Why Do We Gesture When We Speak?

Robert M. Krauss

Columbia University

This is a pre-editing version of a paper published as: Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science* **7**, 54-59.

Why Do We Gesture When We Speak?

Robert M. Krauss[1]

Columbia University

Students of human nature traditionally have considered conversational gestures—unplanned, articulate hand movements that accompany spontaneous speech— to be a medium for conveying semantic information, the visual counterpart of words.[2] Over a century ago, Sir Francis Bacon put the relationship of gesture and language in the form of a simple analogy: "As the tongue speaketh to the ear, so the gesture speaketh to the eye" (Bacon, 1891).

Although the extent to which gestures serve a communicative function is presently a matter of some controversy,[3] there is accumulating evidence that communication is not the only function such gestures serve. Over the past several years my colleagues and I have explored the hypothesis, casually suggested by a remarkably diverse group of writers over the past 60 years, that gestures help speakers formulate coherent speech by aiding in the retrieval of elusive words from lexical memory.

How might gesturing affect lexical retrieval? Human memory employs several different formats to represent knowledge, and much of the content of

[1] Address correspondence to Robert M. Krauss, Department of Psychology, Columbia University, 1190 Amsterdam Ave., Mail Code 5501, New York, NY 10027. E-mail to: rmk@psych.columbia.edu.

[2] Two different types of conversational gestures can be distinguished: *beats,* or as I prefer to call them *motor gestures* (simple, brief, repetitive, movements, that are coordinated with the speech prosody and bear no obvious relation to the semantic content of the accompanying speech) and what I call *lexical gestures* (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). It is the latter type that is the focus of the work reported here. Conversational gestures should be distinguished from the stereotyped hand configurations and movements with specific, conventionalized meanings (e.g. the "thumbs up" sign) that are referred to as *symbolic gestures* or *emblems.* Such gestures often are used to substitute for speech, and clearly serve a communicative function.

[3] See Kendon (1994) and Krauss, Chawla & Chen (1996) for contrasting views.

memory is multiply encoded in more than one representational format. When a concept is activated in one format, it is assumed to activate related concepts in other formats. Our conjecture is that lexical gestures reflect spatio-dynamic features of concepts, and that they participate in lexical retrieval by a process of cross-modal priming.

Evidence for a lexical function of gesture can be found in four kinds of data: (1) differences in the gestures that accompany rehearsed and spontaneous speech; (2) the temporal relation of speech and gesture; (3) the influence of speech content on gesturing; and (4) the effects of preventing speakers from gesturing on speech production.

GESTURE PRODUCTION IN SPONTANEOUS AND REHEARSED SPEECH

Retrieving words from lexical memory is quite different depending on whether one is reciting a speech from memory or speaking spontaneously, and the difference is clearly reflected in the microstructure of speech. For example, pausing is a common occurrence in spontaneous speech and about 60-70 percent of the pauses fall at the juncture between grammatical clauses. Speech that has been memorized contains many fewer pauses, and nearly all of them are interclausal (Butterworth, 1980). Since nonjuncture (i.e., *intra*clausal) pauses often result from difficulties in lexical retrieval, it is not surprising that they are more characteristic of spontaneous speech. Purnima Chawla and I reasoned that if lexical gestures aided in the process of lexical access, we would find more of them, and more nonjuncture pauses, in spontaneous speech than in rehearsed speech (Chawla & Krauss, 1994).

To test this notion we first 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 given to another actor of the same sex, who was asked to portray the original actor in a convincing manner. As has been found in other studies, the conditional probability of a pause being a nonjuncture pause was reliably greater for spontaneous speech than for the rehearsed version of the same speech. Similarly, although the total amount of time speakers spent gesturing did not differ between the spontaneous and rehearsed portrayals, the proportion of time spent making lexical gestures was significantly greater in the spontaneous than in the rehearsed scenes. The conditional probability of nonjuncture silent pauses and the proportion of time a speaker spent making lexical gestures were reliably correlated ($r = .47$). If speakers do use lexical gestures as part of the retrieval process, it would follow that the more hesitant a speaker was, the more lexical gestures he or she would make.

TEMPORAL RELATIONS OF LEXICAL GESTURES AND SPEECH

Gesture and Speech Onsets

If gestures play a role in lexical retrieval, they must stand in a particular temporal relationship to the speech they are presumed to facilitate. For example, it would be difficult to argue that a gesture helped a speaker retrieve a word if the gesture were initiated after the word had been articulated. The word whose retrieval the gesture is hypothesized to enhance is called its *lexical affiliate.* Palmer Morrel-Samuels and I examined the gesture-speech asynchronies (i.e., the onset of the lexical affiliate relative to the onset of the gesture that accompanied it) for 60 lexical gestures drawn from a corpus of speakers describing a variety of pictures and photographs (Morrel-Samuels & Krauss, 1992). As the plot of the distribution of asynchronies (Figure 1) illustrates, all 60 gestures were initiated either prior to or simultaneously with the onset of articulation of the lexical affiliate. The median gesture-lexical affiliate asynchrony was 0.75 s and the mean .99 s (SD = 0.83 s). The smallest asynchrony was 0 s (i.e., gesture and speech were initiated simultaneously) and the largest was 3.75 s. Clearly, gestures precede the words whose retrieval we contend they facilitate.

Insert Figure 1 here

Lexical Retrieval and Gestural Duration

The durations of lexical gestures vary considerably. The average length of the 60 gestures Morrel-Samuels and I examined was 2.49 s (SD = 1.35 s); the briefest was 0.54 s, and the longest 7.71 s. We hypothesized that a gesture's duration should be a function of the time it took the speaker to access its lexical affiliate. Although we can't ascertain the precise moment lexical retrieval occurs, it would have to be before the lexical affiliate is articulated, so we would expect a positive correlation between a gesture's duration and the magnitude of the gesture-lexical affiliate asynchrony.

For the 60 gestures we studied, that correlation is +0.71. The individual data points are plotted in Figure 2. The lighter of the two diagonal lines in that figure is the least-squares regression line; the heavier line below it is the "unit line"—i.e., the line on which all data points would fall if the lexical gesture terminated at the precise moment articulation of the lexical affiliate began. Data points below the unit line represent instances in which the lexical gesture was terminated before articulation of the lexical affiliate began, and points above the line represent instances in which the articulation of the lexical affiliate began before the lexical gesture 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. A lexical gesture's duration is closely related to how long it takes the speaker to access its lexical affiliate.

Insert Figure 2 here

GESTURING AND SPEECH CONTENT

If, as we hypothesize, lexical gestures reflect spatio-dynamic features of concepts, their facilitative effects on retrieval should depend on the conceptual content of what is being said, and we should observe an association between gesturing and conceptual content. Flora Fan Zhang, who at the time the study was done was a student at La Guardia High School in New York City, tested this idea in an experiment she did as part of a Westinghouse Science Talent Search project. She videotaped speakers as they defined twenty common English words (see Figure 3 for the words), and then coded the videotapes for the proportion of time speakers gestured as they spoke.

The twenty words varied greatly in the amount of gesturing that accompanied their definitions, ranging from a high of 44 percent for *under* to a low of 17 percent for *thought* (Figure 3). The words differ on several dimensions, but by having subjects rate them on a variety of scales we were able to extract three factors that accounted for most of the variability: *Activity* (active vs. passive), *Concreteness* (abstract vs. concrete). and *Spatiality* (spatial vs. nonspatial). All three are correlated with the proportion of time the speaker spent gesturing , and a multiple regression model incorporating the 3 factors accounts for nearly 60 percent of the variance in time spent gesturing (r = .768). However most of variance is attributable to the Spatiality factor. The simple correlation of Spatiality with gesture time is not appreciably smaller than the multiple correlation (r= 0.72), and even when the effects of Activity and Concreteness have been partialled out, the correlation between Spatiality and proportion of time spent gesturing is substantial (r= 0.556).

Insert Figure 3 here

Francis Raucher, Yihsiu Chen and I found evidence that is consistent with Flora Zhang's finding in the narratives of undergraduates describing animated action cartoons (Rauscher, Krauss, & Chen, 1996). We first located all of the phrases in the narratives that contained spatial prepositions—about a third of the total—and calculated the mean number of gestures per word associated with those phrases. We then did the same for the remaining two-thirds of the phrases in the narratives. Gesturing during these "spatial content phrases" was nearly five times more frequent than it was for the remaining nonspatial phrases (.498 vs. .101 gestures per word).

EFFECTS OF RESTRICTING GESTURING ON SPEECH

If lexical gestures facilitate lexical retrieval, preventing speakers from gesturing should make lexical retrieval more difficult. In the Rauscher et al., (1996) experiment in which subjects narrated the plots of action cartoons, we increased the difficulty of retrieval by asking subjects to use uncommon words wherever possible *(obscure speech condition),* or to avoid using words containing the letter *c (constrained speech condition),* in addition to a condition in which subjects were allowed to speak normally *(normal speech condition).* These three speech conditions were crossed with a gesture-no gesture condition[4] in a within-subject design, permitting us to compare the effects of not being able to gesture to the effects of conditions that are known to increase the difficulty of lexical retrieval.

---

[4]Subjects were prevented from gesturing under the guise of recording skin conductance from their palms.

Speech production is an on-line process in which several complex cognitive activities must occur in parallel.[5] We can distinguish three stages of the process. Levelt (1989) refers to them as *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 (i.e., an abstract symbol representing the selected word as a semantic-syntactic entity) in the mental lexicon whose meaning matches the content of the preverbal message. Using syntactic information contained in the lemma, the conceptual structure is transformed into a *surface structure.* Then, by accessing word forms stored in lexical memory and constructing an appropriate plan for the utterance's prosody, a phonological encoder transforms this surface structure into a *phonetic plan* (essentially a set of instructions to the articulatory system). The output of the articulatory stage is overt speech, which the speaker monitors and uses as a source of corrective feedback.

Problems in lexical access can be manifested in speech in a number of ways. 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 happens, the speaker may speak more slowly, pause silently, utter a filled pause ("uh," "er, "um," etc.), incompletely articulate or repeat a word, restart the sentence, etc. We expected that preventing speakers from

[5]The precise details of the process are not uncontroversial and several production models have been proposed that differ in significant ways. For present purposes the differences are less important than the similarities. The account we give is based on Levelt (1989), but all of the models with which we are familiar make similar distinctions.

gesturing would exacerbate the problem of producing fluent speech, and hypothesized that it would have an especially adverse impact when the content of speech was spatial.

Speech rate and speech content

We calculated speech rates in words per minute (wpm) during spatial content phrases and elsewhere. The normal, obscure and constrained speech conditions were designed to represent increasing levels of difficulty of lexical access, and a variety of measures indicate that they accomplished that goal.[6] Speakers spoke more slowly in the obscure and constrained speech conditions than they did in the normal condition (98, 87 and 151 words per minute, respectively). They also spoke more slowly when they were not permitted to gesture, but only when the content of speech was spatial (116 vs. 100 words per minute, across the three speech conditions). With other kinds of content, speakers spoke somewhat more rapidly when they could not gesture. The detrimental effects of preventing speakers from gesturing on speech rate seem limited to speech whose conceptual content is spatial.

Dysfluency and speech content

We counted all of the dysfluencies (long and short pauses, filled pauses, incompleted and repeated words, and restarted sentences) in our speakers' narratives, and tallied their frequency per word in spatial content phrases and elsewhere. Not surprisingly, speakers were more dysfluent overall in the obscure and constrained speech conditions than in the natural condition, and they are considerably more dysfluent during SCPs than elsewhere. As Figure 4 illustrates, the effect of preventing gesturing depends on whether the content of

[6]For example, both the mean syllabic length of words and the type-token ratio of narratives were greater in the obscure and constrained conditions than in the normal condition; syllabic length is related to frequency of usage by Zipf's Law, and the type-token ratio (the ratio of the number of different words in a sample to the total number of words) is a commonly-used measure of lexical diversity. Both indices are related to accessibility.

the speech is spatial or not: with spatial content, preventing gesturing increases the rate of dysfluency; with nonspatial content, preventing gesturing has no effect.

Insert Figure 4 here

Juncture and non-Juncture Filled Pauses

A variety of factors can affect how rapidly and fluently people speak. Is there any way to be sure that the adverse effects of preventing gesturing is due specifically to difficulty in lexical retrieval? Perhaps the speech event most directly related to problems in lexical access is the filled pause.[7] Filled pauses can fall either at the boundary between grammatical clauses or within a clause; the former often are called juncture pauses and the latter hesitations or nonjuncture pauses. Although juncture pauses can have a variety of causes, hesitations are believed to be attributable primarily to problems in lexical access (Butterworth, 1980).

We computed the conditional probability of a nonjunctur.e 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 5. Making lexical access more difficult, by requiring speakers to use obscure words or forcing them to avoid words containing a particular letter, increases the relative frequency of nonjuncture filled pauses. Preventing speakers from gesturing had the same effect. Without constraints on lexical selection (i.e., the normal speech condition) and no restriction on gesturing, about 25% of speakers' filled pauses fell within clause boundaries. When they could not gesture, that percentage increased to about 36%. Since the most common cause of nonjuncture filled pauses is

[7]See (Krauss et al., 1996) for a review of evidence supporting this claim.

problems in word finding, these results indicate that preventing speakers from gesturing makes lexical access more difficult, and support the hypothesis that lexical gestures aid in lexical access.

Insert Figure 5 here

HOW DO GESTURES AFFECT SPEECH?

It seems clear that gesturing facilitates the production of fluent speech by affecting the ease or difficulty of retrieving words from lexical memory. What is not clear is exactly how gestures accomplish this. Logically, there are three points in the speech production process at which gestures might play a role. At the conceptualizing stage, gesturing might help the speaker formulate the concept that will be expressed in speech; at the stage of grammatical encoding, information in the gesture could help the speaker map the concept onto a lemma in the lexicon; at the stage of phonological encoding, gesturing could aid in the retrieval of the word form or lexeme. All three could result in slow and dysfluent speech. The evidence, although far from definitive, suggests that the primary effect of gesturing is on retrieval of the word form. The argument is made in detail in Krauss, Gottesman, Chen and Zhang (1997) ; here I'll just point to a couple of relevant findings.

The first type of data comes from neuropsychological investigations of gesturing in patients who have suffered brain damage due to strokes. Patients diagnosed as anomic produce a higher rate of lexical gestures while telling a simple story than either their normal controls or patients whose contra-lateral strokes resulted in visuo-spatial deficits (Hadar, Burstein, Krauss, & Soroker, in press) . Anomic patients have relatively good comprehension and can repeat words and short phrases, but do poorly in tests of object naming. Moreover, Wernike's aphasics (who show deficits in comprehension, but are adequate in repetition) gesture less than anomic aphasics or patients diagnosed as suffering from conduction aphasia (who show fair sentence comprehension, but marked deficits in both object naming and repetition) (Hadar, Wenkert-Olenik, Krauss, & Soroker, in press).

Second, Frick-Horbury and Guttentag, (in press) have found that preventing subjects from gesturing in the Tip-of-the-Tongue (TOT) situation increased the rate of retrieval failures. In the TOT paradigm, subjects are read definitions of uncommon words and try to recall the word form. The definitions by means of which the TOT state is induced are roughly equivalent to the information in the lemma, suggesting that preventing gesturing interferes with retrieval of the word form.

Although it is possible, and perhaps even likely, that what we are calling lexical gestures can affect speech processing at the conceptual stage and during lemma retrieval, the evidence we have at this point indicates that their influence on speech is mediated by their ability to facilitate retrieval of the word form during phonological encoding.

AFTERWORD

Many years ago, my maternal grandfather told me a story about two men in his hometown, Vitebsk, Belorussia, walking down a road on a bitterly cold winter day. One man chattered away animatedly, while other nodded from time to time, but said nothing. Finally, the man who was talking turned to his friend and said: "So, nu, Shmuel, why aren't you saying anything?" "Because," replied Shmuel, "I forgot my gloves." At the time, I didn't see the point of the story. Half a century later it has become a primary focus of my research.

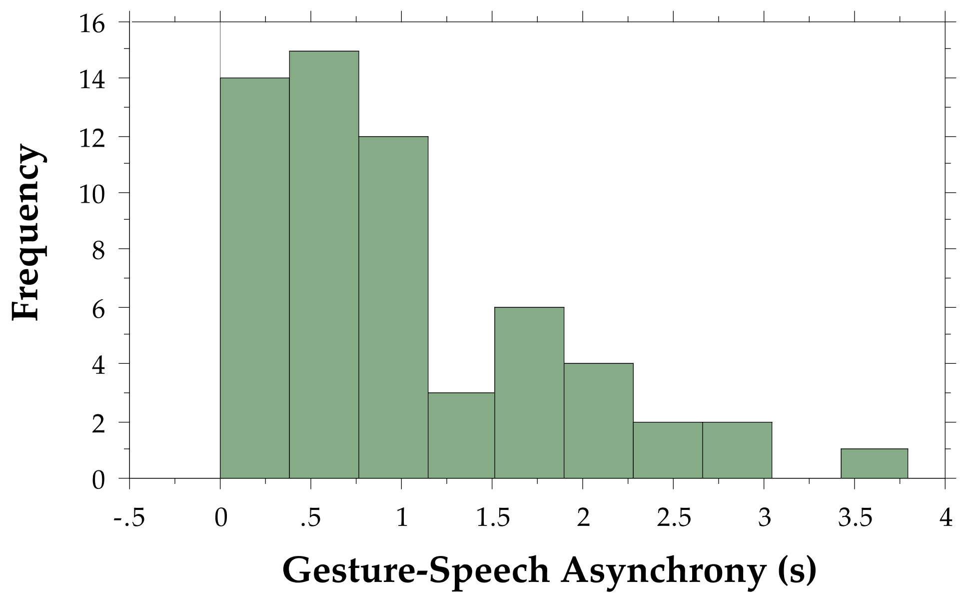

Figure 1

Distribution of Gesture-Speech Asynchronies (onset time of speech minus onset time of gesture) for 60 lexical gestures.

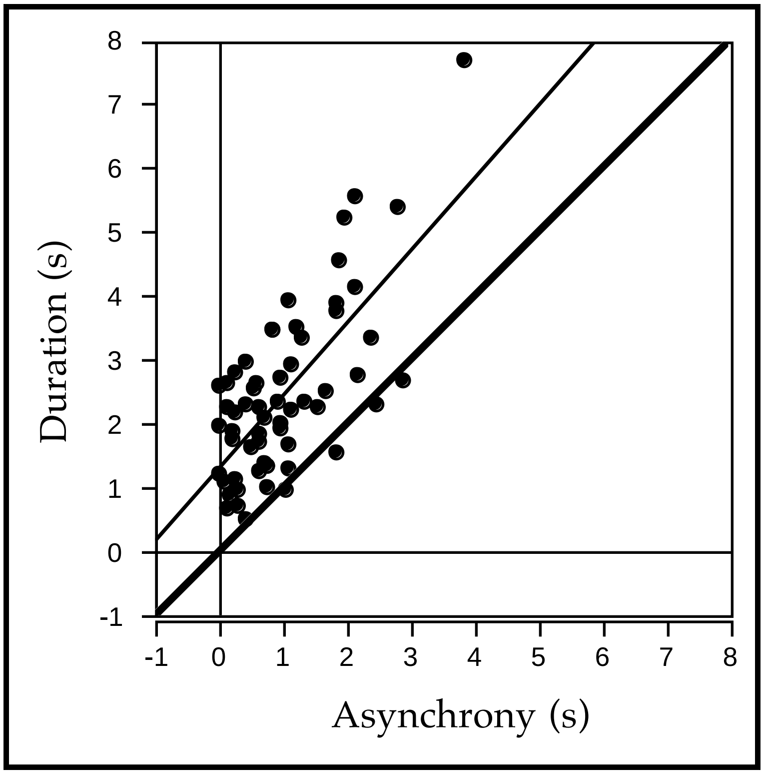

Figure 2

Duration of lexical gestures plotted against speech-gesture 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).

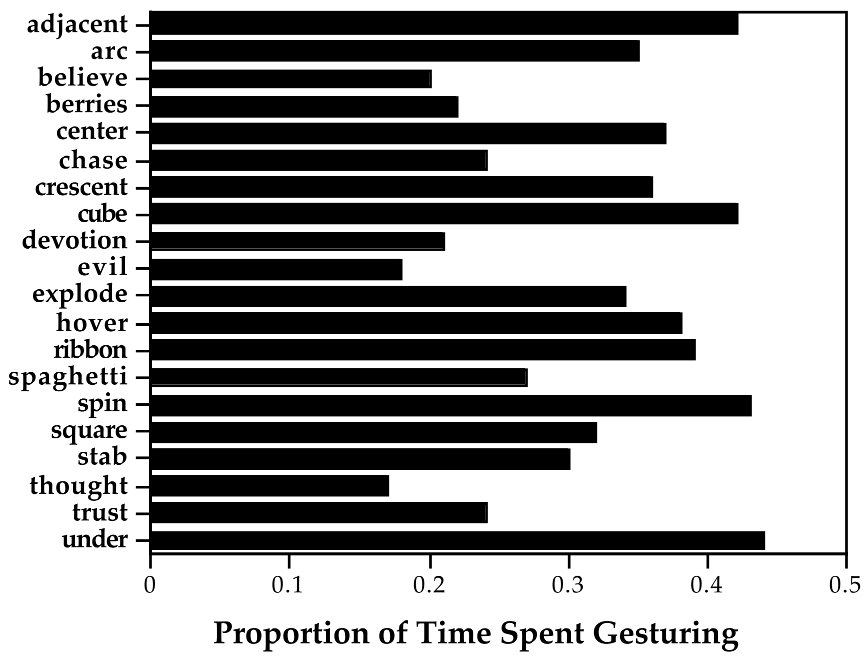

Figure 3

Mean proportion of speaking time spent gesturing during definitions of 20 common words.

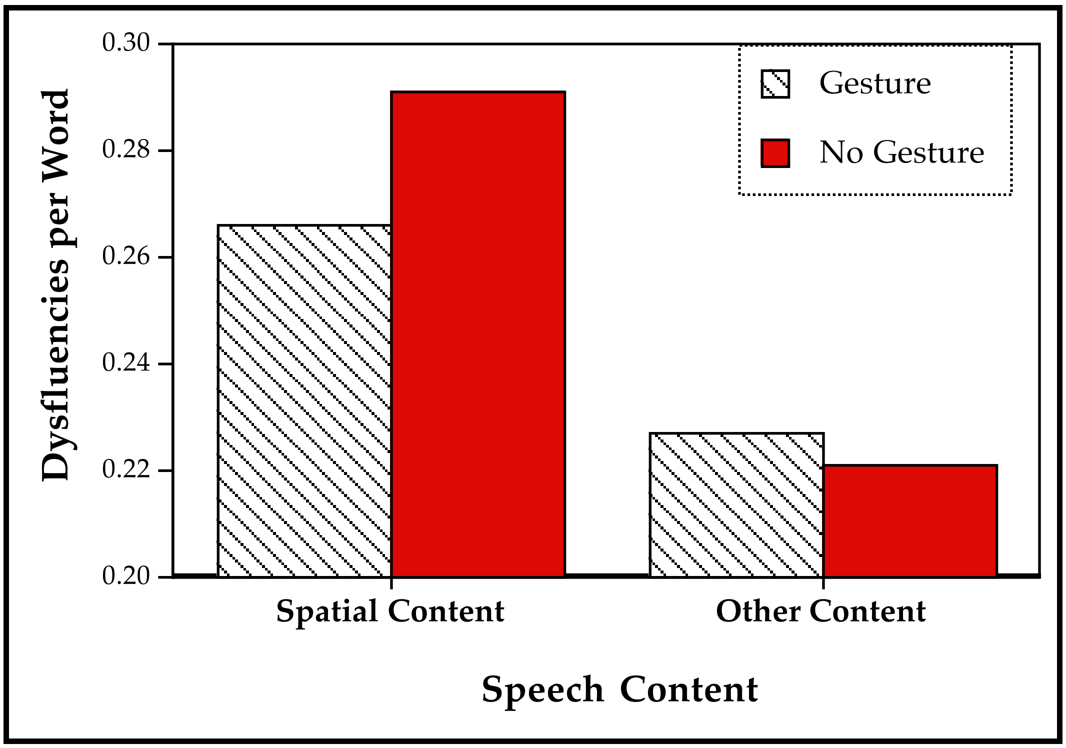

Figure 4

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

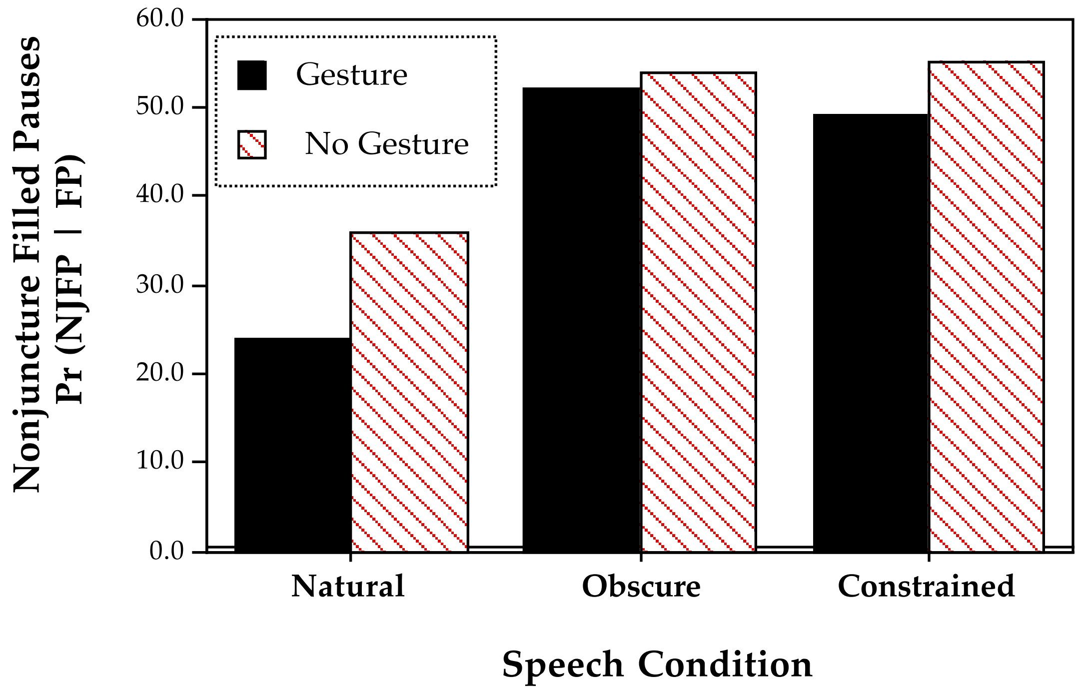

Figure 5

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

References

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

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Speech and talk* (pp. 155-176). London: Academic Press.

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

Frick-Horbury, D., & Guttentag, R. E. (in press). The effects of restricting hand gesture production on lexical retrieval and free recall. *American Journal of Psychology.*

Hadar, U., Burstein, A., Krauss, R. M., & Soroker, N. (. (in press). Ideational gestures and speech: A neurolinguistic investigation. *Language and Cognitive Processes.*

Hadar, U., Wenkert-Olenik, D., Krauss, R. M., & Soroker, N. (in press). Gesture and the processing of speech: Neuropsychological evidence. *Brain and Language.*

Kendon, A. (1994). Do gestures communicate?: A review. *Research on Language and Social Interaction* , **27**, 175-200.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389-450). San Diego, CA: Academic Press.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (1997). Lexical gestures and lexical access: A process model (unpublished paper) [Actually this paper will be published in D. McNeill (Ed.), *Language and gesture: Window into thought and action* . There has been a change of publisher for the volume. It will be listed as *in press* as soon as a new publisher is selected.]

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

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.

Rauscher, F. B., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science* , **7**, 226-231.

ACKNOWLEDGMENTS

Purnima Chawla, Yihsiu Chen. Palmer Morrel-Samuels, Francis Rausher, and Flora Fan Zhang were my collaborators in the studies described here. The research was supported by National Science Foundation grants BNS 86-16131 and SBR 93-10586. I've benefited greatly from discussions of these matters with Sam Glucksberg, Uri Hadar, Julian Hochberg and Lois Putnam.