

Movement Facilitates Speech Production: A Gestural Feedback Model

Ezequiel Morsella and Robert M. Krauss
Columbia University

Address for correspondence:

Robert M. Krauss
Department of Psychology
Columbia University
1190 Amsterdam Avenue
New York, NY, 10027
Fax: (212) 854-3949
e-mail: rmk@psych.columbia.edu

Running Head: Movement Facilitates Speech

Word Counts: Abstract = 129, Body Text = 4,832, References = 1,624.

Movement Facilitates Speech Production: A Gestural Feedback Model

Ezequiel Morsella and Robert M. Krauss
Columbia University

ABSTRACT

Although the hand and arm movements (gestures) that accompany speech traditionally have been regarded as communicative, accumulating evidence suggests that they play a functional role for the speaker. According to the Gestural Feedback Model (GFM), lexical gestures participate in speech production by increasing the semantic activation of words grounded in sensorimotor features, hence facilitating retrieval of the word form. In Experiment 1, the magnitude of muscle activation observed during lexical retrieval was predicted by the target word's concreteness and spatiality. In Exp. 2, more gesturing occurred when participants described visual objects from memory than when the object was visually present. Fluency decreased when speakers were prevented from gesturing. The implications of these findings for the GFM and for theories of semantic representation are discussed.

Movement Facilitates Speech Production: A Gestural Feedback Model

People often perform a variety of movements while engaged in cognitively demanding tasks. An attempt to imagine the shape of an inverted “S,” for example, may be accompanied by averted gaze, a furrowed brow, and gesticulation. Similar sorts of movements accompany mental arithmetic, spatial reasoning, and musical imagery. And in the complex cognitive task we call conversation, there are postural shifts, changes in gaze direction, sweeping movements of the arms, and an elaborate medley of hand and finger movements.

Why do people perform such actions? One possibility is that these ubiquitous behaviors are epiphenomenal—functionally unrelated to the cognitive tasks they accompany. Perhaps due to the tacit acceptance of this view, the functional role of these movements has received relatively little scientific attention. However, the contrary view—that movements play a functional role in cognition—has a long history in psychology. Nearly 75 years ago, Washburn (1928) contended that “...the motor innervations underlying the consciousness of effort are not mere accompaniments of directed thought, but an essential cause of directed thought” (p. 105). Forty years later, Hebb (1968) espoused a similar notion.

Although movements, or the processes involved in their planning and execution, may not be essential to mental functioning, there is considerable evidence that they can facilitate cognitive processing. Some motor events reduce sensory input, and by so doing modulate the kinds of information that enters the perceptual system. Other movements increase sensory input. We will refer to movements that facilitate cognitive functioning by attenuating sensory inputs that would impede performance as *subtractive movements*, and movements that facilitate functioning by selectively increasing introducing or maintaining the activation of information as *additive movements*.

A familiar subtractive movement is the tendency to shift gaze from complex to less complex visual arrays while performing tasks involving language. Gaze aversion in this situation serves to reduce potential informational input that would compete for

the cognitive resources required by the complex task of speech production (see Butterworth, 1978; Beattie, 1980; Glenberg, Schroeder, & Robertson, 1998). In a recent set of experiments, Glenberg et al. (1998) found the tendency to avert gaze to be positively related to the difficulty of the cognitive task at hand, and that averting gaze improves performance in memory tasks. Chiu, Hong & Krauss (unpublished) found that subjects required to fixate visually on their conversational partner's face while describing a route to a destination spoke less fluently than those required to fixate an inanimate object or allowed to look where they chose. Whether a particular signal will impede cognitive processing depends on the task. In a classic series of experiments, Brooks (1968, 1970) demonstrated that "recall of verbal information is most readily disrupted by concurrent vocal activity; recall of spatial information is most readily disrupted by concurrent spatially monitored activity" (Brooks, 1968, p. 349).

In contrast, additive movements facilitate cognition by introducing, or increasing the accessibility of, inputs that enhance a cognitive operation. An example is the (often subvocal) articulation involved in the phonological loop, whereby representations (e.g., of a phone number) are kept in working memory through the active process of rehearsal. In the loop, proprioceptive (or re-afferent) inputs from vocalization or subvocalization are believed to reactivate quickly decaying phonemic representations in a working-memory buffer (Baddeley, 1986; see also Hitch & Burgess, 1999).

The hand-arm actions that often accompany everyday conversation, variously called "representational gestures" (McNeill, Cassell, & McCullough, 1994), illustrators (Ekman & Friesen (1969) and "gesticulations" (Kendon, 1980, 1983), constitute another important class of additive movements. We refer to them "lexical gestures" (Krauss, Chen, & Gottesman, 2000), and they are the focus of our investigation.

Lexical Gestures and Semantic Activation

A lexical gesture is a complex, articulate hand-arm movement of variable length whose form seems related to the ideational content of the conversational speech it accompanies.¹ A diverse group of investigators (DeLaguna 1927; Dobrogaev 1929; Mead 1934; Werner and Kaplan 1963; Freedman and Hoffman 1967; Moscovici 1967; Butterworth & Hadar, 1989) have suggested that gestures facilitate word retrieval from the mental lexicon. There is some empirical evidence to support this view. Compared to those who can gesture, speakers prevented from gesturing have more retrieval

failures in the Tip-of-the-Tongue paradigm (TOT) (Frick-Horbury & Guttentag, 1998), employ less vivid imagery in their speech (Rimé, Schiaratura, Hupet, & Ghysselinckx, 1984), and speak less rapidly and make more speech errors associated with difficulty in lexical retrieval (Rauscher, Krauss, & Chen, 1996). There is also evidence from brain lesion studies: aphasics with speech problems primarily involving word retrieval tend to gesture more than both normal controls and other aphasics whose problems are primarily conceptual (Hadar, Wenkert-Olenik, Krauss, & Soroker, 1998).

Gestural Feedback Model

The available evidence (see Krauss & Hadar, 1999 for recent reviews) indicates that lexical gestures are most likely to occur during retrieval failures in speech with spatial content. We believe that this liaison is more than coincidence, and represents a functional relationship in which lexical gestures serve as additive movements that facilitate the access of words grounded in sensorimotor components. The Gestural Feedback Model (GFM) is an attempt to describe the process by which gestures participate in lexical retrieval. The model is embedded in a framework of assumptions about the speech production process and the nature of semantic representations in general. We will first present a brief overview of lexical retrieval in speech production. Of course, the details of the speech production process are far from settled, and a number of models that differ in significant ways have been proposed (e.g., Butterworth, 1989; Caramazza, 1997; Dell, 1986; Garrett, 1980; Levelt, 1989; MacKay, 1987; Stemberger, 1985). For our purposes the differences are less important than their similarities.

Converging evidence from a variety of sources (reaction-time experiments: Schriefers, Meyer, & Levelt, 1990; error analyses: Fromkin, 1971; Garrett, 1980); and brain lesion studies: Badecker, Miozzo, & Zanuttini, 1995; Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Kay & Ellis, 1987) suggests that there are at least two levels of representation at play during lexical processing in word production. At one level, that of the lexical node, a word is specified by its semantic and syntactic properties. The lexical node cat, for instance, is specified by the features of number (singular vs. plural) and grammatical class (noun), among other syntactic features. The other primary level of representation encodes information about a word's phonology—e.g., that the word cat is composed of the phonemes /k/, /æ/, and /t/, is

monosyllabic, has only one vowel, etc. Consistent with this form of lexical organization, lexical retrieval in word production appears to involve two distinct stages: one stage is devoted to the selection of a word's lexical node and its syntactic features, and the other stage is aimed at retrieving word phonology (Butterworth, 1989; Caramazza, 1997; Dell, 1986; Garrett, 1980; Levelt, 1989; MacKay, 1987; Stemberger, 1985). We will refer to these stages as *lexical node² selection* and *phonological encoding*.

There is little disagreement about two additional assumptions we make concerning lexical retrieval in word production. First, the semantic system activates a cohort of related lexical nodes. In the course of retrieving the word CAT, the lexical nodes for LION, RABBIT, and TAIL also are activated. In normal circumstances the target lexical node – cat, in our example – is selected because it reaches the highest level of activation. Second, selection proceeds in a fixed order: the activation of lexical nodes precedes that of word phonology, though there is evidence that activation can cascade from the lexical node level onto the phonological one before lexical selection has taken place (see Morsella & Miozzo, in press).

Lexical node selection is driven by semantic activation. We hypothesize that lexical gestures: (1) can re-activate the lexical nodes that are richly endowed with sensorimotor features; and (2) that they facilitate retrieval by sustaining the activation of those nodes long enough to allow phonological encoding to take place (see Figure 1).

FIGURE 1 ABOUT HERE

Along with Martin, Ungerleider, and Haxby (2001), we believe that "The semantic representation of an object is composed of stored information about the features and attributes defining that object, including its typical form, color, motion, and the motor movements associated with its use" (p. 1023). Semantic representations of word concepts can be encoded in both propositional and nonpropositional formats, and we assume that words whose retrieval are facilitated by gestures are more likely to be those that have been analogically encoded in what we call sensorimotor features. Such analogical representations are part of a view of mental representation in which

Movement Facilitates Speech

semantic representations are grounded in bodily processes reflecting patterns of interaction with the world (Barsalou, 1999; Glenberg; 1997).

As activators of analogical features, lexical gestures tend to reflect spatial or functional attributes of the target's semantic representation, rather than less tangible, abstract properties—e.g., that the target word is an animal, is warm-blooded, and is not extinct. Lexical search for the word *LEVEL* might be accompanied by a horizontal hand motion with fingers extended and palm facing downward; *CORKSCREW* by a twisting motion. The movement accompanying the search for *LEVEL* has properties that are isomorphic with features of the word-concept; movements accompanying the search for *CORKSCREW* resemble motor events associated with the target's use. We would expect words like “democracy” or “insipid” that lack such tangible properties to be accompanied by movements less often.

We hypothesize that these gestures continually reactivate semantics through feedback from effectors or motor commands, in much the same way that vocal (or subvocal) rehearsal keeps echoic representations active in the phonological loop (Baddeley, 1986; Burgess & Hitch, 1999). This roundabout method of keeping things in mind may be necessary because purposefully-activated mental representations are transient and the process of activating them is effortful (Farah, 1995); hence, they are difficult to hold for the lengthy intervals that often occur in memory search (e.g., in a TOT state).

The GFM predicts that lexical gestures will accompany the retrieval of words grounded in sensorimotor features, and that the more a word's meaning is grounded in sensorimotor features, the more gesturing will accompany its retrieval. The model also predicts that when semantic activation is low, lexical gestures will be invoked to increase semantic activation, and that in the absence of their facilitative role, speech will be less fluent.

In Experiment 1, we tested the first set of predictions by monitoring muscle activation during a lexical retrieval task in which participants tried to name words after reading their definitions. Because concrete words are more likely than abstract words to be grounded in sensorimotor features (Jessen, Heun, Erb, Granath, Klose, Papassotiropoulos, & Grodd 2000; Paivio, 1979), half of the stimulus words stimulus

words we used were concrete and the other half were abstract. The GFM leads us to expect that the amount of muscle activation accompanying retrieval of a word will be a function of the degree to which the word's semantic representation is grounded in sensorimotor features.

In Experiment 2, participants described visual objects that were either visually available or no longer present. This enabled us to test the second set of predictions. Because of their role in sustaining semantic features, we expected lexical gestures to be more common when the to-be-described stimulus was no longer present than they would when features of the stimulus were visually accessible. We also examined the effects of restricting arm movements on the fluency of descriptions.

EXPERIMENT 1

Using an electromyograph, we monitored electro-muscular activity in participants' dominant forearms as they performed a lexical retrieval task that required them to identify a target word from its definition. In most studies of gestures, speakers are videotaped and their gestures are coded from the videotape. This method produces acceptable reliabilities, especially when the coding involves the relatively long durations found in continuous discourse. We chose to monitor muscular activity electromyographically for two reasons: (1) Observational coding can take into account only muscular activation that rises to the level of overt movement; the electromyograph can detect muscular activity below the movement threshold (cf., Cacioppo, Tassinari, & Fridlund, 1990). Such activation is relevant to our model. (2) Trial durations in our word retrieval task are relatively brief (3-45 s in a pilot study). Coding such brief movements is a demanding and somewhat subjective task. The EMG provides an objective measure of the amount of muscular activity that occurs within a specified time window.

METHOD

Participants

Thirty Columbia University students (14 male and 16 female) received \$8 for their participation, which took about forty minutes. All were native English speakers.

Movement Facilitates Speech

An additional 42 undergraduates (20 males and 22 females) received course credit for participating in a word rating study.

Test Materials

A total of 36 low frequency nouns and their definitions (Appendix I) gleaned from previous studies (Burke, MacKay, Worthley, & Wade, 1991; Brown & Nix, 1996; Frick-Horbury et al., 1998; Jones, 1989; Meyer & Bock, 1992; Perfect & Hanley, 1992) were used as stimuli.³ On the Kucera and Francis (1967) written frequency count, the mean frequency of our concrete words was 0.684 ($SD = 0.885$), and for the abstract group it was 2.706 ($SD = 4.701$). The nouns were classified as abstract or concrete *a priori*, according to the criterion of Gorman (1961), and this categorization was corroborated by ratings obtained from an independent sample of judges (see below). Nineteen nouns were classified as concrete, and 17 as abstract.

Apparatus and Layout

The experiment utilized two rooms. One (the experimental room) housed the participant and an experimenter (E1); in the other (the observation room), a second experimenter (E2) monitored and recorded the events occurring in the experimental room. The experimental room contained a video camera, an intercom speaker, and a computer monitor, and two chairs for the participant and experimenter. The participant's chair was fitted with special arm extensions that allowed nine electrode leads to be attached, leaving enough slack to permit some arm movement. The input cable of the electrode leads ran through a port in the wall to a polygraph (Grass Instrument Wide-Band A.C. Pre-Amplifier and Integrator, Model 7P3B, DC Driver Amplifier, Model 7DAG, and Chart Drive, Model 7H 25-60) in the observation room. A video camera in the experimental room was trained on the participant; a second camera in the observation room monitored the EMG chart. Their signals were inputted to a Panasonic 3500 System Switcher (WJ-3500), producing a split screen image showing both the participant and EMG chart that was displayed on a video monitor and recorded on a VCR in the observation room. Stimuli were presented on a monitor connected to a Macintosh computer running PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Procedure

Movement Facilitates Speech

To minimize the likelihood of participants attending to their hand movements, the experiment was described as a study of the relationship between memory and stress--the latter purportedly indexed by the Galvanic Skin Response (GSR) measured by the attached electrodes. Following standard EMG preparations, two active electrodes (Grass Silver/Silver Chloride bipolar electrodes, 5 mm) were placed on the participant's forearm (2 cm apart), on the region of the *m. brachioradialis*, one-third of the arm's distance from the lateral epicondyle of the humerus. A ground electrode was attached to the participant's left ear lobe. Dummy electrodes were attached to the same region of the non-dominant arm and to the right leg. Handedness was determined by asking the participant questions such as, "With which arm do you write, throw a ball, swing a bat, and hold a knife?"

Participants were told they would see the definition of an English word on the monitor, and that their task would be to identify the word. They were given up to 60 s to identify the word, and were told the word if they could not identify it in this amount of time or if they said they did not know it. E1 controlled the presentation of definitions and recorded the accuracy of the response. At the end of the session, participants were debriefed and the true purpose of the research was revealed. None expressed suspicion that arm movements were being studied.

Quantification of Data

To quantify the raw EMG signals, we employed a technique patterned after one developed by Garrity (1977). We divided the trial into 5 s intervals, and measured the amplitude (in μV) of the two largest pen deflections in each 5 s interval. We will refer to the mean of these measures across all response periods within each trial as EMG Amplitude. EMG Amplitudes were quantified blind with respect to the stimulus word. Signals resulting from activity unrelated to the tasks—e.g., scratching, adjusting clothing, or fiddling with the electrodes—were noted on the EMG chart and excluded from the data analysis.

Each trial was coded as correct or incorrect. "Correct" trials were ones on which the participant produced the target word; "incorrect" trials were those on which the participant indicated that he or she didn't know the word, produced a word other than the target word, or on which 60 s passed without a word being produced.

RESULTS

EMG Amplitude

Mean EMG Amplitudes for the 36 target words are shown in Figure 2. We calculated the mean EMG Amplitudes for concrete and for abstract words for each participant, and performed a paired t-test on them. Reliably greater EMG Amplitudes were found during retrieval of concrete words than abstract words ($\underline{M} = 9.87$ vs. $8.06 \mu\text{V}$; $\underline{t}(29) = 2.76$, $p < .01$). For 22 of the 30 participants, EMG Amplitude was greater for concrete words. The same effect was found in a by-item analysis (Concrete $\underline{M} = 9.93$, $\underline{SD} = 2.5$, and Abstract $\underline{M} = 8.07$, $\underline{SD} = 1.3 \mu\text{V}$; $\underline{t}(35) = 3.38$, $p < .01$). The effect of word type varied marginally among participants; the participant \times condition interaction $\underline{F}(1,29) = 1.38$; $p = .09$. Descriptive statistics for these and other findings can be found in Table 1.

 FIGURE 2 ABOUT HERE

We also examined amplitude as a function of whether the participant was able to retrieve the target word. As is shown in Table 1, we found slightly greater EMG Amplitudes on incorrect trials for both abstract and concrete words, but the difference ($-1.5 \mu\text{V}$) was of marginal reliability ($\underline{t}(29) = -1.5$, $p > 0.10$). On average, participants correctly identified about 55% of the target words from their definitions. As expected, abstract words were more difficult to identify than concrete words (63% vs. 46%, respectively; $\underline{t}(29) = 4.93$, $p < .0001$). The fact that greater EMG amplitude was found on incorrect trials should be interpreted with some caution. Our “incorrect” grouping included cases in which participants had no idea what the target word was and cases in which they were in a Tip-of-the-Tongue state that they could not resolve.

Trial Duration

The duration of individual trials varied considerably. As expected, abstract words took longer than concrete words to identify ($\underline{Ms} = 23.6$ vs. 15.5 s, respectively; $\underline{t}(29) = 7.91$, $p < .0001$). Trials on which the target word was not correctly identified were

Movement Facilitates Speech

about four times longer than those on which the participant produced the target word, for both abstract and concrete words.

Correlations among Dependent Measures

Although Trial Duration and Identification Accuracy were highly correlated ($r(35) = -.90$), EMG Amplitudes for individual words were uncorrelated with either Duration or Accuracy ($r(35) = -.041$ and $r(35) = -.027$, respectively).

To ascertain that it was word type and not Duration that accounted for the EMG amplitude effect, partial correlation analyses were carried out. Duration accounts for less than 1% of the variance in EMG amplitude ($\underline{pr}^2 = .01$), when other explanatory variables are taken into account (see below).

Conceptual Properties of Words that Related to Activation

To gain some insight into the semantic properties of words whose retrieval is accompanied by muscular activity, EMG Amplitude was correlated with several word-concept attributes. The values for these attributes were obtained by having another group of participants rate the target words on 8 dimensions. Using a 7-point bipolar scale ranging from "not at all" to "very," they indicated how *spatial*, *concrete*, *active*, *pantomimable*, *familiar*, *drawable*, *manipulable*, and *valuable* each of the 36 word-concepts were.

The four attributes most highly correlated ($r_s > .30$) with EMG Amplitude on Task 1 were concrete ($r = .44$), drawable ($r = .41$), spatial ($r = .38$), and manipulable ($r = .34$). A multiple regression model with the four scales as independent variables accounted for 29% of the variance in EMG Amplitude ($\underline{R} = 0.53$). However, most of that was due to the concrete and spatial scales; removing drawable and manipulable reduced \underline{R}^2 negligibly--from .29 to .27.

EXPERIMENT 2⁴

We videotaped participants as they described drawings that were either visually present or absent. Some of the pictures depicted nameable objects, the others were abstract line drawings. We also varied whether or not participants were allowed to move their arms. A mixed 2x2x2 design was employed with movement restricted or not (Restriction) and stimulus present or absent (Presence) as between-subjects factors,

and nameable vs. unnameable stimuli (Codability) as a within-subjects factor. According to the GFM, the function of lexical gestures is to activate sensorimotor features, leading us to expect more gesturing when semantic activation is low (the stimulus absent condition, compared to the stimulus present condition), and less gesturing accompanying descriptions of unnameable (compared to nameable) objects, because these objects could readily be encoded verbally, without engaging analogical, sensorimotor representations (Paivio, 1979). Finally, we expected speech rate to decrease when participants were not allowed to move their arms, thus denying them the facilitative effects of gesturing.

METHOD

Participants

Seventy-nine Columbia University students (44 male and 35 female) received course credit for their participation. All were native English speakers.

Materials and Apparatus

40 green-on-black line drawings served as stimuli (see Figure 3). Twenty-eight were non-nameable objects based on figures used by Fussell and Krauss (1989a, 1989b), and twelve were line drawings of identifiable objects (e.g., glasses, wrench, armored tank, house, flower, ice cream cone, and hot-air balloon).

FIGURE 3 ABOUT HERE

Again, the experiment utilized two rooms. One (the experimental room) housed the participant and contained two video cameras, an intercom speaker, and a computer monitor for stimulus presentation. One camera was trained on the participant's face and torso, and the other was trained on the computer monitor. In the other room (the observation room) the experimenter monitored and recorded the events occurring in the experimental room. The signal from the video cameras in the experimental room were inputted to a Panasonic 3500 System Switcher (WJ-3500) in the observation room. This produced a split screen image showing both the participant and the computer monitor on a video monitor that was recorded on a VCR.

Procedure

Participants were run individually. The experiment was described to them as a referential communication study. Participants received instructions via the program PsyScope (Cohen et al., 1993) and were told the cover story that their descriptions of forty objects would be tape recorded and played, a week later, to another participant whose task it was to identify the objects from a larger selection after hearing the description. Participants were told that only their face would be video-recorded for the purpose of visually deciphering syllables that were poorly audio-recorded. This was a cover story to ensure that participants would not be self-conscious of their movements. A camera was trained on the participant's head and torso. For the group of participants that was in the Restricted condition, we presented dummy electrodes and told the cover story that we were trying to discover how arousing the description task was by measuring the GSR. The electrodes were put on both forearms, and participants were told that movement of the limbs could ruin the quality of the recordings.

On each trial, the visual stimulus was shown for 10 s. At the end of the 10 s inspection period, a signal appear instructing the participant to describe the stimulus. Participants had up to 45 s to describe the stimulus, after which the display automatically timed out, and the screen read, "Your time has expired. Click the mouse to continue to the next trial." Participants who completed their description before the 45 s timeout could continue onto the next trial by pressing a mouse key. Stimuli were presented in random order. In the stimulus present condition, the stimulus picture remained on the screen during description period. In the stimulus absent condition, the screen was blanked after the inspection period. Participants were fully debriefed at the end of the session. None suspected the true purpose of the study.

Dependent Measures

The primary dependent variable was gesture rate: the proportion of description time during which the participant gestured. This was obtained by examining the video record and counting the number of picture frames that captured lexical gesture, following the procedure of Rauscher et al. (1996). The number of frame were converted to seconds and divided by trial duration. Gestures were coded as "lexical gestures" following previously used criteria (Rauscher et al., 1996). We also coded "motor movements." According to the GFM, only lexical gestures activate semantic

representations. Hence we predict that, unlike lexical gestures, presence or absence of the stimulus should have no effect on the amount of motor movements. Other gestures such as symbolic gestures and adaptors (e.g., scratching the nose and adjusting clothing) were coded and excluded from the analysis.

RESULTS

Movement Analysis

Gesture rate was roughly 20% greater in the stimulus absent ($\underline{M} = 44\%$, $\underline{SD} = 0.28$) compared to the present condition ($\underline{M} = 26\%$, $\underline{SD} = 0.29$), and non-nameable objects were accompanied by more gesturing than nameable objects ($\underline{M} = 40\%$, $\underline{SD} = 0.31$ and $\underline{M} = 29\%$, $\underline{SD} = 0.28$, respectively). The means are plotted in Figure 4. A 2x2 analysis of variance with Presence as a between-subjects factor and Nameable as a within-subjects factor revealed reliable main effects for Presence ($\underline{F}(1, 44) = 4.97$, $\underline{p} = .03$) and Codability ($\underline{F}(1, 44) = 117.9982$, $\underline{p} < 0.001$), with no Presence x Codability interaction ($\underline{F}(1, 44) = 0.68$, $\underline{p} = .41$).

FIGURE 4 AROUND HERE

Motor gestures were rare, occurring about 1% of the time when the object was present, and 0.8% of the time when it was absent ($\underline{F}(1, 44) = 0.42$, $\underline{p} = 0.52$). However, significantly more motor gestures occurred with nameable ($\underline{M} = 0.016$, $\underline{SD} = 0.03$) than with non-nameable objects ($\underline{M} = 0.003$, $\underline{SD} = 0.006$), $\underline{F}(1, 44) = 6.12$, $\underline{p} = .02$. This difference could be a simple consequence of the fact that, because motor and lexical gestures use the same limbs, a high rate of lexical gesturing (in the non-nameable condition) precludes motor gestures from occurring. It is interesting to note that lexical gestures still occurred 26% of the time when the referent was present. If lexical gestures do facilitate lexical node selection, speakers may have needed help retrieving words like “squiggle,” “angle,” and “loop” (the types of words typically used to describe nonsense objects) even when the visual object was present.

Speech Analysis

We also looked at speech rate (syllables per second) as an index of dysfluency. Because each trial could have lasted up to 45 s ($\underline{M} = 26$ s, $\underline{SD} = 12$ s), and because we ran

Movement Facilitates Speech

79 participants, it became unfeasible to calculate the rate speech for all trials and all participants. Instead, we chose to analyze the speech samples of the ten visual-objects with the longest average trial duration ($\underline{M} = 35$ s, $\underline{SD} = 9$ s), with the idea that would be the best place to look for an effect. As expected, these turned out to be non-nameable visual objects that were quite complex. We transcribed all the selected trials in their entirety and obtained the syllable rate by tallying the number of syllables with a hand counter and dividing that number by trial time. Of the 790 trials that were coded, 100 were randomly sampled and coded by two other raters in order to obtain a measure of inter-rater reliability. There was virtually no variability amongst the coding performances of the three raters: the mean correlation coefficient between raters was 1, and the mean absolute difference between raters was less than 1 syllable.

We performed a 2x2 analysis of variance with Presence and Restriction were between-subjects factors. Speech rate (syllables per s) decreased significantly when participants were restricted ($\underline{M} = 2.58$, $\underline{SD} = 0.64$) versus free ($\underline{M} = 2.96$, $\underline{SD} = 0.49$), $\underline{F}(1, 74) = 9.4$, $\underline{p} = 0.003$. Interestingly, speech rate decreased when the object was present ($\underline{M} = 2.69$, $\underline{SD} = 0.58$) rather than absent ($\underline{M} = 2.91$, $\underline{SD} = 0.58$), $\underline{F}(1, 74) = 4.03$, $\underline{p} < .05$. There was no significant interaction between Presence and Restriction ($\underline{F}(1, 74) = 1.3$, $\underline{p} = 0.26$). We did not make any claims *a priori* about the effect of Presence on speech rate, but it is interesting that we found a significant reduction in speech rate when the referent was present. It is possible that visual inspection in this condition competes for the cognitive resources involved in speech production.

DISCUSSION

Evidence from a variety of sources suggests that the movements people make while speaking contribute to speech production by facilitating lexical retrieval (see Krauss, Chen, & Chawla 1996 for review). The Gestural Feedback Model attempts to detail the process by which this is accomplished. The experiments reported here yield new data consistent with the hypothesis that lexical gestures facilitate the retrieval from lexical memory of words that are grounded in sensorimotor features. They do this by maintaining the activation of those features in a working memory buffer until the sought-after word has been retrieved. In Experiment 1, we tested the prediction that muscle activation will accompany the retrieval of words that are grounded in sensorimotor (as compared to abstract) features, and that the more a word's meaning is grounded in such features, the greater will be the muscle activation that accompanies its retrieval. The results supported this prediction. Experiment 2 tested the hypothesis that lexical gestures would be more frequent when semantic activation was low, and that when speakers were prevented from gesturing, speech would be less fluent. The data supported these hypotheses.

In Experiment 1, participants tried to identify words from their definitions. More electromyographic activity in the dominant forearm accompanied words with concrete and spatial semantic content than words with abstract content. The abstract and concrete words differed conceptually on several dimensions, but it was primarily their spatiality and concreteness that accounted for the amount of electromyographic activity observed. Our findings complement those of Rauscher et al. (1996), who found high rates of gesturing in narrative speech during clauses with spatial content. In our feedback model, the semantic features that maintain the activation of target lexical nodes is maintained until phonological encoding has occurred by feedback from effectors or motor commands. Because the meanings of concrete and spatial words tend to be grounded in sensorimotor components, lexical searches involving them are more likely to be accompanied by muscle activation than searches involving abstract words.

We have focused on the role of muscle activation in lexical retrieval. An alternative explanation for the activity we observed might be that it was communicatively intended. Despite the fact that it did not occur in a conversational

context, gesturing could be so deeply ingrained a habit that they accompanied speech even when they were not functional. De Ruiter makes the case for the gestures made when people talk on the telephone:

...[T]he fact that people gesture on the telephone does not rule out the possibility that gestures are communicatively intended. Gesturing could be so intrinsically linked to speaking that it is hard to suppress gesturing when speaking on the telephone. It is conceivable that if gesturing is deeply integrated with the speaking process, the mere fact that the addressee is invisible is not sufficient to cause people to suppress gesturing (de Ruiter, 1998, p. 76).

We find this argument unconvincing. Although people certainly do things from "force of habit," de Ruiter's proposal is inconsistent with other things we know about gestures. For example, gesture restriction selectively impairs the retrieval of speech with spatial content (Rauscher et al., 1996), and gestures can facilitate word retrieval in the TOT state (Frick-Horbury & Guttentag, 1998). If gesture's primary function is to communicate, it is not clear why their inhibition would selectively affect the retrieval of one semantic class, nor why their execution would facilitate the resolution of TOT states. But more generally, holding that gestures are both communicatively intended *and* largely ineffective commits one to a model of communication that runs counter to a modern understanding of how people use language (and other behavior) to communicate (see Clark, 1996 for a discussion of language as a joint activity).

This is not to deny that gestures occasionally may enhance a listener's ability to grasp that speaker's intended meanings. Because they bear some relation to semantic attributes of their lexical affiliates, lexical gestures are capable of conveying information (cf. Krauss, Morrell-Samuels & Colasante, 1991, Experiment 3), and this information may be especially helpful when the verbal content is inadequate (e.g., when the speaker is trying to communicate in a language he or she has not fully mastered). Our claim is not that lexical gestures never convey information; we regard the amount and kinds of information they convey, and the circumstances under which they do it, to be an open question. The only strong claim we make is that they play an important role in the retrieval of lexical items grounded in sensorimotor components.

A significant limitation of Experiment 1 is that the electromyographic output of only one muscle group was monitored. In the experiment, we also observed muscular

Movement Facilitates Speech

activity in other theoretically-important but unrecorded regions (e.g., the nondominant arm). Typically these movements were not general and diffuse; rather, they seemed related to the semantic attributes of the target words, regardless of whether these attributes were spatial (e.g., diamond-shaped movements for trellis and horizontal movements for "bleachers") or functional (e.g., rotational movements of the hand for "skewer").

In Experiment 2, speakers gestured more when describing line drawings of objects from memory than they did when the drawings were visually available. They also gestured more when describing drawings of non-nameable objects than drawings of familiar objects. We also replicated the finding of Rauscher et al. (1996) that restricting gesturing decreases speech rate. The claim that these movements serve to increase semantic activation is supported by the finding that they occur frequently when activation is low, as when visual information about a referent is absent, and when the representations involved are difficult to encode verbally, as with non-nameable objects.

Again, the view that what we call lexical gestures are communicative in origin (e.g., Beattie & Shovelton, 2000) is difficult to reconcile with the findings of Experiment 2. Speakers gestured when they were alone, and gesturing occurred most often when semantic activation was low. From a communicative standpoint it is difficult to explain why gesturing should occur in such circumstances or why low levels of semantic activation should be associated with high levels of gesturing.

What initiates gestural activity? In our model, the early semantic processes that generate input to the speech production process routinely involve non-propositional mental representations. We believe that whenever such representations are activated, the kinds of movements we call lexical gestures are likely to be initiated (Krauss et al., 1999; Krauss, et al., 2000). Such movements serve to sustain the activation of the mental representations until the next stage of the speech production process, phonological encoding, has been completed.

The claim that movements can maintain the activation of semantic features, and that these features can participate in the retrieval process, may be a reflection of how semantic knowledge is distributed across different regions of the brain (Tranel, Damasio, & Damasio 1997; Tranel, Logan, Randall, & Damasio, 1997), including regions subserving motor and perceptual functions (for a review see Gainotti, Silveri,

Movement Facilitates Speech

Daniele, & Giustolisi, 1995; Martin, Wiggs, Ungerleider, & Haxby, 1996). The specific area of the brain in which a particular bit of semantic knowledge is stored seems related to the brain regions involved in its acquisition. Reviewing the brain lesion literature, Giannotti et al. (1995) found support for the hypothesis that action words are localized in motor areas, and object names in areas of sensory integration. Martin et al. (1996) found that naming pictures of tools activates the same premotor area that is activated by imagined hand movements. Presumably such movements are an important component of functional knowledge about the objects. As is evident in the Martin et al. (1996) study, it appears that sensorimotor semantic areas of the brain can be activated in diverse ways. Although at this point we can do little more than speculate about underlying neural mechanisms, we believe that gesturing is one way of activating these embodied representations in areas of the brain that are involved in the execution and proprioception of these movements.

We began our discussion by distinguishing between subtractive and additive movements. We then focused on one type of additive movement, the lexical gesture, and examined its relationship to language production. Of course this is not to say that this is the only cognitive role for hand and arm movements. Goldin-Meadow, Nusbaum, Kelly, and Wagner (2001) have found that some gestures decrease cognitive load during an explanation task, and Alibali, Kita, and Young (2000) have found gestures that facilitate the conceptual processes preceding language production. It is worth noting that, of the additive movements we have identified so far--the lexical gesture and the vocal (or subvocal) activities of the phonological loop--all achieve their function via feedback from the execution of an output process (e.g., a motor command). Whether this is mere coincidence or a phenomenon that could elucidate general principles about how working memory operates, and how movements can affect cognition, deserves further investigation.

ACKNOWLEDGMENTS

Experiment 1 was done by the first author under the supervision of the second author in partial completion of the MA degree at Columbia University. We gratefully acknowledge the advice and comments of Lois Putnam, Robert Remez, James Magnuson, Michele Miozzo, and Robert B. Tallarico, and the assistance of Ari Dollid, Jennifer Kim, Stephen Krieger, Anna Marie Nelson, Anne Ribbers, Lauren Walsh, Adam Waytz, and Jillian White.

TABLE 1

Word Type	EMG Amplitude		Trial Duration		Percent Correct	
	Mean	<u>SD</u>	Mean	<u>SD</u>	Mean	<u>SD</u>
Concrete (n = 19)	9.87	8.25	15.05	15.48	63%	18
Correct	9.25	8.57	7.77	3.03		
Incorrect	10.68	9.78	29.05	12.66		
Abstract (n = 17)	8.06	8.22	23.61	9.04	46%	21
Correct	7.81	7.97	9.26	3.82		
Incorrect	8.22	8.74	37.72	16.07		

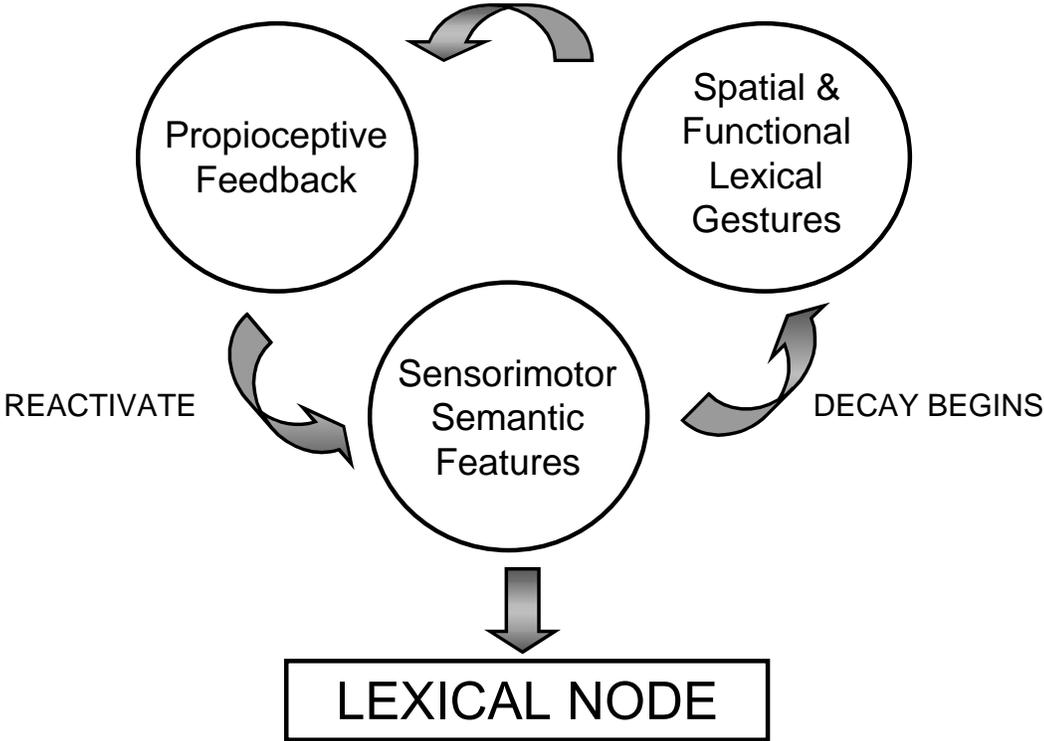


Figure 1: The Gestural Feedback Model

Movement Facilitates Speech

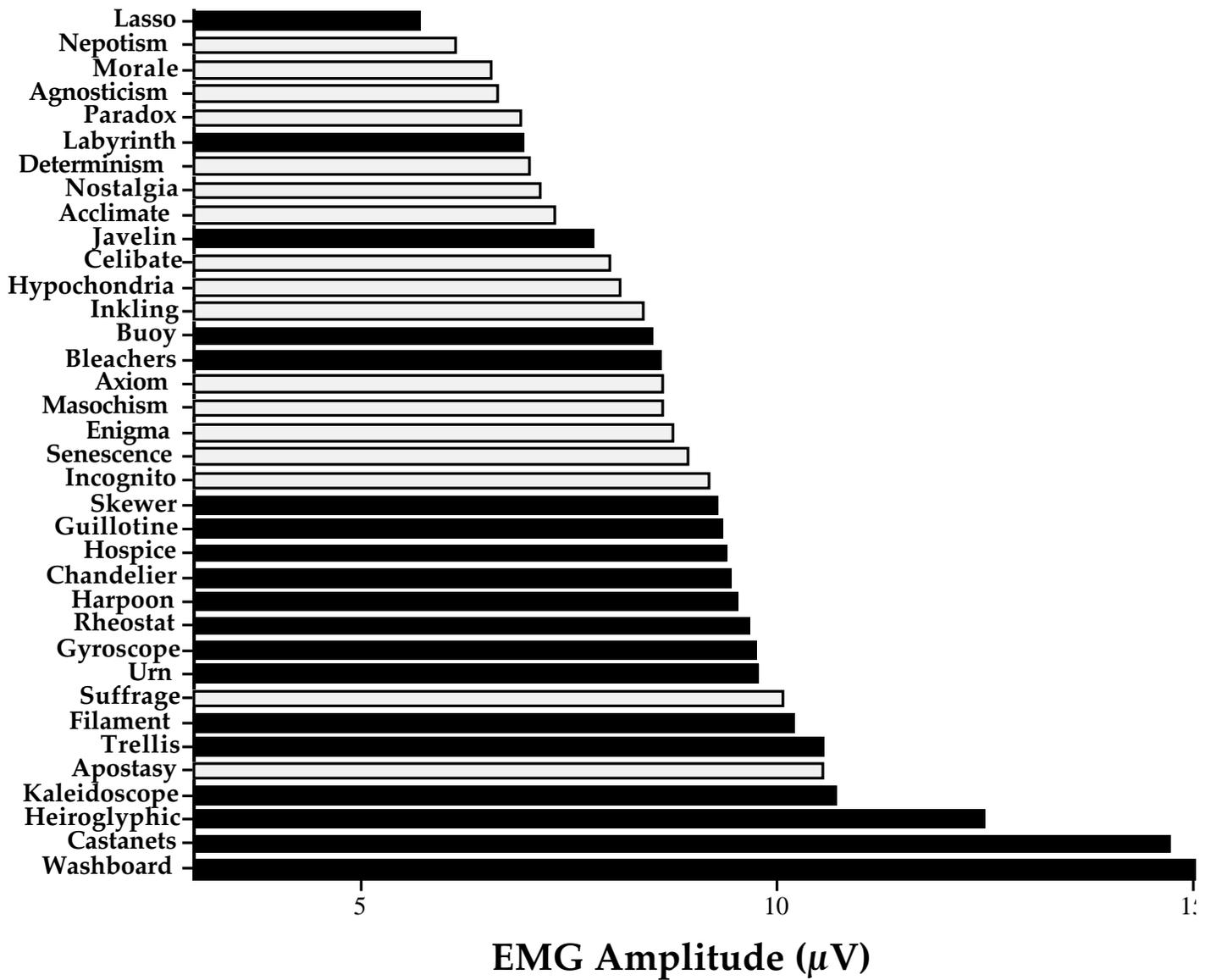


Figure 2. EMG Amplitude (in μV) on Task 1 for the 36 target words (filled bars are concrete words, unfilled bars are abstract words).

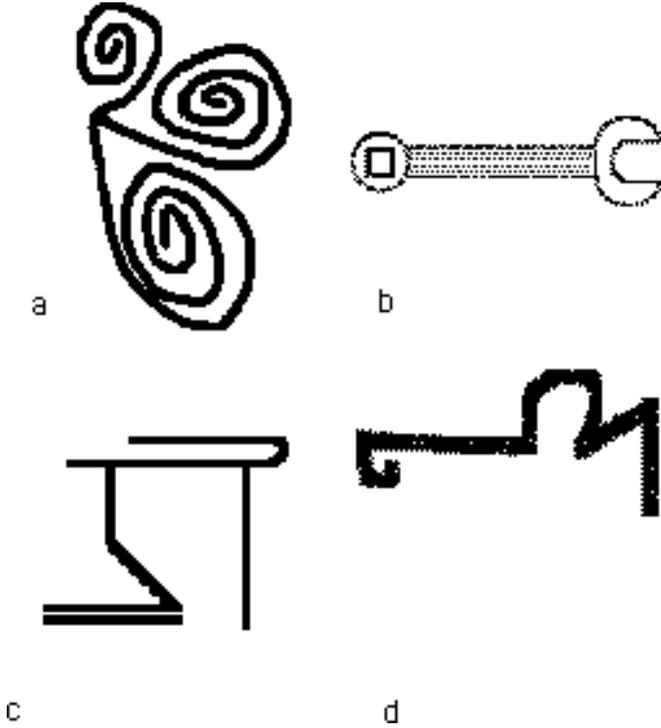


Figure 3: Sample of visual stimuli used in Experiment 2. One nameable (b) and three non-nameable (a, c, and d) objects.

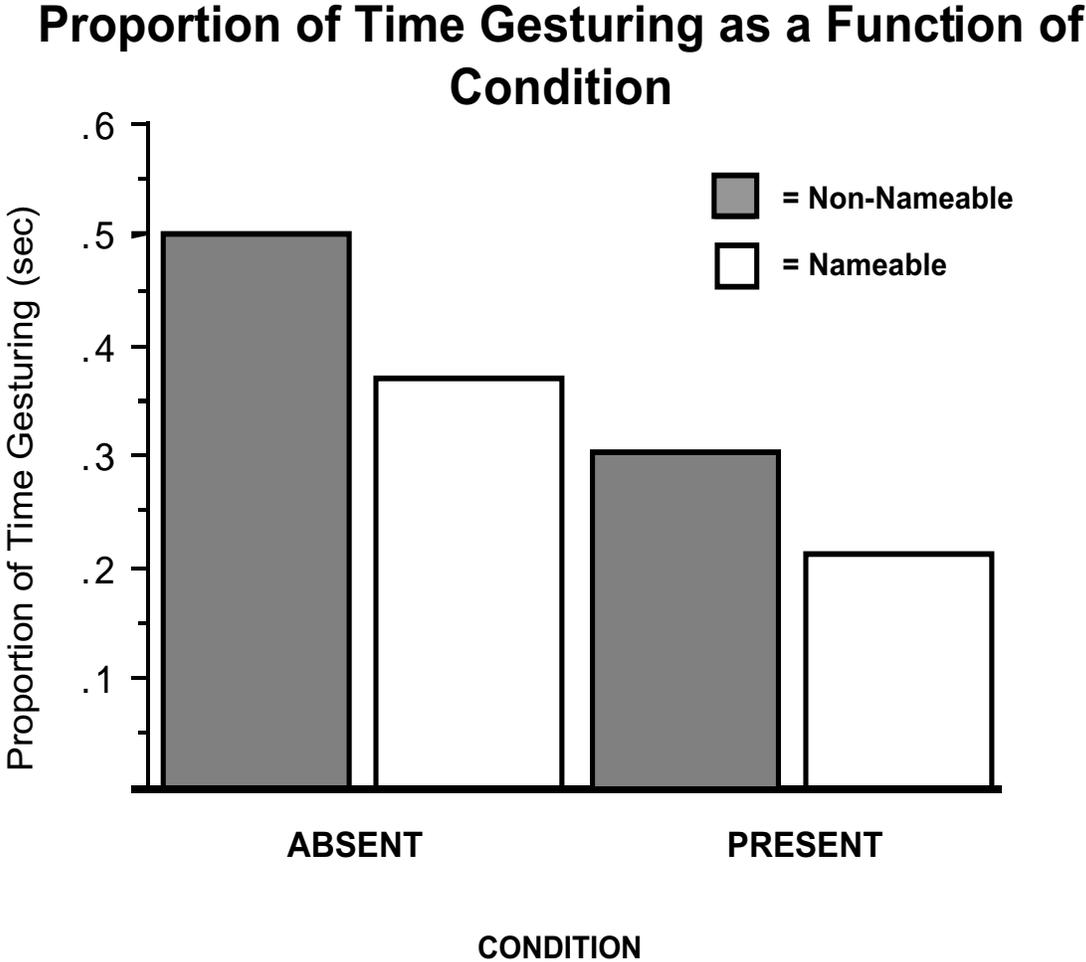


Figure 4: Proportion of time gesturing as a function of condition.

REFERENCES

- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. Language & Cognitive Processes, 15, 593-613.
- Baddeley, A. D. (1986). Working memory. Oxford, England: Oxford University Press.
- Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. Cognition, 57, 193-216.
- Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-560.
- Beattie, G. W. (1980). A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. British Journal of Social Psychology, 20, 243-248.
- Beattie, G. W. & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. British Journal of Psychology, 91, 473-491.
- Brooks, L. R. (1968). Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 22, 349-368.
- Brooks, L. R. (1970). An extension of the conflict between visualization and reading. Quarterly Journal of Experimental Psychology, 22, 91-96.
- Brown, A. S., & Nix, L. A. (1996). Age-related changes in the tip-of-the-tongue experience. American Journal of Psychology, 109, 79-91.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. Psychological Review, 106, 551-581.
- Burke, D. M., Mackay, D. G., Worthley, J. S., & Wade, E. (1991). One the tip-of-the-tongue: What causes word finding failures in young and older adults? Journal of Memory and Language, 30, 542-579
- Butterworth, B. (1978). Maxims for studying conversations. Semiotica, 24, 317-339.
- Butterworth, B. (1989). Lexical access in speech production. In Marslen-Wilson, W (Ed.), Lexical representation and process. Cambridge, MA: MIT Press.

- Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. Psychological Review, *96*, 168-174.
- Caccioppo, J. T., Tassinari, L. G., & Fridlund, A. J. (1990). The skeletomotor system. In J. T. Cacioppo & L. G. Tassinari (Eds.), Principles of psychophysiology: Physical, social, and inferential elements (pp. 325-384). Cambridge, England: Cambridge University Press.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? Cognitive Neuropsychology, *14*, 177-208.
- Chiu, C-y., Hong, Y-y., & Krauss, R.M. (unpublished). Gaze direction and fluency in conversational speech. University of Hong Kong. (Available as downloadable .pdf file at:
<http://www.columbia.edu/~rmk7/PDF/Gaze.pdf>.)
- Clark, H. H. (1996). Using language. Cambridge, UK: Cambridge University Press.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. Behavior Research Methods, Instruments, & Computers, *25*, 257-271.
- de Laguna, G. (1927). Speech: Its function and development. New Haven, CT: Yale University Press.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. Psychological Review, *93*, 283-321.
- De Ruiter, J.-P. (1998). Gesture and speech production. Nijmegen, NI: MPI Series in Psycholinguistics.
- Dittman, A. T. & Llewelyn, L. G. (1969). Body movement and speech rhythm in social conversation. Journal of Personality and Social Psychology, *23*, 283-292.
- Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. Iazykovedenie i Materializm, 105-173.
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal communication: Categories, origins, usage, and coding. Semiotica, *1*, 49-98.
- Farah, M. J. (1995). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), The cognitive neurosciences. Cambridge, MA: MIT Press.

Movement Facilitates Speech

- Feyereisen, P., & deLannoy, J. -D. (1991) Gesture and speech: Psychological investigations. Cambridge, England: Cambridge University Press.
- 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.
- Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. American Journal of Psychology, *111*, 43-62.
- Fromkin, V. A. (1971). The non-anomalous of anomalous utterances. Language, *47*, 27-52.
- 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.
- Gainotti, G., Silveri, M. C., Daniele, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: A critical survey. Memory, *3*, 247-264.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), Language Production. Vol. 1: Speech and Talk. London: Academic Press.
- Garrity, L. I. (1977). Electromyography: A review of the current status of subvocal speech research. Memory and Cognition, *5*, 615-622.
- Glenberg, A. M. (1997). What memory is for. Behavioral & Brain Sciences, *20*, 1-55.
- Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. Memory & Cognition, *26*, 651-658.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesture lightens the load. Psychological Science, *12*, 516-522.
- Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N. (1976). The "tip-of-the-tongue" phenomenon in aphasia. Cortex, *12*, 145-153.

Movement Facilitates Speech

- Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. Journal of Experimental Psychology, *61*, 23-29.
- Hadar, U., Wenkert-Olenik, D., Krauss, R.M., Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. Brain & Language, *62*, 107-126.
- Hebb, D. O. (1968). Concerning imagery. Psychological Review, *75*, 466-477.
- Jessen, F., Heun, R., Erb M., Granath D.-O. Klose, U. Papassotiropoulos, A. & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. Brain and Language, *74*, 103-112.
- Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the tongue phenomenon. Memory & Cognition, *17*, 69-76.
- Kay, J., & Ellis, A. (1987). A cognitive neuropsychological case study of anomia. Brain, *110*, 613-629.
- Kempen, G. & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. Cognition, *14*, 185-209.
- Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In Key (Ed.), pp. 207-227.
- Kendon, A. (1983). Gesture and speech: how they interact. In J. M. Weimann & R. P. Harrison (Eds.), Nonverbal interaction. (pp. 14-35). Beverly Hills, CA: Sage Publishers.
- Krauss, R. M., Chen Y., Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? Advances in Experimental Social Psychology, *28*, 389-450.
- Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), Language and gesture. Cambridge, UK: Cambridge University Press.
- Krauss, R.M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In, R. Campbell & L. Messing (Eds.), Gesture, speech, and sign (93-116). Oxford: Oxford University Press.

Movement Facilitates Speech

- Krauss, R.M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? Journal of Personality and Social Psychology, *61*, 743-754.
- Kucera, H. & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.
- Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: The MIT Press.
- MacKay, D. G. (1987). The organization of perception and action: A theory for language and other cognitive skills. New York: Springer Verlag.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), The new cognitive neurosciences, second Edition, (pp. 1023-1036). Cambridge, MA: The MIT Press
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature, *379*, 649-652.
- McNeill, D. (1992). Hand and mind: What gestures reveal about thought. Chicago: The University of Chicago Press.
- Mead, G. H. (1934). Mind, self, and society. Chicago: University of Chicago Press.
- Meyer, A.S., & Bock, K. (1992). The tip-of-the tongue phenomenon: Blocking or partial activation? Memory & Cognition, *20*, 715-726.
- McNeill, D., Cassell, J., & McCullough, K-E. (1994). Communicative effects of speech-mismatched gestures. Research on Language and Social Interaction, *27*, pp. 223 - 237.
- Morsella, E. & Miozzo, M. (in press). Evidence for a cascade model of lexical access in speech production. Journal of Experimental Psychology: Learning, Memory, and Cognition.
- Moscovici, S. (1967). Communication processes and the properties of language. In Berkowitz (ed.), pp. 226-270.
- Paivio, A. (1979). Imagery and verbal processes.. Hillsdales, New Jersey.: Lawrence Erlbaum Associates, Publishers.

Movement Facilitates Speech

- Perfect, T. J., & Hanley, J. R. (1992). The tip-of-the-tongue phenomenon: Do experimenter-presented interlopers have any effect? Cognition, *45*, 55-75.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in the processing of speech. Psychological Science, *7*, 226-231.
- Rimé, B., Schiaratura, L., Hupet, M., & Ghysseleinckx, A. (1984). Effects of relative immobilization on speaker's nonverbal behavior and on dialogue imagery level. Motivation and Emotion, *8*, 311-325.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in production: Picture-word interference studies. Journal of Memory and Language, *29*, 86-102.
- Stemberger, J. P. (1985). Bound morpheme errors in normal and agrammatic speech: One mechanism or two? Brain and Language, *25*, 246-256.
- Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. Neuropsychologia, *35*, 1319-1327.
- Tranel, D., Logan, C. G., Randall, J. F., & Damasio, A. R. (1997). Explaining category-related effects in the retrieval of conceptual and lexical knowledge for concrete entities: operationalization and analysis of factors. Neuropsychologia, *35*, 1329-1339.
- Washburn, M. F. (1928). Emotion and thought: a motor theory of their relation. In C. Murchison (Ed.). Feelings and emotions: the Wittenberg Symposium. Massachusetts: Clark University Press.
- Werner, H. & Kaplan, B. (1963). Symbol formation. New York: Wiley.
- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. American Journal of Psychology, *114*, 591-600.

FOOTNOTES

¹ Lexical gestures should be distinguished from symbolic gestures or emblems, which are hand-arm signs with conventionalized meanings (e.g., the "thumbs-up" sign), and from motor gestures or beats, which are brief, repetitive co-speech movements that are roughly coordinated with the speech prosody, but seem unrelated to the semantic content.

² In psycholinguistics, this level of representation is sometimes referred to as the "lemma" (Kempen & Huijbers, 1983; Levelt, 1989). We prefer the theoretically more neutral term "lexical node."

³ We chose low-frequency words because they are more difficult to retrieve than high-frequency words, and hence more likely to benefit from the hypothesized facilitatory effects of gesturing.

⁴ In a recently published paper, Wesp, Hesse, Keutmann, and Wheaton (2001) describe an experiment similar to ours in design, and report identical results for one of our main independent variables. Experiment 2 was completed before we were aware of the Wesp et al. paper.