Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 3, 615–622

0278-7393/92/$3.00

Word Familiarity Predicts Temporal Asynchrony of Hand Gestures and Speech

Palmer Morrel-Samuels and Robert M. Krauss
Columbia University

Seventeen Ss were videotaped as they provided narrative descriptions of 13 photographs. Judgments from 129 naive untrained Ss were used to isolate 60 speech-related gestures and their lexical affiliates (i.e., the accompanying word or phrase judged as related in meaning) from these 221 narratives. A computer–video interface measured each gesture, and a 3rd group of Ss rated word familiarity of each lexical affiliate. Multiple regression revealed that gesture onset preceded voice onset by an interval whose magnitude was inversely related to the lexical affiliate's rated familiarity. The lexical affiliate's familiarity was also inversely related to gesture duration. Results suggest that difficulty encountered during lexical access affects both gesture and speech. Familiarity's systematic relations with gesture–speech asynchrony and gesture duration make it unlikely that speech and gesture are produced independently by autonomous modules.

When people talk they often move one or both hands. Casual observers presume that these arm and hand movements—the gestures that accompany spontaneous speech—are related in meaning to what is being said. Yet, despite their ubiquity and the long-standing fascination such behaviors have held for scientists and humanists alike, the functional significance of gesturing for the speaker and the nature of the relation between gesture and speech are far from clear. If speech-related gestures are merely an atavistic vestige of the evolutionary process that resulted in speech (Condillac, 1756/1974; Hewes, 1973; Tylor, 1878) or are essentially random movements that serve to dissipate muscular tension (Dittman & Llewelyn, 1969) and coordinate the speech articulators (Hadar, Steiner, & Rose, 1984), then one should expect that gesture production is essentially independent of the cognitive processes associated with speech.[1] Such a view is consistent with Fodor's (1983, 1985) notion of modularity and requires that speech and gesture be produced by autonomous modules. Alternatively, as some have speculated, gesture production and speech production may be interdependent: Gestures may help a speaker retrieve elusive words from memory (DeLaguna, 1927; Ekman & Friesen, 1972; Feyereisen, 1987; Freedman, 1972; Moscovici, 1967; Werner & Kaplan, 1972), or both gesture and speech may be generated from a single mental representation (McNeill, 1985, 1987).

In this article, we focus on one of the most conspicuous features of speech and speech-related gestures: the temporal coordination between the gesture and the word or phrase that seems to share its meaning. There is general agreement that gestures anticipate speech: Gesture and speech are coordinated temporally such that gesture initiation typically precedes speech onset of the *lexical affiliate,* the word or phrase that accompanies the gesture and seems related to its meaning (Schegloff, 1984). This coordination is thought to reflect either the common origin of gesture and speech (Butterworth & Beattie, 1978; Feyereisen, 1987; Kendon, 1980, 1983; Krauss & Morrel-Samuels, 1988; McNeill, 1985, 1987; Schegloff, 1984) or the autonomy of the independent modules dedicated to their respective production (Levelt, Richardson, & La Heij, 1985).

Temporal Coordination of Speech and Gesture

Levelt et al. (1985) have contrasted two theories of temporal interdependence: an interactive theory positing that feedback

This article is based on a doctoral dissertation submitted to Columbia University by Palmer Morrel-Samuels. The research was made possible by a grant from the National Science Foundation to Robert M. Krauss (BNS 86-16131), by a Biomedical Research Support Grant from Columbia University, and by a Graduate Fellowship from the Psychology Department.

We gratefully acknowledge the comments and suggestions of the members of the dissertation committee, Julian Hochberg and Walter Mischel, as well as the assistance of Christine Colasante, Janna Stubbelfield, Brooks Bays, and Gloria Weaver. The helpful comments of Lois Putnam, Wayne Wickelgren, David McNeill, and two anonymous reviewers were also appreciated.

Correspondence concerning this article should be addressed to Palmer Morrel-Samuels, who is now at EDS Center for Advanced Research, 2001 Commonwealth Boulevard, Ann Arbor, Michigan 48105. Electronic mail may be sent to PalmerMS@UMICHUM.

[1] We use *speech-related gesture* or merely *gesture* to denote those hand movements whose meaning is related to that of the accompanying spontaneous speech. In previous work, approximate synonyms have been illustrators (Ekman & Friesen, 1972), gesticulations (Kendon, 1983), speech-focused movements (Butterworth & Beattie, 1978), referential gestures (Cicone, Wapner, Foldi, Zurif, & Gardner, 1979), and iconix (McNeill, 1985), to name but a few. These are distinguished from gesture language signs—also called emblems (Ekman & Friesen, 1972), autonomous gestures (Kendon, 1983), formal pantomimic gestures (Wiener, Devoe, Rubinow, & Geller, 1972), and gestural signs (Herman, Morrel-Samuels, & Pack, 1990)—whose form and meaning are established by convention so that they can substitute for speech. As we use the term, speech-related gestures are a subcategory of general hand movements and are distinct from gestural signs that supplant speech and from self-adaptors—the acts of contact such as adjusting clothes and scratching (Ekman & Friesen, 1972) that have no apparent relation to the meaning of the accompanying speech.

between gesture and speech allows interdependence throughout motor planning and execution and a ballistic theory based on notions of modularity (Fodor, 1983, 1985), which holds that "the two systems are independent during the phase of motor execution, the temporal parameters having been pre-established in the planning phase" (Levelt et al., 1985, p. 133). Focusing on deictic gestures, gestures that point to an object or a location (Ekman & Friesen, 1972), Levelt et al. conducted an experiment in which subjects pointed to one illuminated light in an array of either two or four light-emitting diodes. While pointing, subjects said nothing, said "this light," or said "this light" when pointing to nearby targets and "that light" when pointing to distant targets. Levelt et al. found that gesture duration was independent of both the number of lights in the array and the number of potential verbal expressions. Although some of their findings were consistent with the interactive view (e.g., silent pointing was initiated faster than voiced pointing), Levelt et al. concluded that the data offered general support for the ballistic view and suggested that deictic gestures and speech are produced by autonomous modules.

However, temporal coordination need not be similar for deictic gestures and speech-related gestures. The two differ in several respects. First, unlike most speech-related gestures, deictic gestures are likely to convey more information than their lexical affiliates (e.g., saying "this one" without pointing can be far less informative than pointing and saying nothing). Second, problems of lexical access in deixis are minimal. There are only a small number of potential deictic expressions, and their frequency of occurrence in the language is high. Although it may be, as Levelt et al. (1985) contended, that deictic gestures and their lexical affiliates are independent during execution, there is little basis for generalizing this conclusion from deictic gestures to speech-related gestures.

According to McNeill (McNeill, 1981, 1985, 1987; McNeill & Levy, 1982), speech-related gestures and lexical affiliates are temporally coordinated because they have a shared computational stage. He (McNeill, 1987) claimed that

> gestures and sentences share a stage of development before speech takes on its socially constituted form; and gestures exhibit this stage without distortion because they lack direct social regulation. Whereas in the case of spoken sentences the morphological and phonological levels stand between mental processes and speech, in the case of gestures there are no such levels that alter the display of mental processes. (p. 210)

On this view, gestures anticipate speech because they do not require morphological and phonological processing.[2] McNeill (1987) went on to say that "the anticipation is very slight, but still significant, for it shows unquestionably that the speaker's thinking is initially in imagistic form" (p. 212).

Although we agree with McNeill (1987) that gesture and speech have a common origin and that they do precede the speech with which they are associated, we differ as to the locus of the common origin and the significance of that temporal coordination. We do not agree that the tendency for gestures to anticipate speech constitutes incontrovertible evidence that the speaker's thinking is imagistic; although we appreciate the analogy between gestures and pictures, that perceived similarity can hardly be taken as proof that cognitive processes themselves are imagistic. In our view, gestures derive not from images (as McNeill would have it) but from a system of motoric representations of concepts, many of which also come to be reflected in speech. Moreover, we propose that the common origin of gesture and speech is at the presemantic level of the speaker-gesturer's communicative intention. According to this view, the communicative intention activates both an abstract propositional representation—a representation that is processed to furnish an utterance (Levelt, 1989)—and a motoric representation that may be reflected in a gesture.[3] Typically, the motoric representation that is activated by the communicative intention will be of a concept (or some elements of a concept) that is expressed in the accompanying speech. Despite psychoanalytic claims others have made to the contrary (e.g., Mahl, 1979), we assume that on most occasions the motoric representation will be more rudimentary (i.e., that it will contain fewer attributes of the communicative intention) than the semantic representation.

However, it is important to note that it need not always be the case that motoric representations are less differentiated than semantic representations. The words that constitute an utterance are a reflection of the speaker's attempt to convey lexically the abstract concepts contained in the underlying proposition. For many reasons (e.g., the real-time processing demands of speech production in conversation, the language's failure to provide an adequate lexical item, and momentary lapses of memory) the individual may be unable (a) to access a lexical entry that precisely expresses the concept represented in the proposition or (b) to construct such an expression out of multiple lexical elements. On such occasions the motoric representation may more adequately convey the sense of the communicative intention (or one specific aspect of that sense). For example, a speaker who wants to characterize someone's public mourning as affectedly exaggerated may be temporarily unable to access *lugubrious* and may substitute for it a word such as *mournful* that expresses only part of the intended sense. On such occasions a gesture might express the intended sense or some aspect of that intended sense (e.g., its melodramatic affectation) more adequately than the word it accompanies. (See Zajonc & Markus, 1982, 1984, for a model of affect and motor representation that is consistent with this example.) Similarly, although words and numerical terms can quantify size, velocity, and shape with a high degree of precision, if a speaker gestures while using a general term (e.g., *curved* to describe a parabola), then the gesture may be more informative than the lexical affiliate. Nevertheless, it seems justifiable to assume that semantic representations are more precisely differentiated than the motoric representations used

[2] The argument that gesturing anticipates speech because the latter requires linguistic processing and the former does not seems to assume that gestures do not require processing to become articulate movements—as though images could be translated directly into hand movements. We find such an assumption implausible. Although gestures, unlike sentences, may be noncompositional, their production still requires an executable motor program.

[3] Although it is not relevant to the present discussion, we also believe that communicative intentions can activate representations in other modalities as well (e.g., acoustic, somatic, and olfactory).

to generate speech-related gestures. The grounds for such an assumption are clear: Given a normal adult lexicon of 20,000 or more distinct entries, it is hard to argue that words are inherently less precise than gestures.

Accordingly, we propose that asynchrony in the onsets of speech and gesture results from the fact that the concepts gestures represent are typically less differentiated than those represented in their lexical affiliates. Hence, we would expect that gesture initiation typically would be apparent before voice onset of the lexical affiliates because search times would be shorter for the former than the latter. Moreover, we would expect that the difference in their onset times (i.e., the size of the interval by which the gesture precedes speech) would be a function of the accessibility of the lexical affiliate: The more accessible the lexical affiliate, the closer its onset would be to that of the accompanying gesture.

Although different indices of accessibility furnish contradictory results (Gernsbacher, 1984), word familiarity often predicts performance on recognition tasks requiring lexical assess (Johns & Swanson, 1988; McCann, Besner, & Davelaar, 1988; Narine & Widner, 1988; Paivio & O'Neill, 1970). For example, subjects have longer response latencies when they are required to access unfamiliar words from memory to complete a word recognition task (Gernsbacher, 1984). A similar effect operates during recall: Pauses in spontaneous speech typically precede unusual words that cannot be easily guessed from the written transcript (Goldman-Eisler, 1958, 1968), presumably because speakers pause as they seek unusual and unfamiliar words in memory. Butterworth and Beattie (1978) claimed that speech-related gestures are initiated during such production pauses (Henderson, Goldman-Eisler, & Skarbek, 1966), an observation that is consistent with the interactive view proposed above. If our assumptions regarding speech and gesture production are justified, and if it is warranted to use familiarity as an index of accessibility, then temporal coordination should be a function of the lexical affiliate's rated familiarity. Specifically, we should see unfamiliar words being accompanied by greater asynchrony between gesture initiation and speech onset.

Method

Eliciting and Measuring Hand Movements

Materials

Eleven undergraduates rated 20 color slides depicting a variety of subjects (e.g., landscapes, abstract art, buildings, people, and machines) on 11 semantic differential scales (Jenkins, Russell, & Suci, 1958; Osgood, Suci, & Tannenbaum, 1957) selected to represent the three major semantic differential factors: Evaluation, Activity, and Potency. From these ratings, a mean score for each slide on each of the three dimensions was computed. On the basis of these scores, we chose as stimuli 13 slides whose ratings were well distributed along each of the three major semantic differential factors.

Subjects

Seventeen undergraduates (9 men and 8 women) were recruited to take part in an experiment on communication. Their participation (as well as that of the other subjects described below) fulfilled a course requirement. Two additional subjects were excluded from the study because they correctly guessed the experiment's purpose during a postsession interview.

Procedure

The 13 stimulus photographs were each projected for 70 s in random order onto a screen 18 ft away; the subject viewed each slide and described it to a female confederate positioned so that she could see the subject but not the projection screen. During these narratives the subject was videotaped by two cameras, one providing a front view and the other a side view; both views were combined side-by-side on one screen and were accompanied by an audible audio track. From the videotape, a computer–video interface (Morrel-Samuels & Krauss, 1990) was used to identify the onset and termination of every hand movement that traveled 1.5 inches or more in 0.8 s.[4] The same system furnished measurements of duration, velocity, and spatial extent for all hand movements exceeding the criterion. Tests showed these measures had adequate reliability and validity (Morrel-Samuels & Krauss, 1990).

Apparatus

Videotapes were made using two video cameras (Panasonic Model WV240P) positioned on perpendicular axes 18 ft from the subject. Their output was fed to a special effects generator (Panasonic Model WJ–540A) by which the two images were juxtaposed and fed to a U-matic format videocassette recorder (Sony VO–2800) along with the output of a video frame numberer (Berkely Viatronics) that generated a visible number on each video frame. A low-impedance microphone (Beyer dynamic) was used for recording the audio track. Timing pulses generated by an Apple II+ microcomputer were recorded on one of the videotape's audio channels using an optimized Basic program (Signal Maker) written for this system and based on a similar application (Krauss, Morrel-Samuels, & Hochberg, 1988).

Hand movements were measured from the videotape by an optimized compiled Basic program (Video Interface) written for this application; it used a parallel interface card (John Bell Engineering Model JBE 6522) that functioned as a timer. Movements were measured by tracking the video image of the speaker's hands (in real time) with a computer-generated cursor controlled by a joystick (Kraft Model D4). Tracking was done twice for each hand, once for the side view and once for the front view. A schematic drawing of the equipment configuration and additional details of the system can be found elsewhere (Morrel-Samuels & Krauss, 1990).

Selecting and Analyzing Speech-Related Gestures

Materials

The 221 videotaped narratives (17 subjects describing 13 slides) furnished a corpus of 2,328 hand movements that exceeded a threshold of 1.5 inches per 0.8 s. Verbatim transcripts were made of each

[4] This criterion was chosen because it represented a reasonable compromise given the resolution of our video equipment and the constraints of our computer's memory. Smaller gestures and movements would have been difficult to see under the interface's cursor, and a shorter sampling interval would have required the collection of more XYZ points than the computer's memory could accommodate.

narrative; they included hesitations, repairs, filled pauses, and other dysfluencies but were punctuated to facilitate reading.

Subjects

A total of 129 untrained naive subjects (77 men and 52 women) participated in the study. Subjects understood their task was to identify speech-related gestures in the videotaped narratives.

Procedure

Identifying gestures and lexical affiliates. Subjects worked in groups of 10, positioned so they could see and hear the videomonitor without seeing each other. Seventeen sessions were held, and at each session, narratives from more than one speaker were presented. Some subjects participated in more than one session. At the beginning of each session the process by which the narratives were generated was described, and the original 13 slides were displayed briefly. The 10 subjects were then given transcripts of the narrative they were about to see and were told that their task was to underline places in their transcript where the meaning of a word or phrase was related to the meaning of the gesture. No definition of "related to" was provided, and subjects who asked were instructed to use whatever criterion seemed appropriate to them. They then watched and listened to the narratives, with the videotape being stopped at major phrase boundaries. The marked transcripts were compiled, and any word or phrase selected by 8 or more of the judges was counted as a potential lexical affiliate. A total of 193 of these were identified in the 221 narratives.

Attributing lexical affiliates to gestures. The original videotapes were then reviewed independently by a research assistant and one of the authors (P. Morrel-Samuels). From the pool of potential lexical affiliates, 116 were attributed to a unilateral hand movement by both judges; of these 116, 80 satisfied the criterion of being attributed to one and only one gesture that was measured in its entirety.[5] To assess the accuracy of these measurements, both judges furnished estimations of duration and spatial extent for each of these 80 gestures. Because a few gestures seemed to be undermeasured or overmeasured by the computer interface, and because a power analysis (Cohen & Cohen, 1983) suggested that only 60 stimuli would be necessary to avoid a Type II error, the final phase of the experiment was restricted to the 60 gestures having the smallest discrepancy between estimated extent and measured extent. For these 60 gestures there was a high correlation between estimated spatial extent and measured extent, $r(60) = .83$, $p < .0001$; independent estimations and measurements of duration were also highly correlated, $r(60) = .96$, $p < .0001$.

Evaluating the lexical affiliate. Seventeen additional subjects (8 women and 9 men) read each lexical affiliate and rated its familiarity on a 1- to 7-point scale ranging from *familiar* to *unfamiliar* (Gernsbacher, 1984). In 28 cases the lexical affiliate contained more than one word (e.g., *in front, right side, come up, floating around, off her shoulders*); the lexical affiliate was a single word (e.g., *covered, protruding, partitioned, edge, foggy,* or *surrounding*) in 32 cases.

The moment of voice onset for each lexical affiliate was identified independently by the 2 judges from slow-motion analysis of the narratives' soundtrack. Onset times were indexed to the incrementing numbers that appeared on each video frame, and to the timing pulses created by the computer interface, so that gestures and speech could be specified by a common time base.

Results

The primary dependent variable is what we call *gesture–speech asynchrony*: the difference between the onset times of

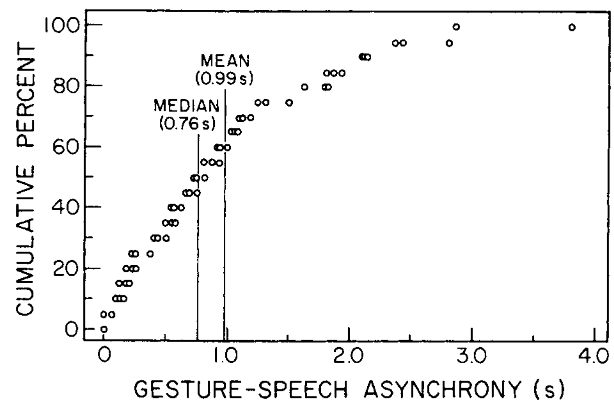

Figure 1. Cumulative distribution of asynchronies for 60 gesture-lexical affiliate pairs. (A positive value indicates the gesture preceded the speech.)

the gesture and its lexical affiliate. The larger the asynchrony, the greater the interval by which the gesture precedes the lexical affiliate. Asynchrony ranged from 0 to 3.8 s in a positively skewed distribution ($M = 0.99$ s; $SD = 0.83$ s; $Mdn = 0.76$; Kolomogorov's $D = 0.14$, $p < .01$; skewness = 1.14; kurtosis = 1.14). Consistent with earlier research (Butterworth & Beattie, 1978; Schegloff, 1984), no negative values were observed—that is, no gesture was preceded by its lexical affiliate (Figure 1).

To test the hypothesis that gesture–speech asynchrony varies systematically with word accessibility, we calculated a multiple regression that included the gesture's spatial extent and the number of syllables as covariates, with mean rated familiarity of the lexical affiliate as the predictor variable.[6] Thirty percent of the variance in gesture–speech asynchrony was accounted for by this model, $F(3, 56) = 7.89$, $p < .0002$. Only spatial extent and rated familiarity made significant unique contributions after all other effects were partialed out (see Table 1). A leverage plot (or partial residual plot) shows the relation that exists between familiarity and asynchrony when all other effects are partialed out (see Figure 2).

When the duration of the 60 speech-related gestures was entered as the dependent variable in an otherwise identical

[5] Specifically, the sample excluded 12 gestures that continued after the narrative's termination or traveled out of the cameras' fields of view, 20 gestures whose measurements merged with those of a contiguous self-adaptor (i.e., a hand movement to scratch the skin or adjust clothing; Ekman & Friesen, 1972), and 4 gestures containing a moment of inactivity long enough to require formating as two distinct motions. All bilateral gestures were excluded because the measurements from each hand might be nonidentical, a fact that would impede or prevent analysis and interpretation.

[6] Spatial extent was treated as a covariate on the grounds that the structure and function of small gestures may be distinct from those of large gestures (Ekman & Friesen, 1972). Although number of syllables can have little effect on memory search rate in some conditions (Clifton & Tash, 1973), we follow Church's (1987) recommendation that it be partialed from statistical analyses on the grounds that polysyllabic words tend to be less familiar (Zipf, 1935) and may entail unique demands during speech planning.

Table 1
Multiple Regression Analysis of Gesture–Speech Asynchrony and Gesture Duration

	Gesture–speech asynchrony				Gestural duration			
Variable	B value	*SE*	Partial *F*	*p*	B value	*SE*	Partial *F*	*p*
Number of syllables	−.133	.081	2.70	*ns*	−.007	.021	0.12	*ns*
Spatial extent	.027	.006	21.28	<.0001	.008	.002	30.74	<.0001
Rated familiarity	−.356	.172	4.28	<.05	−.142	.044	10.33	<.01

Note. For all *F* tests, $df = 1, 56$; $n = 60$. *SE* = standard error.

multiple regression, the model accounted for 42% of the variance, $F(3, 56) = 13.47$, $p < .0001$. A leverage plot (Figure 3) shows the relation between the two variables after covariates are partialed out. Again, both spatial extent and rated familiarity made unique contributions to variance in the dependent variable, as is shown in the right-hand columns of Table 1. It is not surprising, given this pattern of results, that gesture duration and gesture–speech asynchrony are highly correlated, $r(60) = .71$, $p < .0001$: The larger the interval by which a gesture precedes speech, the longer the duration of the gesture.[7] Inspection of the scatterplot of the two variables (Figure 4) reveals that durations are virtually always larger than their associated asynchronies; note that all but four of the points lie above the unit line. On average, gestures terminated 1.5 s after the initiation of their lexical affiliates (*SD* = 0.97 s). The mean asynchronous interval for the four atypical gestures was −0.14 s and the largest asynchrony (−0.26 s) was relatively small. These data show that the gestures in this corpus rarely terminate before speech onset of the lexical affiliate.

Discussion

Our data indicate that gestures are synchronized with speech and that they are initiated before or simultaneously with (but not after) the onset of their lexical affiliates. The magnitude of the gesture–speech asynchrony decreases with the affiliate's familiarity: The more familiar the lexical affiliate, the smaller the asynchrony. The familiarity of the lexical affiliate is also related to the gesture's duration: The more familiar the lexical affiliate, the more brief the duration of the associated gesture.

The finding that the lexical affiliate's judged familiarity predicts both gesture–speech asynchrony and gestural duration supports earlier claims that gesture production is tied to features of lexical access (Butterworth & Beattie, 1978; Kendon, 1972, 1975; McNeill, 1987). In such previous work no systematic attempt was made to account for the variability in asynchrony. Butterworth and Beattie (1978) invoked Hick's (1952) law to explain the temporal precedence of gestures; they reasoned that if the mental lexicon contains 20,000 to 30,000 entries and the set of potential gestures is much smaller, then searching the repertoire of gestures will require less time than searching the mental lexicon. However, Butterworth and Beattie's explanation fails to account for our data: If words take longer to access simply because there are more of them, then the magnitude of the latency between gestures and words should be driven entirely by stochastic processes and should include at least a small number of negative values. Such probabilistic models cannot easily account for the fact that the asynchrony of these independently selected gestures varied systematically with familiarity and the fact that those gestures never began after, and rarely terminated before, word onset. For modular models to accommodate these findings, a set of additional assumptions would be necessary. For example, modular models might adopt the assumptions that speech planning always requires more time than gesture planning and that gestures are the functional equivalent of filled pauses (such as "er" and "uhm") that can be terminated at will after lexical selection and speech articulation have run their course. However, because gestures can involve a great number of muscle groups between the shoulder girdle and the phalanges, and because speech-related gestures must necessarily have semantic content attributed to them (unlike filled pauses), neither assumption seems justified here. Moreover, even if these or similar assumptions can be invoked to account for our evidence of temporal coordination, such models would still be faced with the need to explain why gesture duration

Figure 2. Leverage plot showing partialed relationship between gesture–speech asynchrony and mean rated familiarity. (The ordinate plots the residuals for asynchrony when predicted by all independent variables except familiarity; the abscissa plots the residuals for familiarity when predicted by those same variables. The least-squares regression line is also plotted.)

[7] In the preceding multiple regression, as recommended for temporal variables (Kenny, 1979), duration was submitted to a log transform. For the zero-order correlation, untransformed data were used.

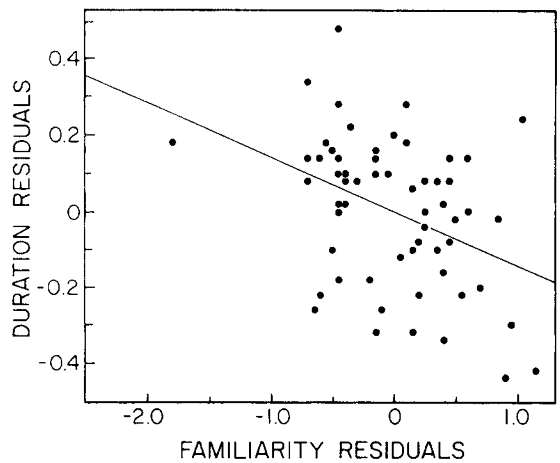

Figure 3. Leverage plot showing partialed relationship between gesture duration and mean rated familiarity. (The ordinate plots the residuals for gesture duration when predicted by all independent variables except familiarity; the abscissa plots the residuals for familiarity when predicted by those same variables. The regression line is also plotted.)

should vary systematically with the familiarity of the lexical affiliate.

The fact that the duration of a gesture also is a function of its lexical affiliate's familiarity is consistent with the hypothesis that these behaviors aid in the retrieval of elusive words from lexical memory. Also consistent with this notion is the substantial correlation between duration and gesture–speech asynchrony. If gestures aid word finding, one would expect them not to be terminated before lexical access had been achieved. Of course, it is impossible to ascertain the precise point at which a sought-after lexical item has been found, but it necessarily must be before it is spoken. Hence, it is relevant that almost all of the gestures in our corpus had durations greater than the asynchronous interval, even for cases in which the intervals were 3 s or longer. In only 4 out of 60 instances did the gesture terminate before articulation of the lexical affiliate had begun.

As was noted earlier, the possibility that gestures play a facilitative role in speech production has been suggested by several theorists past and present. However, thus far, empirical support for the hypothesis has been weak. According to Schlauch (1936), Dobrogaev, a Soviet psychologist, found that preventing speakers from gesturing resulted in decreased fluency, impaired articulation, and reduced vocabulary size, but insufficient detail is presented to evaluate the claim. Rimé, Schiaratura, Hupet, and Ghysselinckx (1984) found that the vividness of imagery in speech decreased when a speaker's hands, arms, legs, feet, and head were immobilized; effects on other aspects of speech production were negligible. There is also suggestive evidence that speakers gesture more when they are constrained to use an unfamiliar language (Marcos, 1976, 1978; Sainsbury & Wood, 1977). On the other hand, Graham and Heywood (1975) reported that preventing speakers from gesturing had no consequential effects on their speech. Given the limited power of their test (only 6 speakers were sampled) and the fact that some reliable differences in fluency were found (viz., under some conditions, subjects who were prevented from gesturing paused a greater proportion of the total speaking time), their conclusion strikes us as premature.

In our view, the hypothesis that gestures do facilitate speech production is at least plausible. As a motoric representation of a concept (or an aspect of a concept) that the speaker is attempting to express verbally, the gesture may serve as a cross-modal prime that helps the speaker access a lexical entry expressing the intended sense. If gestures did serve such a priming function, we would expect them to precede the words with which they are associated and we would also expect them not to be terminated before the lexical affiliate had been accessed. As was indicated earlier, the gestures in our corpus conform to both of these expectations. General support for such an hypothesis comes from work showing that gestures can function as retrieval cues for subjects attempting to recall previously presented lexical affiliates (Woodall & Folger, 1985), although caution is warranted because the effect does not necessarily extend to tests of recognition that require distinguishing one gesture from another (Krauss, Morrel-Samuels, & Colasante, 1991). Nevertheless, lack of facilitation during a recognition test in which subjects see gestures produced by another speaker does not rule out facilitation for the actual speaker during active speech production.

The relation between gestural duration and the familiarity of the lexical affiliate poses a problem for the ballistic view of gesture and speech production (Levelt et al., 1985), at least insofar as speech-related gestures are concerned. If gesture and speech were independent during the execution phase, a gesture's duration would have to be determined before execution by either a subsystem or an autonomous module. As our data indicate, not only are gestures initiated before their lexical affiliates, but it is not typical for a gesture to terminate before articulation of the lexical affiliate has begun. Thus, to

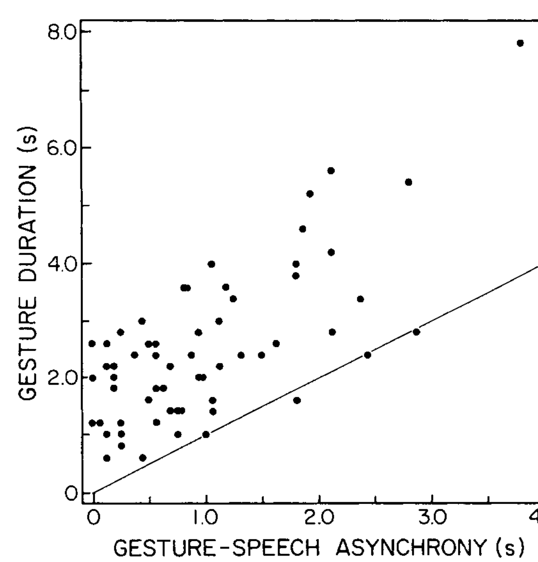

Figure 4. Gesture duration plotted against gesture–speech asynchrony. (The unit line is also shown.)

plan a gesture of sufficient duration, the speaker would have to know in advance how long lexical access will take. For this to occur without feedback between the two modules, one would have to assume either that long-duration gestures are associated with unfamiliar words or that the speaker somehow can predict and accommodate the amount of time lexical access will require. Neither assumption strikes us as particularly plausible. Even if we were to adopt one of these unlikely alternatives, neither alone would provide a satisfactory explanation for the systematic relation that word familiarity has with both gesture–speech asynchrony and gesture duration. In our view it is reasonable to accept the underlying assumptions adopted here and to interpret the data as coherent support for an interactive model linking speech production and gesture production.

References

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. T. Smith (Ed.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 344-360). New York: Plenum Press.

Church, K. W. (1987). Phonological parsing and lexical retrieval. Spoken word recognition [Special issue]. *Cognition, 25,* 53–69.

Cicone, M., Wapner, W., Foldi, N., Zurif, E., & Gardner, H. (1979). The relation between gesture and language in aphasic communication. *Brain and Language, 8,* 324–349.

Clifton, C., & Tash, J. (1973). Effect of syllabic word length on memory-search rate. *Journal of Experimental Psychology, 99,* 231–235.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlational analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Condillac, E. B. de (1974). *An essay on the origin of human knowledge: Being a supplement to Mr. Locke's essay on the human understanding.* New York: AMS Press. (Original work published 1756).

DeLaguna, G. (1927). *Speech: Its function and development.* New Haven, CT: Yale University Press.

Dittman, A. T., & Llewelyn, L. G. (1969). Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology, 23,* 283–292.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication, 22,* 353–374.

Feyereisen, P. (1987). Gestures and speech, interactions and separations: A reply to McNeill (1985). *Psychological Review, 94,* 493–498.

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

Fodor, J. A. (1985). Precis of the modularity of mind. *The Behavioral and Brain Sciences, 8,* 1–42.

Freedman, N. (1972). The analysis of movement behavior during the clinical interview. In A. W. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 153–175). Elmsford, NY: Pergamon Press.

Gernsbacher, M. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity, orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology, 10,* 96–106.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* San Diego, CA: Academic Press.

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5,* 185–189.

Hadar, U., Steiner, T., & Rose, C. (1984). The relationship between head movements and speech dysfluencies. *Language and Speech, 27,* 333–342.

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech, 9,* 207–216.

Herman, L., Morrel-Samuels, P., & Pack, A. (1990). Bottlenosed dolphin and human recognition of veridical and degraded video displays of an artificial gestural language. *Journal of Experimental Psychology: General, 119,* 215–230.

Hewes, G. (1973). Primate communication and the gestural origins of language. *Current Anthropology, 14,* 5–24.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11–26.

Jenkins, J., Russell, W., & Suci, G. (1958). An atlas of semantic profiles for 360 words. *American Journal of Psychology, 71,* 688–699.

Johns, E., & Swanson, L. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 180–190.

Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–216). Elmsford, NY: Pergamon Press.

Kendon, A. (1975). Gesticulation, speech and the gesture theory of language origin. *Sign Language Studies, 9,* 349–373.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Relationship of verbal and nonverbal communication* (pp. 207–227). The Hague, The Netherlands: Mouton.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, CA: Sage.

Kenny, D. (1979). *Correlation and causality.* New York: Wiley.

Krauss, R. M., & Morrel-Samuels, P. (1988, February). *Some things we do and don't know about hand gestures.* Paper presented at the meeting of the American Association for the Advancement of Science, Boston, MA.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61,* 743–754.

Krauss, R., Morrel-Samuels, P., & Hochberg, J. (1988). Videologger: A computerized multichannel event recorder for analyzing videotapes. *Behavior Research Methods, Instruments, & Computers, 20,* 37–40.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

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

Mahl, G. (1979). Body movement, ideation, and verbalization during psychoanalysis. In S. Weitz (Ed.), *Nonverbal communication: Readings with commentary* (2nd ed., pp. 168–181). New York: Oxford University Press.

Marcos, L. (1976). Bilinguals in psychotherapy: Language as an emotional barrier. *American Journal of Psychotherapy, 30,* 552–560.

Marcos, L. (1978). Hand movements in relation to the encoding process in bilinguals. *Dissertation Abstracts International, 38,* 5276B.

McCann, R., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 14,* 693–706.

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

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

McNeill, D. (1987). *Psycholinguistics: A new approach.* New York: Harper & Row.

McNeill, D., & Levy, E. (1982). Conceptual representations in language activity and speech. In R. Jarvella & W. Klein (Eds.), *Speech, place and action* (pp. 271–295). New York: Wiley.

Morrel-Samuels, P., & Krauss, R. (1990). Cartesian analysis: A computer-video interface for measuring gestures without physical contact. *Behavior Research Methods, Instruments & Computers, 22,* 466–470.

Moscovici, S. (1967). Communication processes and the properties of language. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* San Diego, CA: Academic Press.

Narine, J., & Widner, R. (1988). Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 694–699.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Paivio, A., & O'Neill, B. (1970). Visual recognition thresholds and dimensions of word meaning. *Perception and Psychophysics, 8,* 273–275.

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue level. *Motivation and Emotion, 8,* 311–325.

Sainsbury, P., & Wood, E. (1977). Measuring gesture: Its cultural and clinical correlates. *Psychological Medicine, 7,* 63–72.

Schegloff, E. (1984). On some gestures' relation to speech. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversational analysis* (pp. 266–296). Cambridge, England: Cambridge University Press.

Schlauch, M. (1936). Recent Soviet studies in linguistics. *Science and Society, 1,* 157.

Tylor, E. B. (1878). *Researches into the early history of mankind.* London: John Murry.

Werner, H., & Kaplan, B. (1972). *Symbol formation.* New York: Wiley.

Wiener, M., Devoe, S., Rubinow, S., & Geller, J. (1972). Nonverbal behavior and nonverbal communication. *Psychological Review, 79,* 185–214.

Woodall, W., & Folger, J. (1985). Nonverbal cue context and episodic memory: On the availability and endurance of nonverbal behaviors as retrieval cues. *Communication Monographs, 52,* 319–333.

Zajonc, R. B., & Markus, H. (1982). Affective and cognitive factors in preferences. *Journal of Consumer Research, 9,* 123–131.

Zajonc, R. B., & Markus, H. (1984). Affect and cognition: The hard interface. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotion, cognition, and behavior* (pp. 73–102). Cambridge, England: Cambridge University Press.

Zipf, G. (1935). *The psycho-biology of language: An introduction to dynamic philology.* Boston: Houghton Mifflin.

Received March 29, 1991
Revision received October 17, 1991
Accepted October 21, 1991 ■