NONVERBAL BEHAVIOR AND NONVERBAL COMMUNICATION:

WHAT DO CONVERSATIONAL HAND GESTURES TELL US?

ROBERT M. KRAUSS, YIHSIU CHEN, AND PURNIMA CHAWLA

Columbia University

This is a pre-editing version of a chapter that appeared in M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389-450). San Diego, CA: Academic Press..

NONVERBAL BEHAVIOR AND NONVERBAL COMMUNICATION:

WHAT DO CONVERSATIONAL HAND GESTURES TELL US?

ROBERT M. KRAUSS, YIHSIU CHEN, AND PURNIMA CHAWLA
Columbia University

1. THE SOCIAL PSYCHOLOGICAL STUDY OF NONVERBAL BEHAVIOR

1.1 Nonverbal behavior as nonverbal communication

Much of what social psychologists think about nonverbal behavior derives from a proposal made more than a century ago by Charles Darwin. In *The expression of the emotions in man and animals* (Darwin,1872), he posed the question: Why do our facial expressions of emotions take the particular forms they do? Why do we wrinkle our nose when we are disgusted, bare our teeth and narrow our eyes when enraged, and stare wide-eyed when we are transfixed by fear? Darwin's answer was that we do these things primarily because they are vestiges of *serviceable associated habits* — behaviors that earlier in our evolutionary history had specific and direct functions. For a species that attacked by biting, baring the teeth was a necessary prelude to an assault; wrinkling the nose reduced the inhalation of foul odors; and so forth.

But if facial expressions reflect formerly functional behaviors, why have they persisted when they no longer serve their original purposes? Why do people bare their teeth when they are angry, despite the fact that biting is not part of their aggressive repertoire? Why do they wrinkle their noses when their disgust is engendered by an odorless picture? According to Darwin's intellectual heirs, the behavioral ethologists (e.g., Hinde, 1972; Tinbergen, 1952), humans do these things because over the course of their evolutionary history such behaviors have acquired communicative value: they provide others with external evidence of an individual's internal state. The utility of such information generated evolutionary pressure to select sign behaviors, thereby schematizing them and, in Tinbergen's phrase, "emancipating them" from their original biological function.[1]

1.2 Noncommunicative functions of nonverbal behaviors

So pervasive has been social psychologists' preoccupation with the communicative or expressive aspects of nonverbal behaviors that the terms nonverbal behavior and nonverbal communication have tended to be used

---

[1]See Fridlund (1991) for a discussion of the ethological position.

interchangeably.[2] Recently, however, it has been suggested that this communicative focus has led social psychologists to overlook other functions such behaviors serve. For example, Zajonc contends that psychologists have been too quick to accept the idea that facial expression are primarily expressive behaviors. According to his "vascular theory of emotional efference" (Zajonc, 1985; Zajonc, Murphy, & Inglehart, 1989) , the actions of the facial musculature that produce facial expressions of emotions serve to restrict venous flow, thereby impeding or facilitating the cooling of cerebral blood as it enters the brain. The resulting variations in cerebral temperature, Zajonc hypothesizes, promote or inhibit the release of emotion-linked neurotransmitters, which, in turn, affect subjective emotional experience. From this perspective, facial expressions do convey information about the individual's emotional state, but they do so as an indirect consequence of their primary, noncommunicative function.

An analogous argument has been made for the role of gaze direction in social interaction. As people speak, their gaze periodically fluctuates toward and away from their conversational partner. Some investigators have interpreted gaze directed at a conversational partner as an expression of intimacy or closeness (cf., Argyle & Cook, 1976; Exline, 1972; Exline, Gray, & Schuette, 1985; Russo, 1975) . However, Butterworth (1978) argues that gaze direction is affected by two complex tasks speakers must manage concurrently: planning speech, and monitoring the listener for visible indications of comprehension, confusion, agreement, interest, etc. (Brunner, 1979; Duncan, Brunner, & Fiske, 1979) . When the cognitive demands of speech planning are great, Butterworth argues, speakers avert gaze to reduce visual information input, and, when those demands moderate, they redirect their gaze toward the listener, especially at places where feedback would be useful. Studies of the points in the speech stream at which changes in gaze direction occur, and of the effects of restricting changes in gaze direction (Beattie, 1978; Beattie, 1981; Cegala, Alexander, & Sokuvitz, 1979) , tend to support Butterworth's conjecture.

1.3 Interpersonal and intrapersonal functions of nonverbal behaviors

Of course, nonverbal behaviors can serve multiple functions. Facial expression may play a role in affective experience—by modulating vascular blood flow as Zajonc has proposed or through facial feedback as has been suggested by Tomkins and others (Tomkins & McCarter, 1964)—and at the same time convey information about the expressor's emotional state. Such communicative effects could involve two rather different mechanisms. In the first place, many nonverbal behaviors are to some extent under the individual's control, and can be produced voluntarily. For example, although a smile may be a normal accompaniment of an affectively positive internal state, it can at least to some degree be produced at will. Social norms, called "display rules," dictate that one exhibit at least a moderately pleased expression on certain social occasions.

---

[2] For example, the recent book edited by Feldman and Rimé (1991) reviewing research in this ara is titled *Fundamentals of Nonverbal Behavior,* despite the fact that all of the nonverbal behaviors are discussed in terms of the role they play in communication (see Krauss (1993) .

Kraut (1979) found that the attention of others greatly potentiates smiling in situations that can be expected to induce a positive internal state. In the second place, nonverbal behaviors that serve noncommunicative functions can provide information about the noncommunicative functions they serve. For example, if Butterworth is correct about the reason speakers avert gaze, an excessive amount of gaze aversion may lead a listener to infer that the speaker is having difficulty formulating the message. Conversely, the failure to avert gaze at certain junctures, combined with speech that is overly fluent, may lead an observer to infer that the utterance is not spontaneous.

Viewed in this fashion, we can distinguish between *inter*personal and *intra*personal functions that nonverbal behaviors serve. The interpersonal functions involve information such behaviors convey to others, regardless of whether they are employed intentionally (like the facial emblem) or serve as the basis of an inference the listener makes about the speaker (like dysfluency). The intrapersonal functions involve noncommunicative purposes the behaviors serve. The premise of this chapter is that the primary function of conversational hand gestures (unplanned, articulate hand movements that accompany spontaneous speech) is not communicative, but rather to aid in the formulation of speech. It is our contention that the information they convey to an addressee is largely derivative from this primary function.

2. GESTURES AS NONVERBAL BEHAVIORS

2.1 A typology of gestures

All hand gestures are hand movements, but not all hand movements are gestures, and it is useful to draw some distinctions among the types of hand movements people make. Although gestural typologies abound in the literature, there is little agreement among researchers about the sorts of distinctions that are necessary or useful. Following a suggestion by Kendon (1983) , we have found it helpful to think of the different types of hand movements that accompany speech as arranged on a continuum of lexicalization—the extent to whch they are "word-like." The continuum is illustrated in Figure 1.

---

insert Figure 1 about here

2.1.1 Adapters

At the low lexicalization end of the continuum are hand movements that tend not to be considered gestures. They consist of manipulations either of the person or of some object (e.g., clothing, pencils, eyeglasses)—the kinds of scratching, fidgeting, rubbing, tapping, and touching that speakers often do with their hands. Such behaviors are most frequently referred to as *adapters* (Efron, 1941/1972; Ekman & Friesen, 1969b; Ekman & Friesen, 1972) . Other terms that have been used are *expressive movements* (Reuschert, 1909) , *body-focused movements* (Freedman & Hoffman, 1967) , *self-touching gestures* (Kimura, 1976) , *manipulative gestures* (Edelman & Hampson, 1979) , *self manipulators* (Rosenfeld, 1966, and *contact acts* (Bull & Connelly, 1985) . Adapters are not gestures as that term is usually understood. They are not perceived as communicatively intended, nor are they perceived to be meaningfully related to the speech they accompany, although they may serve as the basis for dispositional inferences (e.g., that the speaker is nervous, uncomfortable, bored, etc.). It has been suggested that adapters may reveal unconscious thoughts or feelings (Mahl, 1956; Mahl, 1968) , or thoughts and feelings that the speaker is trying consciously to conceal (Ekman & Friesen, 1969a; Ekman & Friesen, 1974) , but little systematic research has been directed to this issue.

2.1.2 Symbolic gestures

At the opposite end of the lexicalization continuum are gestural signs—hand configurations and movements with specific, conventionalized meanings—that we will call *symbolic gestures* (Ricci Bitti & Poggi, 1991) . Other terms that have been used are *emblems* (Efron, 1941/1972) , *autonomous gestures* (Kendon, 1983), *conventionalized signs* (Reuschert, 1909), *formal pantomimic gestures* (Wiener, Devoe, Rubinow, & Geller, 1972), *expressive gestures* (Zinober & Martlew, 1985), and *semiotic gestures* (Barakat, 1973). Familiar symbolic gestures include the "raised fist," "bye-bye," "thumbs-up," and the extended middle finger sometimes called "flipping the bird." In contrast to adapters, symbolic gestures are used intentionally and serve a clear communicative function. Every culture has a set of symbolic gestures familiar to most of its adult members, and very similar gestures may have different meanings in different cultures (Ekman, 1976) . Subcultural and occupational groups also may have special symbolic gestures that are not widely known outside the group. Although symbolic gestures often are used in the absence of speech, they occasionally accompany speech, either echoing a spoken word or phrase or substituting for something that was not said.

2.1.3 Conversational gestures

The properties of the hand movements that fall at the two extremes of the continuum are relatively uncontroversial. However there is considerable disagreement about movements that occupy the middle part of the lexicalization continuum, movements that are neither as word-like as symbolic gestures nor as devoid of meaning as adapters. We refer to this heterogeneous set of hand movements as *conversational gestures.* They also have been called *illustrators* (Ekman & Friesen, 1969b; Ekman & Friesen, 1972, *gesticulations* (Kendon, 1980; Kendon, 1983) , and *signifying signs* (Reuschert, 1909) . Conversational gestures are hand movements that accompany speech, and seem related to the speech they accompany. This apparent relatedness is manifest in three ways: First, unlike symbolic gestures, conversational gestures don't occur in the absence of speech, and in conversation are made only by the person who is speaking. Second, conversational gestures are temporally coordinated with speech. And third, unlike adapters, at least some conversational gestures seem related in form to the semantic content of the speech they accompany.

Different types of conversational gestures can be distinguished, and a variety of classification schemes have been proposed (Ekman & Friesen, 1972; Feyereisen & deLannoy, 1991; Hadar, 1989a; McNeill, 1985). We find it useful to distinguish between two major types that differ importantly in form and, we believe, in function.

Motor movements

One type of conversational gesture consists of simple, repetitive, rhythmic movements, that bear no obvious relation to the semantic content of the accompanying speech (Feyereisen, Van de Wiele, & Dubois, 1988) . Typically the hand shape remains fixed during the gesture, which may be repeated several times. We will follow Hadar (1989a; Hadar & Yadlin-Gedassy, 1994) in referring to such gestures as *motor movements;* they also have been called “batons” (Efron, 1941/1972; Ekman & Friesen, 1972 and “beats” (Kendon, 1983; McNeill, 1987). Motor movements are reported to be coordinated with the speech prosody and to fall on stressed syllables (Bull & Connelly, 1985; but see McClave, 1994) , although the synchrony is far from perfect.

Lexical movements

The other main category of conversational gesture consists of hand movements that vary considerably in length, are nonrepetitive, complex and changing in form, and, to a naive observer at least, appear related to the semantic content of the speech they accompany. We will call them *lexical movements,* and they are the focus of our research.[3]

3. INTERPERSONAL FUNCTIONS OF CONVERSATIONAL GESTURES

3.1 Communication of semantic information

Traditionally, conversational hand gestures have been assumed to convey semantic information.[4] "As the tongue speaketh to the ear, so the gesture

---

[3]A number of additional distinctions can be drawn. Butterworth and Hadar (1989) and Hadar and Yadlin-Gedassy (1994) distinguish between two types of lexical movements: conceptual gestures and lexical gestures; the former originate at an earlier stage of the speech production process than the latter. Other investigators distinguish a category of *deictic* gestures that point to individuals or indicate features of the environment; we find it useful to regard them as a kind of lexical movement. Further distinctions can be made among types of lexical movements (e.g., *iconic* vs. *metaphoric* (McNeill, 1985; McNeill, 1987)), but for the purposes of this chapter the distinctions we have drawn will suffice.

[4]By semantic information we mean information that contributes to the utterance's "intended meaning"(Grice, 1969; Searle, 1969). Speech, of course, conveys semantic information in abundance, but also may convey additional information (e.g., about the speaker's emotional state, spontaneity, familiarity with the topic, etc.) through variations in voice quality, fluency, and other vocal properties. Although such information is not, strictly speaking, part of the speaker's intended meaning, it nonetheless may be quite informative. See Krauss and Fussell (in press) for a more detailed discussion of this distinction.

speaketh to the eye" is the way the 18th century naturalist Sir Francis Bacon (1891) put it. One of the most knowledgeable contemporary observers of gestural behavior, the anthropologist Adam Kendon, explicitly rejects the view that conversational gestures serve no interpersonal function—that gestures "…are an automatic byproduct of speaking and not in any way functional for the listener"— contending that

> ...gesticulation arises as an integral part of an individual's *communicative* effort and that, furthermore, it has a direct role to play in this process. Gesticulation…is important principally because it is employed, along with speech, in fashioning an effective utterance unit (Kendon 1983, p. 27, *Italics* in original).

3.1.1 Evidence for the "gestures as communication" hypothesis

Given the pervasiveness and longevity of the belief that communication is a primary function of hand gestures, it is surprising that so little empirical evidence is available to support it. Most writers on the topic seem to accept the proposition as self evident, and proceed to interpret the meanings of gestures on an *ad hoc* basis (cf., Birdwhistell, 1970).

The experimental evidence supporting the notion that gestures communicate semantic information comes from two lines of research: studies of the effects of visual accessibility on gesturing, and studies of the effectiveness of communication with and without gesturing (Bull, 1983; Bull, 1987; Kendon, 1983) . The former studies consistently find a somewhat higher rate of gesturing for speakers who interact face-to-face with their listeners, compared to speakers separated by a barrier or who communicate over an intercom (Cohen, 1977; Cohen & Harrison, 1972; Rimé, 1982) . Although differences in gesture rates between face-to-face and intercom conditions may be consistent with the view that they are communicatively intended, it is hardly conclusive. The two conditions differ on a number of dimensions, and differences in gesturing may be attributable to factors that have nothing to do with communication—e.g., social facilitation due to the presence of others (Zajonc, 1965). Moreover, all studies that have found such differences also have found a considerable amount of gesturing when speaker and listener could not see each other, something that is difficult to square with the "gesture as communication" hypothesis.

Studies that claim to demonstrate the gestural enhancement of communicative effectiveness report small, but statistically reliable, performance increments on tests of information (e.g., reproduction of a figure from a description; answering questions about an object on the basis of a description) for listeners who could see a speaker gesture, compared to those who could not (Graham & Argyle, 1975; Riseborough, 1981; Rogers, 1978) . Unfortunatey, all the studies of this type that we have found suffer from serious methodological shortcomings, and we believe that a careful assessment of them yields little support for the hypothesis that gestures convey semantic information. For example, in what is probably the soundest of these studies, Graham and Argyle had speakers describe abstract line drawings to a small audience of listeners who then tried to reproduce the drawings. For half of the descriptions, speakers were allowed to gesture; for the remainder, they were required to keep their arms folded. Graham and Argyle found that audiences of the non-gesturing speakers reproduced the figures somewhat less accurately. However, the experiment does not control for the possibility that speakers who were allowed to gesture produced better *verbal* descriptions of the stimuli, which, in turn, enabled their audiences to reproduce the figures more accurately. For more detailed critical reviews of this literature, see Krauss, Morrel-Samuels and Colasante (1991) and Krauss, Dushay, Chen and Rauscher (in press).

3.1.2 Evidence inconsistent with the "gestures as communication" hypothesis

Other research has reported results inconsistent with the hypothesis that gestures enhance the communicative value of speech by conveying semantic information. Feyereisen, Van de Wiele, and Dubois (1988) showed subjects videotaped gestures excerpted from classroom lectures, along with three possible interpretations of each gesture: the word(s) in the accompanying speech that had been associated with the gesture (the *correct response);* the meaning most frequently attributed to the gesture by an independent group of judges (the *plausible response);* a meaning that had been attributed to the gesture by only one judge (the *implausible response).* Subjects tried to select the response that most closely corresponded to the gesture's meaning. Not surprisingly the "plausible response" (the meaning most often spontaneously attributed to the gesture) was the one most often chosen; more surprising is the fact that the "implausible response" was chosen about as often as the "correct response."

Although not specifically concerned with gestures, an extensive series of studies by the British Communication Studies Group concluded that people convey information just about as effectively over the telephone as they do when they are face-to-face with their co-participants (Short, Williams, & Christie, 1976; Williams, 1977).[5] Although it is possible that people speaking to listeners they cannot see compensate verbally for information that ordinarily would be conveyed by gestures (and other visible displays), it may also be the case that the contribution gestural information makes to communication typically is of little consequence.

Certainly reasonable investigators can disagree about the contribution that gestures make to communication in normal conversational settings, but insofar as the research literature is concerned, we feel justified in concluding that the communicative value of these visible displays has yet to be demonstrated convincingly.

[5] More recent research has found some effects attributable to the lack of visual access (Rutter, Stephenson & Dewey, 1981; Rutter, 1987), but these effects tend to involve the perceived social distance between communicators, not their ability to convey information. There is no reason to believe that the presence or absence of gesture *per se* is an important mediator of these differences.

3.2 Communication of Nonsemantic Information

Semantic information (as we are using the term) involves information relevant to the intended meaning of the utterance, and it is our contention that gestures have not been shown to make an important contribution to this aspect of communication. However, semantic information is not the only kind of information people convey. Quite apart from its semantic content, speech may convey information about the speaker's internal state, attitude toward the addressee, etc., and in the appropriate circumstances such information can make an important contribution to the interaction. Despite the fact that two message are identical semantically, it can make a great deal of difference to passengers in a storm-buffeted airplane whether the pilot's announcement "Just a little turbulence, folks—nothing to worry about" is delivered fluently in a resonant, well-modulated voice or hesitantly in a high-pitched, tremulous one, (Kimble & Seidel, 1991).

It is surprising that relatively little consideration has been given to the possibility that gestures, like other nonverbal behaviors, are useful communicatively because of nonsemantic information they convey. Bavelas, Chovil, Lawrie, and Wade (1992) have identified a category of conversational gestures they have called *interactive gestures* whose function is to support the ongoing interaction by maintaining participants' involvement. In our judgment, the claim has not yet been well substantiated by empirical evidence, but it would be interesting if a category of gestures serving such functions could be shown to exist.

In Section 4 we describe a series of studies examining the information conveyed by conversational gestures, and the contribution such gestures make to the effectiveness of communicatiion.

4. GESTURES AND COMMUNICATION: EMPIRICAL STUDIES

Establishing empirically that a particular behavior serves a communicative function turns out to be a less straightforward matter than it might seem.[6] Some investigators have adopted what might be termed a *interpretive* or *hermeneutic* approach, by carefully observing the gestures and the accompanying speech, and attempting to infer the meaning of the gesture and assign a communicative significance to it (Bavelas et al., 1992; Birdwhistell, 1970; Kendon, 1980; Kendon, 1983; McNeill, 1985; McNeill, 1987; Schegloff, 1984) .

We acknowledge that this approach has yielded useful insights, and share many of the goals of the investigators who employ it; at the same time, we believe a method that relies so heavily on an investigator's intuitions can yield misleading results because there is no independent means of corroborating the observer's inferences. For a gesture to convey semantic information, there must be a relationship between its form and the meaning it conveys. In interpreting

---

[6]Indeed, the term *communication* itself has proved difficult to define satisfactorily (see (Krauss & Fussell, in press) for a discussion of this and related issues).

the gesture's meaning, the interpeter relates some feature of the gesture to the meaning of the speech it accompanies. For example, in a discussion we videotaped, a speaker described an object's position relative to another's as "…a couple of feet behind it, maybe oh [pause], ten or so degrees to the right." During the pause, he performed a gesture with palm vertical and fingers extended, that moved away from his body at an acute angle from the perpendicular. The relationship of the gesture to the conceptual content of the speech seems transparent; the direction of the gesture's movement illustrates the relative positions of the two objects in the description. However, direction was only one of the gesture's property. In focusing on the gesture's direction, we ignored its velocity, extent, duration, the particular hand configuration used— all potentially meaningful features— and selected the one that seemed to make sense in that verbal context. In the absence of independent corroboration, it's difficult to reject the possibility that the interpretation is a construction based on the accompanying speech that owes little to the gesture's form. Without the accompanying speech, the gesture may convey little or nothing; in the presence of the accompanying speech, it may add little or nothing to what is conveyed by the speech. For this reason, we are inclined to regard such interpretations as a source of hypotheses to be tested rather than useable data.

Moreover, because of differences in the situations of observer and participant, even if such interpretations could be corroborated empirically it's not clear what bearing they would have on the communicative functions the gestures serve. An observer's interpretation of the gesture's meaning typically is based on careful viewing and re-viewing of a filmed or videotaped record. The naive participant in the interaction must process the gesture on-line, while simultaneously attending to the spoken message, planning a response, etc. The fact that a gesture contained relevant information would not guarantee that it would be accessible to an addressee.

What is needed is an independent means of demonstrating that gestures convey information, and that such information contributes to the effectiveness of communication. Below we describe several studies that attempt to assess the kinds of information conversational gestures convey to naive observers and the extent to which gestures enhance the communicativeness of spoken messages.

4.1 The semantic content of conversational gestures

For a conversational gesture to convey semantic information, it must satisfy two conditions. First, the gesture must be associated with some semantic content; second, that relationship must be comprehensible to listeners. "Gestionaries" that catalog gestural meanings do not exist; indeed, we lack a reliable notational system for describing gestures in some abstract form. So is not completely obvious how one establishes a gesture's semantic content. Below we report three experiments that use different methods to examine the semantic content of gestures.

4.1.1 The semantic content of gestures and speech

One way to examine the semantic content of gestures is to look at the meanings naive observers attribute to them. If a gesture conveys semantic content related to the semantic content of the speech that accompanies it, the meanings observers attribute to the gesture should have semantic content similar to that of the speech. Krauss, Morrel-Samuels and Colasante (1991, Expt. 2) showed subjects videotaped gestures and asked them to write their impression of each gesture's meaning. We will call these *interpretations.* We then had another sample of subjects read each interpretation, and rate its similarity to each of two phrases. One of the phrases had originally accompanied the gesture, and the other had accompanied a randomly selected gesture.

The stimuli used in this and the next two experiments were 60 brief (M = 2.49 s) segments excerpted from videotapes of speakers describing pictures of landscapes, abstractions, buildings, machines, people, etc. The process by which this corpus of gestures and phrases were selected is described in detail in (Krauss et al., 1991; Morrel-Samuels, 1989; Morrel-Samuels & Krauss, 1992) , and will only be summarized here. Naive subjects, provided with transcripts of the descriptions, viewed the videotapes sentence by sentence. After each sentence, they indicated (1) whether they had seen a gesture, and (2) if they had, the word or phrase in the accompanying speech they perceived to be related to it. We will refer to the words or phrases judged to be related to a gesture as the gesture's *lexical affiliate.* The 60 segments whose lexical affiliates were agreed upon by 8 or more of the 10 viewers (and met certain other technical criteria) were randomly partitioned into two sets of 30, and edited in random order onto separate videotapes.

Six subjects (3 males and 3 females) viewed each of the 60 gestures, without hearing the accompanying speech, and wrote down what they believed to be its intended meaning. The tape was paused between gestures to give them sufficient time to write down their interpretation. Each of the 60 interpretations produced by one interpreter was given to another subject (judge), along with two lexical affiliates labeled "A" and "B." One of the two lexical affiliates had originally accompanied the gesture that served as stimulus for the interpretation; the other was a lexical affiliate that had accompanied a randomly chosen gesture. Judges were asked to indicate on a six point scale with poles labeled "very similar to A" and "very similar to B" which of the two lexical affiliates was closer in meaning to the interpretation.

On 62% of the trials (s.d. = 17%) judges rated the gesture's interpretation to be closer in meaning to its original lexical affiliate than to the lexical affiliate of another gesture. This value is reliably greater than the chance value of .50 ($t_{(59)} = 12.34$, $p < .0001$). We also coded each of the 60 lexical affiliates into one of four semantic categories: *Locations* (e.g., "There's another young girl to the woman's *right,"* "passing it *horizontally* to the picture"[7]); *Actions* (e.g., "rockets or bullets

---

[7] The italicized words are those judged by subjects to be related in meaning to the meaning of the gesture.

flying out," "seems like it's going to *swallow them up*"); *Objects* (e.g., "scarf or *kerchief* around her head", "actual *frame* of the window and the Venetian blind"); and *Descriptions* (e.g., "one of those *Pointillist* paintings," "which is *covered* with paper and books").[8] Accuracy varied reliably as a function of the lexical affiliate's semantic category ($F_{(3,56)} = 4.72$, $p < .005$). Accuracy was greatest when the lexical affiliates were Actions (73 percent), somewhat lower for Locations (66 percent) and considerably lower for Object Names and Descriptions (57 and 52 percent, respectively). The first two means differ reliably from 50 percent ($t_{(56)} = 5.29$ and 4.51, respectively, both ps < .0001); the latter two do not (ts < 1).

Gestures viewed in isolation convey some semantic information, as evidenced by the fact that they elicit interpretations more similar in meaning to their own lexical affiliates than to the lexical affiliates of other gestures. The range of meanings they convey seem rather limited when compared to speech. Note that our gestures had been selected because naive subjects perceived them to be meaningful *and* agreed on the words in the accompanying speech to which they were related. Yet interpretations of these gestures, made in the absence of speech, were judged more similar to their original lexical affiliates at a rate that was only 12% better than chance. The best of our six interpreters (i.e., the one whose interpretations most frequently yielded the correct lexical affiliate) had a success rate of 66%; the best judge/interpreter combination achieved an accuracy score of 72%. Thus, although gestures may serve as a guide to what is being conveyed verbally, it would be difficult to argue on the basis of these data that they are a particularly effective guide. It needs to be stressed that our test of communicativeness is a relatively undemanding one-- i. e., whether the interpretation enabled a judge to discriminate the correct lexical affiliate from a randomly selected affiliate that, on average, was relatively dissimilar in meaning. The fact that, with so lenient a criterion, performance was barely better than chance undermines the plausibility of the claim that gestures play an important role in communication when speech is fully accessible.

4.1.2 Memory for gestures

An alternative way of exploring the kinds of meanings gestures and speech convey is by examining how they are represented in memory (Krauss et al., 1991, Experiments 3 and 4). We know that words are remembered in terms of their meanings, rather than as strings of letters or phonemes (get ref). If gestures convey meanings, we might likewise expect those meanings to be represented in memory. Using a recognition memory paradigm, we can compare recognition accuracy for lexical affiliates, for the gestures that accompanied the lexical affiliates, and for the speech and gestures combined. If gestures convey information that is different from the information conveyed by the lexical affiliate, we would expect that speech and gestures combined would be better recognized than either speech or gestures separately. On the other hand, if gestures simply convey a less rich version of the information conveyed

---

[8]The 60 LAs were distributed fairly equally among the four coding categories (approximately 33, 22, 22 and 23 percents, respectively), and two coders working independently agreed on 85 percent of the categorizations (k = .798).

by speech, we might expect adding gestural information to speech to have little effect on recognition memory, compared to memory for the speech alone.

The experiment was run in two phases: a Presentation phase, in which subjects saw and/or heard the material they would later try to recognize; a Recognition phase, in which they heard and/or saw a pair of segments, and tried to select the one they had seen before. We examined recognition in three modality conditions: an *audio-video condition,* in which subjects attempted to recognize the previously exposed segment from the combined audio and video; a *video-only condition,* in which recognition was based on the video portion with the sound turned off; and an *audio-only condition,* in which heard the sound without seeing the picture. We also varied the Presentation phase. In the *single channel* condition, the 30-segments presented in the same way they would later be recognized (i.e., sound only if recognition was to be in the audio-only condition, etc.). In the *full channel* condition, all subjects saw the audio-visual version in the Presentation phase, irrespective of their Recognition condition. They were informed of the Recognition condition to which they had been assigned, and told they would later be asked to distinguish segments to which they had been exposed from new segments on the basis of the video portion only, the audio portion only, or the combined audio-video segment. The instructions stressed the importance of attending to the aspect of the display they would later try to recognize. About 5 min after completing the Presentation phase, all subjects performed a forced-choice recognition test with 30 pairs of segments seen and/or heard in the appropriate recognition mode.

A total of 144 undergraduates, 24 in each of the 3 x 2 conditions, served as subjects. They were about equally distributed between males and females.

The means for the six conditions are plotted in Figure 2. Large effects were found for recognition mode ($F_{(2,33)} = 40.23$, $p < .0001$), presentation mode ($F_{(1,33)} = 5.69$, $p < .02$), and their interaction ($F_{(2,33)} = 4.75$, $p < .02$). Speech accompanied by gesture was no better recognized than speech alone ($F < 1$). For the audio-only and audio-video conditions, recognition rates are virtually identical in the two presentation mode conditions, and hearing speech in its gestural context did not improve subsequent recognition. However, in the video-only condition there were substantial differences in performance across the two initial presentations. Compared to subjects who initially saw only the gestures, subjects who had viewed gestures and simultaneously heard the accompanying speech were subsequently *less* likely to recognize them. Indeed, their mean recognition rate was only about ten percent better than the chance level of 50 percent, and the difference in video-only recognition accuracy between the two experiments (.733 vs. .610) is reliable ($F_{(1, 33)} = 14.97$, $p <$ .0001).

Conversational gestures seen in isolation appear not to be especially memorable, and, paradoxically, combining them with the accompanying speech makes them significantly less so. We believe that subjects found it difficult to recognize gestures they had seen in isolation a few minutes earlier because the gestures had to be remembered in terms of their physical properties rather than their meanings. Why then did putting them in a communicative context make them more difficult to recognize? Our hypothesis is that subjects used the verbal context to impute meanings to the gestures, and used these meanings to encode the gestures in memory. If the meanings imputed to the gesture were largely a product of the lexical affiliate, they would be of little help in the subsequent recognition task. The transparent meaning a gesture has when seen in the context of its lexical affiliate may be illusory—a construction deriving primarily from the lexical affiliate's meaning.

---

insert Figure 2 about here

4.1.3 Sources of variance in the attribution of gestural meaning

Our hypothesized explanation for the low recognition accuracy of gestures initially seen in the context of the accompanying speech is speculative, because we have no direct way of ascertaining the strategies our subjects employed when they tried to remember and recognize the gestures. However, the explanation rests on an assumption that is testable, namely, that the meanings people attribute to gestures derive mainly from the meanings of the lexical affiliates. We can estimate the relative contributions gestural and speech information make to judgments of one component of a gesture's meaning: its semantic category. If the gesture's form makes only a minor contribution to its perceived meaning, remembering the meaning will be of limited value in trying to recognize the gesture.

To assessg this, we asked subjects to assign the gestures in our 60 segments to one of four semantic categories (Actions, Locations, Object names and Descriptions) in one of two conditions: a *video-only* condition, in which they saw the gesture in isolation or an *audio-video* condition, in which they both saw the gesture and heard the accompanying speech. Instructions in the audio-video condition stressed that it was the meaning of the gestures that was to be categorized. Two additional groups of subjects categorized the gestures' lexical affiliates—one group from the audio track and the other from verbatim transcripts. From these four sets of judgments, we were able to estimate the relative contribution of speech and gestural information to this component of a gesture's perceived meaning. Forty undergraduates, approximately evenly divided between males and females, served as subjects—ten in each condition (Krauss et al., 1991, Expt. 5).[9]

---

insert Table 1 about here

---

[9] The subjects in the Transcript condition were paid for participating. The remainder were volunteers.

Our experiment yields a set of four 4 x 4 contingency tables displaying the distribution of semantic categories attributed to gestures or lexical affiliates as a function of the semantic category of the lexical affiliate (Table 1) The primary question of interest here is the relative influence of speech and gestural form on judgments of a gesture's semantic category. Unfortunately, with categorical data of this kind there is no clear "best" way to pose such a question statistically.[10] One approach is to calculate a multiple regression model using the 16 frequencies in the corresponding cells of video-only, audio-only and transcript tables as the independent variables, and the values in the cells of the audio + video table as the dependent variable. Overall, the model accounted for 92 percent of the variance in the cell frequencies of the audio + video matrix ($F_{(3,12)} = 46.10$; $p < .0001$); however, the contribution of the video-only matrix was negligible. The ß-coefficient for the Video-only matrix is -.026 ($t = .124$, $p < .90$); for the audio-only condition, ß = .511 ($t = 3.062$, $p < .01$) and for the transcript condition ß = .42 ($t = 3.764$, $p < .003$). Such an analysis does not take between-subject variance into account. An alternative analytic approach employs multiple analysis of variance (MANOVA). Each of the four matrices in Table 1 represents the mean of ten matrices—one for each of the ten subjects in that condition. By treating the values in the cells of each subject's 4x4 matrix as 16 dependent variables, we can compute a MANOVA using the four presentation conditions as a between-subjects variable. Given a significant overall test, we could then determine which of the six between-subjects conditions contrasts (i.e., Audio + Video vs. Audio-only, Audio + Video vs. Transcript, Audio + Video vs. Video-only, Audio-only vs. Transcript, Audio-only vs. Video-only, Transcript vs. Video-only) differ reliably. Wilk's test indicates the presence of reliable differences among the four conditions ($F_{(36,\ 74.59)} = 6.72$, $p < .0001$). F-ratios for the six between-condition contrasts are shown in Table 2. As that table indicates, the video-only condition differs reliably from the audio + video condition, and from the audio-only and transcript conditions as well. The latter two conditions differ reliably from each other, but not from the audio + video condition.

---

insert Table 1 about here

Both analytic approaches lead to the conclusion that judgments of a gesture's semantic category based on visual information alone are quite different from the same judgments made when the accompanying speech is accessible. What is striking is that judgments of a gesture's semantic category made in the presence of its lexical affiliate are not reliably different from judgments of the lexical affiliate's category made from the lexical affiliate alone. Unlike the regression analysis, the MANOVA takes the within-cell variances into account, but it does not readily yield an index of the proportion of variance accounted for by each of the independent variables.

---

[10] Because many cells have very small expected values, a log-linear analysis would be inappropriate.

---

insert Table 2 about here

Taken together, the multiple regression and MANOVA analyses lead to a relatively straightforward conclusion: At least for the 60 gestures in our corpus, when people can hear the lexical affiliate their interpretation of the gesture's meaning (as that is reflected in its semantic category) is largely a product of what they hear rather than what they see. Both analyses also indicate that the audio-only and transcript condition contribute unique variance to judgments made in the audio + video condition. Although judgments made in the audio-only and transcript conditions are highly correlated (r (15) = .815, $p < .0001$), the MANOVA indicates that they also differ reliably. In the regression analysis, the two account for independent shares of the audio + video variance. Because the speech and transcript contain the same *semantic* information, these results suggest that such rudimentary interpretations of the gesture's meaning take paralinguistic information into account.

4.2 Gestural contributions to communication

The experiments described in the previous section attempted, using a variety of methods, to assess the semantic content of spontaneous conversational hand gestures. Our general conclusion was that these gestures convey relatively little semantic information. However, any conclusion must be tempered by the fact that there is no standard method of assessing the semantic content of gestures, and it might be argued that our results are simply a consequence of the imprecision of our methods. Another approach to assessing the communicativeness of conversational gestures is to examine the utility of the information they convey. It is conceivable that, although the semantic information gestures convey is meager quantitatively, it plays a critical role in communication, and that the availability of gestures improve a speaker's ability to communicate. In this section we will describe a set of studies that attempt to determine whether the presence of conversational gestures enhances the effectiveness of communication.

If meaningfulness is a nebulous concept, communicative effectiveness is hardly more straightforward. We will take a functional approach: communication is effective to the extent that it accomplishes its intended goal. For example, other things being equal, directions to a destination are effective to the extent that a person who follows them gets to the destination. Such an approach makes no assumptions about the message's form: how much detail it contains, from whose spatial perspective it is formulated, the speech genre it employs, etc. The sole criterion is how well it accomplishes its intended purpose. Of course, with this approach the addressee's performance contributes to the measure of communicative effectiveness. In the example, the person might fail to reach the destination because the directions were insufficiently informative or because the addressee did a poor job of following them. We can control for the variance attributable to the listener by having several listeners respond to the same message.

The procedure we use is a modified referential communication task (Fussell & Krauss, 1989a; Krauss & Glucksberg, 1977; Krauss & Weinheimer, 1966). Reference entails using language to designate some state of affairs in the world. In a referential communication task, one person (the speaker or encoder) describes or designates one item in an array of items in a way that will allow another person (the listener or decoder) to identify the target item. By recording the message the encoder produces we can present it to several decoders and assess the extent to which it elicits identification of the correct stimulus.

4.2.1 Gestural enhancement of referential communication

We conducted three experiments to examine the extent to which access to conversational gestures enhanced the communicative effectiveness of messages in a referential communication task (Krauss et al., in press). The experiments were essentially identical in design. What we varied from experiment to experiment was the content of communication, by varying the nature of the stimuli that encoders described. In an effort to examine whether the communicative value of gestures depended upon the spatial or pictographic quality of the referent, we used stimuli that were explicitly spatial (novel abstract designs), spatial by analogy (novel synthesized sounds), and not at all spatial (tastes).

Speakers were videotaped as they described the stimuli either to listeners seated across a small table *(Face-to-Face* condition), or over an intercom to listeners in an adjoining room *(Intercom* condition). The videotaped descriptions were presented to new subjects (decoders), who tried to select the stimulus described. Half of these decoders both saw and heard the videotape, the remainder only heard the soundtrack. The design permits us to compare the communicative effectiveness of messages accompanied by gestures with the effectiveness of the same messages without the accompanying gestures. It also permits us to examine the communicative effectiveness of gestures originally performed in the presence of another person (hence potentially communicatively intended) with gestures originally performed when the listener could not see the speaker.

Novel Abstract Designs

For stimuli we used a set of 10 novel abstract designs taken from a set of designs previously used in other studies (Fussell & Krauss, 1989a; Fussell & Krauss, 1989b). A sample is shown in Figure 3. 36 undergraduates (18 males and 18 females) described the designs to a same-sexed listener who was either seated face-to-face across a small table or over an intercom to a listener in another room. Speakers were videotaped via a wall-mounted camera that captured an approximately waist-up frontal view.

---

insert Figure 3 about here

To construct stimulus tapes for the Decoder phase of the experiment, we drew 8 random samples of 45 descriptions (sampled without replacement) from the 360 generated in the Encoder phase, and edited each onto a videotape in random order. 86 undergraduates (32 males and 54 females) either heard-and-saw one of the videotapes *(Audio-Video* condition), or only heard its soundtrack *(Audio-only* condition) in groups of 1-5.

The mean proportions of correct identifications in the four conditions are shown in the left panels of Table 3 . As inspection of that table suggests, accuracy does not vary reliably as a function of decoder condition ($F(1, 168) = 1.21$, $p = .27$). Decoders were no more accurate identifying graphic designs when they could both see and hear the person doing the describing than they were when they could only hear the describer's voice. A reliable effect was found for encoder condition ($F(1, 168) = 5.72$, $p = .02$). Surprisingly, decoders were more somewhat accurate identifying the designs from descriptions that originally had been given in the intercom decoding condition. However, regardless of the encoding condition, being able to see the encoder did not affect the decoder's accuracy either positively or negatively; the Encoder x Decoder interaction was not significant ($F(111, 168) = 1.61$, $p= .21$).

---

insert Table 3 about here

Novel Sounds

The same 36 undergraduates who described the novel designs also listened to 10 pairs of novel sounds using headphones, and described one sound from each pair to their partner. Half of the encoders described the sounds to a partner seated in the same room; for the remainder their partner was located in a nearby room. The sounds had been produced by a sound synthesizer, and resembled the sorts of sound effects found in a science fiction movie. Except for the stimulus, conditions for the Encoding phase of the two experiments were identical. From these descriptions, 6 stimulus tapes, each containing 60 descriptions selected randomly without replacement were constructed. 98 paid undergraduates, 43 males and 55 females, served as decoders, serving in groups of 1-4. They either heard (in the Audio-Only condition) or viewed and heard (in the Audio-Video condition) a description of one of the synthesized sounds, then heard the two sounds, and indicated on a response sheet which of the two sounds matched the description.

As was the case with the graphic designs, descriptions of the synthesized sounds made in the Intercom encoding condition elicited a somewhat higher level of correct identifications than those made Face-to-Face ($F(1,168) = 10.91$, $p < .001$). However, no advantage accrued to decoders who could see the speaker in the video, compared to those who could only hear the soundtrack. The means for the audio-video and audio-only conditions did not differ significantly ($F(1,168) = 1.21$, $p = .27$), nor did the Encoder x Decoder interaction ($F < 1$). The means for the 4 conditions are shown in the right panel of Table 3.

Tea Samples

As stimuli, we used 8 varieties of commercially-available tea bags that would produce brews with distinctively different tastes. 36 undergraduates, approximately evenly divided between males and females, participated as encoders. They were given cups containing two tea samples, one of which was designated the target stimulus. They tasted each, and were videotaped as they described the target to a same-sex partner so it could be distinguished from its pairmate. Half of the encoders described the sample in a face-to-face condition, the remainder in an intercom condition. From these videotaped descriptions, two videotapes were constructed each containing 72 descriptions, half from the face-to-face condition, the remainder from the intercom condition. 43 undergraduates (20 males and 23 females) either heard or heard-and-saw one of the two videotapes. For each description, they tasted two tea samples and tried to decide which it matched.

---

insert Table 4 about here

Overall identification accuracy was relatively low (M = .555; SD = .089) but better than the chance level of .50 ($t(85) = 5.765$, $p < .0001$). The means and standard deviations are shown in Table 4. As was the case for the designs and sounds, ANOVA of the proportion of correct identifications revealed a significant effect attributable to encoding condition, but for the tea samples the descriptions of face-to-face encoders produced a slightly, but significantly, *higher* rate of correct identifications than those of intercom encoders ($F(1,41) = 5.71$, $p < .02$). Nevertheless, as in the previous two experiments, no effect was found for decoding condition, or for the encoding x decoding interaction (both Fs < 1).

Thus, in none of the three experiments did we find the slightest indication that being able to see a speaker's gestures enhanced the effectiveness of communication, as compared simply to hearing the speech. Although the logic of statistical hypothesis testing does not permit positive affirmation of the null hypothesis, our failure to find differences can't be attributed simply to a lack of power of our experiments. Our 3 experiments employed considerably more subjects, both as encoders and decoders, than is the norm in such research. By calculating the statistical power of our test, we can estimate the Least Significant Number (LSN)—i.e., the number of subjects that would have been required to reject the null hypothesis with a = .05 for the audio-only vs. audio-video contrast, given the size of the observed differences. For the Novel Designs it is 548; for the Sounds it is 614. The LSNs for the encoder condition x decoder condition interactions are similarly large: 412 and 7677 for Designs and Sounds, respectively.

Nor was it the case that speakers simply failed to gesture, at least in the first two experiments. On average, speakers gestured about 14 times per minute when describing the graphic designs and about 12 times per minute when describing the sounds; for some speakers, the rate exceeded 25 gestures per minute. Yet no relationship was found between the effectiveness with which a message communicated and the amount of gesturing that accompanied it. Given these data, along with the absence of a credible body of contradictory results in the literature, it seems to us that only two conclusions are plausible: either gestural accompaniments of speech do not enhance the communicativeness of speech in settings like the ones we studied, or the extent to which they do so is negligible.

4.2.2 Gestural enhancement of communication in a nonfluent language

Although gestures may not ordinarily facilitate communication in settings such as the ones we have studied, it may be the case that they do so in special circumstances—for example, when the speaker has difficulty conveying an idea linguistically. Certainly many travelers have discovered that energetic pantomiming can make up for a deficient vocabulary, and make it possible to "get by" with little mastery of a language. Dushay (1991) examined the extent to which speakers used gestures to compensate for a lack of fluency, and whether the gestures enhanced the communicativeness of messages in a referential communication task.

His procedure was similar to that used in the studies described in Section 4.2.1. As stimuli he used the novel figures and synthesized sounds employed in those experiments, and the experimental set-up was essentially identical. His subjects (20 native-English-speaking undergraduates taking their fourth semester of Spanish) were videotaped describing stimuli either face-to face with their partner (a Spanish-English bilingual) or communicating over an intercom. On half of the trials they described the stimuli in English, and on the remainder in Spanish. Their videotapes of their descriptions were edited and presented to eight Spanish/English bilinguals, who tried to identify the stimulus described. On half of the trials, they heard the soundtrack but did not see the video portion (*audio-only* condition), and on the remainder they both heard and saw the description (*audio-visual* condition).

Speakers did not use more conversational gestures when describing the stimuli in Spanish than in English.[11] When they described the novel figures their

---

[11]Although this was true of conversational gestures, overall the rate for all types of gestures was slightly higher when subjects spoke Spanish. The difference is accounted for largely by what Dushay called "groping movements" (repetitive, typical circular bilateral

gesture rates in the two languages were identical, and when they described the synthesized sound their gesture rate in Spanish was slightly, but significantly, lower. Moreover, being able to see the gestures did not enhance listeners' ability to identify the stimulus being described. Not surprisingly, descriptions in English produced more accurate identifications than descriptions in Spanish, but for neither language (and neither stimulus) type did the speaker benefit from seeing the speaker.

4.3 Gestures and the communication of nonsemantic information

The experiments described above concerned the communication of *semantic* information, implicitly accepting the traditionally-assumed parallelism of gesture and speech. However, semantic information is only one of the kinds of information speech conveys. Even when the verbal content of speech is unintelligible, paralinguistic information is present that permits listeners to make reliable judgments of the speaker's internal affective state (Krauss, Apple, Morency, Wenzel, & Winton, 1981; Scherer, Koivumaki, & Rosenthal, 1972; Scherer, London, & Wolf, 1973) . Variations in dialect and usage can provide information about a speaker's social category membership (Scherer & Giles, 1979) . Variations in the fluency of speech production can provide an insight into the speaker's confidence, spontaneity, involvement, etc. There is considerable evidence that our impressions of others are to a great extent mediated by their nonverbal behavior (DePaulo, 1992). It may be the case that gestures convey similar sorts of information, and thereby contribute to participants' abilities to play their respective roles in interaction.

4.3.1 Lexical movements and impressions of spontaneity

Evaluations of spontaniety can can affect the way we understand and respond to others' behavior. Our research on spontaneity judgments was guided by the theoretical position that gestures, like other nonverbal behaviors, often serve both intrapersonal and interpersonal functions. An interesting characteristics of such behaviors is that they are only partially under voluntary control. Although many self-presentational goals can be achieved nonverbally (DePaulo, 1992), the demands of cognitive processing constrains a speaker's ability to use these behaviors strategically (Fleming & Darley, 1991). Because of this, certain nonverbal behaviors can reveal information about the cognitive processes that underlie a speaker's utterances. Listeners may be sensitive to these indicators, and use them to draw inferences about the conditions under which speech was generated.

Chawla and Krauss (1994) studied subjects' sensitivity to nonverbal cues that reflect processing by examining their ability to distinguish between

---

movements of the hands at about waist level), which were about seven times more frequent when subjects spoke Spanish. Groping movements seemed to occur when speakers were having difficulty recalling the Spanish equivalent of an English word. Different instances of groping movements vary little within a speaker, although there is considerable variability from speaker to speaker, so it is unlikely that they are used to convey information other than that the speaker is searching for a word.

spontaneous and rehearsed speech. Subjects either heard (audio condition), viewed without sound (video condition), or heard and saw (audio-video condition) 8 pairs of videotaped narratives, each consisting of a spontaneous narrative and its rehearsed counterpart. The rehearsed version was obtained by giving the transcript of the original narrative to a professional actor of the same sex who was instructed to prepare an authentic and realistic portrayal.[12] Subjects were shown the spontaneous and the rehearsed versions of each scene and tried to identify the spontaneous one. They also were asked to rate how real or spontaneous each portrayal seemed to be. Comparing performance of subjects in the three presentation conditions allowed us to assess the role of visual and vocal cues while keeping verbal content constant.

In the audio-video presentation condition, subjects correctly distinguished the spontaneous from the rehearsed scenes 80% of the time. In the audio and video conditions, accuracy was somewhat lower (means = 66% and 60%, respectively) although in both cases it was reliably above chance level of 50%. The audio-video condition differed reliably from the audio and video conditions, but the latter two conditions did not.

Subjects evidenced some sensitivity to subtle nonverbal cues that derive from differences in the way spontaneous and rehearsed speech are processed. A scene's spontaneity rating in the audio-video condition was significantly correlated with the proportion of time the speaker spent making lexical movements, and with the conditional probability of non juncture pauses . Given that nonjuncture pauses and lexical movements both reflect problems in lexical access, and that the problems of lexical access are much greater in spontaneous speech than in posed or rehearsed speech, we would expect that these two behaviors would be reliable cues in differentiating spontaneous and rehearsed speech. Interestingly, the subjects' judgments of spontaneity were not related to the total amount of time spent gesturing or to the total number of pauses in the speech.

Unfortunately, we were not able to get any direct corroboration of our hypothesis from subjects' descriptions of what cues they used to make their judgments. It appears that subjects use nonverbal information in complex ways that they are unable to describe. Our subjects appeared to have no insight into the cues they had used and the processes by which they had reached their judgments. Their answers were quite confused and no systematic trends could be found from these open ended questions.

The results of this experiment are consistent with our view that gestures convey nonsemantic information that could, in particular circumstances, be quite useful. Although our judges's ability to discriminate spontaneous from rehearsed scenes was far from perfect, especially when they had only visual information to work with, our actors portrayals may have been unusually artful; we doubt that portrayals by less skilled performers would have been as convincing. Of course, our subjects viewed the scenes on videotape, aware that

---

[12]Additional details on this aspect of the study are given in Section 6.1.1.

one of them was duplicitous. In everyday interactions, people often are too involved in the situation to question others' authenticity.

5. INTRAPERSONAL FUNCTIONS: GESTURES AND SPEECH PRODUCTION

An alternative to the view of gestures as devices for the communication of semantic information focuses on the role of gestures in the speech production process.[13] One possibility, suggested several times over the last 50 years by a remarkably heterogeneous group of writers, is that gestures help speakers formulate coherent speech, particularly when they are experiencing difficulty retrieving elusive words from lexical memory (DeLaguna, 1927; Ekman & Friesen, 1972; Freedman, 1972; Mead, 1934; Moscovici, 1967; Werner & Kaplan, 1963) , although none of the writers who have made the proposal provide details on the mechanisms by which gestures accomplish this. In an early empirical study, Dobrogaev (1929) reported that preventing speakers from gesturing resulted in decreased fluency, impaired articulation and reduced vocabulary size.[14]

More recently, three studies have examined the effects of preventing gesturing on speech. Lickiss and Wellens (1978) found no effects on verbal fluency from restraining speakers' hand movements, but it is unclear exactly which dysfluencies they examined. Graham and Heywood (1975) compared the speech of the six speakers in the Graham and Argyle (1975) study who described abstract line drawings and were prevented from gesturing on half of the descriptions. Although statistically significant effects of preventing gesturing were found on some indices, Graham and Heywood conclude that "… elimination of gesture has no particularly marked effects on speech performance" (p. 194). Given their small sample of speakers and the fact that significant or near-significant effects were found for several contrasts, the conclusion seems unwarranted. In a rather different sort of study, Rimé, Schiaratura, Hupet and Ghysselinckx (1984) had speakers converse while their head, arms, hands, legs, and feet were restrained. Content analysis found less vivid imagery in the speech of speakers who could not move.

Despite these bits of evidence, support in the research literature for the idea that gestures are implicated in speech production, and specifically in lexical access, is less than compelling. Nevertheless, this is the position we will take. To

---

[13] Another alternative, proposed by Dittmann and Llewellyn (1977) , is that gestures serve to dissipate excess tension generated by the exigencies of speech production. Hewes (1973) has proposed a theory of the gestural origins of speech in which gestures are seen as vestigial behaviors with no current function—a remnant of human evolutionary history. Although the two theories correctly (in our judgment) emphasize the connection of gesturing and speech, neither is supported by credible evidence and we regard both as implausible.

[14] Unfortunately, like many papers written in that era, Dobrogaev's includes virtually no details of procedure, and describes results in qualitative terms (e.g., "Both the articulatory and semantic quality of speech was degraded"), making it impossible to assess the plausibility of the claim. We will describe an attempt to replicate the finding in Section 6.

understand how gestures might accomplish this, it is necessary to consider the process by which speech is produced.[15]

5.1 Speech production

Although several different models of speech production have been proposed, virtually all distinguish three stages of the process. We will follow Levelt (1989) in calling them *conceptualizing, formulating,* and *articulating.* Conceptualizing involves, among other things, drawing upon declarative and procedural knowledge to construct a communicative intention. The output of the conceptualizing stage—what Levelt refers to as a *preverbal message*—is a conceptual structure containing a set of semantic specifications. At the formulating stage, the preverbal message is transformed in two ways. First, a grammatical encoder maps the to-be-lexicalized concept onto a lemma in the mental lexicon (i.e., an abstract symbol representing the selected word as a semantic-syntactic entity) whose meaning matches the content of the preverbal message, and, using syntactic information contained in the lemma, transforms the conceptual structure into a *surface structure.* Then, a phonological encoder transforms the surface structure into a *phonetic plan* (essentially a set of instructions to the articulatory system) by accessing word forms stored in lexical memory and constructing an appropriate plan for the utterance's prosody. The output of the articulatory stage is overt speech. The process is illustrated schematically by the structure in the shaded portion of Figure 4 below.

5.2 Gesture production

The foregoing description of speech production leaves out many details of what is an extremely complex process, and many of these details are matters of considerable contention. Nevertheless, there is reason to believe the account is essentially correct in its overall outline (see Levelt, 1989 for a review of the evidence). Unfortunately, we lack even so rudimentary a characterization of the process by which conversational gestures are generated, and, because there is so little data to constrain theory, any account we offer must be regarded as highly speculative.

Our account of the origins of gesture begins with the representation in short term memory that comes to be expressed in speech. For convenience, we will call this representation the *source concept.* The conceptual representation outputted by the conceptualizer that the grammatical encoder transforms into a linguistic representations will incorporate only some of the source concept's features. Or, to put it somewhat differently, in any given utterance only certain aspects of the source concepts will be relevant to the speaker's communicative intention. For example, one might recall the dessert served the previous evening's meal, and refer to it as the *cake.* The particular cake represented in memory had a number of properties (e.g., size, shape, flavor, etc.) that are not part of the semantic of the word *cake.* Presumably if these properties were

---

[15] In discussions of speech production, "gesture" often is used to refer to what are more properly called *articulatory gestures*—i.e., linguistically significant acts of the articulatory system. We will restrict our use of the term to hand gestures.

relevant to the communicative intention, the speaker would have used some lexical device to express them (e.g., *a heart-shaped cake).* [16]

Our central assumption is that lexical movements are made up of representations of the source concept, expressed motorically. Just as the linguistic representation often does not incorporate all of the features of the source concept, lexical movements reflect these features even more narrowly. The features they incorporate are primarily spatio-dynamic.

5.2.1 A gesture production model

We have tried to formalize some of our ideas in a model parallelling Levelt's speech production model that is capable of generating both speech and conversational gestures. Although the diagram in Figure 4 is little more than a sketch of one way of structuring such a system, we find it useful because it suggests some of the mechanisms that might be necessary to account for the ways that gesturing and speaking interact.

The model requires that we make several assumptions about memory and mental representation:

(1) Human memory employs a number of different formats to represent knowledge, and much of the contents of memory is multiply encoded in more than one representational format.

(2) Activation of a concept in one representational format tends to activate related concepts in other formats.

(3) Concepts differ in how adequately (i.e., efficiently, completely, accessibly, etc.) they can be represented in one or another format.

(4) Some representations in one format can be translated into the representational form of another format (e.g., a verbal description can give rise to a visual image, and vice-versa).

None of these assumptions is particularly controversial, at least at this level of generality.

---

insert Figure 4 about here

---

[16]The difference between the source concept and the linguistic representation is most clearly seen in reference, where the linguistic representation is formulated specifically to direct a listener's attention to some thing, and typically will incorporate only as much information as is necessary to accomplish this. Hence one may refer to a person as "The tall guy with red hair," employing only a few features of the far more complex and differentiated conceptual representation.

5.2.2 Lexical movements

We follow Levelt in assuming that inputs from working memory to the conceptualizing stage of the speech processor must be in propositional form. However, much of the knowledge that represents the source concept is multiply encoded in both propositional and nonpropositional representational formats, or is encoded exclusively in a nonpropositional format. In order to be reflected in speech, nonpropositionally encoded information must be "translated" into propositional form.

Our model posits a spatial/dynamic feature selector that transforms information stored in spatial or dynamic formats into a set of *spatial/dynamic specifications.* How this might work can be illustrated in a hypothetical example. Consider the word "vortex." Conceptually, a state of affairs that might be called "a vortex" would include spatial elements like size and form, and such dynamic elements as rate and path of motion. The conceptual structure also includes other elements—that the *vortex* is composed of a liquid, that it is capable of drawing things into it, and so forth. Let us say that the linguistic representation *vortex* includes the elements (1) movement, (2) circular, (3) liquid, (4) mass, (5) in-drawing. A speaker having thus conceptualized a state of affairs, and wanting to convey this characterization, would search the lexicon for an entry that incorporates the relevant features, and ultimately arrive at the word *vortex.* At the same time the spatial/dynamic feature selector might select the elements of motion and circularity, and transform them into a set of spatio/dynamic specifications—essentially abstract properties of movements.[17] These abstract specifications are, in turn, translated by a motor planner into a *motor program* that provides the motor system with a set of instructions for executing the lexical movement.

5.2.3 Motor Movements

Relatively little is known about the origins of motor movements or the functions they serve. Unlike lexical movements, which we believe have a conceptual origin, motor movements appear to be a product of speech production. Their coordination with the speech prosody suggests they could not be planned before a phonetic plan for the utterance had been established. As is illustrated in the speech production portion of the diagram in Figure 4, the phonetic plan is a product of the phonological encoder. Hence, by this account, motor movements originate at a relatively late stage of the speech production process. We hypothesize that a prosodic feature detector reduces the output of the phonological encoder to a relatively simple set of *prosodic specifications* marking the timing, and, perhaps, the amplitude, of primary stresses in the speech. A motor planner translates the specification into a *motor*

[17]We can only speculate as to what such abstract specifications might be like. One might expect concepts incorporating the feature FAST to be represented by rapid movements, and concepts incorporating the feature LARGE to be represented by movements with large linear displacements. However, we are unaware of any successful attempts to establish systematic relations between abstract dimensions of movement and dimensions of meaning. For a less-than-successful attempt, see Morrel-Samuels (1989).

program—expressing the cadence of stressed syllables in terms of the periodicity of strokes of the gesture, and the loudness of the stressed syllables in terms of the gesture's amplitude.

5.3 Integration of speech and gesture production

Up to this point in our discussion, the processes of lexical movement production and speech production have proceeded independently, and at least one view of gestures holds that the two processes are essentially autonomous (Levelt, Richardson, & La Heij, 1985). However, our contention is that both lexical movements and motor movements play a role in speech production, and indeed that this is their primary function.

5.3.1 Lexical movements and lexical access

We believe that the primary effect of lexical movements is at the stage of grammatical encoding, where they facilitate access to the lemmas contained in the mental lexicon. A lexical movement represents some of the same conceptual features as are contained in the preverbal message. In large part, lexical search consists of an attempt to retrieve from the lexicon entries that satisfy the preverbal message's specifications. In the example of VORTEX described above, the conceptual features of motion and circularity are represented both in the lexical movement and in the meaning of the word *vortex*, which is the movement's lexical affiliate. We hypothesize that when the speaker is unable to locate the lemma for a lexical entry whose meaning matches the specifications of the preverbal message, motorically-represented features in the lexical movement aid in the process of retrieval by serving as a cross-modal prime. Hadar and Butterworth (1993) have suggested that the gestural representation serves to "hold" the conceptual properties of the sought-for lexical entry in memory during lexical search, and this seems plausible to us.

In our model of the process (Figure 4) we have represented the role that lexical movements play in lexical access by showing the output of the muscle system (an implementation of the motor program produced by the motor planner) affecting the grammatical encoder. Our assumption is that the motorically represented features of the lexical movement are apprehended proprioceptively—i.e., that it is the lexical movements, rather than the program, that prime the grammatical encoder. The diagram also shows the output of the phonological encoder affecting the motor planner. Such a path is necessary in order for the gesture production system to know when to terminate the gesture. If the function of gestures is to facilitate lexical access, there is no reason to produce them once the sought-for lexical item has been retrieved, and the gesture system needs some way of knowing this.

5.3.2 Lexical movements and conceptualizing

Some theorists have proposed that the conversational gestures we are calling lexical movements also can influence speech production at the conceptualizing stage. Hadar and Yadlin-Gedassy (1994) draw a distinction between lexical gestures and conceptual gestures. Conceptual gestures could be used by speakers to frame the conceptual content that will become part of the intended meaning of the utterance. It is not unreasonable to think that some gestures serve this function. Speakers trying to describe the details of a highly-practiced motor act (e.g., tying shoelaces, hitting a backhand in tennis) often will perform the act in a schematic way as they construct their description. Presumably doing so helps them work out the steps by which the act is accomplished, and this may be necessary because the motoric representation better represents the content (e.g., is more complete, more detailed, etc.) than an abstract or propositional representation.

Although the idea that gestures aid in conceptualizing is not implausible, we are unaware of any relevant empirical evidence pro or con. As a practical matter it is not clear how one would go about distinguishing conceptual gestures from lexical movements. Hadar and Yadlin-Gedassy (1994) suggest that the gestures that occur during hesitations are likely to be lexical gestures, and we agree (see Section 6.1.1), but lexical access can take place during juncture pauses, too, so there is no reason to believe that gestures at junctures are exclusively, or even predominantly, conceptual. Hadar, Burstein, Krauss & Soroker (1995) speculate that conceptual gestures will tend to be less iconic than lexical gestures.[18] However, because the iconicity of many lexical movements is difficult to establish (see Section 7.2.1), the distinction at best will be one of degree. Hence, for reasons of parsimony and because we do not have a principled way of distinguishing the two kinds of gestures, we have not included conceptual gestures (or movements) in our model, and simply assume that the conversational gestures in our corpora are either lexical movements or motor movements.

5.3.3 Motor movements and speech prosody

The function of motor movements is obscure, despite the fact that, along with other coverbal behaviors (e.g., head movements, eye blinks, etc.), they are ubiquitous accompaniments of speech. It has been claimed that motor movements along with other coverbal behaviors aid the speaker in coordinating the operation of the articulators (cf., Hadar, 1989b). We know of no data relevant to this claim, but it strikes us as plausible. In our schematic diagram we have represented this by showing the output of the motor system feeding the articulators, as well as producing motor movements.[19] Probably it is unnecessary to note that our account of the origins and functions of motor movements is highly speculative, and we include it only for completeness. None of the data we present will be relevant.

---

[18]This will be so, according to Hadar et al. because "…iconicity is usually determined by reference to a particular word: the lexical affiliate. Since there are mediating processes between conceptual analysis and lexical retrieval, there is a higher probability that the eventually selected word will fail to relate transparently to the conceptual idea which shaped the gesture. This will result in judging the gesture 'indefinite.'"

[19]On the other hand, Heller, Miller, Reiner, Rupp & Tweh (1995) report that motor movements are more frequent when speakers are discussing content that is novel rather than familiar. If such a relationship could be established, it might follow that motor movements are affected by events occurring at the conceptual stage of processing.

In the next section we will describe the results of several studies of the role gestures play in speech production.

6. GESTURES AND LEXICAL ACCESS: EMPIRICAL STUDIES

In this section we will examine evidence relevant to our conjecture that gesturing serves to facilitate speech production— specifically that the conversational gestures we term lexical movements help the speaker access entries in the mental lexicon. The data will be drawn from several studies, and focuses on four aspects of the speech-gesture relationship: gesture production in rehearsed vs. spontaneous speech, the temporal relation of speech and gesture, the influence of speech content on gesturing, and the effect on speech production of preventing a speaker from gesturing.

6.1 Gesturing in spontaneous and rehearsed speech

6.1.1 Lexical movements in spontaneous and rehearsed speech

The microstructures of spontaneously generated speech and speech that has been rehearsed reveal that lexical access presents different problems under the two conditions. For example, while 60-70% of the pauses in spontaneous speech are found at grammatical clause junctures *(juncture pauses),* speech read from a prepared text (where neither planning nor lexical access is problematic) contains many fewer pauses, with nearly all of them falling at grammatical junctures (Henderson, Goldman-Eisler, & Skarbek, 1965) . Repeatedly expressing the same content produces similar results. Butterworth and Thomas (described in Butterworth, 1980) found that in subjects' initial descriptions of a cartoon, 59% of their pausing time fell at clause boundaries, but by the seventh (not necessarily *verbatim)* repetition they spent considerably less time pausing, and 85% of the pauses that did occur fell at clause boundaries. Nonjuncture pauses (i.e., pauses that fall within the body of the clause) are generally believed to reflect problems in lexical access, and the fact that they occur so infrequently in read or rehearsed speech is consistent with the conclusion that lexical access is relatively unproblematic.

As part of the experiment described in Section 4.3.1, Chawla and Krauss (1994) videotaped professional actors spontaneously answering a series of questions about their personal experiences, feelings, and beliefs. Their responses were transcribed and turned into “scripts” that were then given to another actor of the same sex along with instructions to portray the original actor in a convincing manner. Videotapes of both the spontaneous original responses and the rehearsed portrayals were coded for the frequency of lexical movements. Although the overall amount of time spent gesturing did not differ between the spontaneous and rehearsed portrayals, the proportion of time spent making lexical movements was significantly greater in the spontaneous than in the rehearsed scenes ($F(1,12) = 14.14$, $p = .0027$). Consistent with other studies, the conditional probability of a pause being a nonjuncture pause (i.e., Probability (Nonjuncture Pause | Pause)) was significantly greater for spontaneous speech ($F(1, 12) = 7.59$, $p = .017$). The conditional probability of nonjuncture silent pauses and the proportion of time a speaker spent making lexical movements were reliably correlated ($r(14) = .47$, $p < .05$). If speakers use lexical movements as part of the lexical retrieval process, it would follow that the more hesitant a speaker was, the more lexical movements he or she would make.

6.2 Temporal Relations of Gesture and Speech

If lexical movements facilitate lexical access, their location in the speech stream relative to their lexical affiliates would be constrained. We can specify a number of constraints on the temporal relations of gesture and speech that should be observed if our hypothesis is correct.

6.2.1 Relative onsets of lexical movements and their lexical affiliates

It would make little sense to argue that a gesture helped a speaker produce a particular lexical affiliate if the gesture were initiated after the lexical affiliate had been articulated, and it has been known for some time that gestures tend to precede, or to occur simultaneously with, their lexical affiliates (Butterworth & Beattie, 1978) . Morrel-Samuels and Krauss (1992) carefully examined the time course of the 60 lexical movements described in Section 4.1.1 relative to the onsets of their lexical affiliates. All 60 were initiated either simultaneously with or prior to the articulation of their lexical affiliates. The median gesture-lexical affiliate asynchrony (the interval by which the movement preceded the lexical affiliate) was 0.75 s and the mean .99 s (SD = 0.83 s). The smallest asynchrony was 0 s (i.e., movement and speech were initiated simultaneously) and the largest was 3.75 s. The cumulative distribution of asynchronies is shown in Figure 5.

---

insert Figure 5 about here

6.2.2 Speech-gesture asynchrony and lexical affiliate accessibility

Although all of the lexical movements in our 60 gesture corpus were initiated before or simultaneously with the initiation of their lexical affiliates, there was considerable variability in the magnitude of the gesture-speech asynchrony. Our hypothesis that these gestures aid in lexical access leads us to expect at least some of this variability to be attributable to the lexical affiliate's accessibility. Other things being equal, we would expect lexical movements to precede inacessible lexical affiliates by a longer interval than lexical affiliates that are highly accessible.

Unfortunately, there is no direct way to measure a lexical entry's accessibility, but there is some evidence that accessibility is related to familiarity (Gernsbacher, 1984) . We had 17 undergraduates rate the familiarity of the 60 lexical affiliates, 32 of which were single words and the remainder two- or three-word phrases, on a seven-point scale,[20] and hypothesized that the asynchrony for a given lexical affiliate would be negatively correlated with its rated familiarity. The correlation, although in the predicted direction, was low $r(60) = -.16$, $F(1, 58) = 1.54$, $p > .20$. However, a multiple regression model that included the gesture's spatial extent (i.e., the total distance it traversed), and the lexical affiliate's syllabic length and rated familiarity accounted for thirty percent of the variance in asynchrony ($F(3,56) = 7.89$, $p < .0002$). The right panel of Figure 6 is a partial residual plot (sometimes called a leverage plot) showing the relationship of famili7arity to gesture-speech asynchrony, after the effects of spatial extent and syllabic length have been removed; the correlation between the two variables after partialling is -.27 ($F(1,58) = 4.44$, $p < .04$) The left panel shows the relationship before partialling. Consistent with our hypothesis, familiarity accounts for some of the variability in asynchrony, although, perhaps not surprisingly, the relationship is affected by a number of other factors.

---

insert Figure 6 about here

It is also possible to manipulate access, by restricting the types of words the speaker can use. Rauscher, Krauss and Chen (in press) videotaped subjects describing the plots of animated action cartoons to a partner. The difficulty of lexical access was independently varied either by requiring subjects to use obscure words (obscure speech condition) or to avoid using words containing the letter *C*.(constrained speech condition); these were contrasted with a normal speech condition. Their narratives were transcribed and all of the phrases containing spatial prepositions were identified. We call these spatial content phrases (SCPs), and they comprised about 30% of the phrases in the corpus. Gesture rates were calculated by dividing the amount of time the speaker spent gesturing during SCPs by the number of words in SCPs; the same was done for time gesturing during nonspatial phrases. As lexical access was made more difficult, the rate of gesturing increased in SCPs but not elsewhere; the interaction of spatial content x speech condition produced a statistically reliable effect ($F(2, 80) = 6.57$, $p < .002$). The means are shown in Figure 7.

---

insert Figure 7 about here

6.2.3 Gestural duration and lexical access

The durations of lexical movements vary considerably. In the 60 gesture corpus examined by Morrel-Samuels and Krauss, the average lexical movement lasted for 2.49 s (SD = 1.35 s); the duration of the briefest was 0.54 s, and the

[20]In cases where the lexical affiliate contained more than one word, the familiarity rating of the least familiar word was assigned to the lexical affiliate as a whole.

longest 7.71 s. What accounts for this variability? If, as we have hypothesized, a lexical movement serves to maintain conceptual features in memory during lexical search, it should not be terminated until the speaker has articulated the sought-for lexical item. Of course, there is no way we can ascertain the precise moment lexical access occurs, but it would have to occur before the lexical affiliate is articulated, so we would expect a lexical movement's duration to be positively correlated with the magnitude of the lexical movement-lexical affiliate asynchrony.

In their 60 gesture corpus, Morrel-Samuels and Krauss found this correlation to be +0.71 ($F(1,58) = 57.20$, $p \leq .0001$). The data points are plotted in Figure 8. The heavier of the two lines in that figure is the "unit line"—i.e., the line on which all data points would fall if the lexical movement terminated at the precise moment that articulation of the lexical affiliate began; the lighter line above it is the least-squares regression line. Data points below the unit line represent cases in which the lexical movement was terminated before articulation of the lexical affiliate began, and points above the line represent cases in which the articulation of the lexical affiliate began before the lexical movement terminated. Note that all but three of the 60 data points fall on or above the unit line, and the three points that fall below the unit line are not very far below it. It seems quite clear that a lexical movement's duration is closely related to how long it takes the speaker to access its lexical affiliate, as our model predicts. This finding poses a serious problem for the "modular" view of the relation of the gesture and speech production system proposed by Levelt et al. (1985) which claims that "…the two systems are independent during the phase of motor execution, the temporal parameters having been preestablished in the planning phase" (p. 133). If gesture and speech were independent during the execution phase, the lexical movement's duration would have to be specified prior to execution, and in order to plan a gesture of sufficient duration, it would be necessary for the speaker to know in advance how long lexical access will take. It's not clear to us how a speaker could know this.

---

insert Figure 8 about here

6.3 Gesturing and Speech Content

If our assumption that lexical movements reflect spatio-dynamic representations in non-propositional memory is correct, we should be able to observe an association between gesturing and the conceptual content of speech. We are aware of very few systematic attempts to relate gesturing and speech content, and the data we have are less than conclusive.

6.3.1 Gesturing and description type

Abstract graphic designs

The messages describing novel graphic designs and synthesized sounds obtained in the experiments by Krauss, et al. (in press) were coded into categories of description types, and the rate of gesturing associated with these description types was examined. For the novel designs, we used a category system developed by Fussell and Krauss (1989a) for descriptionsof these figures that partitions the descriptions into three categories: *Literal descriptions,* in which a design was characterized in terms of its geometric elements — as a collection of lines, arcs, angles, etc.; *Figurative descriptions,* in which a design was described in terms of objects or images it suggested; *Symbol descriptions*, in which a design was likened to a familiar symbol, typically one or more numbers or letters.[21] When a message contained more than one type of description (as many did), it was coded for the type that predominated. Overall, about 60% of the descriptions were coded as figurative, about 24% as literal and the remaining 16% as symbols.

For the descriptions of the graphic designs, a one-way ANOVA was performed with description type (literal, figurative or symbol) as the independent variable and gesture rate as the dependent variable to determine whether gesturing varied as a function of the kind of content. A significant effect was found $F\ (2, 350) = 4.26$, $p = .015$. Figurative descriptions were accompanied by slightly more gestures than literal descriptions; both were accompanied by more gestures than were the symbol descriptions (14.6 vs. 13.7 vs. 10.6 gestures per m, respectively). Both figurative and literal descriptions tended to be formulated in spatial terms. Symbol descriptions tended to be brief and static—essentially a statement of the resemblance.

Sound descriptions

We tried to adapt the coding scheme used in the content analysis of the graphic design descriptions for the analysis of the sound descriptions, but the sound analogs of those categories did not adequately capture the differences in the way the synthesized sounds were described, and it was necessary to develop a five-category system that was considerably more complicated than the one used for the pictures. The first three categories referred to straightforward acoustic dimensions: *pitch, intensity,* and *rate of periodicity.* The fourth category, *object sound,* described the stimulus sound in terms of some known sound source, most often a musical instrument. Finally, the fifth category contrasted

---

[21]Some abridged examples of the three types of descriptions are: Literal: "On the right-hand side there's an angle, about a forty-five degree angle, that seems to form an arrow pointing up towards the top. Then at the top of that, at the pointing point of that angle, there's a 180 degree horizontal line that then goes into a part of a rectangle…"; Figurative: "It's sort of like a bird. Reminds me of pictures I've seen of like the phoenix, rising up to regenerate or whatever…"; Symbol: "…this looks like a Greek letter psi, and looks like somebody wrote this letter with their eyes closed, so it's like double lined all over the place…"

elements of the sound in terms of *background and foreground*.[22] For graphic design descriptions, it was relatively easy to determine the category type that predominated, but the overwhelming majority of the sound descriptions employed more than one category or dimension, and often no single type clearly prevailed. For this reason we had to resort to a less satisfactory multiple-coding scheme in which each description received a score of 0-10 on all five categories. Scores were constrained to sum to 10 for any description, with the value for each category representing its relative contribution to the total description.

Because the coding scheme used for the sound descriptions did not assign each description to one and only one category, it was not possible to perform the same analysis on them. Instead, we computed correlations between a description's score on each of the five coding categories and the rate of gesturing that accompanied it. This was done separately for each category. Only the object sound category was significantly associated with gesture rate (r (329) = -0.14, p = .012); the more one of the sounds was likened to a sound made by some object, the lower was the rate of gesturing that accompanied it. The correlations for the four other description types were small (all rs ≤ 0.07) and nonsignificant.

6.3.2 Lexical movement and spatial content

In the Rauscher et al., (in press) experiment described in Section 6.2.2, gesture rates were calculated separately for phrases with spatial content (SCPs) and phrases with other content. Overall, gesture rates were nearly 5 times higher SCPs than elsewhere (.498 vs. .101), and the difference was highly reliable $F(1, 40) = 204.5$, $p < .0001$). The means are plotted in Figure 7 above.

6.4 Effects of Restricting Gesturing on Speech

If lexical movements help in the process of lexical access. it's not unreasonable to suppose that preventing a speaker from gesturing would make lexical access more difficult, and that this would be directly reflected in less fluent speech. The experiment by Rauscher et al. (in press) referred to in Section 6.3.2 crossed three levels of lexical access difficulty (obscure, constrained and normal speech conditions) with a gesture-no gesture condition. Subjects were prevented from gesturing under the guise of recording skin conductance from their palms.

---

[22] Some abridged examples of descriptions incorporating these three dimensions are: Pitch: "The one you want here is the is the higher pitched one. It's a vibrating thing that increases. It ascends the scale…" Intensity: "What I perceive are two largely similar tones, the difference between the two being one is louder than the other. The one I would like you to select is the loudest of the two tones." Rate of periodicity: "Listen for the frequency of certain intervals and the one your looking for is slower . You're going to have…a certain number of notes played and then they'll repeat. So the one you are looking for, the intervals will be much slower." Object sound: "Sound two sounds more like someone is playing an electric organ or something fairly musically." Background/foreground: "This one …almost sounds like one tone, and then in the background you can just barely hear, sort of … like a tick-tock or something like that, whereas the other one, is more of two separate things, going on at the same time…"

This permitted assessment of the effects of not gesturing on several speech indices.

6.4.1 Speech rate and speech content

We know that lexical movements are more likely to occur when the conceptual content of speech is spatial (see Section 6.3.2). If our hypothesis that they enhance lexical access is correct, the detrimental effects of preventing a speaker from gesturing should be particularly marked for speech with spatial content. Rauscher et al. (in press) calculated their subjects' speech rates in words per minute (wpm) during spatial content phrases and elsewhere in the normal, obscure and constrained speech conditions. The speech conditions were designed to represent increasing levels of difficulty of lexical access, and it can be shown that they accomplished that goal.[23] Speakers spoke significantly more slowly in the obscure and constrained speech conditions than they did in the normal condition ($F(2, 80) = 75.90$, $p < .0001$). They also spoke more slowly when they were not permitted to gesture, but only in SCPs $F(1, 40) = 13.91$, $p < .001$); with nonspatial content, speakers spoke somewhat more rapidly when they could not gesture. Means for the 3 x 2 x 2 conditions are shown in Figure 9. It seems clear that the detrimental effects of preventing speakers from gesturing on fluency are limited specifically to speech whose conceptual content is spatial.

The fact that the effects of gesturing is restricted to speech with spatial content also reduces the plausibility of an alternative explanation for the results of this experiment. It might be argued that keeping one's hands still while talking is "unnatural" and requires cognitive effort, and that our results simply reflect diminished processing capacity due to having to remember not to gesture. However, such a "cognitive overload" explanation fails to account for the fact that the deleterious effects of preventing gesturing is specific to speech with spatial content. When the content of speech is nonspatial, speech rate *increases* slightly when gesturing is not allowed.

---

insert Figure 9 about here

6.4.2 Dysfluency and speech content

Problems in lexical access are a common cause of speech errors. Since speech production is an on-line process in which conceptualizing, formulating and articulating must occur in parallel, it is not unusual for a speaker to experience momentary difficulty locating a lexical item that will fulfill the semantic specifications set out at an earlier stage of the process. When this

---

[23]For example, both mean syllabic length of words in the latter two conditions and the type-token ratio (TTR) were greater in the latter two conditions than in the normal condition; syllabic length is related to frequency of usage (Zipf, 1935) and the TTR (the ratio of the number of different words in a sample [types] to the total number of words [tokens]) is a commonly-used measure of lexical diversity. Both are related to accessibility.

happens, the speaker may pause silently, utter a filled pause ("uh," "er, "um," etc.), incompletely articulate or repeat a word, restart the sentence, etc.

Rauscher et al. counted the total number of dysfluencies (long and short pauses, filled pauses, incompleted and repeated words, and restarted sentences) that occurred in spatial content phrases, and divided that number by the number of words in SCPs in that narrative; they did the same for dysfluencies that occurred in phrases without spatial content. These values were then subjected to a 2 (gesture condition) x 3 (speech condition) x 2 (content: spatial vs. nonspatial) ANOVA. Results paralleled those found for speech rate. Significant main effects were found for speech condition and content. Subjects were more dysfluent overall in the obscure and constrained speech conditions than in the natural condition (F (2, 78) = 38.32, $p < .0001$), and they are considerably more dysfluent during SCPs than elsewhere (F (1, 39) = 18.18, $p < .0001$). The two variables also interact significantly (F (2, 78) = 11.96, $p < .0001$). Finally, a significant gesture x speech condition x content interaction (F (2, 78) = 4.42, $p < .015$) reflects the fact that preventing gesturing has different effects on speech depending on whether its content is spatial or nonspatial. With spatial content, preventing gesturing increases the dysfluency rate, and with nonspatial content preventing gesturing has no effect. The means are shown in Figure 10.

---

insert Figure 10 about here

6.4.3 Filled pauses

Preventing speakers from gesturing negatively affects their ability to produce fluent speech when the content of that speech is spatial. However, a variety of factors can affect speech and dysfluency rates. Is it possible to ascertain whether this adverse effect is due specifically to the increased difficulty speakers experience accessing their lexicons when they cannot gesture, rather than some other factor? The measure that seems most directly to reflect difficulty in lexical access is the filled pause. Schachter, Christenfeld, Ravena and Bilous (1991) argue that filled pause rate is a consequence of the size of the lexicon from which words are selected; the filled pause rate in college lectures is predicted by the lecture's TTR.[24] A high TTR implies that more alternatives are being considered in lexical selection, which, by Hick's Law (Hick, 1952; Lounsbury, 1954) , should make lexical choice more difficult. A pause (filled or otherwise) can fall either at the boundary between grammatical clauses or within a clause. The former are often called juncture pauses and the later hesitations or nonjuncture pauses. Although juncture pauses have a variety of causes, nonjuncture pauses are believed to be attributable primarily to problems in lexical access (Butterworth, 1980).

---

[24]See fn. 23.

Rauscher et al. computed a 2 (gesture condition) x 3 (speech condition) ANOVA using as dependent variable the conditional probability of a nonjuncture filled pause (i.e., the probability of a nonjuncture filled pause, given a filled pause) in spatial content phrases. The means are plotted in Figure 11 below. Significant main effects were obtained for both speech condition (F (2, 80) = 49.39, $p < .0001$) and gesture condition (F (1, 40) = 8.50, $p < .006$). Making lexical access more difficult, by requiring speakers to use obscure words or forcing them to avoid words containing a particular letter, increased the proportion of nonjuncture filled pauses in their speech. Preventing speakers from gesturing had the same effect. With no constraints on speaking (the normal/gesture condition), about 25% of the filled pauses were nonjuncture (intraclausal). When subjects could not gesture, that number was increased to about 36%. Since nonjuncture filled pauses are most likely due to problems in word finding, these results indicate that preventing speakers from gesturing makes lexical access more difficult, and support the hypothesis that lexical movements aid in lexical access.

---

insert Figure 11 about here

---

7. General Discussion

It probably is fair to say that our work raises as many questions as it has answered. In the next section we will consider some of the important issues that remain unresolved.

7.1 How do gestures facilitate speech?

Taken as a whole, our data support the contention that lexical gestures facilitate lexical access. However, many of the details about the process by which gesturing affects speaking remain obscure and underspecified at the theoretical level. Among them are the following:

7.1.1 Gesturing and speech content

Although our data suggest that the generation of words and phrases with spatial content are affected by lexical gesture production, the limited nature of the corpora we have examined needs to be kept in mind, Without systematically sampling content areas, we can only speculate about the relationship of content and gesturing. If our theory of the origins of lexical movements is correct, we would expect that, along with spatial content, words expressing motion and action also would be activated by gestural cross modal priming. But it is only fair to admit that our present understanding of the properties of words that are likely to be accompanied by gestures is quite rudimentary.

Complicating matters is the possibility that many of the concepts that come to be represented in speech have spatio-dynamic features, and that these features may or may not be relevant to the discourse at a given time. This can come about in a variety of ways. For example, the same lexical item can be used both to designate a category of objects and a particular member of that category—"The *cake* was served by the bride" vs. "Coffee and *cake* is a usually served at the end the meeting." When the reference is to a particular cake *("the cake"),* the object in question will have a definite shape, but the generic *cake* leaves this feature unspecified. Cakes can be round, square, oblong, heart-shaped, etc., and this may or may not be part of the concept CAKE as it functions in the speaker's communicative intention.

Now imagine a speaker saying "He was completely surprised when a waiter put the birthday cake in front of him," accompanying the utterance with an encircling gesture. The presence of the definite article implies that the speaker had a specific cake in mind, and we would infer from the shape of the accompanying gestural movement that that cake was round. If English had different words for round and not-round cakes, or if like Navaho it obligatorily inflected concrete nouns for shape, the cake's shape would have been conveyed by the utterance regardless of its relevance to the speaker's communicative intention. Because English does not have such grammatical devices available, the speaker would have to employ a more complex expression to convey that information ("a round cake," "a heart-shaped cake," etc.). Consistent with a Gricean perspective (Grice, 1969, 1975), we would not expect speakers to do this unless shape was relevant to their communicative intention.

This brings us to the theoretical question. If shape is not a relevant feature of the word-concept *cake,* how could a gesture reflecting shape enhance lexical access? Yet it seems likely that speakers do indeed perform such gestures. Three possibilities occur to us: (1) It is possible that not all of what we are calling lexical gestures actually play a role in lexical access—that only gestures representing features incorporated in the lexical item's lemma serve this function. (2) Alternatively, the gesture may be thought of as comunicatively intended—a way for the speaker to convey information that is not part of the verbal message. This, we take it, is what Kendon means when he contends that gesturing "… is employed, along with speech, in fashioning an effective utterance unit" (Kendon 1983, p. 27). (3) Finally, it is possible that gestures of the sort we have described have an effect at the conceptualizing stage of speech production, acting as a kind of "motoric" imagery, helping the speaker retrieve from declarative memory the specific object or situation that will become part of the communicative intention. We will pursue this issue further in our discussion of "conceptual gestures" below.

7.1.2 Activation and termination of lexical movements

The model represented in Figure 4 presents a general view of the way gesture production and speech production systems interact. Within the broad outlines of this architecture, there are a number of ways the process could be implemented, and they have somewhat different theoretical implications. For example, the model is not specific about the mechanism that generates lexical movements—i.e., whether they are triggered by problems in lexical retrieval, or simply occur as an automatic product of the process that produces speech. One possibility (that might be called a *failure activitation* mechanism) is that lexical movements are elicited by the inability to retrieve a sought-for lexical entry: difficulties encountered in lexical access generate signals to the motor planner to initiate the motor program that results in a lexical movement; when lexical access is achieved, the absence of the signal terminates the lexical movement. Alternatively, the motor programs might be initiated independently of the speech processor's state, and their execution truncated or aborted by retrieval of a lemma that satisfies the specifications of the preverbal message. This might be termed a *retrieval-deactivation* mechanism. In this latter case, the gesture system must receive feedback from the speech processor when lexical retrieval has been accomplished in order to terminate the gesture.

7.1.3 "Conceptual" vs. "lexical" gestures

Although our model stresses the importance of gesturing for lexical access, it probably is the case as others have argued (Butterworth & Hadar, 1989; Hadar & Yadlin-Gedassy, 1994) that gestures also play a role at the conceptualizing stage of speech production. Speakers sometimes seem to use gestures to frame the contents of their communicative intentions, especially when the conceptual content relates to some overlearned motor act. Unfortunately, as we have noted above, distinguishing between lexical and conceptual gestures on formal grounds is unlikely to be satisfactory, and systematic study of the functions of conceptual gestures probably will require experiments that manipulate the conceptual content of speech.

7.1.4 Gestural representation of abstract concepts

It is not too difficult to imagine how gesturing might play a role in the production of speech that is rich in spatial or dynamic information. Many features of spatial or dynamic concepts can be depicted gesturally, and there are a number of potential ways the information contained in such lexical movements could aid the speaker. But people also gesture when their speech deals with such abstract matters as justice, love, finances and politics, and it is not always obvious how conceptual content of this sort can be represented gesturally. McNeill (1985) deals with this problem by distinguishing between *iconic* and *metaphoric* gestures. An iconic gesture represents its meaning pictographically; the gesture bears a physical resemblance to what it means. For metaphoric gestures, the connection between form and meaning is less direct, or at least less obvious. As McNeill puts it:

> Metaphoric gestures exhibit images of abstract concepts. In form and manner of execution, metaphoric gestures depict the vehicles of metaphors...The metaphors are independently motivated on the basis of cultural and linguistic knowledge (p. 356).

Essentially McNeill's contention is that metaphoric gestures are produced in the same way that linguistic metaphors are generated. However, we lack a satisfactory understanding of the processes by which linguistic metaphors are produced and comprehended is sketchy at best (cf. Glucksberg, 1991; in press), so to say that such gestures are visual metaphors may be little more than a way of saying that their iconicicity is not apparent.

An alternative view is that many abstract concepts that are not spatial or dynamic *per se* incorporate spatial or dynamic features. We often use such spatial and dynamic terms to refer to states of affairs in the world that are spatial or dynamic only in a figurative or nonliteral sense. For example, we use terms like *above* or *below* to refer to individuals' positions in social organizations, implicitly formulating social hierarchies in spatial terms. Such phrases as *He grasped the idea* and *Time moved slowly* would be anomalous if taken literally (Clark, 1973; Jackendorf, 1985). The process by which terms from one domain of experience are extended to other domains is called *semiotic extension* (McNeill, 1979).

7.1.5 Functions of motor movements

Although there is evidence that lexical movements play a role in lexical access, it is not clear what function is served by the other ubiquitous type of conversational gesture—motor movements. Theorists have proposed a number of quite different functions: they exert a structuring influence in the control of discourse production (Rimé, 1982); they disambiguate syntax (McClave, 1991); they serve as “extranarrative comments” (McNeil & Levy, 1982); they reflect the organization of the discourse (Kendon, 1980, 1983). The evidence offered thus far in support of these proposals is, in our judgment, inconclusive.

Neurolinguistic evidence underscores the differences between lexical and motor movements. For example, motor movements show a compensatory increase in Broca’s, but not in Werneicke’s, aphasia; the opposite is the case for lexical movements (Hadar, 1991). We have speculated that the two kinds of gestural movements are generated at different stages of the speech production process (Section 5.2.3). Motor movements appear more closely related to the motor aspects of speech production than to its conceptual content, and like Hadar we see similarities between them and other "coverbal" behaviors, such as head movements. Hadar's (1989a, 1989b) proposal that coverbal behaviors serve to coordinate the activities of the articulators seems plausible, although we know of no relevant evidence.

7.1.6 The significance of individual differences in gesturing

Although we have not presented the data here, in all of our studies we have observed substantial individual differences in the rate at which speakers gesture. For example, in the referential communication experiment involving novel abstract designs (described in Section 4.2.1), gesture rates across speakers ranged from 1.0 to 28.1 gestures per minute. These differences are reasonably consistent across content; the correlation between mean rates describing graphic designs and synthesized sounds was $r = 0.775$ ($p < .001$). In the Rauscher et al. (in press) study, gesture rates ranged from 0.5 to 30 per minute. Gestural frequency and form also are said to vary markedly from culture to culture, although the evidence for the claim is largely anecdotal. In an early study, Efron (1941/1972) reported his impressions of differences in the form and amplitude of conversational gestures of Italian and Jewish immigrants in New York. Certainly the belief that ethnic groups differ greatly in how frequently and energetically they gesture is common.

What accounts for inter-individual and inter-cultural differences in gesturing? One possibility is that they are stylistic, with no particular significance for the individual's cognitive functioning. To draw an analogy with linguistic variation, spoken languages differ in a variety of ways—they employ different speech sounds, different prosodic patterns, place the voice differently in the vocal range, etc. Similar sorts of variability can be observed among speakers of the same language. Generally speaking, these variations are not thought to have great cognitive significance, and the same may be true of inter-cultural or inter-individual differences in gesturing.

However, it is intriguing to speculate on the possibility that differences in the quantity and quality of gestures reflect differences in underlying cognitive processes. Unfortunately the evidence available at this point is far too sketchy to permit more than conjecture. In the dataset from the Rauscher et al. (in press) experiment, the single variable that best predicted individual differences in gesturing was speech rate; the more rapidly a speaker spoke, the more he or she gestured ($r(127) = 0.55$, $p < .0001$). The density of the information conveyed by each of the narratives from that experiment was assessed by counting the number of "idea units" per word (Butterworth & Beattie, 1978). Speech rate and idea rate were orthogonal ($r = 0.05$). The density of idea units in a narrative was modestly (but significantly) correlated with the amount of gesturing that accompanied it ($r(127) = .18$, $p < .04$). Using a median split, we divided the narratives into those that were informationally dense (high idea unit rate) and those that were informationally sparse (low idea unit rate) groups, and then recomputed the correlations of speech rate and gesturing within-group. For the informationally sparse narratives, speech rate accounted for more than 50% of the variability in gesturing ($r(62) = 0.72$, $p < .0001$), while in the informationally dense narratives it accounted for considerably less ($r(58) = 0.42$, $p < .001$). Similarly although dysfluency rate is uncorrelated with idea unit rate overall, the relationship within the two groups differs markedly (low idea rate group: $r(123) = -0.54$, $p < .0001$; high idea rate group: $r(122) = -0.04$), These data (and others like it) point to the possibility that gestures may serve different strategic purposes for different speakers. A high rate of gesturing may mean quite different things in information-rich and information-sparse narratives and, perhaps, for speakers who habitually speak succinctly and those whose speech is more discursive.

Such strategies might derive from different ways speakers habitually conceptualize the world. For example, a person can be described as "*grasping* the point" or "*understanding* the point." In both cases, it is clear that the reference is to an act of comprehension, and *grasping* might be thought of as a metaphoric description of comprehending. Are *grasping* and *understanding* simply synonyms in this context, conveying the same intended meaning, or does the fact that the speaker metaphorically likens comprehension to a physical act reveal something about the way comprehension is conceived? We know that gesturing is associated with spatial content, and believe that it is likely to accompany motoric content as well. What we do not know is whether "*grasping* the point" and "*grasping* the rope" are equally likely to be accompanied by gestures.

It may seem farfetched, and certainly we know of no data that is relevant, but is it possible some people gesture more than others because they habitually think about the world in spatial or motoric terms? In the Rauscher et al. experiment, subjects differed systematically in the extent to which they used spatial language. In each narrative, we calculated the percentage of phrases that were spatial. It ranged from 10 to 53 percent, with a median of 31 percent. Each subject contributed six narratives, and we would expect the number greater than and less than 31 percent to be roughly equal if subjects are not consistent in their use of spatial language. In fact the value of χ^2 for the 2 (above/below median) x 41 (subjects) contingency table was 127.21 ($p = .001$). These difference are not attributable to content, since subjects were describing the same six "Roadrunner" cartoons. Apparently subjects differed in the extent to which they conceptualized the events of those cartoons in spatial terms, and this was reflected in their speech. It also was reflected in their gesturing. A multiple regression with percent of spatial phrases and speech rate as independent variables accounted for one-third of the variance in the proportion of speaking time during which the speaker gestured ($r(127) = .578$, $p < .0001$).

7.1.7 The role of gesturally-conveyed nonsemantic information

Most studies of the communicative value of conversational gestures have focused on the way they help to convey a message's intended meaning—what we are calling semantic information—with little consideration of the possibility that they may be a rich source of other kinds of information. Our data have lead us to conclude that the amounts of semantic information conversational gestures typically convey is small and, except under special circumstances, probably insufficient to make an important contribution to listener comprehension. At the same time, we recognize that communicative interchanges do not end with the addressee's identification of the speaker's communicative intention. The response to a communicative act often takes into account the addressee's perception of the speaker's *perlocutionary intention*—i.e., the result the communicative act is designed to accomplish (Krauss & Fussell, in press). When a used car salesman represents a battered jalopy as having been owned by a retired teacher who drove it only on Sundays to and from church, certainly the addressee will comprehend the salesman's intended meaning, but any response to the assertion is likely to be tempered by the addressee's perceptions of what it was intended to achieve.

It has been suggested that nonverbal behaviors can play a role in such judgments, and in this way make an important contribution to communication, Nonverbal behaviors, or certain aspects of nonverbal behavior, can provide information about the individual's internal state that is independent of the message's intended meaning. Discrepancies between this information and the information in the message are a likely source of attributions about the communicator's perlocutionary intentions (cf., DePaulo, 1992). The term that often has been used to refer to this process is "nonverbal leakage" (Ekman & Friesen, 1969a, 1974), but that may represent too narrow a view of the process. The nonverbal behaviors that form the basis of perceptions of state are intrinsic parts of the communicative act. We expect some pitch elevation in the voices of people responding to a stress-inducing question, but when they respond the same way to a neutral question, we are likely to seek an explanation (Apple, Streeter & Krauss, 1979).

As the research described in Section 4.3.1 indicates, subjects can discriminate spontaneous from rehearsed versions of the same narrative from viewing the video track, without hearing the accompanying sound (Chawla & Krauss, 1994). suggesting that the visual information contains cues relevant to the speaker's spontaneity. Although spontaneous and rehearsed speakers gestured equally often, a greater proportion of the spontaneous speakers' gestures were lexical movements. Judgments of spontaneity were reliably correlated with the proportion of time the speakers spent making lexical movements, but not with the total time spent gesturing, suggesting that lexical movements served as a cue to spontaneity. However, despite this correlation, raters seldom mentioned gesturing as a one of the cues they used to judge spontaneity. We appear to have an intuitive appreciation of the way lexical movements relate to spontaneous speech production, just as we intuitively understand the significance of dysfluency for speech production. It remains to be seen what other sorts of inferences listener/viewers can draw from their partner's gestural behavior.

7.2 A summing up: What *DO* conversational hand gestures tell us?

On closer examination, the chapter's subtitle *(What do conversational hand gestures tell us?)* reveals itself to be pragmatically ambiguous. It can be interpreted in at least three different ways, depending upon who is taken to be the referent of *us,* and what is understood as the implicit indirect object ("What do they tell us *about what?").*

7.2.1 How do gestures contribute to comprehension?

On one interpretation, the *us* refers to the addressee, and the *about what* to the information the gesture conveys to that addressee. It is the question that traditionally has been asked about gestures and focuses on their *inter*personal function—how do they contribute to our understanding of the speaker's message? Our brief answer is that in the situations we have studied they contribute relatively little. Contrary to Edward Sapir's familiar aphorism, gestures do not seem to constitute "…an elaborate and secret code that is written nowhere, known to none, and understood by all" (Sapir, 1949, p. 556) . Certainly it is true that our methods for measuring the amount of semantic information gestures convey are indirect and crude. Nevertheless, such evidence as we have indicates that the amount of information conversational gestures convey is very small— probably too small relative to the information conveyed by speech to be of much communicative value. Could there be special circumstances in which conversational gestures are especially useful? Certainly one can imagine that being the case, but at this point we have little understanding of what the circumstances might be or precisely how the gestures might contribute to comprehension.

There is, however, some evidence that gestures can convey *nonsemantic* information, and it is not too difficult to think of circumstances in which such information could be useful. Here, the study of speech and gestures overlaps with the study of person perception and attribution processes, because gestures, in their cultural and social context, may enter into the process by which we draw conclusions about people—their backgrounds, their personalities, their motives and intentions, their moods and emotions, etc. Further, since the significance of gestures can be ambiguous, it is likely that our beliefs and expectations about the speaker-gesturer will affect the meanings and consequences we attribute to the gestures we observe.

Another way of pursuing this question is to ask how gesturing affects the way listeners process verbal information. Do gestures help engage a listener's attention? Do they activate imagistic or motoric representations in the listener's mind? Do they become incorporated into representations that are invoked by the listener when the conversation is recalled? One hypothesis, currently being tested in our laboratory, is that gestures facilitate the processes by which listeners construct mental models of the events and situations described in a narrative. Communication has been defined as the process by which representations that exist in one person's mind come to exist in another's (Sperber & Wilson, 1986). If our hypothesis is correct, gestures may affect the nature of such representations, and thus contribute importantly to at least some kinds of communication.

7.2.2 How does gesturing affect speech?

On a second construal, the question *What do conversational hand gestures tell us?* concerns the *intra*personal functions of gesture—here, the role they play in speech production. It might be paraphrased "How does gesturing affect us when we speak? The *us* in this interpretation is the speaker, and the *about what* has to do with the ideas the speaker is trying to articulate in speech. Our response to this question is that gestures are an intrinsic part of the process that produces speech, and that they aid in the process of lexical access, especially when the words refer to concepts that are represented in spatial or motoric terms. They "tell us" about the concepts underlying our communicative intentions that we seek to express verbally. In this way, conversational gestures may indirectly serve the function conventionally attributed to them. That is, they may indeed enhance the communicativeness of speech, not by conveying information that is apprehended visually by the addressee, but by helping the speaker formulate speech that more adequately conveys the communicative intention.

7.2.3 What can we learn from studying conversational gestures?

In a third interpretation the *us* refers to those of us who study behaviors like gestures, and the *about what* refers to the process by which speech and gesture are generated. Our response to this question is the most speculative, but in some ways it is to us the most interesting. It involves the ways we represent and think about the experienced world, and the ways such representations come to be manifested in speech when we communicate.

Considering the functions of conversational gestures reminds us that although linguistic representations derive from propositional representations of experience, not all mental representation is propositional. Spatial knowledge and motoric knowledge may have their own representational formats, and some components of emotional experience seem to be represented somatically. These representations (perhaps along with others) will be accessed when we recall and think about these experiences. However, when we try to convey such experiences linguistically, we must create new representations of them, and there is some evidence that so doing can change how we think about them. For example, describing a face makes it *more* difficult to recognize that face subsequently (Schooler & Engsteller-Schooler, 1990), and this "verbal overshadowing" effect, as it has been termed, is not limited to representations of visual stimuli (Schooler, Ohlsson, & Brooks, 1993; Wilson, Lisle, Schooler, Hidges, Klaaren, & LaFleur, 1993; Wilson & Schooler, 1991). Linguistic representations may contain information that was not part of the original representations, or omit information that was. It is possible that gestures affect the internal representation and experience of the conceptual content of the speech they accompany, much as facial expressions are believed to affect the experience of emotion.

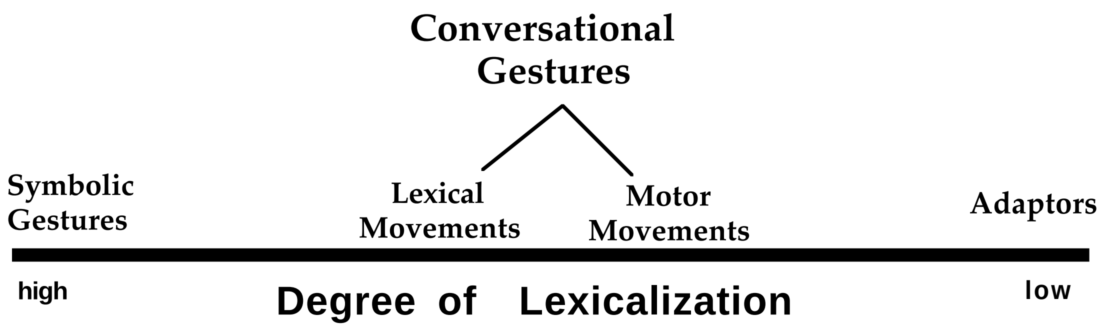
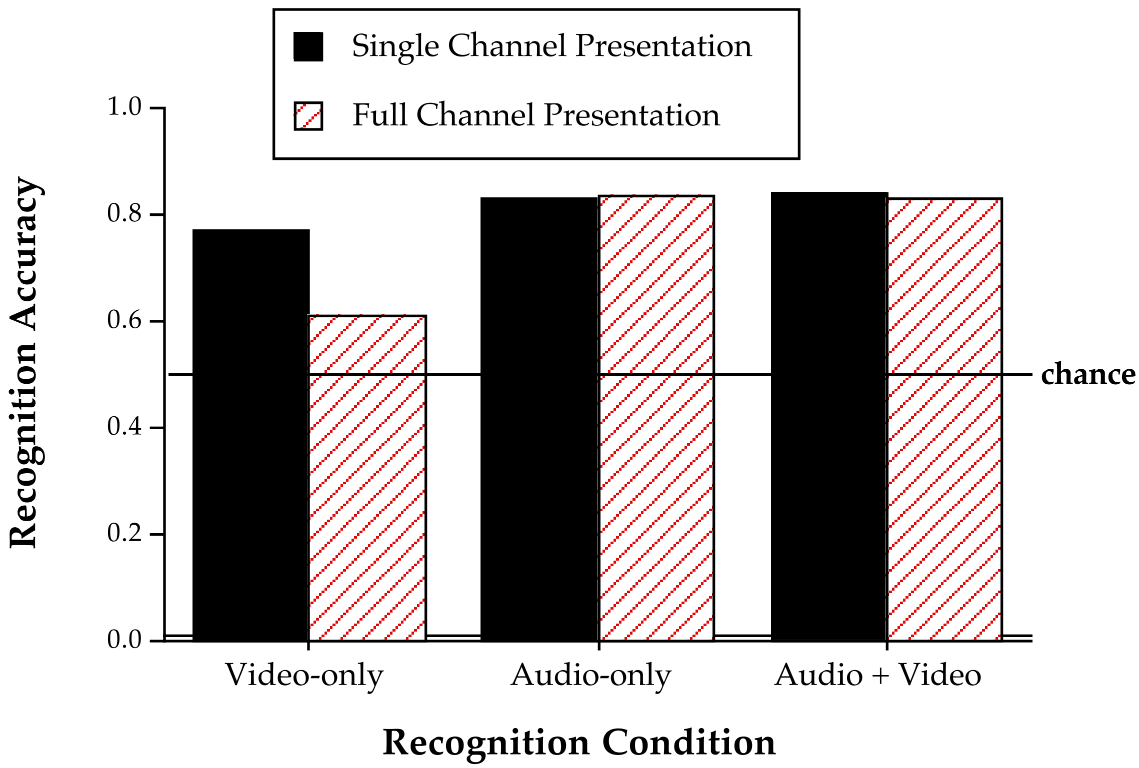

Conversational Gestures
Symbolic Gestures
Lexical Movements
Motor Movements
Adaptors
high
Degree of Lexicalization
low

Figure 1



Single Channel Presentation
Full Channel Presentation
1.0
0.8
0.6
0.4
0.2
0.0
Recognition Accuracy
chance
Video-only
Audio-only
Audio + Video
Recognition Condition

Figure 2

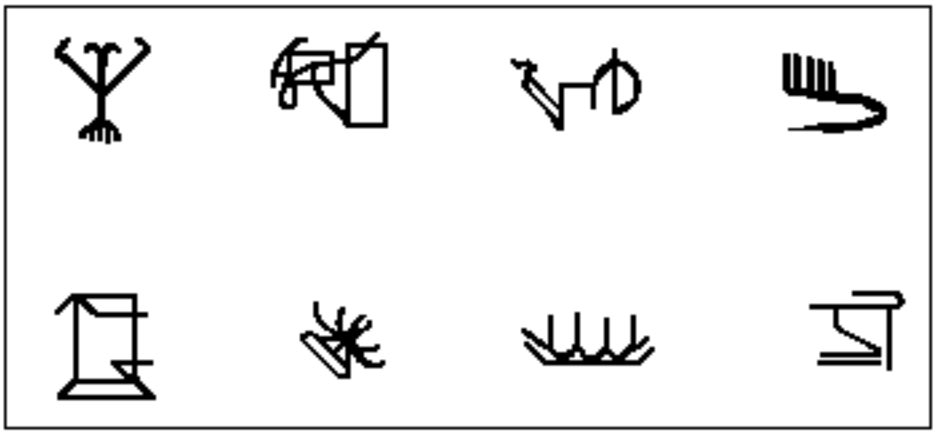


Figure 3

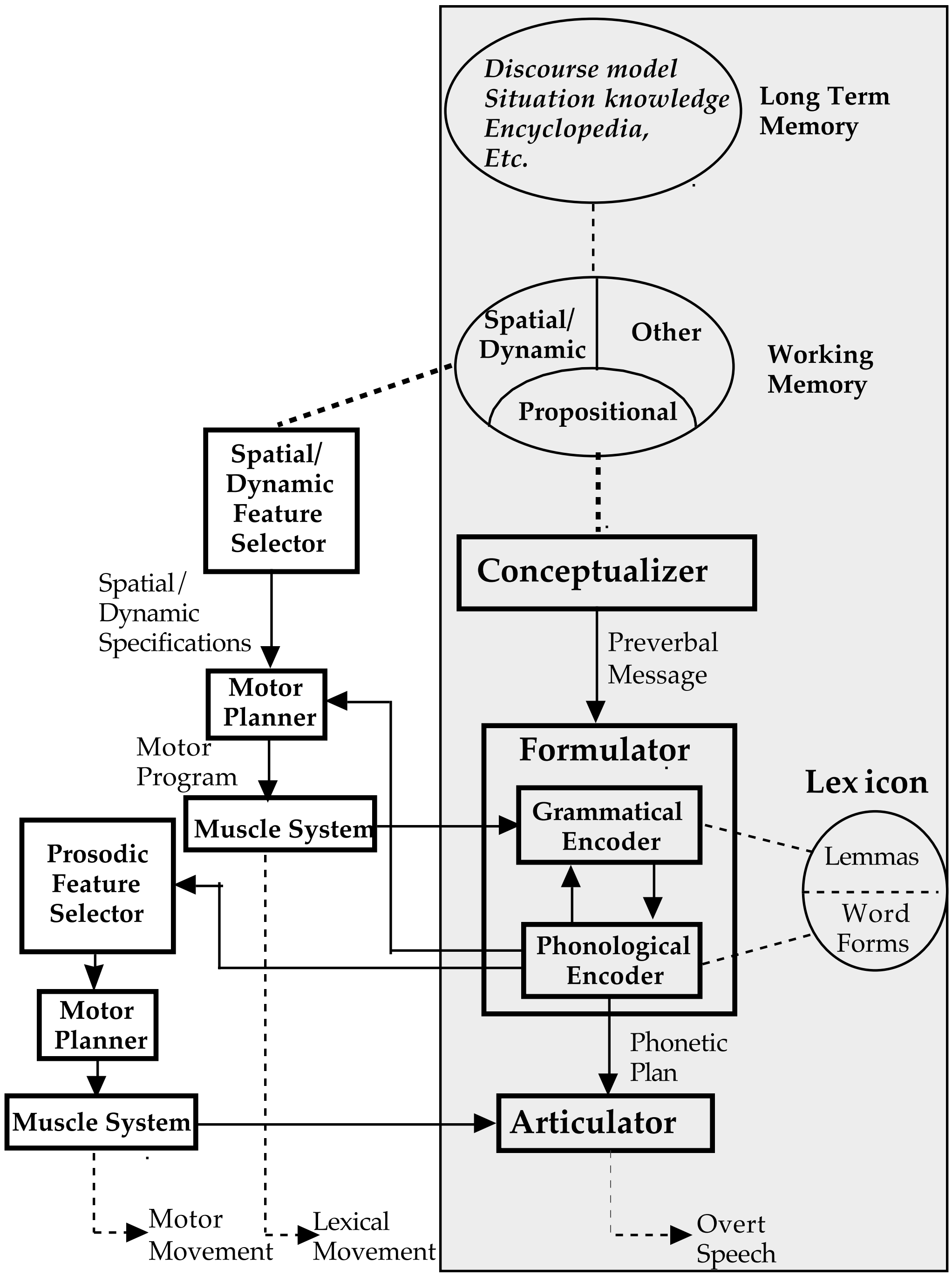

Discourse model
Situation knowledge
Encyclopedia,
Etc.
Long Term Memory
Spatial/ Dynamic
Other
Propositional
Working Memory
Spatial/ Dynamic Feature Selector
Spatial/ Dynamic Specifications
Conceptualizer
Preverbal Message
Motor Planner
Motor Program
Muscle System
Formulator
Lex icon
Grammatical Encoder
Lemmas
Prosodic Feature Selector
Word Forms
Phonological Encoder
Motor Planner
Phonetic Plan
Muscle System
Articulator
Motor Movement
Lexical Movement
Overt Speech

Figure 4

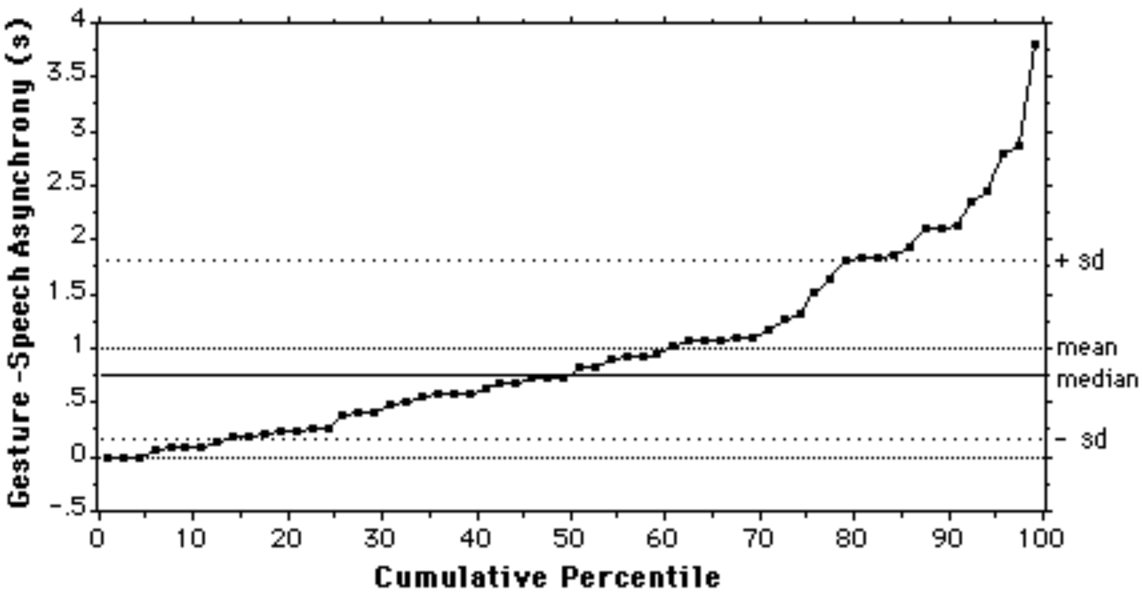
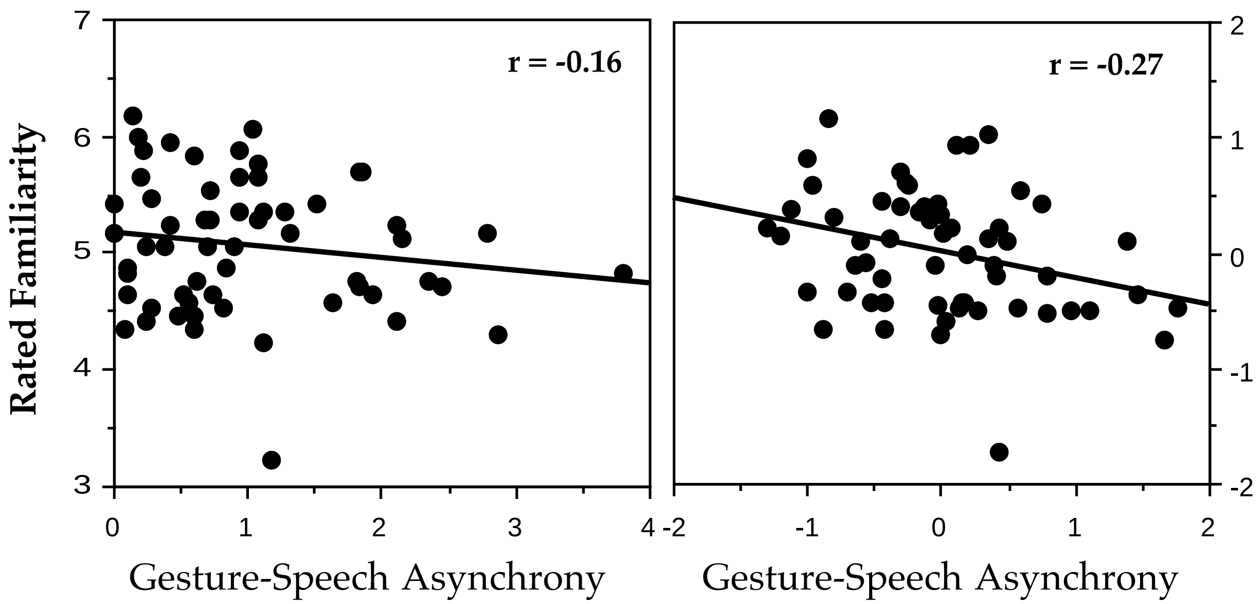

Gesture -Speech Asynchrony (s)
4
3.5
3
2.5
2
1.5
1
.5
0
-.5
0
10
20
30
40
50
60
70
80
90
100
Cumulative Percentile
+ sd
mean
median
- sd

Figure 5



Rated Familiarity
r = -0.16
r = -0.27
7
6
5
4
3
2
1
0
-1
-2
0
1
2
3
4
-2
-1
0
1
2
Gesture-Speech Asynchrony
Gesture-Speech Asynchrony

Figure 6

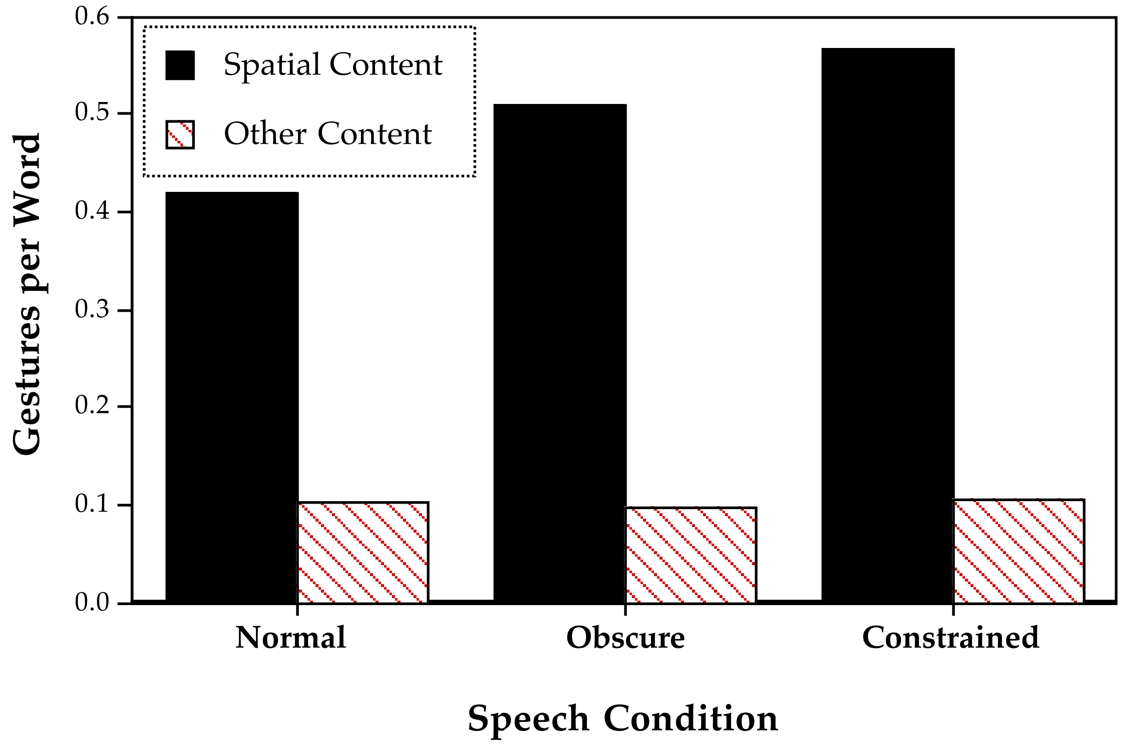

0.6
0.5
0.4
0.3
0.2
0.1
0.0
Spatial Content
Other Content
Gestures per Word
Normal
Obscure
Constrained
Speech Condition

Figure 7

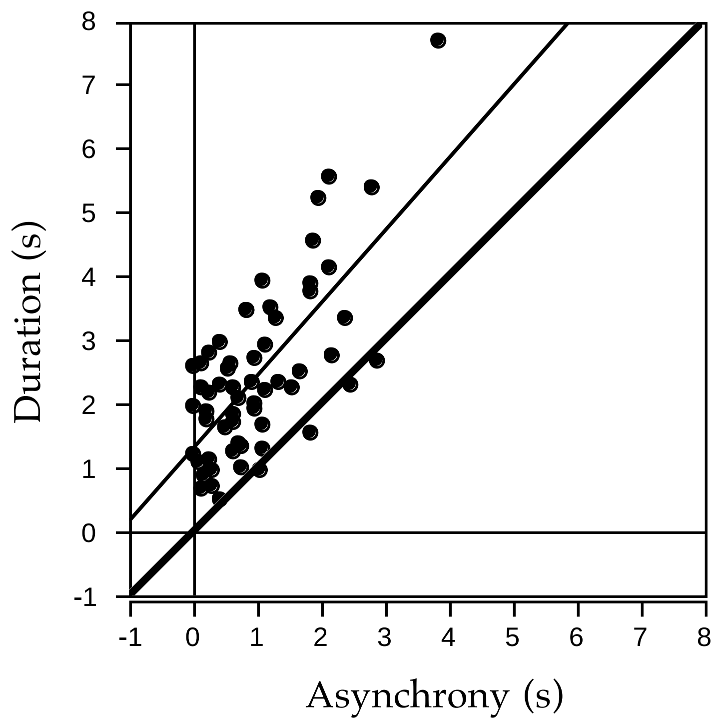


Figure 8

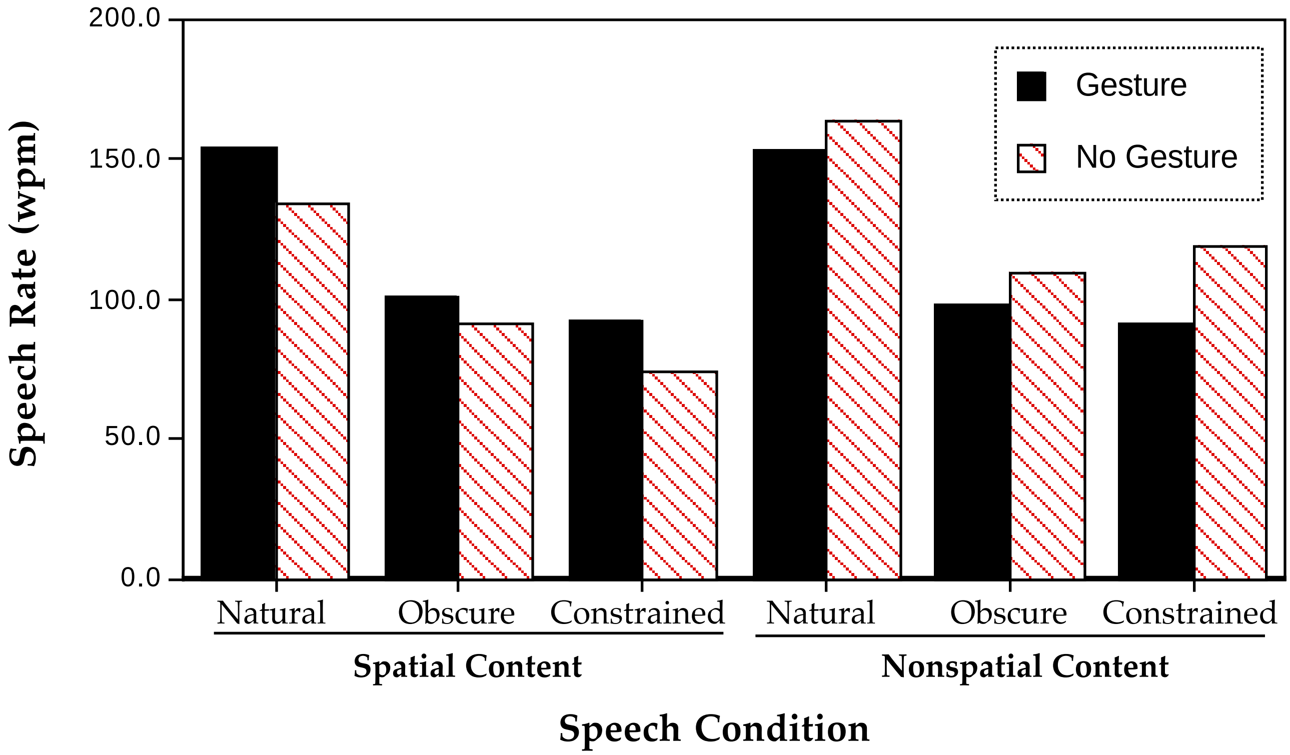

Speech Rate (wpm)
200.0
150.0
100.0
50.0
0.0
Gesture
No Gesture
Natural
Obscure
Constrained
Natural
Obscure
Constrained
Spatial Content
Nonspatial Content
Speech Condition

Figure 9

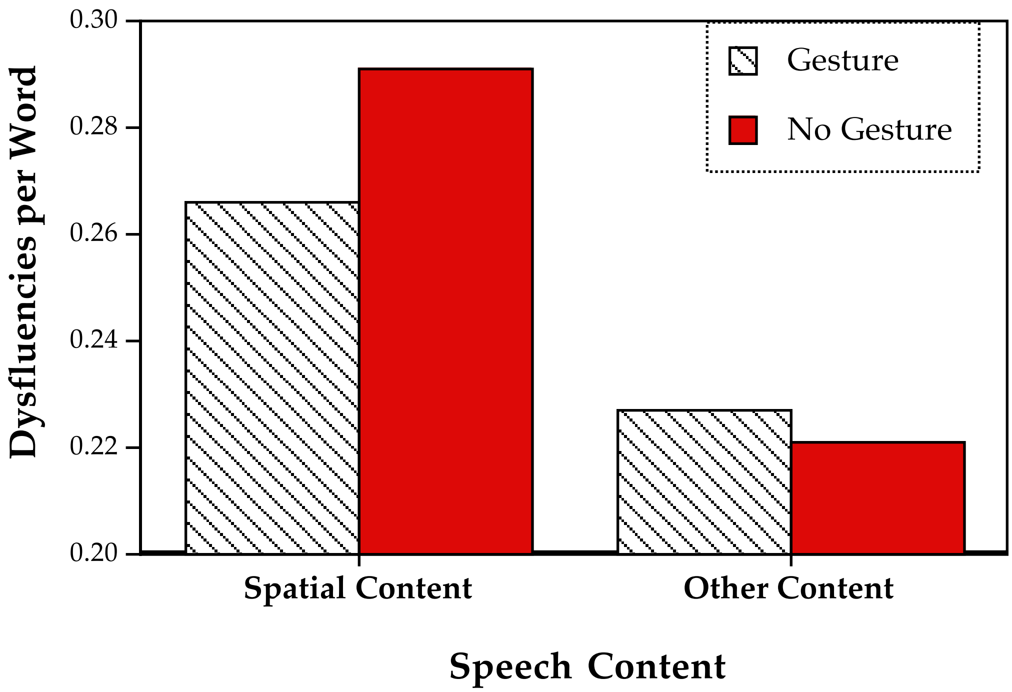

0.30
0.28
0.26
0.24
0.22
0.20
Dysfluencies per Word
Gesture
No Gesture
Spatial Content
Other Content
Speech Content

Figure 10

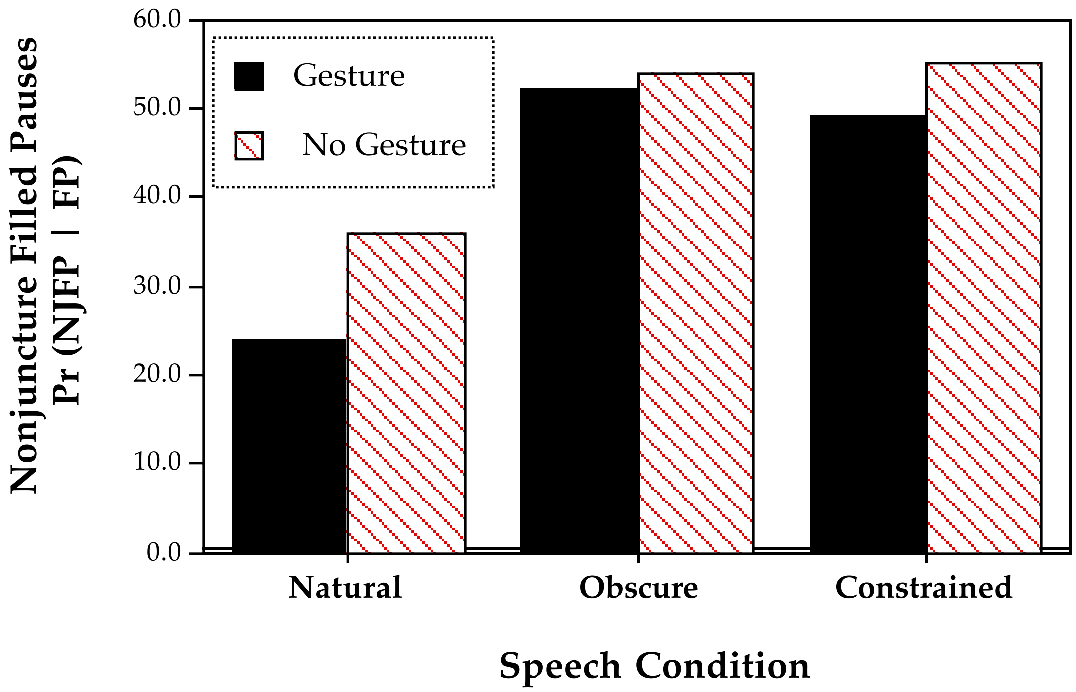

Nonjuncture Filled Pauses
Pr (NJFP | FP)
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Gesture
No Gesture
Natural
Obscure
Constrained
Speech Condition

Figure 11

Semantic Category Judged from

a. Audio + Video

Semantic Category of Lexical Affiliate		A	L	O	D
	A	91	7	5	27
	L	32	110	6	52
	O	4	28	52	46
	D	12	31	12	85
	Σ	139	166	85	210
	%	23.2	27.7	14.2	35.0

b. Video-0nly

	A	L	O	D	%
A	54	13	20	43	21.7
L	55	44	28	73	33.3
O	28	22	32	48	21.7
D	26	21	20	73	23.3
Σ	163	100	100	237	
%	27.2	16.7	16.7	39.5	

c. Audio-only

Semantic Category of Lexical Affiliate		A	L	O	D
	A	69	9	6	46
	L	18	100	16	66
	O	3	5	47	75
	D	9	21	30	80
	Σ	99	135	99	267
	%	16.5	22.5	16.5	44.0

d. Transcript

	A	L	O	D	%
A	91	1	2	36	21.7
L	28	107	11	54	33.3
O	0	0	109	21	21.7
D	5	20	6	109	23.3
Σ	124	128	128	220	
%	20.7	21.3	21.3	36.7	

Table 1

Subjects' assignments of gestures or speech to semantic category as a function of the semantic category of the lexical affiliate, shown for judgments made from (a) audio + video, (b) video-only, (c) audio-only and (d) transcript. (A = actions; L = locations; O = object names; and D = descriptions.) N = 600 (10 subjects x 60 judgments) per matrix

	A+V	Aud	Tran	Vid
Audio+Video	--	1.61	3.14	6.52**
Audio-only	--	--	5.33*	17.6***
Transcript	--	--	--	12.85***
Video-only	--	--	--	--

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .0001$

Table 2

Value of multivariate *F*-ratios (Wilk's λ) for between-condition contrasts. All contrasts with 12,7 df.

Encoding Condition	Decoding Condition: Designs, Audio-Only	Designs, Audio-Video		Sounds, Audio-Only	Sounds, Audio-Video	
Intercom	.696	.692	.694	.669	.685	.677
Face-to-Face	.614	.667	.642	.635	.644	.639
	.655	.679		.652	.664	

Table 3

Accuracy of identification (proportion correct) of novel figures and sounds as a function of encoding and decoding condition.

Encoding Condition	Decoding Condition: Audio-Only	Decoding Condition: Audio-Video	
Intercom	.528	.541	.535
Face-to-Face	.586	.566	..575
	.557	.554	

Table 4

Accuracy of identification of tea samples (proportion correct) as a function of encoding and decoding condition. (Values in parentheses are standard deviations.)

FIGURE CAPTIONS

Figure 1: A continuum of gesture types.

Figure 2: Recognition accuracy (proportion correct) for videotaped segments from video-only, audio-only and audio-video presentations.

Figure 3: A sample of the novel graphic designs used as stimuli.

Figure 4: A model of speech and gesture production processes. Boxes represent processing components; circle and ellipses represent knowledge stores. The speech-production (shaded) section is adapted directly from Levelt (1989); components of the gesture-production section were suggested by a rather different global architecture proposed informally by Jan Peter de Ruiter.

Figure 5: Cumulative distribution of lexical movement-lexical affiliate asynchronies.

Figure 6: Mean familiarity ratings for the 60 lexical affiliates plotted against its lexical movement-lexical affiliate asynchrony (in s). Left panel shows relationship before spatial extent and number of syllables have been partialled and right panel shows relationship after partialling .

Figure 7: Gesture rate (time spent gesturing / number of words in phrase) in spatial content phrases and elsewhere in the natural, obscure and constrained speech conditions.

Figure 8: Duration of lexical movement plotted against lexical movement-lexical affiliate asynchrony (both in s). The heavier line is the unit line; the lighter line above it is the least-squares regression line (see text for explanation).

Figure 9: Speech rate (words per minute) in the natural, obscure, and constrained speech conditions for spatial and nonspatial content when subjects could and could not gesture.

Figure 10: Dysfluency rates (number of long and short pauses, filled pauses, incompleted and repeated words, and restarted sentences per word) in gesture and no gesture conditions for spatial and nonspatial content

Figure 11: Conditional probability of nonjuncture filled pause (Pr (NonJ FP | FP) in three speech conditions when subjects could and could not gesture.

ACKNOWLEDGMENTS

We gratefully acknowledge the helpful advice, comments and suggestions of Susan Fussell, Uri Hadar, Julian Hochberg, Lois Putnam, and Mark Zanna. Christine Colasante, Robert Dushay, Palmer Morrel-Samuels and Frances Rauscher were part of in the research program described here, and their contributions are indicated in conjunction with the specific studies in which they participated. The research and preparation of this report were supported by National Science Foundation grants BNS 86-16131 and SBR 93-10586 to the first author. Correspondence should be addressed to the first author at Department of Psychology, 1190 Amsterdam Avenue, 402B Schermerhorn Hall, Columbia University, New York, NY 10027. E-mail: rmk@paradox.psych.columbia.edu

REFERENCES

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, Eng.: Cambridge University Press.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, **37**, 715-727.

Bacon, F. (1891). *The advancement of learning, Book 2.* (4 ed.). London: Oxford University Press.

Barakat, R. (1973). Arabic gestures. *Journal of Popular Culture,* **6**, 749-792.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, **15**, 469-489.

Beattie, G. W. (1978). Sequential patterns of speech and gaze in dialogue. *Semiotica*, **23**, 29-52.

Beattie, G. W. (1981). A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. *British Journal of Social Psychology*, **20**, 243-248.

Birdwhistell, R. L. (1970). *Kinesics and context*. Philadelphia: University of Pennsylvania Press.

Brunner, L. J. (1979). Smiles can be backchannels. *Journal of Personality and Social Psychology*, 37, 728-734.

Bull, P. (1983). *Body movement and interpersonal communication*. London: Wiley.

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior*, **9**, 169-187.

Bull, P. E. (1987). *Gesture and posture*. Oxford, Eng.: Pergamon Press.

Butterworth, B. (1978). Maxims for studying conversations. *Semiotica*, **24**, 317-339.

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Speech and talk*. London: Academic Press.

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*. New York: Plenum.

Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stage. *Psychological Review*, **96**, 168-174.

Cegala, D. J., Alexander, A. F., & Sokuvitz, S. (1979). An investigation of eye gaze and its relation to selected verbal behavior. *Human Communications Research*, **5**, 99-108.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. **30**, 580-601.

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.

Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* . New York: Academic Press.

Cohen, A. A. (1977). The communicative functions of hand illustrators. *Journal of Communication*, **27**, 54-63.

Cohen, A. A., & Harrison, R. P. (1972). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, **28**, 276-279.

Darwin, C. R. (1872). *The expression of the emotions in man and animals*. London: Albemarle.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Review*, **111**, 203-243.

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

Dittmann, A. T. (1977). The role of body movement in communication. In A. W. Siegman & S. Feldstein (Ed.), *Nonverbal behavior and nonverbal communication* . Hillsdale, NJ: Erlbaum.

Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. *Iazykovedenie i Materializm*, 105-173.

Duncan, S. J., Brunner, L. J., & Fiske, D. W. (1979). Stratey signals in face-to-face interaction. *Journal of Personality and Social Psychology*, **37**, 301-313.

Dushay, R. D. (1991). *The association of gestures with speech: A reassessment*. Unpublished doctoral dissertation Dissertation, Columbia University.

Edelman, R., & Hampson, S. (1979). Changes in non-verbal behavior during embarrassment. *British Journal of Social and Clinical Psychology*, **18**, 385-390.

Efron, D. (1941/1972). *Gesture, race and culture*. The Hague: Mouton (first edition 1941).

Ekman, P. (1976). Movements with precise meanings. *Journal of Communication*, **26**, 14-26.

Ekman, P., & Friesen, W. V. (1969a). Nonverbal leakage and clues to deception. *Psychiatry*, **32**, 88-106.

Ekman, P., & Friesen, W. V. (1969b). The repertoire of nonverbal communication: Categories, origins, usage, and coding. *Semiotica,*, **1**, 49-98.

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

Ekman, P., & Friesen, W. V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology*, **29**, 288-298.

Exline, R. V. (1972). Visual interaction: The glances of power and preference. In J. K. Cole (Ed.), *Nebraska symposium on motivation (Vol. 19)* (pp. 163-206). Lincoln: University of Nebraska Press.

Exline, R. V., Gray, D., & Schuette, D. (1985). Visual behavior in a dyad as affected by interview content and sex of respondent. *Journalof Personality and Social Psychology*, **1**, 201-209.

Feldman, R. S., & Rimé, B. (1991). *Fundamentals of nonverbal behavior*. New York: Cambridge University Press,

Feyereisen, P., & deLannoy, J.-D. (1991). *Gesture and speech: Psychological investigations*. Cambridge: Cambridge University Press.

Feyereisen, P., Van de Wiele, M. □., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive*, **8**, 3-25.

Fleming, J. H., & Darley, J. M. (1991). Mixed messages: The multiple audience problem and strategic social communication. **9**, 25-46.

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

Freedman, N., & Hoffman, S. (1967). Kinetic behavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Perceptual and Motor Skills*, **24**, 527-539.

Fridlund, A. J. (1991). *Darwin's anti-Darwinism in The Expression of the Emotions in Man and Animals* . University of California.

Fussell, S., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, **25**, 203-219.

Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, **19**, 509-526.

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

Glucksberg, S. (1991). Beyond literal meanings: The psychology of allusion. *Psychological Science*, **2**, 146-152.

Glucksberg, S. (in press). How metaphors work. In A. Ortony (Ed.), *Metaphor and thought (2nd edition)* . Cambridge, Eng.: Cambridge University Press.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, **10**, 57-67.

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

Grice, H. P. (1969). Utterer's meaning and intentions. *Philosophical Review*, **78**, 147-177.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Ed.), *Syntax and semantics: Speech acts* . New York: Academic Press.

Hadar, U. (1989a). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology*, **8**, 221-228.

Hadar, U. (1989b). Gestural modulation of speech production: The role of head movement. *Language and Communication*, **9**, 245-257.

Hadar, U. (1991). Speech-related body movement in aphasia: Period analysis of upper arm and head movement. *Brain and Language*, **41**, 339-366.

Hadar, U., Burstein, A., Krauss, R.M. & Soroker, N. (1995). Visual imagery and word retrieval as factors in the generation of ideational gestures in brain-damaged and normal speakers. *Unpublished paper, Tel Aviv University.*

Hadar, U., & Butterworth, B. (1993). *Iconic gestures, imagery and word retrieval in speech.* Unpublished manuscript, Tel Aviv University.

Hadar, U., & Yadlin-Gedassy, S. (1994). Conceptual and lexical aspects of gesture: Evidence from aphasia. *Journal of Neurolinguistics,* **8**, 57-65.

Heller, J. F., Miller, A. N., Reiner, C. D., Rupp, C., & Tweh, M. (1995). Beat gesturing during conversation: Markers for novel information. Poster presented at meetings of the Eastern Psychological Association.

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). Temporal patterns of cognitive activity and breath control in speech. **8**, 236-242.

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.

Jackendorf, R. (1985). *Semantics and cognition*. Cambridge, MA: MIT Press.

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

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

Kimura, D. (1976). The neural basis of language qua gesture. In H. Whitaker & H. A. Whitaker (Ed.), *Studies in neurolinguistics* . New York: Academic press.

Krauss, R. M. (1993). Nonverbal behaviors *à la carte* (review of Robert S. Feldman & Bernard Rimé (Eds.), "Fundamentals of nonverbal behavior). **38**, 507-508.

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology*, **40**, 312-320.

Krauss, R. M., Dushay, R. A., Chen, Y., & Bilous, F. (in press). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology.*

Krauss, R. M., & Fussell, S. R. (in press). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Ed.), *Social psychology: A handbook of basic principles* . New York: Guilford.

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American,* **236**, 100-105.

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. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, **4**, 343-346.

Kraut, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, **37,** 1539-1553.

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

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

Lickiss, K. P., & Wellens, A. R. (1978). Effects of visual accessibility and hand restraint on fluency of gesticulator and effectiveness of message. *Perceptual and Motor Skills,* **46**, 925-926.

Lounsbury, F. G. (1954). Transitional probability, linguistic structure, and systems of habit-family heirarchies. In C. E. Osgood & T. Sebeok (Ed.), *Psycholinguistics: A survey of theory and research problems* . Bloomington, Indiana: University of Indiana Press.

Mahl, G. (1956). Disturbances and silences in the patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology*, **53**, 1-15.

Mahl, G. F. (1968). Gestures and body movement in interviews. In J. Schlien (Ed.), *Research in psychotherapy* . Washington, DC: American Psychological Association.

McClave, E. Z. (1991). *Intonation and gesture*. Unpublished doctoral dissertation, Georgetown University.

McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research,* **23**, 45-66.

McNeill, D. (1979). *The conceptual basis of language*(. Hillsdale, NJ: Erlbaum.

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.

Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.

Morrel-Samuels, P. (1989). *Gesture, word and meaning: The role of gesture in speech production and comprehension*. Unpublished doctoral dissertation, Columbia University.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **18**, 615-623.

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

Rauscher, F. B., Krauss, R. M., & Chen, Y. (in press). Conversational hand gestures, speech and lexical access: The role of lexical movements in speech production. *Psychological Science.*

Reuschert, E. (1909). *Die gebardensprache die taubstummen*. Leipzig: von University Press.

Ricci Bitti, P. E., & Poggi, I. A. (1991). Symbolic nonverbal behavior: Talking through gestures. In R. S. Feldman & B. Rimé (Ed.), *Fundamentals of nonverbal behavior* (pp. 433-457). New York: Cambridge University Press.

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal behaviour. *European Journal of Social Psychology*, **12**, 113-129.

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

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, **5**, 172-183.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behaviors within utterances. *Human Communication Research*, **5**, 54-62.

Rosenfeld, H. (1966). Instrumental affiliative functions of facial and gestural expressions. *Journal of Personality and Social Psychology*, **4**, 65-72.

Russo, N. F. (1975). Eye contact, interpersonal distance, and the equilibrium theory. *Journal of Personality and Social Psychology*, **31**, 497-502.

Rutter, D. (1987). *Communicating by telephone.* Oxford, Eng.: Pergamon Press.

Rutter, D. R., Stephenson, G. M., & Dewey, M. E. (1981). Visual communication and the content and style of communication. *British Journal of Social Psychology*, **20**, 41-52.

Sapir, E. (1949). The unconscious patterning of behavior in society. In D. Mandelbaum (Ed.), *Selected writing of Edward Sapir in language, culture and personality* (pp. 544-559). Berkeley, CA: University of California Press.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology\,* **60**, 362-367.

Schegloff, E. (1984). On some gestures' relation to speech. In J. M. Atkinson & J. Heritage (Ed.), *Structures of social action* . Cambridge: Cambridge University Press.

Scherer, K. R., & Giles, H. (1979). *Social markers in speech*. Cambridge: Cambridge University Press,

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotion fromcontent-masked speech. *Journal of Psycholinguistic Research*, **1**, 269-285.

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, **7**, 31-44.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, **22**, 36-71.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, **122**, 166-183.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Chichester, Eng: Wiley.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*(. Cambridge, MA: Harvard University Press.

Tinbergen, N. (1952). Derived activities: Their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology*, **27**, 1-32.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, Monograph Supplement 1*, **18**, 119-158.

Werner, H., & Kaplan, B. (1963). *Symbol Formation*. New York: Wiley.

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

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, **84**, 963-976.

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hidges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Journal of Personality and Social Psychology,* **19**, 331-339.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, **60**, 181-192.

Zajonc, R. B. (1965). Social facilitation. *Science*, ,

Zajonc, R. B. (1985). Emotion and facial efference: A theory reclaimed. *Science*, **228**, 15-21.

Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Review*, **96**, 395-418.

Zinober, B., & Martlew, M. (1985). Developmental change in four types of gesture in relation to acts and vocalizations from 10 to 21 months. *British Journal of Developmental Psychology*, **3**, 293-306.

Zipf, G. K. (1935). *The psychobiology of language*. New York: Houghton-Mifflin.