless significant ($F(1,19) = 26.8$, $p < .0001$).

It was suggested above that the subject might achieve a greater degree of synchronization between pointing and voicing in the more relaxed off-line condition. Such a trend would be reflected in the difference between voice onset time and apex time being less dependent on the experimental conditions, and in particular the distance of the LED. However, an analysis of variance carried out on the values of T_E–T_V' (Table 1) showed a pattern similar in most respects to that of the on-line condition. There was a significant effect of distance, with speech occurring 25 milliseconds after the apex in the case of near LEDs, but only 2 milliseconds after in the case of far ones. This shift in relative timing in going from near to far was of the same magnitude and in the same direction as in the on-line case, where speech preceded the apex by 41 milliseconds in the case of near LEDs and by 66 milliseconds in the case of far ones. At the same time, as in the on-line case, the overall figures conceal marked differences between the situations in the two fields. Thus, in the contralateral field, the difference in the apex–speech interval between near (−4 milliseconds) and far (42 milliseconds) LEDs was substantial and significant ($F(1,19) = 38.6$, $p < .0001$), whereas there was no noticeable difference in the ipsilateral field, the corresponding figures being −46 and −47 milliseconds, respectively. In the ipsilateral field, therefore, there is virtually full compensation for distance. The close similarity between off-line and on-line results makes it unlikely that the *extent* to which speech and gesture synchronize as a function of distance is very dependent on the speed of the response.

However, the relative timing of apex and voice onset was rather different in going from the on-line to the off-line condition. Table 1 shows clearly that apex tended to precede voice onset in the off-line condition (by 14 milliseconds on the average), but to follow it in the on-line one (by 53 milliseconds). An analysis of variance of T_E–T_V' values over the two conditions showed that this difference was significant ($F(1,19) = 10.6$, $p = .0042$). However, the on-line versus off-line factor showed no interaction with either distance, hand, or field, confirming the earlier observation that the pattern of results is essentially the same for the two conditions. The one significant main effect suggests that, when instructed to respond immediately, speakers were more successful in speeding up speech onset than in reducing execution time. The extent to which speeding up occurred under the different experimental conditions can easily be determined from the values given in Table 1. The average ratio of on-line to off-line movement execution times was .75 (with a range over subjects of only .04). For T_V', the corresponding ratio was .63 (range: .04), indicating that the time compression factor takes different values for voice onset and movement execution. The importance of this finding is that there is, apparently, no single optimal synchronization of deictic word relative to deictic gesture. If, as seems to have been the case here, speed of

TABLE 1

EXPERIMENT 1: MOVEMENT EXECUTION DURATION (T_E) AND VOICE ONSET LATENCY (T_V') WHEN REFERRING TO NEAR AND FAR LEDS IN THE LEFT AND THE RIGHT VISUAL FIELD (IN ms)

		Left field		Right field	
Condition	Hand	Far	Near	Near	Far
Off-line	Left				
	T_E	524	462	536	658
	T_V'	562	509	537	617
	Difference	−38	−47	−2	41
	Right				
	T_E	675	552	478	514
	T_V'	632	559	522	569
	Difference	43	−7	−44	−55
On-line	Left				
	T_E	388	356	408	491
	T_V'	365	331	343	391
	Difference	24	26	64	100
	Right				
	T_E	490	409	357	387
	T_V'	385	354	337	354
	Difference	106	55	20	33

execution of the gesture was the limiting factor in the compression process, the speaker could have chosen to time voice onset in accordance with the same compression ratio in order to achieve coincidence with apex. The fact that this did not happen suggests that the two systems adjust their parameters relatively independently, which corresponds better with the ballistic than with the interactive view.

Finally, it may be noted that the extent to which subjects succeeded in achieving correspondence of voice onset and apex, if indeed that was their aim, varied considerably from one individual to another. Thus, in the off-line condition voice onset followed apex by 14 milliseconds on the average, but the standard deviation (over 20 subjects) was no less than 100 milliseconds. In the on-line condition, where voice onset led the apex by 53 milliseconds, the standard deviation was 114 milliseconds.

Although voicing occurred later in the contralateral field than in the ipsilateral, it occurred earlier in relation to the apex, the relative shift being significant in both the on-line case, where it amounted to 55 milliseconds ($F(1,19) = 100.2$, $p < .0001$) and the off-line case where it was 65 milliseconds ($F(1,19) = 69.8$, $p < .0001$). This finding suggests that there may also be other criteria which speakers try to satisfy. A first possibility is that a speaker tries to align the deictic word with the moment of maximum speed; Jeannerod (1981) found that for grasping gestures this moment is reached at about one-third of the execution time. This criterion, however, would make no communicative sense; it carries almost no information for the listener about which target light is intended. The maximum speed moment cannot be very salient for the person across the table, and it will in many cases come too early to distinguish the targets uniquely; there is no comparison to the communicative saliency of the moment and position of apex. Another criterion could be that the pointing finger should be directed at the referent LED. As may be seen from Figure 1, this alignment of finger and referent involves less rotation of forearm and hand in the contralateral field than in the ipsilateral. Hence, the directionality criterion may be met at an earlier point in the movement in the contralateral field than in the ipsilateral.

Discussion

The findings of this experiment show that the timings of gesture and voice onset covary to a significant extent. It is an open question, at this stage, whether the transfer of information between the systems takes place exclusively in the planning phase, in accordance with the ballistic theory, or whether there is feedback during execution as well. It is not just the timing of gesture *initiation* which covaries with voice onset; over and above this voice onset covaries with the duration of gesture *execution*. This excludes a particularly strong version of the ballistic theory, namely, that the speech system has access only to temporal parameters of the gesture planning phase.

A strong version of the interaction theory, predicting absolute synchronization of pointing and voicing, is similarly excluded. While there was evidence for such an alignment in the ipsilateral field, it clearly does not hold in the contralateral. Moreover, the relative timing is dependent on the speed of reaction, insofar as voice onset time and movement execution time, when speeded up, are not equally compressed and to that extent show some degree of independence.

It should be noted that the observed alignment between gesture and speech may be brought about by either unidirectional or bidirectional interaction. That is, the onset of speech may adapt to parameters of the gesture and/or the time course of the gesture may be specified so as to achieve alignment with the deictic utterance. The next experiment was designed to determine the direction of the speech–gesture dependency, as well as the phase(s) in which the dependency is established.

Experiment 2

The most likely form of interaction underlying the covariance between gesture and speech observed in the previous experiment is that voicing adapts to gesture but not conversely. In order to test the hypothesis that the direction of adaptation is from speech to gesture, it is necessary to compare the condition in which gesture and speech accompany each other (the GS condition), as in Experiment 1, with both a gesture-only (G) and a speech-only (S) condition. If adaptation is in the direction posited (i.e., speech to gesture), one would expect to find that the speed of gestural response is independent of whether or not it is accompanied by speech. At the same time, one should find that the speed of voicing *is* affected by the presence or absence of an accompanying gesture. On the other hand, the inverse unidirectional relationship, namely, adaptation of gesture to speech, is less likely to obtain, though the possibility that the interaction is two way, with the parameters of gesture being affected by the presence of a speech response, as well as speech onset being sensitive to the presence of a gestural response, should not be excluded. These alternatives may be evaluated by comparing subjects' responses under the three conditions GS, G, and S.

At the same time, the phase in which the alignment of speech and gesture is established will be investigated by comparing the timing of the apex and voice onset with that of movement initiation.

It should be added that, strictly speaking, the S condition cannot be realized. The present study is concerned with situations in which a deictic gesture is obligatory. When no hand gesture is made, the speaker will still direct his gaze or head toward the target LED; there will always be some form of pointing. But speakers also direct their gaze in conditions *with* hand gesture; they always look at the target LED indicated. The real difference between GS and S conditions is, therefore, one between multiple hand plus gaze gesture and single gaze gesture.

Method

Procedure. The need for the subject to be able to uniquely specify the referent in the speech-only condition without relying

solely on gaze and head turns dictated a modification to the procedure used in Experiment 1. What this entailed was a reduction in the number of stimulus LEDs from 4 to 2, one near and one far, so that the problem of reference could then be solved by the use of the expressions "dit lampje" ("this light") and "dat lampje" ("that light"), respectively. However, in order to be able to once again compare the situations in ipsi- and contralateral fields, it was necessary for trials to be blocked by field. In performing the speech-only condition (S), the subject was asked to rest his finger on the push-button throughout (pressing it when ready to proceed to the next trial). In the dual GS condition, the subjects' task was similar to that in the on-line condition of the previous experiment. Finally, in the gesture-only condition (G) the subject was simply asked to point at the stimulus LED as soon as it came on. In each condition the experimenter sat across the table, and "noted down" the subject's response.

Subjects performed six experimental series, a left-field and a right-field one in each of the three main conditions. Each series consisted of 20 test trials, 10 for each LED, in quasirandom order. In order to accustom the subject to the new situation, four practice trials preceded each block of 20. Half of the subjects began with three series in the left field, while the others started with three in the right. The order of the three series within a field was counterbalanced over subjects, with the left-field order for each individual being the same as the right. In the two conditions involving manual pointing (GS and G), only the right hand was used to perform the gesture. In all other respects, the method was identical to that of Experiment 1.

Subjects. Given the above procedure, there were 12 possible ways of ordering the six experimental series, and this dictated that, in order to achieve counterbalancing, the number of subjects should be a multiple of 12. In view of the low within-subject variance observed in Experiment 1, it was thought that the minimum number of 12 would suffice. The 12 subjects, 5 males and 7 females, were all right handed.

Results

Figure 4 shows the average values for movement initiation time (T_I), apex time (T_A) and voice onset (T_V) under the three different conditions. Figure 5 gives the corresponding (linear) distances moved by the finger in the two conditions which involved pointing, G and GS. It is clear from Figure 5 that, as in Experiment 1, more extended pointing movements are made to far LEDs than to near ones, and Figure 4 shows that this pattern is reflected in the corresponding values of apex time, T_A. In the GS condition movements to far LEDs took 83 milliseconds longer from the moment of flashing than did those to near LEDs; in the G condition the difference was 79 milliseconds. Separate analyses of variance for these two conditions showed the differences to be significant in each case (GS: $F(1,11) = 33.2$, $p < .001$; G: $F(1,11) = 43.5$, $p < .0001$). As was found in Experiment 1, ipsilateral (in this experiment, right field) apex times were shorter than contralateral ones. The difference was 61 milliseconds for the GS condition ($F(1,11) = 24.9$, $p < .001$), and 64 milliseconds for the G condition ($F(1,11) = 24.3$, $p < .001$).

Before presenting a comparison of the GS, G, and S conditions, it is pertinent to consider whether the results of the GS condition show the same pattern as those of the on-line condition of the previous experiment. As far as voice onset, T_V, is concerned, the effect of distance found in Experiment 1 was also found in the GS condition of the present experiment. Overall, voice onset was 47 milliseconds later for far LEDs than for near LEDs ($F(1,11) = 14.8$, $p < .01$), but the effect was stronger in the ipsilateral field, where the difference was 61 milliseconds ($p < .01$, t test, one tailed) than in the contralateral, where it was only 34 milliseconds ($p < .05$, t test, one tailed). On average, T_V led T_A by 23 milliseconds,

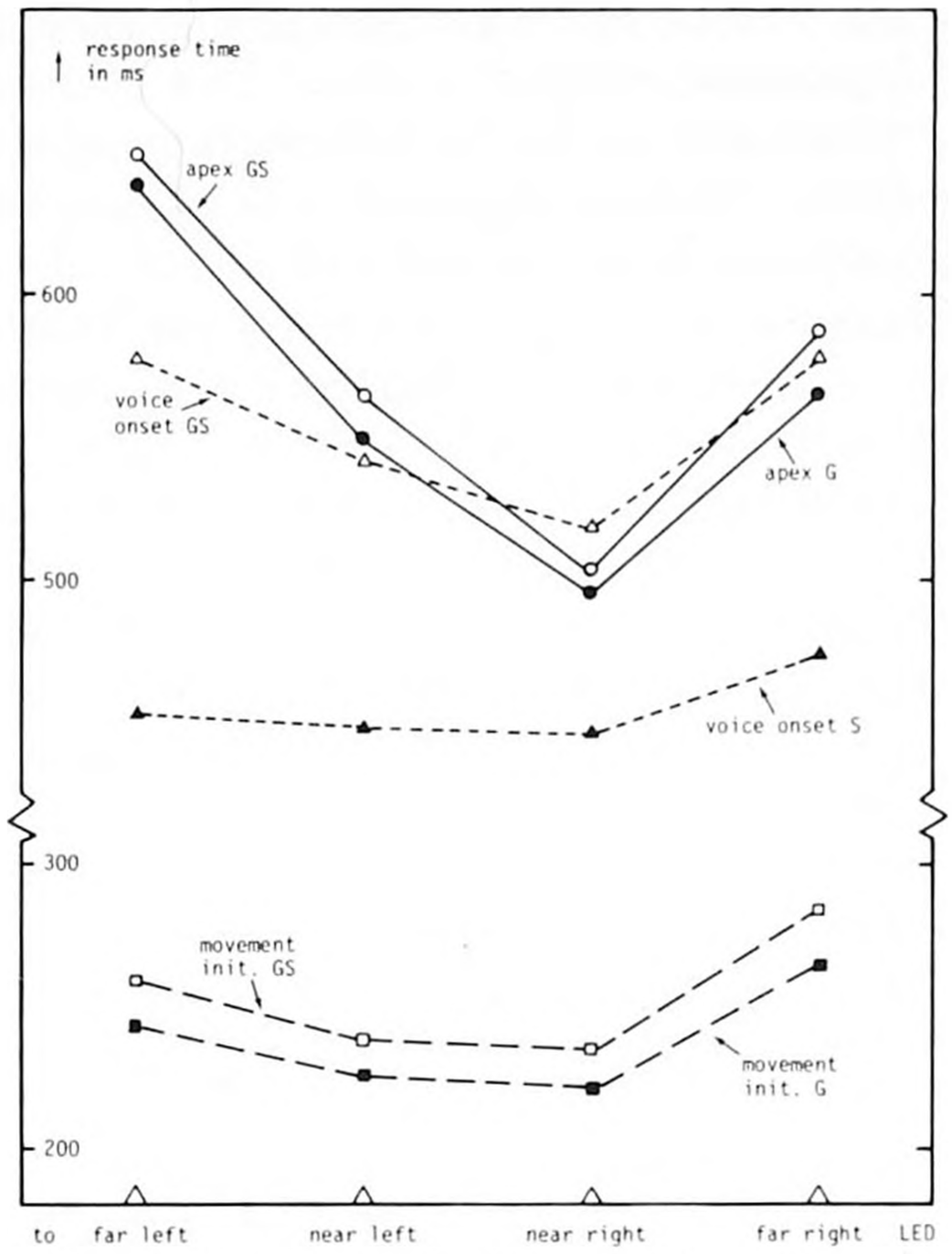

FIG. 4. Experiment 2: Response times for movement initiation, apex and voice onset in referring to four LEDs. Right hand data.

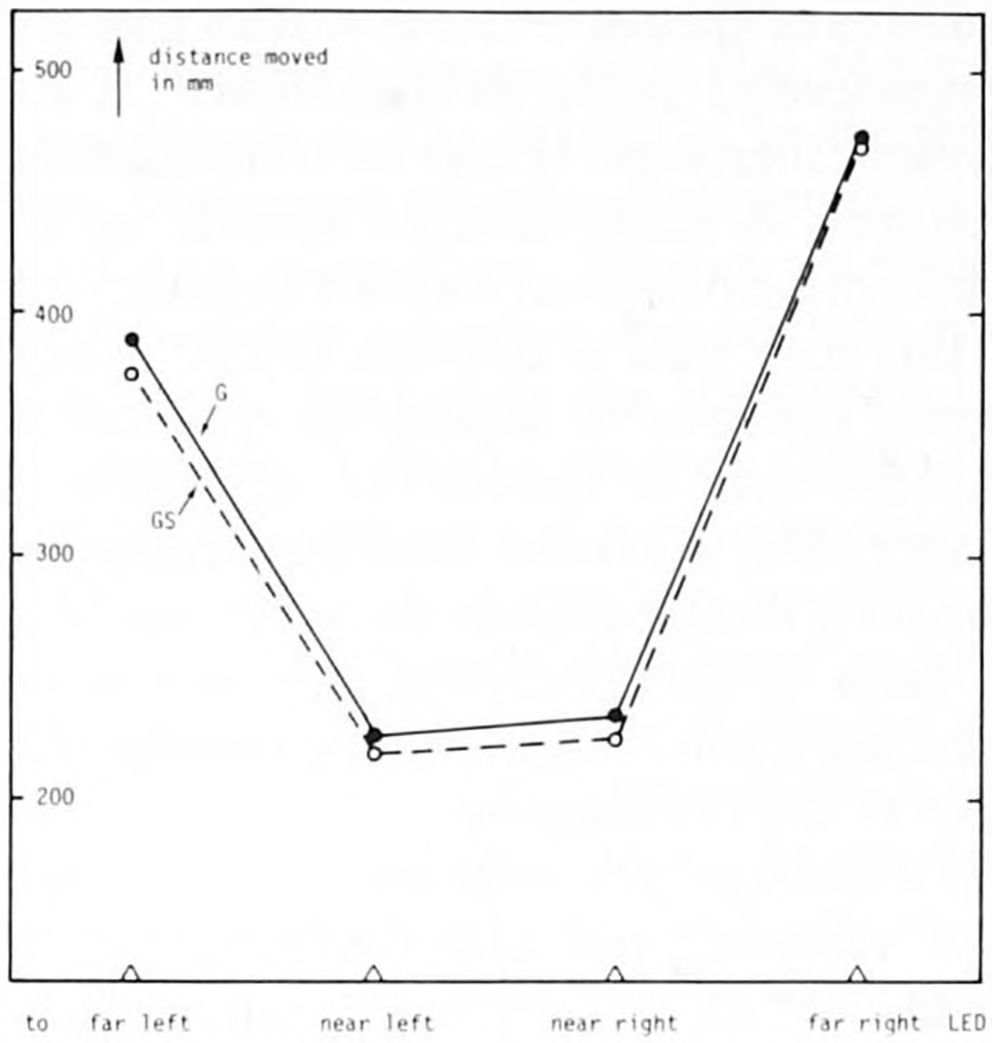

FIG. 5. Experiment 2: Distance moved by the right hand from movement initiation to apex of gesture. Gesture alone (G) and combined gesture and speech (GS) conditions.

though as in Experiment 1 voice onset was later relative to apex in the ipsilateral field, where, on the average, it actually lagged by 2 milliseconds, than in the contralateral where it led by 48 milliseconds. This difference between the two fields is significant ($F(1,11) = 13.1$, $p < .01$).

The variation of T_A–T_V with distance within the two fields exhibits a similar pattern to that found in Experiment 1. Thus, overall in going from near to far LEDs, T_A–T_V increased by a significant margin of 35 milliseconds ($F(1,11) = 9.1$, $p < .05$). However, relative invariance of T_A–T_V with distance was found in the ipsilateral field, where the difference was only 21 milliseconds ($t(11) = 1.64$, n.s.), compared with the contralateral field where it amounted to 50 milliseconds ($t(11) = 3.56$, $p < .01$). In all important respects, therefore, the results for the GS condition of the present experiment are in accord with those of the on-line condition in Experiment 1.

As suggested in the discussion of Experiment 1, the reason that invariance holds in the ipsilateral field but not in the contralateral field and that speech tends to occur later relative to the apex in the former might be that the subject tries to line up the pointing finger with the target LED. This is realized earlier in the contralateral than in the ipsilateral field, and it is more difficult to achieve in the case of the far LED in the ipsilateral field (see Figure 1).

Turning next to a comparison of the three main conditions, the first question that will be considered is whether the gestural system is affected by the planning and execution of voicing. To this end, the timing of gestures in the gesture-only condition was compared with their timing when accompanied by speech. Treating the two conditons, G and GS, as two levels of the same factor, an analysis of variance was performed with T_A as the dependent variable. This analysis showed that, although apex times were slowed by 14 milliseconds when accompanied by speech, the effect was nonsignificant ($F(1,11) = 2.1$, $p = .18$). Although there were also highly significant effects of field ($F(1,11) = 27.9$, $p < .001$) and distance ($F(1,11) = 43.0$, $p < .0001$), neither of these factors showed a significant interaction with the presence or

absence of speech ($F(1,11) = 0.08$ and 0.12, respectively). As far as apex time T_A is concerned, therefore, it can be concluded that gesturing is unaffected by speech, as if it were an autonomous module indeed.

But apex time is the sum of two components, the time to initiate the gesture, T_I, and the time to execute it, T_E. Does the preparation of speech have any discernible effect on these components individually? Is it, more in particular, the case that movement initiation is slower in the GS condition than in the G condition?

An analysis of variance with T_I as dependent variable, and with GS and G as two levels of the same factor, showed that movement initiation was significantly delayed, in the presence of speech, by an interval of 14 milliseconds ($F(1,11) = 6.1$, $p = .03$). The magnitude of the delay corresponded exactly to the nonsignificant 14-millisecond delay in apex time noted above. This result suggests that the planning of speech does indeed delay the initiation of movement, but has no effect on the subsequent execution stage. In other words, the entire pointing motion is somewhat delayed, but its duration is unaffected by the presence of speech. This is, clearly, supportive of the ballistic theory; the preparation of speech slightly affects the preparation of gesture during the planning phase, but after movement initiation the execution of the gesture is ballistic and follows the same time course whether or not it is accompanied by speech.

For the sake of completeness it should be noted that T_I was not significantly different for movements to the ipsilateral and contralateral fields. There was, however, a significant effect of distance in that, over both GS and G conditions, pointing to a far LED was initiated 32 milliseconds later than pointing to a near LED ($F(1,11) = 13.5$, $p < .01$). This effect of distance was stronger in the ipsilateral (right) field (46 milliseconds) than in the contralateral field (19 milliseconds) ($F(1,11) = 9.3$, $p = .01$). Neither field nor distance showed a significant interaction with the speech versus nonspeech factor.

The final issue to be considered is whether, as was expected, voice onset is dependent on the presence of gesture. It is immediately apparent from Figure 4 that voice onset latencies differed considerably in the GS and S conditions. The curve for the GS condition is not only substantially higher, indicating later onset, than the one for the S condition, but their shapes are also markedly different. Thus, the voice onset curve for the speech-only condition is essentially flat, whereas the curve for the GS condition shows the characteristic U shape also observed in Experiment 1.

These impressions are confirmed by an analysis of variance with GS and S as two levels of the same factor, and T_V as a dependent variable. Voice onset was later in the GS than in the S condition by 99 milliseconds ($F(1,11) = 25.5$, $p < .001$), and the difference was more marked for far than for near LEDs ($F(1,11) = 4.8$, $p = .05$). A separate analysis of variance for the S condition alone showed that neither field nor distance affected voice onset to a significant degree; the curve for the S condition can indeed be considered as flat. This finding is especially relevant for the present experiments because it shows that variations in voice onset time as a function of LED distance are unlikely to be due to differences in detection latencies for central versus peripheral stimuli; such differences would have been apparent in the results of the present S condition.

Discussion

It is quite clear that the covariation of speech and gesture observed in the present experiment can be largely explained in terms of the adaptation of speech to gesture, rather than the converse. We found one small but significant exception; movement initiation was somewhat slower in the GS condition than in the G conditions, which suggests that the preparation of speech to some extent interferes with the

planning of gesture. At this point a comparison may be made with the findings of a study by Holender (1980) in which subjects responded to a visually presented stimulus letter either by naming it (speech only), pressing a corresponding key (manual only), or both (speech plus manual). There were four alternative stimulus letters, and hence four different names and four alternative keys. These conditions are similar to our S, G, and GS conditions, respectively. Holender found that the manual reaction time, comparable to our movement initiation time, was the same in the manual-only as in the dual condition, whereas we found a small 14-millisecond difference between the G and GS conditions. Voice onset, however, was markedly delayed in Holender's dual condition, as compared to his speech-only condition. This finding corresponds well to our results concerning voice onset in the GS and S conditions. In other words, the largely unidirectional effect of hand movement on speech, observed in our experiment was also apparent in Holender's data. At the same time, Holender managed to create the inverse effect of speech planning on manual latencies by instructing the subject to give the vocal response first and as fast as possible in the dual task. This instruction could be followed, but severely delayed both the vocal and the manual response. The manual response was given as much as 125 milliseconds after the vocal response, whereas in the previous experiment the former had preceded the latter by about 80 milliseconds. When subjects were instructed to synchronize the two responses, though with less emphasis placed on speed, they were simply not able to do it; the delay between manual and vocal response was still no less than 70 milliseconds. This finding led Holender to conclude that "when used together, these processors compete for a common processing capacity pool." It is, presumably, in order to minimize such competition that subjects space the manual and vocal responses.

There is a certain amount of evidence, though it is not conclusive, that competition for common resources is also at stake in our pointing tasks. There is, in the first place, systematic spacing of movement initiation and voice onset. Second, we found that movement initiation was slightly, but significantly delayed in the GS condition, as compared to the G condition. This may have been caused by resource competition, though other explanations are possible. A third piece of evidence concerns the relative speed of reaction in the on-line condition of Experiment 1, where there were four LEDs and the same deictic expression was used for each, and the GS condition of the present experiment, where there were only two LEDs, but the number of deictic expressions was increased to two. As a result of these changes the pointing response might have been easier to plan in Experiment 2, and the speech more difficult. If the planning of speech requires resources over and above what is available during the gesture planning phase, one would expect voicing to lead the apex by a shorter interval in Experiment 2 than in Experiment 1. In fact, the respective intervals were 23 and 53 milliseconds (but their difference was not significant). Movement initiation, on the other hand, should be slower in Experiment 1 than in 2, because there were more alternatives to choose from in the former case. This argument is supported by the finding that the average value of T_I for the right hand in Experiment 1 (289 milliseconds) was significantly higher than in Experiment 2 (234 milliseconds) ($p < .01$, t test). These are arguments for the hypothesis that the two response systems compete for common resources in the planning phase. If this competition is *limited* to the planning phase, it is consonant with the ballistic theory. If, however, it extends into the execution phase it can only be handled by the interaction theory. In the next experiment this issue of competition for common resources will be studied further by systematically varying the number of verbal and gestural alternatives.

The main results of the present experiment can be summarized as follows. At least for the tasks used here, the interdependency between speech and language is almost completely unidirectional. Speech onset time depends to a substantial degree on the gesture made, but the execution of the gesture is completely independent of whether it is accompanied by speech or not. The presence of speech affects gestural timing only in the planning phase, the initiation of the movement occurring significantly later, by a matter of 14 milliseconds, when speech is present that when it is absent. It was suggested that the latter effect is due to competition for common resources between speech and gesture in the gesture planning phase. So far the results are in agreement with the ballistic theory which limits information exchange between the two response systems to the planning phase. It should be noted, however, that the results of the present experiment do not tell us whether the parameters of voice onset are set during the preparation or during the execution of gesture. Only the former alternative conforms with the tenets of the ballistic theory. This issue will be further analyzed in Experiment 3, and explicitly put to the test in Experiment 4.

EXPERIMENT 3

If there exists competition for common resources between the speech and gesture systems during the planning stage, or even during execution, one would expect to see interaction effects in the pattern of latencies. A simple example may clarify this point. Assuming that initiation of the hand movement is delayed until both gesture and speech have been prepared, then three cases can be distinguished. The first is that the preparatory phases of the two processes take place in parallel without interference or recourse to common resources. In this case, which amounts to full independence, completion of the slower of the two processes determines the moment of movement initiation. Let us assume further that the preparation time varies slightly, but systematically with the number of alternatives for the channel concerned (i.e., number of LEDs to be indicated or the number of deictic terms to be used). If one assumes that, in every instance, the preparation time for gesture is large by comparison with that for speech (given the converse assumption the following argument holds *mutatis mutandis*), only the number of gestural alternatives will have an effect on movement initiation time, since preparation for speech is always completed before preparation for gesture. Hence, the number of alternative deictic terms in the task will have no effect on movement initiation latency, nor will there be an interaction effect between this factor and the number of gestural alternatives.

The second case is that in which the preparatory stages for gesture and speech are organized in fully serial fashion, with the former preceding the latter. In this case not only will the number of gestural alternatives be reflected in the movement initiation latencies, but the number of verbal alternatives will as well. The two effects, moreover, will be additive, and consequently there will be no statistical interaction between the two factors.

In the third case, where the two systems are not operating in a fully serial fashion and compete for common resources, one would expect to find statistical interaction. This case is intermediate to the first and second. Thus, when processing load is low, the two systems can operate more or less in parallel, as in case one, with movement initiation times being relatively unaffected by the number of verbal alternatives. Under high load conditions, however, the operation will have to become more serial in nature, as in the second case above, in order not to exceed the capacity of the processing resources. Consequently, the number of verbal alternatives will be seen to affect overall reaction times, but in a manner which depends on the processing load imposed by the number of gestural alterna-

tives. The resulting interaction, moreover, is likely to be superadditive, insofar as the switch to serial operation is most likely to occur when both systems are coping with a high number of alternative responses.

It should be added that movement initiation need not await *full* preparation of gesture and speech, and that the final stage of planning of either or both responses may take place after movement initiation, as predicted by the interaction theory. This state of affairs would be apparent if the factors of verbal and gestural number of alternatives show interaction effects in the gesture execution times or in the voice onset times, both measured from movement initiation.

The present experiment was designed to investigate the effect on speech and gesture reaction times of varying the number of response choices available to each system, their potential interaction and superadditivity. A further aim was to distinguish the phases in which effects arise by comparing movement initiation, voice onset, and apex times, as weli as durations of movement execution.

Method

Subjects. There were 12 subjects, 6 male and 6 female, all of whom were right handed.

Procedure. The number of gestural alternatives was either two or four. The four-LED condition was the same as the on-line condition in Experiment 1, with the four LEDs distributed over right and left visual fields (4RL). The realization of the two-LED condition was less straightforward. In order to make it comparable to the four-LED condition, the same four LEDs had to be employed, but using just two of them at a time. This requirement was met by partitioning the two-LED condition into four blocks: (i) the two left-field LEDs (2L), (ii) the two right-field LEDs (2R), (iii) the two "near" LEDs, one from each field, creating a narrow functional field (2N), and (iv) the two "far" LEDs, creating a wide functional field (2W). Each of these blocks consisted of 20 test trials, and was preceded by eight practice trials. For the four-LED condition (4RL), two such blocks of 20 test trials were employed in order to achieve greater comparability in the number of trials per LED between two- and four-LED conditions. Hence, in the test phase of the experiment, a subject received six blocks of 20 experimental trials, each preceded by eight practice trials. In order to help reduce learning effects, this test phase was preceded by a practice phase consisting of the same six blocks, but with only eight trials per block. The order of the six blocks, both in the practice phase and in the test phase of the experiment, was fully counterbalanced over subjects, but in such a way that the two 4RL blocks were always separated by two two-LED blocks of trials. As a result, each condition occurred equally often in each of the six order positions. There was a further restriction on the ordering of blocks, discussed shortly.

The number of alternatives for speech was either one: "dat lampje" ("that light") or two: "dat lampje" ("that light") and "dit lampje" ("this light"). In the two-alternative situation a subject was instructed to indicate a "near" LED by means of "dit lampje," and a "far" LED by "dat lampje." This convention would not, of course, have worked for the 2N condition in which both LEDs were "near" or the 2W condition in which both were "far." Accordingly, the ipsilateral LED was indicated by "this light" and the contralateral one by "that light," which is fairly natural. The two levels of the number of speech alternatives factor were presented on different days, about 1 week apart. Half the subjects began with the one-alternative condition, and performed the two-alternative task a week later. The other half received the conditions in the reverse order.

Reference was made above to a further constraint on the ordering of blocks. The restriction arose as a result of the need to ensure that the deictic term used to refer to

a particular light ("this light," "that light") did not change in going from one block to the next. For instance, when two speech alternatives were used, the left field near LED was indicated by "that light" in the 2N condition, but "this light" in the 2L condition. Such blocks were never presented in immediate succession. Subjects always used their right hand in pointing, and they were asked to respond as soon as the LED came on.

Movement analysis. In this experiment and the following one, changes were made in the methods employed to compute the movement initiation time T_I and the apex time T_A. Whereas before these times were determined by computations carried out on the incremental distance function INCD, they were now performed directly on the distance function derived from the x- and y-coordinates (disregarding variation in the z-direction), that is, the linear distance of the pointing finger to its starting position. Thus T_I was defined as the time, measured from LED onset, at which the distance function first exceeded a predetermined threshold (which could be adjusted to take account of the noise level in each subject's record), while T_A was simply the time at which the distance function reached 99% of its maximum value. The distance function, being inherently less noisy than the INCD function, allowed a somewhat greater degree of accuracy in the measurement of these times. Moreover, the first method would not have been suitable for use in Experiment 4 where there were trials on which the arm was restrained in the course of the movement, the result of which was to introduce minima into the INCD function before the extremum of the movement was reached.

Results

Table 2 shows the mean values of T_I, T_A, and T_V under the different combinations of conditions. Before considering the main issues, namely, the effects of the numbers of verbal and gestural alternatives, it is of interest to look at the field and distance effects. In fact, they were quite similar to those obtained in the GS condition of the previous experiment and the on-line condition of Experiment 1. The apex time T_A was significantly greater, by 96 milliseconds, in the case of far LEDs than in the case of near LEDs ($F(1,11) = 183.1$, $p < .0001$). It was also significantly greater, by 58 milliseconds, for movements made in the contralateral (left) field, as compared to those made in the ipsilateral (right) field ($F(1,11) = 54.6$, $p < .0001$). There was also evidence of an interaction, in that the difference between apex times to near and far LEDs was somewhat greater in the contralateral field than in the ipsilateral ($F(1,11) = 8.7$, $p < .05$). This pattern of results was reflected in the corresponding distances traveled by the pointing finger, namely, 225 millimeters (near left), 391 millimeters (far left), 232 millimeters (near right), and 481 millimeters (far right).

Voice onset latency T_V was 62 milliseconds shorter for near than for far LEDs ($F(1,11) = 61.6$, $p < .0001$), and 15 milliseconds shorter in the ipsilateral field than in the contralateral field ($F(1,11) = 9.9$, $p < .01$).

Finally, movement initiation T_I was significantly earlier, by 39 milliseconds ($F(1,11) = 249.2$, $p < .0001$) for near than for far LEDs. At the same time, there was an interaction between distance and field ($F(1,11) = 12.5$, $p < .01$), such that the difference between initiation times to near and far LEDs was more pronounced in the ipsilateral field than in the contralateral. Thus the effects of field and distance on the timing of gesture and voice are very similar to those found in Experiments 1 and 2, we refrain from presenting a graph of these results which hardly differed from those shown in Figures 2 and 4.

The main question of this experiment concerns the effects of the numbers of gestural and verbal alternatives on the timing of the various phases of the response (namely, T_I, T_A, T_E, and T_V). The nature of

TABLE 2

EXPERIMENT 3: MOVEMENT INITIATION (T_I), APEX (T_A), AND VOICING (T_V) LATENCIES (IN ms) FOR REFERRING TO NEAR AND FAR LEDs IN TWO-LED AND FOUR-LED CONFIGURATIONS IN RELATION TO THE NUMBER OF CHOICES OF DEICTIC EXPRESSION (UPPER HALF: ONE; LOWER HALF: TWO)

Deictic expressions	LED array		Left field		Right field	
			Far	Near	Near	Far
"that light"	4RL	T_I	235	211	202	260
		T_A	683	573	531	614
		T_V	608	562	550	609
	2R	T_I			197	238
		T_A			525	594
		T_V			557	603
	2L	T_I	225	196		
		T_A	655	559		
		T_V	612	545		
	2N	T_I		188	189	
		T_A		553	502	
		T_V		538	521	
	2W	T_I	233			257
		T_A	690			622
		T_V	628			601
"this light"	4RL	T_I	269	242	247	281
versus		T_A	714	613	566	646
"that light"		T_V	672	606	596	652
	2R	T_I			221	251
		T_A			542	611
		T_V			555	600
	2L	T_I	248	220		
		T_A	695	592		
		T_V	654	579		
	2N	T_I		210	206	
		T_A		580	524	
		T_V		558	542	
	2W	T_I	251			260
		T_A	689			614
		T_V	632			608

these effects is illustrated in Figure 6, which shows how the relative timing of movement initiation, pointing apex, and voice onset was modified in going from one verbal alternative to two, and from two gestural alternatives to four.

In what follows, the results pertaining to the gestural planning phase will be considered first, and then we will present an analysis of the phase of gesture execution.

The moment of movement initiation is affected by both the number of gestural and the number of verbal alternatives. Movement initiation T_I was on average 19 milliseconds longer in the four-LED condition than in the two-LED conditions ($F(1,11) = 51.1$, $p < .0001$), and 25 milliseconds longer when there was a choice between two verbal alternatives ("this, that") than when there was just one ("that") ($F(1,11) = 17.0$, $p < .01$). Moreover, there was a significant interaction between these two factors ($F(1,11) = 6.5$, $p < .05$). This interaction is "superadditive" in that the difference in T_I

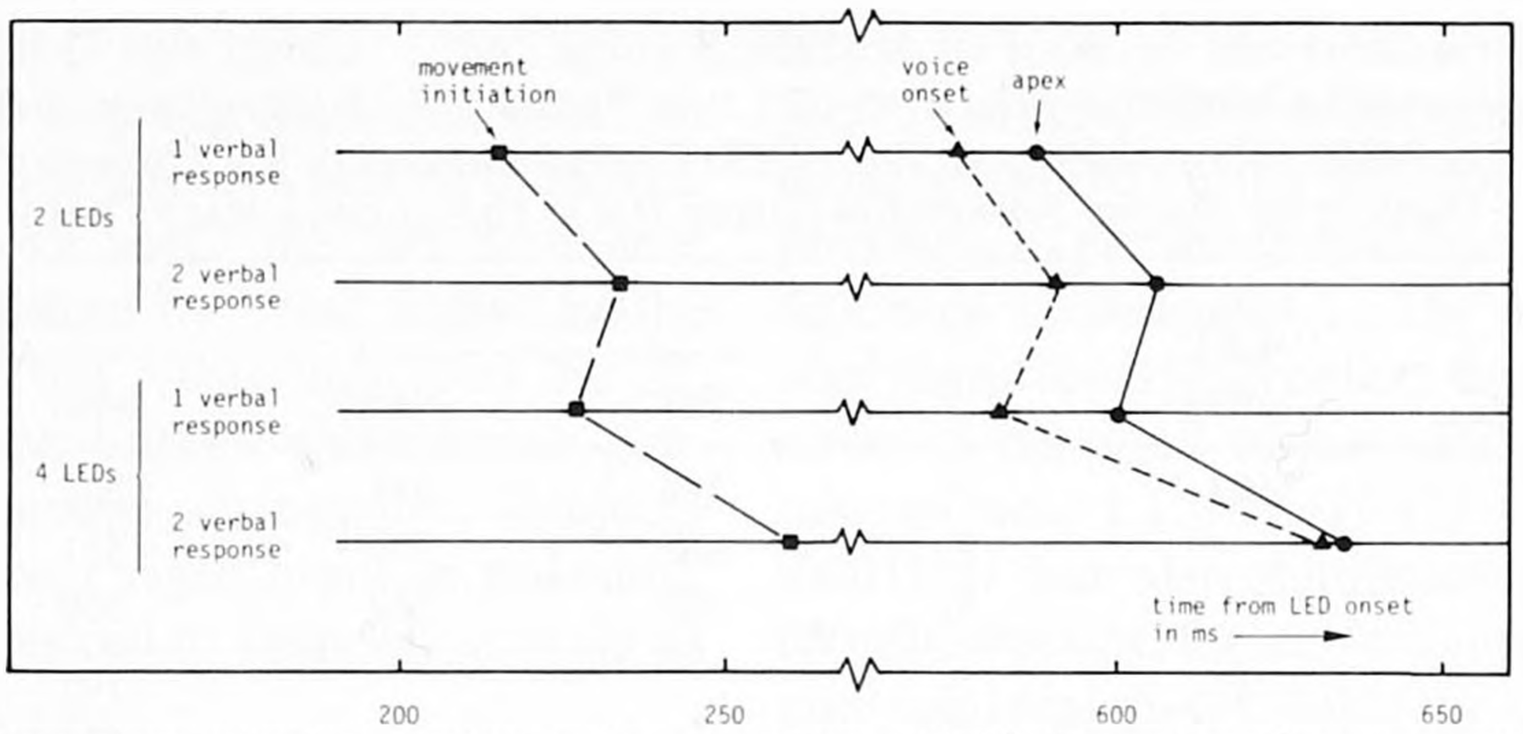

FIG. 6. Experiment 3: The relative timing of movement initiation, apex of gesture, and voice onset in referring to one out of two or four LEDs, by means of a single or of two verbal alternatives.

between the conditions with one and two verbal alternatives is larger (by 15 milliseconds) in the four-LED than in the two-LED condition, precisely the pattern of results one would expect if there is competition for common resources between speech and gesture up to the point at which movement is initiated.

In order to evaluate the respective merits of the ballistic and the interaction theories, it is necessary to view the above finding in relation to what occurs during the phase of execution.

Apex time T_A varied significantly with the number of gestural alternatives, being 21 milliseconds greater, on average, in the four-LED condition, than in the two-LED condition ($F(1,11) = 17.6$, $p < .01$). Although T_A also varied with the number of verbal alternatives, the apex in the two-alternative condition being 26 milliseconds later than in the one alternative, the difference failed to reach significance ($F(1,11) = 4.1$, $p < .07$). Neither was there a significant interaction between the number of gestural and the number of verbal alternatives ($F(1,11) = 3.7$, $p = .08$), although the data suggest the presence of a 16-millisecond margin of "superadditivity." Since the absolute differences at the level of T_A are almost identical to those at the level of T_I, the obvious conclusion is that T_E, movement execution time, is completely insensitive to the main variables of this experiment, the numbers of gestural and verbal response alternatives.

In order to confirm this interpretation, a further analysis of variance was carried out with movement execution time ($T_E = T_A - T_I$) as a dependent variable. This analysis gave significant effects for distance, visual field, and their interaction, but the effects of the two number of alternative factors and their interaction were negligible. (In fact, the four means under the different combinations of levels were 372, 373, 373, and 375 milliseconds!) So far, the situation is quite similar to that found in the previous experiment insofar as there is competition for resources up to the point of movement initiation, the resulting prolongation of this phase being transferred in a simple additive fashion to the apex, without any increase in the duration of the movement execution.

Lastly, what effects can be seen on the timing of voice onset, T_V? In going from two to four gestural alternatives voice onset was delayed by a significant 23 milliseconds ($F(1,11) = 27.3$, $p < .001$). Voice onset was also later, by 32 milliseconds, when there were two verbal alternatives than when there was only one, though in this case the effect was not significant ($F(1,11) = 2.2$, $p = .17$). At the same time there was a significant interaction between the two factors ($F(1,11) = 9.3$, $p < .05$). It exhibits superadditivity, the difference in voice onset between the conditions with one and two verbal alternatives being 34 milliseconds greater when there were four gestural alternatives than when there were only two.

The above effects might be accounted for

in terms of gesture *execution* influencing the timing of voice onset, as would be predicted by the interaction theory, but this explanation is not the only possible one. If voice onset displays a fixed temporal relation to the initiation of gesture, independent of the two factors investigated here, one would conclude that no voice delay is introduced during the *execution* of the movement. In order to clarify this issue, a further analysis of variance was performed, with voice onset referred to movement initiation (i.e., $T_V' = T_V - T_I$) as the dependent variable. Although neither the number of gestural nor the number of verbal alternatives showed a significant effect, the interaction between the two did ($F(1,11) = 8.4$, $p < .05$). The degree of superadditivity (19 milliseconds) was, of course, just the difference between the 34 millisecond superadditivity found in the case of T_V and the 15-millisecond superadditivity for T_I. This finding suggests, therefore, that there is a certain amount of competition for resources during that part of the speech preparation phase which overlaps the early stage of movement execution. The data shown in Table 2 reveal that the increased delay in speech in the condition involving two verbal alternatives and four referent LEDs is largely the result of what happens in the most effortful pointing movement, namely, that to the far LED in the contralateral field. This small, but significant superadditive effect cannot be accounted for by the ballistic theory.

Discussion

The picture which emerges from the results of this experiment is fairly clear. First of all, the latency of movement initiation is affected by both the number of gestural and the number of verbal alternatives. Moreover, the two factors interact in a superadditive fashion, indicating that planning of the two components of the response to some extent takes place in parallel, and that there is a degree of competition for common resources at least up to the point of movement initiation. The pattern of latencies seen at this point is almost exactly reproduced, but for the addition of a constant term, at the apex, the duration of movement execution showing no further effects of either variable or their interaction. In other words, the gesture system behaves in a fully ballistic fashion once the pointing finger has been released. Whether the execution of a gesture, when there are four alternatives available, makes greater demands on processing resources than when there are two is an open issue, though if one simply compares execution times, no such effect is in evidence. One could argue that the present experiment failed to show such an effect because subjects always gestured *as if* there were four alternatives, using the same strategy in all six blocks of trials. This is, however, unlikely, since the number of gestural alternatives did significantly affect movement initiation times.

As far as the timing of voice onset is concerned, the situation is somewhat more complicated. The pattern of speech onset times is not given simply by the addition of a constant term to the corresponding movement initiation times. In fact, what was found was a small, but significant increase in the degree of superadditivity, indicating that the planning of speech is subject to a certain amount of interference from the execution of the movement. The effect is restricted to the gestural movement requiring the most effort, that is, the one to the far LED in the contralateral field; for this limited case, the speech system cannot be said to be fully insensitive to the execution of gesture.

It should be noted that this finding can be interpreted in different ways. On the one hand, there may be increased competition for resources during the most effortful pointing movement. But one could also argue that in the four-LED, two speech alternatives situation the speaker takes greater pains to ensure that the moment of voice onset coincides with the apex of gesture. Especially the longest pointing mo-

tion, that is, the one to the far contralateral LED, gives the greatest opportunity for information to be fed back to the speech system. At this stage, however, the extent of feedback from gesture execution to speech and its organization in time is not clear. The final experiment was designed to shed some light on this issue. We will artificially impede the execution of the pointing gesture and study the effect thereof on speech onset. This interference was arranged to occur at unpredictable moments in the movement execution phase.

Experiment 4

The previous experiments have demonstrated a marked degree of adaptation of speech to gesture. It was, moreover, found that the planning of speech and gesture showed a small but significant degree of interference in the phase of gesture preparation, and possibly extending into the early phase of execution for cases where the gesture is especially effortful. The interaction was interpreted in terms of competition for common resources between the speech and the gesture systems. These resources are, evidently, used in preparation of the gestural and verbal responses, and when coordination is achieved, the parameters of this adaptation may well have been determined during the phase in which competition was observed. There is, on the other hand, the theoretical possibility that adaptation of speech onset parameters to gesture also occurs beyond the phase of gestural planning in which resource competition takes place. The moment of voice onset may be determined by feedback accumulated over the whole or part of the gesture execution phase, without such a process necessarily leading to superadditive effects of the kind observed in the previous experiment. What sort of evidence, over and above superadditivity, would support the latter view, the interaction theory?

In order to evaluate the interaction theory in the most direct way possible, we developed a means of detecting feedback from gesture execution to speech. The procedure consisted of mechanically impeding the pointing movement at unpredictable moments during its execution, and assessing what the consequences of this interference for voice onset latency were. Although one might be able to demonstrate the *feasibility* of feedback during the execution phase in this way, it should be noted that doing so is by no means sufficient to support the conclusion that such feedback *actually* controls the timing of voice onset in the case of normal uninterrupted pointing. The present experiment should therefore be treated as an attempt to define the bounds of interaction; feedback from gesture to speech in normal uninterrupted pointing will certainly not go beyond the temporal limits found in this experiment, nor is it likely to be of greater magnitude.

It is thought that the results of the experiment will also have a bearing on the *mechanism* of gesture execution, and the role that feedback plays within this system. Several models have been proposed, attempting to explain how the motor system achieves control of a movement such as that involved in gesture. According to the *impulse timing model,* planning of the movement consists in the preparation of a temporally organized string of nerve impulses which is transmitted to the musculature of arm and hand when the movement is performed. In terms of the *feedback model* the difference between the desired and actual state of the movement is visually or proprioceptively monitored and a control signal fed back to the innervatory mechanism, which maintains its output to the muscles until the error becomes negligible. The third view, which is embodied in the so-called *mass-spring model* proposes that the apex position is defined by a state of equilibrium which is determined by the ratio of torques between agonist and antagonist muscles. Once these parameters have been set, no further central control or feedback is necessary for the musculature to realize the corresponding position of the

limb. The foregoing alternatives are extensively discussed by Schmidt and McGown (1980), who also present experimental evidence for mass-spring control in a situation not unlike the pointing task of the present experiment.

Method

Apparatus. The experiment involved modification of the load characteristics of the gesturing arm in the course of its movement. This requirement was fulfilled by a piece of apparatus which basically consisted of a suspended mass attached by means of a cord running over a system of pulleys to the subject's wrist. After passing over the pulleys the cord was brought to the vicinity of the subject's wrist (in the rest position) by means of an eye (situated 17 centimeters "to the south" and 11 centimeters "to the east" of the push button, see Figure 1). As a consequence, the force exerted by the mass, when opposed by the arm, acted in a direction which, for movements to the near LED, was horizontal, and roughly parallel to the *y*-axis, and for movements to the far LED was at about 45 degrees (clockwise) to the *y*-axis. The amount of slack which was taken up in the cord before the arm encountered resistance could be varied in steps of about 1.5 centimeters over the whole range of movements. The force needed to just set the mass in motion was about 1600 grams. There was a small residual force of 140 grams made up of the weight of the cord and the frictional resistance of the pulley system, which had to be overcome to just set the system in motion when no mass was applied.

Subjects. Fourteen subjects, ten male and four female, participated in the experiment. All were right handed.

Procedure. There were four experimental conditions. Two of these involved loaded movements in which the load was applied at either the beginning of the movement (LB) or halfway through it (LHW). As a control condition, and also in order to reduce, as far as possible, the extent to which subjects were prepared for a load to be applied, 10 trials under each of the loaded movement conditions were combined with 20 in which no load at all was applied (NL). The order of presentation was randomized over the three conditions within a single block. Throughout the running of this block the subject's wrist remained connected to the load application system, and consequently, movements in the No-Load condition were in fact opposed by the residual force of 140 grams mentioned above.

To provide a basis for comparison with previous experiments, subjects also performed a series of trials without the load application system connected. These free movement (FM) trials were presented in two blocks of 20 trials each, all involving right-handed movements in the right field, 10 to the near LED and 10 to the far LED. The two loaded movement blocks were alternated with these two free movement blocks, half the subjects beginning with one type, and half with the other. For those subjects who were presented with the free movement block first, there were 4 practice trials preceding it, and then a further 8 practice trials preceding the first loaded movement block (2 LB, 2 LHW, and 4 NL trials). For those who performed the loaded movement blocks first, there were just 8 practice trials preceding that block.

In order to accommodate variations in the span of gestural movement from one subject to another, it was necessary to incorporate a means of adjusting the operation of the load application apparatus for each individual, so as to ensure that the load was applied at the correct point in the movement. The load application point could be adjusted by means of a sliding beam which altered the path length of the cord, and thus the amount of slack to be taken up before the resistance of the mass was felt. For the Load Beginning condition, LB, the sliding beam was adjusted to give 3 centimeters of slack, that is, 3 centimeters of free movement from the eye to the point

at which the resistence to the load was encountered. This provision made it impossible for the subject to detect, simply from the tension of the cord when the hand was in the rest position, whether a load was about to be applied or not. (Subjects were instructed to keep the wrist against the eye.) In the case of the Load Halfway condition, LHW, separate calibrations were carried out for movements to near and far LEDs. With the cord attached, but with the apparatus set to give unlimited free movement, the subject was requested to make, and hold, a gestural movement to the LED concerned, the distance moved by the wrist from its rest position was measured and the setting of the apparatus was determined for which the point of load application was exactly halfway to the movement extremum. In practice, the distance covered by the wrist under load application conditions was less than when unloaded, though the difference was only of the order of 1 centimeter.

In all conditions subjects were instructed to respond as quickly as possible to the LED being turned on, and to use the deictic expression "dat lampje" ("that light") for both the near and the far LED. One of the experimenters sat across the table, and "noted down" each response.

Results

Considering first the condition which most closely resembled those of the previous experiments, namely, that involving free movements (FM), an analysis of variance of the values of T_A, T_I, and T_V gave results which were also similar to the earlier ones. Apex time T_A was longer (by 55 milliseconds) for the far LED than for the near LED ($F(1,13) = 23.7$, $p < .001$), and there was a corresponding increase in voice onset time, T_V, of 47 milliseconds ($F(1,13) = 15.4$, $p < .01$). Thus, as was found for the ipsilateral field in earlier experiments, there was almost complete adaptation of voice onset to apex. There was also agreement with earlier findings in respect to movement initiation time which was longer (by 24 milliseconds) for the far LED than for the near one ($F(1,13) = 26.4$, $p < .001$). Thus the results of the FM condition replicate those of earlier experiments in all essential respects.

Before discussing the results pertaining to the "loaded" conditions NL, LB, and LHW we will illustrate what effect load application had on the trajectory of the gesture. Figure 7 gives six plots from one subject, showing the movement trajectories to near and far LEDs in the three "loaded" conditions.

Though these particular plots are quite characteristic of the patterns of movement displayed by other subjects as well, there was also considerable variation between subjects, particularly in the case of the LHW condition, where the point of load application was individually calibrated. The graphs for the LB and LHW condition clearly show the sudden effect of load on the trajectory of the movement. In order to allow a precise analysis of the experimental results, we developed a method of estimating from the Selspot traces the moment at which the force was applied. For this purpose the acceleration/deceleration graphs and the displacement graphs of each individual gesture were computed, for each of the three spatial coordinates. The moment of load application could be determined by visual inspection of these curves, which showed characteristic abrupt changes when the weight was applied. The reliability of this method was computed by comparing the independent determinations by two judges (the second and third author) for a set of 80 traces (from two subjects, to near and far LEDs, in the LB and LHW conditions). Winer's (1961) reliability coefficient was $r = .85$.

The analysis over all trials of all subjects in the LB and LHW conditions revealed that the experimental manipulation of load application had worked as intended. For the Load Beginning condition the mean ap-

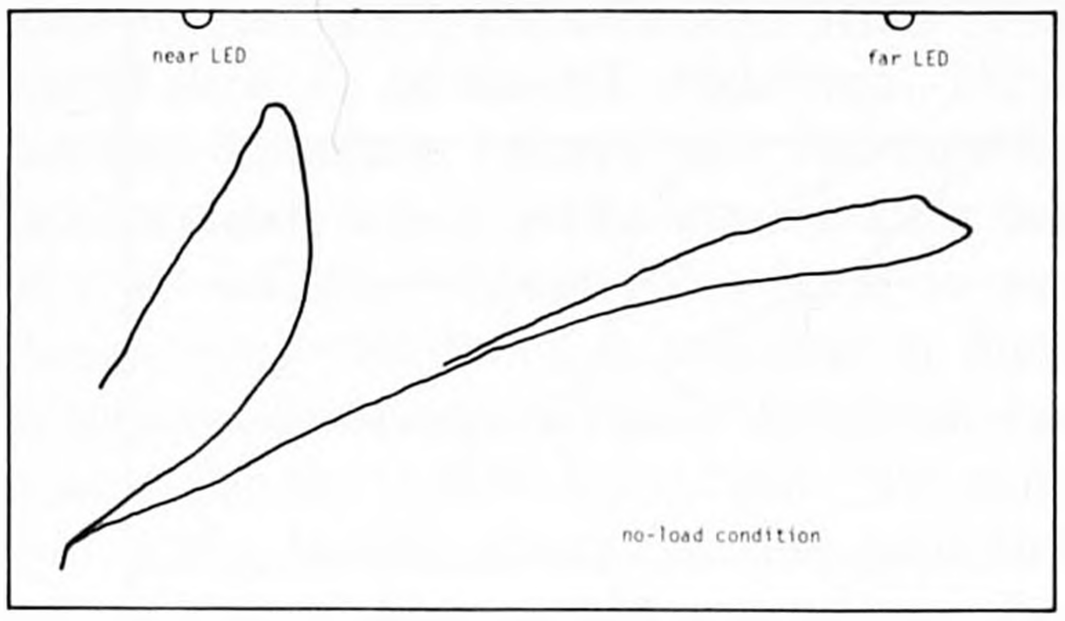

FIG. 7. Experiment 4: Trajectories of a single subject's right hand pointing gestures to near and far referent LEDs in the ipsilateral field for the No-Load, Load Beginning, and Load Half-Way conditions.

plication moments from movement initiation were 89 and 91 milliseconds for the near and the far LED; for the Load Halfway condition these numbers were 143 and 207 milliseconds, respectively. The difference between the latter two values is due to the individual calibration applied.

The main purpose of the present experiment was to determine whether, in the course of *execution* of the gesture, feedback can affect the moment of speech onset, and if so, within what time frame such feedback operates. Some data bearing on this question are presented in Figure 8.

The figure shows that apex times increased in going from the NL to LB to LHW condition, that is, the later the retarding force was applied, the later the apex was reached. Up to a point speech onset did follow the delay of movement in the Load Beginning condition (LB), but it failed to do so in the Load Halfway condition (LHW). A series of analyses of variance were carried out on the loaded movement data, with T_I, T_A, and T_V as dependent variables. The first of these showed that movement initiation time T_I was the same under all three conditions, NL, LB, and LHW, but as in the FM condition, pointing gestures to the far LED were initiated later than those to the near LED (by 22 milliseconds, $F(1,13) = 14.2$, $p < .01$). With respect to apex time T_A, the effect of load condition was such that, in going from the NL to LB to LHW condition, there was a progressive and significant increase in magnitude, the respective values being 714, 826, and 870 milliseconds ($F(2,26) = 69.7$, $p < .0001$). Moreover, all three pairwise comparisons were significant at the 0.001 level. The three load conditions showed a variation in speech onset T_V, which just failed to reach significance ($F(2,26) = 3.2$, $p = .058$). The average voice onset times for NL, LB, and LHW were 776, 816, and 796 milliseconds, respectively. Further pairwise analyses of these values showed that the 40-millisecond increase in T_V in going from NL to LB was significant ($F(1,13) = 7.7$, $p < .05$), but that neither of the other two comparisons were. The same pattern emerges when we take voice onset from movement initiation $T_V' = T_V - T_I$, as the dependent variable; there was a significant increase amounting to 33 milliseconds in going from the NL to LB condition ($F(1,13) = 5.6$, $p < .05$), but not with respect to either of the other pairwise comparisons. In other words, one may conclude that there is some adaptation of speech onset to the prolongation of gesture execution when the load is applied near the beginning of the gesture, but not when the

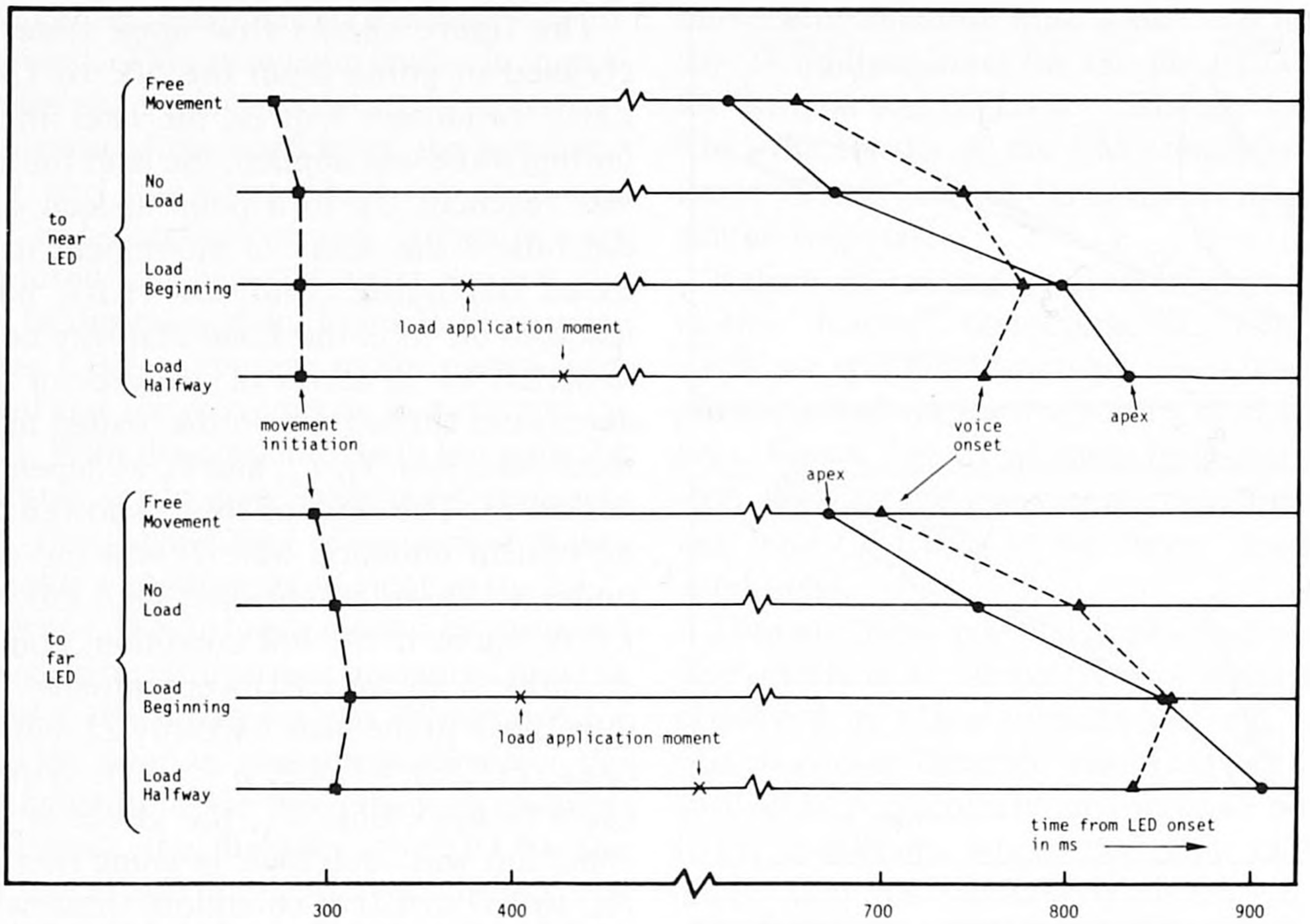

FIG. 8. Experiment 4: The relative timing of movement initiation, load application, apex of gesture, and voice onset in referring to near and far LEDs under four experimental conditions: Free Movement, No-Load, Load Beginning, and Load Half-Way.

retarding force is applied halfway through the movement. In the latter case there was apparently insufficient time for feedback to come into effect.

There could, of course, be a simple explanation of the finding that feedback did not come into play in the LHW condition, namely, that speech had already been released by the time the load was applied. Inspection of the data, however, showed that this order of events occurred in only 5 out of 140 cases. Hence, the conclusion can be that there exists a point in time before the onset of speech beyond which feedback can no longer influence the course of speech timing. Furthermore, this juncture lies somewhere between the points in time at which the load is applied in the LB and LHW conditions. In other words, there is a "dead" period just before the onset of speech where the system is blind to information from the gesture system, that is, where the speech system operates in ballistic fashion. The purpose of the following analysis is to estimate the average extent of this period.

How much leeway did the subject have in the LB condition from the moment of load application to the planned moment of speech onset? Apparently enough to adapt the latter to some degree after the load was sensed. We do know the moment of load application for each gesture. But we do not know the moment for which voice onset was originally planned or projected by the subject. However, it is possible to estimate this moment when we assume that on a No-Load trial, voice onset does indeed occur at the projected moment. The average voice onset time T_V when referring to a LED in the NL condition is, therefore, an unbiased estimate of the projected voice onset time in a loaded movement trial, since the subject cannot foresee whether on a given trial a load will be applied or not. Accordingly, the leeway from load application to projected moment of speech can be estimated by subtracting the load appli-

cation time in the LB condition from voice onset time T_V in the NL condition. These leeway times are, on average, 369 and 404 milliseconds when indicating the near and the far LEDs, respectively. And, as was found, some feedback is possible in these intervals. But no evidence of feedback was found for the LHW condition. We computed the leeway times for this condition according to the same procedure. The values are 215 and 296 milliseconds for pointing to the near and far LEDs, respectively.

We may conclude that the "dead" period before speech onset within which feedback from the gesture system is no longer possible has an average value between 300 and 370 milliseconds. Detailed analyses per subject confirmed this estimate; it also confirmed that most subjects adapted better on trials where they had more leeway.

The above findings may be summarized as follows. Speakers attempt to adapt voice onset to the apex of their gesture, even when the latter is impeded in the course of execution. A significant amount of adaptation was achieved in the Load Beginning condition, where the execution of the gesture was, on the average, prolonged by about 110 milliseconds and the corresponding shift in voice onset was about 40 milliseconds. Hence, the degree of adaptation was some way short of the theoretical optimum. No significant adaptation occurred in the Load Halfway condition. The less time there is available between load application and the projected moment of voice onset, the less successful is the subject in adapting voice to apex. The minimal time a subject needs for adapting voice onset is, on the average, between 300 and 370 milliseconds. Given the fact that the average latency from movement initiation to voice onset was 363 milliseconds in the Free Movement condition of the present experiment and 358, 321, and 361 milliseconds in the comparable conditions of Experiments 1 through 3, one can conclude that, when the pointing gesture is unhampered, speech becomes ballistic almost immediately upon the initiation of gesture. For these more natural situations, therefore, the ballistic theory must be close to accurate.

The Control of Gesture Execution

Finally, the results are evaluated in relation to the question of how the motor system controls the execution of a gestural movement. As was mentioned above, three alternative models may be considered, namely, the impulse timing model, the feedback model, and the mass-spring model. If the system operates in accordance with an *impulse timing model* one would expect a loaded gesture to be less extensive than a free gesture, because the preplanned motor program is considered to contain a complete description of muscle innervation as it develops over time. The outcome of this motor program in terms of movement extent and duration will depend on the resistance encountered by the moving limb. The same program will not carry the limb as far when a retarding force is applied as when it is unimpeded. It should be noted, however, that the *duration* of the gesture would be expected to be about the same in the free as the loaded condition, because the duration of muscular activity is predetermined by the program.

The *feedback model* predicts that apex position will be no different for loaded and free movements, since motor execution is assumed to continuously adapt to incoming visual and/or kinesthetic information. Moreover, a loaded gesture would be expected to be of longer duration than a free gesture, because it would require additional motor activity to reach the target position.

Finally, in the case of the *mass-spring model,* one would expect to find a difference in both apex position and timing as between free and loaded gestures, since the torques of the muscles involved will be affected by the load. As a result, the preset point of equilibrium will be reached earlier in the loaded than in the free condition, and

the limb will have traveled a shorter distance.

An evaluation of the three models was carried out by analyzing the movement execution times and the distances traveled by the pointing finger in the Load and the No-Load conditions. These data are summarized in Table 3.

Consider first the pattern of execution times T_E. There are large and significant increases in T_E, of the order of 20–40%, in going from the No-Load to the Load Beginning and Load Halfway conditions. This observation holds for both the near and far referent LEDs, and rules out the impulse timing model, which predicts no effect of load on execution time.

It should be possible to determine which of the two remaining models, the feedback and the mass-spring model, holds on the basis of the distance data. As may be seen from Table 3, the distance traveled was hardly affected by load, a finding which would be consistent with the assumption of the feedback but not the mass-spring model. There were, nevertheless, slight effects of load on distance traveled which should not be ignored. The extent of gestures to the near LED was significantly less (t test, $p < .01$), by 11 millimeters, in the Load Beginning condition than in the No-Load condition. When the comparison was with the Load Halfway condition the corresponding difference was 17 millimeters ($p < .01$). At the same time there were no significant differences in distances traveled to the far LED. In other words, the mass-spring model does not obtain, at least in a pure form, for gestures to the far LED, though it would appear to have some validity for gestures to the near LED. The latter are, apparently, less open to the influence of feedback than gestures to the far LED, presumably because there is less time for corrective action to be taken during a short gesture than during a long one.

There is a further point to be considered in relation to this issue. The distance traveled by the limb may not be the only relevant parameter in evaluating the relative merits of the mass-spring model. In executing a loaded movement the pointing finger may, as we found, travel the same distance as in the nonloaded case, but still arrive at a different position in space. Such an outcome would indicate that feedback has not been completely effective. Table 4 shows the degree to which the apex positions of loaded movements deviate from the corresponding positions in the No-Load condition. The table gives the mean displacement in the x and y directions, that is, the extent to which the apex position in the loaded condition has shifted to the right of and away from the subject in relation to the apex position in the corresponding No-Load condition. It also gives the apex displacements "as the crow flies," that is, in terms of the Pythagorean sum of the x- and y-deviations. It should be noted that these latter values (and the x/y-values on which they are based) were computed by subject and condition; as a result, the average values *over* subjects in the table do *not*

TABLE 3

EXPERIMENT 4: MEAN GESTURE EXECUTION TIME AND DISTANCE TRAVELED IN NO-LOAD, LOAD BEGINNING, AND LOAD HALFWAY CONDITIONS

Condition	No-Load		Load Beginning		Load Halfway	
LED	Near	Far	Near	Far	Near	Far
Execution time (in ms)	393	450	512**	541**	549**	602**
Distance traveled (in mm)	251	502	240*	498	234*	496

* Difference with respect to corresponding No-Load condition significant at $p < .01$ level; ** $p < .001$ level.

show the Phythagorean relation to the average x- and y-values.

The table shows that there were highly significant deviations "as the crow flies" for all four Load/LED conditions, the apex positions shifting under load by as much as 1.5 to 2.5 centimeters. These shifts were largest in the Load Halfway condition, and their direction depended on which LED was indicated. Thus for the near LED they were toward the subject, that is, the distance traveled was smaller, as already seen in Table 3. For the far LED the displacement was away from the subject and to the left, that is, "northwest" from the unloaded apex position. These patterns can also be observed in the sample traces of Figure 7.

For present purposes it is not necessary to analyse the foregoing results in more detail; it simply suffices to note that the significant displacements of apex position in the loaded conditions suggest that performance conforms with the mass-spring rather than the feedback model. That there is at least *some* feedback, however, is evident from the fact that the deviations are smaller the earlier the load is applied. This should not make any difference for the mass-spring model. The data apparently result from some combination of mass-spring and feedback processes. Finally, it is of interest to observe that where feedback is most in evidence, namely, in the case of the far LED in the Load Beginning condition, we also find the greatest degree of adaptation of voice onset to apex (see Figure 8). It is tempting to conclude that timely feedback enables a speaker to produce voicing and gesture in a coordinated way.

Discussion

The general objective of the present study was to gain more insight into processes underlying the coordination between speech and gesture. In particular, the question was addressed as to how far the execution of speech and gesture can be regarded as being organized in a modular or ballistic fashion. Do the motor systems controlling gesture and speech operate as Leibnizian monads, their observed synchrony simply being the consequence of a harmony preestablished in the planning phase? Or is it rather the case that the two systems, windows open, interact throughout the planning and execution of a coordinated action? The first alternative we termed the *ballistic theory*, the second one the *interaction theory*.

The study was limited to an analysis of deictic expressions, in which a demonstrative term such as "this" or "that" was produced in conjunction with a pointing gesture to the intended referent. The experimental procedure called for the speaker to

TABLE 4
Experiment 4: Deviation of Apex Position in Loaded Conditions from Apex Position in No-Load Condition (in cm)

Condition	Load Beginning		Load Halfway	
LED	Near	Far	Near	Far
Deviation along x-axis	−0.41	−0.61	−1.14	−1.45
Standard deviation	1.02	1.48	1.00	2.01
Sign. level (t test)	<.20	<.20	<.01	<.05
Deviation along y-axis	−1.16	0.73	−1.34	1.39
Standard deviation	0.70	0.89	0.71	1.24
Sign. level (t test)	<.001	<.02	<.001	<.01
Deviation as the crow flies	1.59	1.62	2.02	2.57
Standard deviation	0.66	1.02	0.68	1.69
Sign. level (t test)	<.001	<.001	<.001	<.001

indicate to the listener which of a set of lights was momentarily illuminated, by pointing to the light and/or by saying "this light" or "that light." Detailed analyses of the timing of voice onset, movement initiation, and apex of gesture in four experiments demonstrate, first, a degree of synchronization between voicing and pointing; for an extended gesture, such as that made to a relatively distant target, or one in the speaker's contralateral field, speech onset occurs later than in the case of a gesture to a conveniently located nearby target. The delay in speech onset, however, is absent when the same target is indicated without hand gesture. Synchronization is apparently achieved in a particular manner; speech adapts to gesture, but gesture is only marginally affected by speech. This marginal effect is, moreover, found only at the point of movement initiation. The pointing movement is initiated slightly earlier in the absence of speech or in a situation where each target is indicated by the same verbal expression (e.g., "that light") by comparison with one where the targets are distinguished (e.g., "this light"/"that light"). There is, apparently, some competition for common resources between speech and gesture systems at the stage just prior to movement initiation.

The experimental findings show, second, that once the pointing movement had been initiated, gesture and speech operate in almost modular fashion. Neither a variation in the number of verbal alternatives (one versus two), nor the complete absence of speech, affects the execution of the pointing motion once it has been initiated. The gesture has become ballistic. There is, nevertheless, evidence to suggest that feedback from gesture to speech *can* come into play during the first milliseconds of gesture execution when one tests the limits of the system. By retarding the arm immediately after movement initiation, partial adaptation of voice onset to apex can be observed, amounting to about 30% of the interval by which the execution phase is prolonged. However, when the gesture is retarded at its halfway point, adjustment of speech timing is no longer possible. Thus at some moment prior to speech onset, the speech system becomes ballistic as well. This point in time shows considerable individual variation, but is estimated to occur, on average, between 300 and 370 milliseconds prior to the projected time of voice onset. Given that the average latency from gesture initiation to voice onset is around 350 milliseconds in the unhampered on-line conditions of Experiments 1 through 4, the temporal window within which feedback from gesture to speech can come into play is thus quite small or nonexistent. It is, therefore, doubtful whether in the normal unimpeded case any feedback is operative during the phase of movement execution.

Given these findings, the Leibnizian view turns out to be very nearly correct. The normal case appears to be that *speech and deictic gesture are interactive in the planning phase, but well-nigh ballistic in the execution phase*.

It should be emphasized that we did not make the general claim that gesturing itself is insensitive to feedback during its execution. The gesture data suggested that some visual and/or kinesthetic feedback occurs during the execution of motion. More precisely, the mass-spring model alone cannot give an adequate account of performance in a situation where the gesture is retarded immediately after movement initiation. It is worth noting that Jeannerod and Biguer (1981; see also Jeannerod, 1981) found a very similar state of affairs when they studied the time course of grasping movements. They consist of an arm reaching and a hand grasping component; each of these components are sensitive to visual feedback, but they do not interact during execution.

Finally, some comments are in order concerning the extent to which these findings may be generalized. The situation investigated in this study was highly restricted given the general question of how speech

and gesture are synchronized. A first restriction was our deliberate choice of a class of referring expressions which could be expected to exhibit a high degree of synchrony with the accompanying gesture. The findings on the whole confirmed this expectation and, moreover, showed that synchronization was largely established in the planning phase. Would the same be true when the relationship between the two processes is less direct as, for instance, in the case of Ekman and Friesen's (1969) "illustrators"? In such cases precise synchronization of gesture and speech is probably of less consequence for communicative effectiveness than in the case of deixis, where speech and gesture are more tightly coordinated in the interest of drawing the interlocutor's attention to a particular referent. There would be no reason to expect a greater degree of interaction between the speech and gesture systems in cases of this indirect kind, and in particular not during the execution phase of movement.

A second restriction of the present study concerned the nature of the experimental task. There is a large variety of situations in which a speaker can elect to make a deictic gesture. From the point of view of synchronization of gesture and speech, an important distinction is between situations in which a speaker indicates a transitory event as soon as it occurs (e.g., the traffic light turning green, the train arriving, or a person appearing), and situations in which reference is made to a more permanent target. The results of Experiment 1, which attempted to simulate these two types of situations (the "on-line" and the "off-line") suggest that this distinction is not of great consequence, at least in relation to the question of how far voice succeeds in adapting to gesture for targets at different locations.

A third restriction in the present tasks is that reference is made to a single target only. There are, however, situations where multiple deictic reference is made (e.g., "here and there"). How is the programming of multiple pointing movements organized in such cases? Are the parameters of the complex pointing gesture set before the motion is released? They are unlikely to be, particularly when there are several referents ("here, and there, and there, and there"), in which case a gesture may be planned during the execution of the previous one, so that planning and execution run "in tandem," as it were. It would be premature to generalize the present findings to these or other complex cases of deixis. And the same can be said with respect to complex iconic gestures (cf. McNeill, 1981), where one part of the gesture can relate to one word and another part to another word in the utterance.

References

Arbib, M. (1981). Perceptual structures and distributed motor control. In S. R. Geiger (Ed.), *Handbook of physiology. Sect. 1, The nervous system, Part 2*. Bethesda: American Physiological Society.

Brown, D. A. (1969). Advanced methods for the calibration of metric cameras. *Proceedings 1969 Symposium on Computational Photogrammetry. Syracuse, New York*. Falls Church: American Society of Photogrammetry.

Condon, W. S., & Ogston, W. D. (1971). Speech and body motion synchrony of the speaker–hearer. In D. L. Horton & J. J. Jenkins (Eds.), *Perception of language*. Columbus, Ohio: Merrill.

Efron, D. (1972). *Gesture, race and culture*. The Hague: Mouton. (Reprinted from *Gesture and environment*, 1941, New York: King's Crown Press).

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 49–98.

Fodor, J. A. (1983). *The modularity of mind. An essay on faculty psychology*. Cambridge, Mass: MIT Press.

Freedman, N. (1972). The analysis of movement behavior during the clinical interview. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication*. New York: Pergamon.

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.

Jeannerod M. (1981). Intersegmental coordination during reaching at natural visual objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, N.J.: Erlbaum.

Jeannerod, M., & Biguer, B. (1981). Visuomotor

mechanisms in reaching within extrapersonal space. In D. Ingle, M. Goodale, & R. Mansfield (Eds.), *Advances in the analysis of visual behavior*. Boston: MIT Press.

KENDON, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal communication and language*. The Hague: Mouton.

MCNEILL, D. (1979). *The conceptual basis of language*. Hillsdale, N.J.: Erlbaum.

MCNEILL, D. (1981). Action, thought and language. *Cognition*, **10**, 201–208.

SCHMIDT, R. A., & MCGOWN, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. *Journal of Motor Behavior*, **12**, 149–161.

WINER, B. J. (1961). *Statistical principles in experimental design*. New York: McGraw–Hill.

WOLTRING, H. J. (1980). Planar control in multicamera calibration for 3-D gait studies. *Journal of Biomechanics*, **13**, 39–48.

(Received February 2, 1984)
(Revision received August 23, 1984)