Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 101 (2007) 234–245
Brain and Language
www.elsevier.com/locate/b&l

How iconic gestures enhance communication: An ERP study

Ying Choon Wu, Seana Coulson *

Cognitive Science Department 0515, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA

Accepted 5 December 2006
Available online 12 January 2007

Abstract

EEG was recorded as adults watched short segments of spontaneous discourse in which the speaker's gestures and utterances contained complementary information. Videos were followed by one of four types of picture probes: cross-modal related probes were congruent with both speech and gestures; speech-only related probes were congruent with information in the speech, but not the gesture; and two sorts of unrelated probes were created by pairing each related probe with a different discourse prime. Event-related potentials (ERPs) elicited by picture probes were measured within the time windows of the N300 (250–350 ms post-stimulus) and N400 (350–550 ms post-stimulus). Cross-modal related probes elicited smaller N300 and N400 than speech-only related ones, indicating that pictures were easier to interpret when they corresponded with gestures. N300 and N400 effects were not due to differences in the visual complexity of each probe type, since the same cross-modal and speech-only picture probes elicited N300 and N400 with similar amplitudes when they appeared as unrelated items. These findings extend previous research on gesture comprehension by revealing how iconic co-speech gestures modulate conceptualization, enabling listeners to better represent visuo-spatial aspects of the speaker's meaning.

Keywords: Gesture; N400; N300; Semantic integration; Language comprehension; Object recognition; Conceptual integration; Embodiment; ERP; Meaning; Simulation

1. Introduction

Co-speech gestures provide a channel for speakers to express additional information related to their communicative intent. While uttering, "It's actually a double door," for example, a speaker may indicate the shape of a Dutch rather than French style door with the configuration of his hands (see Fig. 1). A number of behavioral studies suggest that gestures such as these play a beneficial role in communication. Listeners rely on speakers' gestures to disambiguate communicative intent in cases where understanding may be impeded—due to noise in the speech signal, for example (Rogers, 1978; Thompson & Massaro, 1986, 1994), or due to additional inferential processing engendered by indirect requests (Kelly, 2001; Kelly, Barr, Church, & Lynch, 1999). Listeners also exhibit a more accurate understanding of instructions and narratives when the speaker's accompanying gestures are visible (Beattie & Shovelton, 1999b, 2002; Graham & Argyle, 1975; Morford & Goldin-Meadow, 1992; Singer & Goldin-Meadow, 2005; Valenzeno, Alibali, & Klatzky, 2003). However, see Krauss, Dushay, Chen, and Rauscher (1995) and Goldin-Meadow and Sandhofer (1999) and for an alternative view.

These findings suggest that some properties of gestures may activate semantic information related to the content of the talk in progress. However, little is known about the cognitive and neural processes mediating this remarkable feat of multi-modal integration. Given growing interest in the role of motor mirroring systems in action comprehension (Rizzolatti & Arbib, 1998), the study of gesture may provide cognitive neuroscience with a further venue for understanding the relationship between sensori-motor and higher order conceptual processing.

It has been proposed by McNeill and others that during comprehension, speech and gesture are integrated into a common underlying conceptual representation. He writes,

* Corresponding author. Fax: +1 858 534 1128.
E-mail address: coulson@cogsci.ucsd.edu (S. Coulson).

0093-934X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2006.12.003

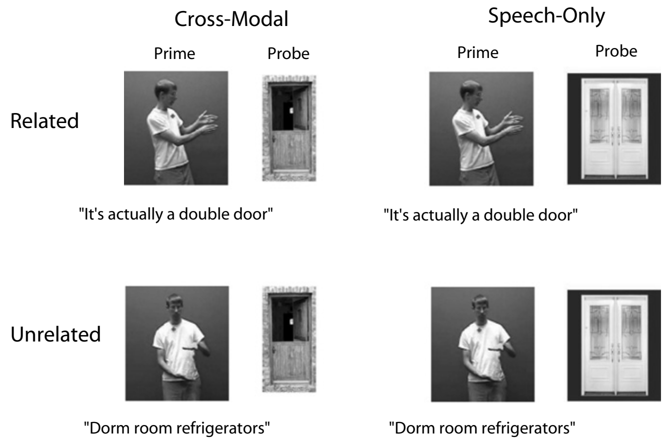

Fig. 1. Experimental design. Cross-modal and speech-only picture probes occurred in related and unrelated trials.

"…listeners, after a brief delay, cannot tell whether information came to them in gesture or in speech, the two having become unified (p. 17) (McNeill, 1998)." In support of this idea, a number of studies have investigated the comprehension of discourse in which the speakers' gestures express something different from their words (as in the Dutch door example). By assessing listeners' own accounts of what they had understood (Alibali, Flevares, & Goldin-Meadow, 1997; Cassell, McNeill, & McCullough, 1999; Goldin-Meadow, Wein, & Chang, 1992; Kelly & Church, 1998), or their responses on questionnaires (Goldin-Meadow & Sandhofer, 1999), it has been demonstrated that listeners are sensitive to information made available in both modalities.

The goal of the present study is to investigate how speech and gesture affect real-time interpretation processes. Previous behavioral research has demonstrated that when information is presented to listeners only through gesture, but not directly in speech, it is nevertheless accessible in long-term memory (for review, see Goldin-Meadow, 2003). However, little is known about the encoding processes whereby gesture-based information enters memory systems. Further, semantic activations induced by co-speech gestures have only begun to be investigated. The present study addresses the cognitive and neural processes mediating speech–gesture integration.

Recent research involving event-related potentials (ERPs) has begun to shed light on this question. ERPs represent dynamic voltage fluctuations that derive from synaptically generated current flow within patches of neural tissue. Tiny signals detectable at the scalp (on the order of microvolts) are amplified and digitized, yielding a record of on-going brain activity in the form of an electroencephalogram (EEG). By averaging portions of EEG recorded in synchrony with the presentation of a specific class of stimuli, it is possible to draw inferences about cognitive processes engaged by that type of stimulus. Because scalp-recorded potentials typically reflect contributions from a number of different neural sources, it is necessary to average event-related responses across many trials in order to cancel out random noise introduced by background neural activity. The resulting ERP waveform can be analyzed as a series of positive- and negative-going deflections (commonly referred to as components) which are characterized by their amplitude, time course and distribution across scalp electrode sites.

A component particularly relevant to semantic processing is the N400, which was discovered during early research on language processing (Kutas & Hillyard, 1980). Kutas and Hillyard recorded ERPs to the last word of sentences that either ended congruously (as in (1)), or incongruously (as in (2)).

(1) I take my coffee with cream and sugar.
(2) I take my coffee with cream and dog.

By averaging the signal elicited by congruous and incongruous sentence completions, respectively, these investigators were able to reveal systematic differences in the brain's electrical response to these stimulus categories occurring approximately 400 ms after stimulus onset. Subsequent research has shown that N400 components are generated whenever stimulus events induce semantic or conceptual processing. As such, many investigators have used the N400 component of the brain waves as a dependent variable in psycholinguistic experiments (for review, see Kutas, Federmeier, Coulson, King, & Muente, 2000).

To investigate the effect of gestures on language comprehension, Kelly, Kravitz, and Hopkins (2004) recorded ERPs elicited by spoken words articulated in synchrony with gestures that were either congruent and incongruent with word meanings. Stimuli were constructed by videotaping an actor as he gestured to either a tall, thin glass or a short, wide dish in front of him while saying one of four speech tokens—namely, *tall*, *thin*, *short* or *wide*. Gestures indicated the location of these two items, and also depicted either the height or width of their referent. In the matching condition, the actor's speech corresponded with both the object as well as the spatial dimension indicated in gesture. In the complementary condition, the speech token described a different dimension of the referent from that depicted by the gesture (e.g. *tall* uttered in accompaniment with a gesture indicating the thin diameter of the glass). In the mismatch condition, the speech token corresponded to one object while the gesture corresponded to the other. Finally, in the no gesture condition, speech was presented alone.

Results yielded early effects of gesture congruency (between 100 and 352 ms), with mismatching and complementary stimuli eliciting relative to other conditions larger P1 and P2 components, which reflect auditory sensory processing. N400-like effects were also observed at bilateral temporal electrode sites, with mismatch trials eliciting more negative ERPs than all other conditions around 450 ms post-stimulus. These findings suggest that gesture congruency affects both early sensory as well as higher order semantic processing of words.

Other studies have approached the neuro-cognitive underpinnings of gesture comprehension by measuring ERPs elicited by gestures themselves. Besides words, the