CENTER FOR RESEARCH IN LANGUAGE

August 2005 **Vol. 17, No. 2**

The Newsletter of the Center for Research in Language, University of California, San Diego, La Jolla CA 92093-0526
Tel: (858) 534-2536 • E-mail: editor@crl.ucsd.edu • WWW: http://crl.ucsd.edu/newsletter

• • •

FEATURE ARTICLE

Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures

Ying Choon Wu

Department of Cognitive Science
University of California, San Diego

EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition (distributed via the World Wide Web) and welcome response from friends and colleagues at UCSD as well as other institutions. Please visit our web site at `http://crl.ucsd.edu`.

SUBSCRIPTION INFORMATION

If you know of others who would be interested in receiving the newsletter, you may add them to our email subscription list by sending an email to `majordomo@crl.ucsd.edu` with the line `"subscribe newsletter <email-address>"` in the body of the message (e.g., `subscribe newsletter jdoe@ucsd.edu`). Please forward correspondence to:

John Lewis and Jenny Staab, Editors
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (858) 534-2536 • E-mail: editor@crl.ucsd.edu

Back issues of this newsletter are available on our website. Papers featured in recent issues include the following:

Analyzing Semantic Processing Using Event-Related Brain Potentials
Jenny Shao, Northwestern University
Helen Neville, University of Oregon
Vol. 11, No. 5, December 1998

Blending and Your Bank Account: Conceptual Blending in ATM Design
Barbara E. Holder
Department of Cognitive Science, UCSD
Vol. 11, No. 6, April 1999

Could Sarah Read the Wall Street Journal?
Ezra Van Everbroeck
Department of Linguistics, UCSD
Vol. 11, No. 7, November 1999

Introducing the CRL International Picture-Naming Project (CRL-IPNP)
Elizabeth Bates, et al.
Vol. 12, No. 1, May 2000

Objective Visual Complexity as a Variable in Studies of Picture Naming
Anna Székely
Eotvos Lorand University, Budapest
Elizabeth Bates
University of California, San Diego
Vol. 12, No. 2, July 2000

The Brain's Language
Kara Federmeier and Marta Kutas
Department of Cognitive Science, UCSD
Vol. 12, No.3, November 2000

The Frequency of Major Sentence Types over Discourse Levels: A Corpus Analysis
Frederic Dick and Jeffrey Elman
Department of Cognitive Science, UCSD
Vol. 13, No.1, February 2001

A Study of Age-of-Acquisition (AoA) Ratings in Adults
Gowri K. Iyer, Cristina M. Saccuman, Elizabeth A. Bates, and Beverly B.Wulfeck
Language & Communicative Disorders, SDSU & UCSD and Center for Research in Language, UCSD
Vol. 13, No. 2, May 2001

Syntactic Processing in High- and Low-skill Comprehenders Working under Normal and Stressful Conditions
Frederic Dick, Department of Cognitive Science, UCSD
Morton Ann Gernsbacher, Department of Psychology, University of Wisconsin
Rachel R. Robertson, Department of Psychology, Emory University
Vol. 14, No. 1, February 2002

Teasing Apart Actions and Objects: A Picture Naming Study
Analia L. Arevalo
Language & Communicative Disorders, SDSU & UCSD
Vol. 14, No. 2, May 2002

The Effects of Linguistic Mediation on the Identification of Environmental Sounds
Frederic Dick , Joseph Bussiere and Ayşe Pınar Saygın
Department of Cognitive Science and Center for Research in Language, UCSD
Vol. 14, No. 3, August 2002

On the Role of the Anterior Superior Temporal Lobe in Language Processing: Hints from Functional Neuroimaging Studies
Jenny Staab
Language & Communicative Disorders, SDSU & UCSD
Vol. 14, No. 4, December 2002

A Phonetic Study of Voiced, Voiceless, and Alternating Stops in Turkish
Stephen M. Wilson
Neuroscience Interdepartmental Program, UCLA
Vol. 15, No. 1, April 2003

New corpora, new tests, and new data for frequency-based corpus comparisons
Robert A. Liebscher
Cognitive Science, UCSD
Vol. 15, No.2; December 2003

The relationship between language and coverbal gesture in aphasia
Eva Schleicher
Psychology, University of Vienna & Cognitive Science, UCSD
Vol. 16, No. 1, January 2005

In search of Noun-Verb dissociations in aphasia across three processing tasks
Analía Arévalo, **Suzanne Moineau**
Language and Communicative Disorders, SDSU & UCSD, Center for Research in Language, UCSD
Ayşe Saygin
Cognitive Science & Center for Research in Language, UCSD
Carl Ludy
VA Medical Center Martinez
Elizabeth Bates
Cognitive Science & Center for Research in Language, UCSD
Vol. 17, No. 1, March 2005

CRL Newsletter, Vol. 17 No. 2, August 2005

Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures

Ying Choon Wu

Department of Cognitive Science
University of California, San Diego

Abstract

In conversation, communication transpires through both talk and gesture. While describing a platter, a speaker might trace the outline of its shape in the air. Does understanding such iconic gestures engage processes similar to those recruited in the comprehension of pictures, photographs, and other image-based representations of the visual world? Previous research using event-related potentials (ERPs) has demonstrated that the second member of an unrelated picture pair results in enhanced negative going deflections of the ERP waveform around 300 ms (N300) and 400 ms (N400) post-stimulus onset as compared to responses elicited by related picture probes. To test whether the semantic analysis of gestures elicits similar effects, we recorded the electroencephalogram (EEG) from 16 healthy adults as they viewed spontaneously produced iconic co-speech gestures preceded by congruous and incongruous contexts. Gestures were presented either dynamically in the form of short, soundless video-clips or statically in the form of freeze frames extracted from gesture videos. N400-like effects were observed in response to both gesture types. Static gesture stills also elicited an N300 effect. These findings demonstrate that understanding gestures involves processes similar to those underlying the comprehension of other meaningful representations, including words and pictures.

Introduction

Little is known about the underlying mechanisms mediating the comprehension of co-speech iconic gestures. Given the close rhythmic and semantic correspondence between gesture and speech, as well as overlap in their developmental trajectories through childhood (Bates & Dick, 2002), researchers have proposed that these modalities comprise complementary modes of signification, mediated by more general capacities for thought and symbolization. Gestures afford the opportunity to encode global, holistic relations, which contrast with the analytic, linearly segmentable properties of speech (McNeill, 1992).

One promising venue for investigating this theory experimentally is through event-related potentials (ERPs). Via scalp recordings of dynamic voltage fluctuations time-locked to the onset of stimulus presentation, researchers have discovered similar patterns of brain response during the semantic analysis of image and language based representations. In particular, both pictures and words elicit a negative-going deflection of the ERP waveform around 400 ms (N400) post-stimulus onset, with enhanced amplitude in response to contextually incongruent items (Kutas & Federmeier, 2000). This finding points to commonalities in the underlying processes mediating the comprehension of both imagistic and linguistic representation. However, N400 effects elicited by pictures tend to be larger over the front of the head, while N400 effects produced by words tend to be larger at centro-parietal sites. This topographical difference is consistent with overlapping, but non-identical neural systems responsible for picture and word comprehension.

The present study will address the possibility that understanding iconic co-speech gestures recruits processes similar to those implicated in the comprehension of other kinds of image-based representations, including line drawings, photographs, and video. Constructing meaning from images is proposed to involve a variety of inter-dependent sub-processes, whereby visual features are integrated into a higher order structural description, possibly involving three dimensional geometric primitives such as geons (Biederman, 1987). Perceptually-based structural descriptions are matched with object representations in long term memory, making accessible associated semantic knowledge, such as the object's name. Electrophysiological research has uncovered two correlates of image-based semantic analysis. Specifically, in pictorial priming paradigms contextually incongruent pictures have been shown to elicit an anterior negativity peaking around 300 ms after stimulus onset (N300), as well as a more broadly distributed negativity peaking approximately 400 ms post-stimulus (N400) (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999).

The N400 response to incongruous images has been related to the "classic" N400 elicited by linguistic stimuli. A well-studied ERP component, the lexical N400 is generally thought to index the degree of semantic fit between a word and its preceding context (Coulson & Federmeier, 2003, in press). Expected sentence-final endings typically elicit little or no N400, for example, while unexpected completions elicit an N400 component with a large amplitude (Kutas & Hillyard, 1980, 1984). Likewise, sentences completed with unexpected pictures also result in a large N400 response as compared to expected ones (Federmeier & Kutas, 2001, 2002; Ganis, Kutas, & Sereno, 1996; Nigam, Hoffman, & Simons, 1992).

In addition to sentential congruity, both the word and picture N400 are sensitive to different degrees of relatedness, with larger amplitudes for the second item of an unrelated pair (Kutas & Hillyard, 1988; McPherson & Holcomb, 1999). Just as pseudo-words elicit larger N400s than unrelated words (Holcomb, 1988), unrecognizable images elicit larger N400s than recognizable ones (McPherson & Holcomb, 1999). Further, the amplitude of both the word and the picture N400 is modulated by the global, discourse-level coherence of a word or picture within a story context (van Berkum, Hagoort, & Brown, 1999; West & Holcomb, 2002). The similarities in the time course and functional characterization of the word and picture N400 have led to the suggestion that the neural system involved in understanding pictures may function similarly to the system responsible for understanding language.

However, the N300, which has not been found in response to words, also suggests some processes specific to image comprehension. Like the N400, the N300 is modulated by contextual congruity (Hamm, Johnson, & Kirk, 2002). Yet, the amplitude of the N400 exhibits sensitivity to gradations in relatedness between an image and its preceding context, with highly related items eliciting little or no N400, moderately related items eliciting intermediate-sized N400, and unrelated items eliciting even more enhanced N400. By contrast, in a study involving image pairs with graded degrees of associative relatedness (e.g. highly related: toothbrush and toothpaste; moderately related: ketchup and mustard; unrelated: lion and stoplight), the amplitude of the N300 reflected differentiation only for related and unrelated items, but not for moderately and highly related ones. That is, N300 amplitude reflects whether items are related, but not the strength of the relation, since moderately related and highly related items have equally enhanced amplitude relative to unrelated items. Further, N300 effects tend to be largest over anterior electrode sites, while N400 effects are more broadly distributed (McPherson & Holcomb, 1999). On the basis of these functional and spatial differences, researchers have proposed that the picture N300 and N400 index different aspects of the semantic processing of images. In particular, the N300 has been proposed to reflect the process of matching perceptual input with image-based representations in long term memory (Schendan & Kutas, 2002; West & Holcomb, 2002), while the N400 family of potentials has been related to the integration of semantic information irregardless of input modality.

Previous studies have tested the hypothesis that understanding iconic gestures is mediated by some of the same processes involved in the comprehension of other kinds of meaningful representations, including pictures and words (Wu & Coulson, 2004). A corpus of spontaneous, iconic co-speech gestures was produced by videotaping an individual describing cartoon segments and then using the recording to create short, soundless video clips of the speaker's gestures. The speaker was told that the experimenters were creating stimuli for a memory experiment and was unaware of the intent to elicit spontaneous gestures. Trials were constructed by pairing the original cartoon segments with either congruous or incongruous gesture clips. EEG was recorded from twenty four healthy adults as they viewed 160 paired cartoon-gesture sequences, and ERPs were time-locked to the onset of gesture presentation. Both congruous and incongruous gestures elicited a negative going deflection of the waveform peaking around 450 ms after stimulus onset (gesture N450), with enhanced negativity for incongruous items. Because of their similarities in time course and sensitivity to contextual congruity, the gesture N450 was construed as a member of the N400 class of negativities, and proposed to index the integration of gesturally-based semantic information with preceding context. No N300 effects were observed.

The absence of N300 in preceding experiments may be interpreted as an indication that understanding gestures do not involve the kinds of image-specific activations indexed by the N300. Alternatively, it is also possible that due to the visual complexity of dynamic gestures as compared to static representations of objects, the processes indexed by the N300 are delayed in onset during gesture comprehension, overlapping with those indexed by the N450. Visually simpler "static gesture snapshots" may yield earlier, more discernible N300-like activity, in keeping with the idea that gestural representations engage image based processes in much the same manner as pictorial and photographic ones.

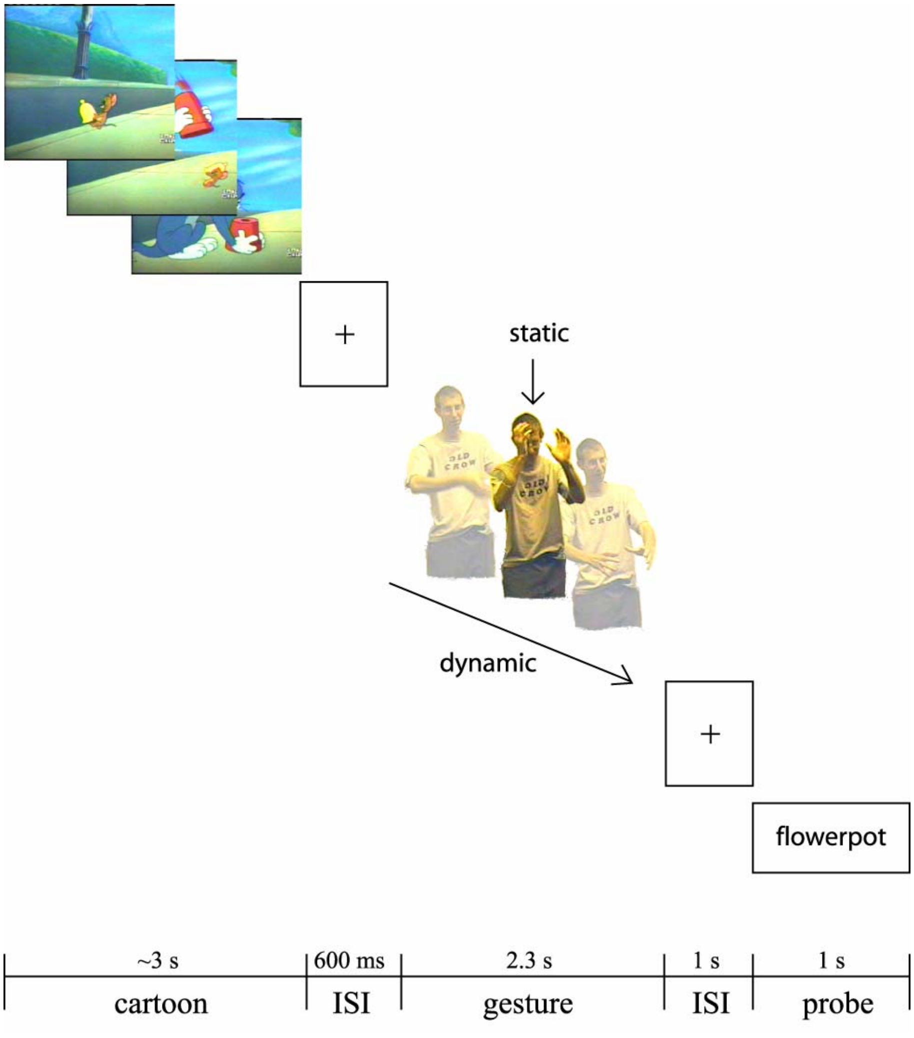

Figure 1. Trials included a short cartoon segment followed by either a dynamic or static gesture and then a probe word.

The goal of the present study was to assess ERPs elicited in response to static freeze frames extracted from dynamic gesture video clips. To create static gestures, we isolated a single frame (in the form of a jpeg file) from the video sequence of each gesture. Typically, jpegs were extracted just at the onset of the meaningful phase of movement – that is, the stroke (Kendon, 1972) phase of the gesture – or during a pause between strokes, in the case of iterative or complex gestures, in order to avoid blurring resulting from rapid motion of the hands. Although static representations of iconic gestures contain considerably less information than dynamic ones, jpegs were selected which preserved important semiotic cues, such as hand shape, body configuration, and hand location. Thus, understanding these gesture "snapshots" is likely to involve similar visual analysis and integration processes mediating the understanding of full gestures.

Trials were constructed by pairing cartoon contexts with congruent and incongruent dynamic and static gestures, with each gesture following the presentation of a cartoon. Related or unrelated probe words followed each cartoon-gesture pair, and participants were asked to judge whether or not the probe was related to the preceding stimuli. This procedure provided a way to assess participant comprehension without requiring explicit decisions regarding the stimulus of interest. Participants were presented with one of four lists, which were constructed such that across lists, each static and dynamic gesture appeared as a congruous and incongruous item, and no items were repeated within lists. In this way, each stimulus served as its own control.

EEG was recorded from 16 healthy adults as they viewed cartoon-gesture pairs and as they classified subsequent probe words as related or unrelated to the preceding content (see Figure 1). ERPs were time-locked to the onset of gestures, and congruency effects on the N300 and N400 were assessed by measuring the mean amplitude of the ERPs between 300 and 400 ms post-stimulus and 400 and 600 ms post-stimulus, respectively. More negative ERPs elicited by incongruous items relative to congruous ones in these time windows would be consistent with findings from other investigations using image-based media, suggesting that comprehending gestures recruits cognitive resources similar to those engaged during the comprehension of pictures.

Results

Behavior

On average, participants classified 87% of related words and 94% of unrelated words correctly. A 2x2 repeated measures ANOVA with the factors of Word Relatedness (Related, Unrelated) and Gesture Type (Static, Dynamic) revealed that this difference was statistically reliable (F(1,15)=11, p=.005), suggesting a preference for the "unrelated" response. No difference in mean accuracy classifying probe words following dynamic and static gestures was observed (F=.65, n.s.), nor were any interactions between Gesture Type and Word Relatedness (F=.05, n.s.).

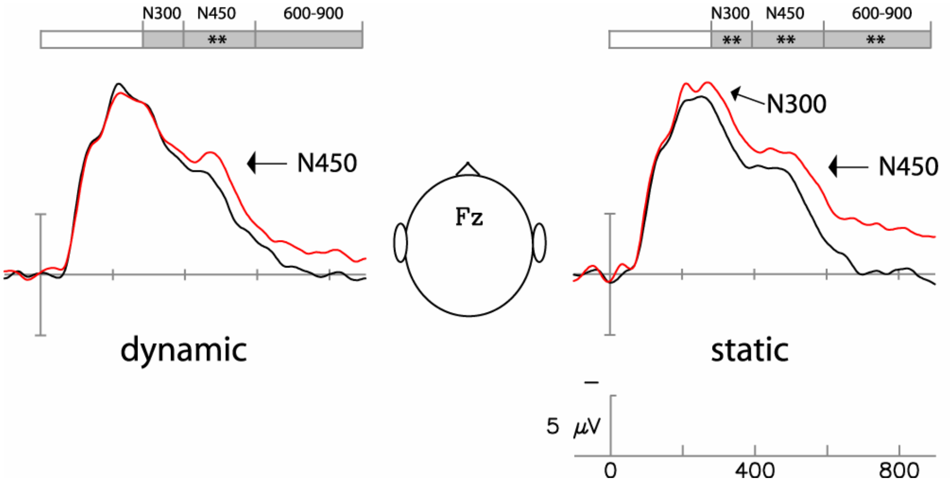

Figure 2. ERPs time-locked to the onset of static and dynamic gestures. Note that negative is plotted up.

Event Related Potentials

Figure 2 shows ERPs recorded at the frontal midline electrode site (Fz), time-locked to the onset of static and dynamic gestures. For all trials, a large, negative-going onset potential can be observed, peaking around 240 ms after stimulus onset, followed by a second negative-going deflection of the waveform peaking around 450 ms post-stimulus. In the case of dynamic gestures, effects of gesture congruency can be observed from around 400 ms to the end of the epoch (900 ms), with enhanced negativity for incongruous items relative to congruous ones. In the case of static gestures, congruency effects begin earlier, around 230 ms post-stimulus, again with incongruous items continuing to elicit more negative ERPs until the end of the epoch.

To assess the effect of gesture congruency on the N300, the mean amplitude of the ERPs was measured between 300 and 400 ms post-stimulus and subjected to a 2x2 repeated measures ANOVA with the factors of Congruency (congruent, incongruent) and Gesture Type (static, dynamic), as well as Electrode Site (29 levels). An interaction between Congruency and Gesture Type (F(1, 15)=5.7, p<.05) licensed follow-up comparisons of congruency effects within static and dynamic gestures individually. For static gestures, incongruous trials resulted in more negative ERPs (F(1,15)=6, p<.05), while for dynamic gestures, gesture congruency did not reliably modulate ERP amplitudes (F=.16, n.s.) during this epoch.

To assess the effect of gesture congruency on the gesture N450, the same analysis was applied to the mean amplitude of ERPs measured between 400 and 600 ms post-stimulus. A main effect of Congruency – collapsed across both types of gesture – was obtained (F(1,15)=12.5, p<.005), qualified by an interaction with Electrode Site (F(28,420)=3, p<.05). The distribution of the N450 effect was investigated with follow-up analyses at midline, medial and lateral electrode strips. At the midline, a trend toward a Congruency x Posteriority interaction (F(6,90)=3, p=.07) was driven by the increased congruency effect over frontal electrode sites relative to posterior ones for both static and dynamic gestures. For both gesture types, the congruency effect was also larger over right hemisphere sites than over left hemisphere ones (Congruency x Hemisphere: Lateral (F(1,15)=4.9, p<.05)).

Finally, comparisons of the mean amplitude of ERPs elicited between 600 and 900 ms post-stimulus yielded a main effect of Congruency (F(1,15)=13.6, p<.005), as well as an interaction between Congruency and Gesture Type (F(1,15)=5.2, p<.05). This interaction was found to be driven by the absence of reliable congruency effects for dynamic gestures within this time window (F=1.7, n.s.), in contrast to robust effects obtained for static gestures (Congruency Main Effect (F(1,15)=26, p<.0005)): Congruency by Electrode Site (F(28,420)=4.25, p<.01). For dynamic gestures, post hoc analyses revealed a main effect of congruency at midline electrodes only (F(1,15)=5.3, p<.05). The static gesture congruency effect was larger over frontal electrode sites (Congruency x Posteriority: Midline (F(6,90)=5.3, p<.05); Lateral: (F(3,45)=4.4, p<.05), and exhibited a trend toward enhancement over right hemisphere electrode sites relative to left hemisphere ones (Congruency x Hemisphere: Lateral (F(1,15)=3.5, p=.08)).

Discussion

To investigate the online processing of gestures, we recorded ERPs time-locked to the onset of spontaneously produced, iconic gestures preceded by congruent and incongruent contexts. In addition to congruency, the dynamicity of gestures was also manipulated: dynamic gestures consisted of short video clips, while static gestures were comprised of single still images extracted from each dynamic gesture stream. Between 400 and 600 ms post-stimulus, all gestures elicited a negative deflection of the ERP waveform (gesture N450), with incongruous items yielding enhanced negativity relative to congruous ones. Incongruous static gestures also elicited an enhanced, frontally focused negativity peaking between 300 and 400 ms post-stimulus (N300). Both types of gestures elicited a late congruency effect (600-900 ms), which was much larger, more robust, and more broadly distributed for static items.

The gesture N450 effect described here corroborates outcomes reported in previous studies using this paradigm (Wu & Coulson, 2004). It is hypothesized to index semantic integration processes similar to those indexed by the N400 observed in response to the semantic analysis of pictures, and analogous to those underlying the classic N400 elicited by verbal stimuli. Like the picture N400, the gesture N450 was broadly distributed, with largest effects over frontal electrode sites, and was not visible at occipital sites (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999; West & Holcomb, 2002). Further, the time course of the N450 effect observed in the present study was in keeping with that of N400 activity reported in experiments involving visually complex scenes (West & Holcomb, 2002), or videographic stimuli (Sitnikova, Kuperberg, & Holcomb, 2003).

The new finding produced by the present study is the modulation of the N300 by gesture congruency. The N300 is a frontally distributed negativity elicited by a wide range of static image types, including line drawings (Barrett & Rugg, 1990; Federmeier & Kutas, 2001, 2002; Hamm et al., 2002), photographs of common objects (McPherson & Holcomb, 1999), photographs of conventionally meaningful hand shapes (such as "thumbs up") (Gunter & Bach, 2004), and complex scenes (West & Holcomb, 2002). The N300 is sensitive to gradations of categorical (Federmeier & Kutas, 2002; Hamm et al., 2002), but not associative (McPherson & Holcomb, 1999) relatedness, and has been proposed to index an early object identification process whereby the structural properties of a percept are matched with a basic-level category representation in long term memory[1]. This idea is supported by the finding that partially complete line drawings elicit a frontal negativity peaking around 350 ms (N350) with enhanced amplitude for unidentified as compared to identified items (Schendan & Kutas, 2002). If the N300 indexes the process of matching percepts with stored image-based representations, then stimuli that are difficult to identify should elicit enhanced N300 due to the increased search space engendered by the greater range of possible matches.

In the present study, the finding of increased N300 in response to contextually incongruous static gestures suggests that understanding gestural representations recruits image specific processes. If participants had understood the stimuli merely as photographs of a man, then no differences in ERPs elicited by congruous and incongruous trials would be expected. However, the congruency effects reported here suggest that individuals construed the configurations of the speaker's body and hands as semiotic, and integrated them with other contextually active information. Further, the N300 congruency effect indicates that "seeing" the semantic relations depicted by gestures recruits processes similar to those involved in parsing visual features contained in line drawings or photographs into meaningful objects.

It could be argued that the present results do not reflect integration at the semantic level. Rather, they may be driven by much simpler feature matching operations, whereby participants merely observed perceptual similarities between components of the speaker's gesture and components of the cartoon. However, this proposal is not likely to be the case for a number of reasons. First, in approximately half of the trials, the speaker's gestures depict features of objects, including the slant of a dump truck bed, the width and path of a stream of gunpowder, the wobbling of a pole, and the falling of a candle. Often, his hands or extended index fingers are used to trace the outline of objects, such as a platter or a panel. In all of these instances, the gestures bear no direct resemblance to their referents; rather, the capacity for gestural signification derives from selective structural correspondences. The speaker's arm does not resemble a candle, for example; however, by bending at the elbow and extending his forearm and hand upward, the speaker enacts the original vertical orientation of the candle in the cartoon. Similarly, although iterative curving motions produced by his index fingers look nothing like the platter depicted in a scene where a character enters the room carrying a meal, these movements can nevertheless be integrated to represent an oval shape that can be construed as denoting the shape of the platter.

In other cases, where the speaker's gestures depict actions performed in the cartoon, correspondences are still largely categorical rather than perceptually similar. Aside from obvious differences in setting, the identities of the agents, and the presence of various instruments, objects, and characters in the cartoons as compared to the gestures, actions in the cartoons were often shown in profile, while the speaker faced the camera as he gestured (see Figure 1). Moreover, the cartoon characters' actions often involved rapid, wide-ranging motions (e.g. leaping, diving, flying) that exhibited different visual characteristics from the gestural enactments. Finally, the fact that static gestures, which bore no resemblance to cartoons along movement based dimensions, nevertheless elicited ERPs similar to those obtained in response to full gestures suggests that participants were integrating features of gestures with other contextually active information rather than performing direct perceptual matching.

If the N300 indexes image specific processes critical to gesture comprehension, the question arises why

---

[1] Federmeier & Kutas (2002) propose visual similarity between categorically-related items to explain the discrepancy between the graded N300 relatedness effect obtained in their own experiment as compared with the binary responsivity to relatedness reported in McPherson & Holcomb (1999). "Whereas we used categorically-related stimuli, which will often tend to share visual features, McPherson and Holcomb's study included associatively-related items (e.g. hamburger and french fries), which may be less similar at a visual level. p. 745"

congruency effects within this time window are observed only in response to static and not dynamic gestures. One explanation is that in response to moving, visually complex stimuli, processes indexed by the N300 might become activated slightly later than they would in response to static objects, overlapping with processes indexed by the N450. In fact, it is possible that the N300 and N450 do not reflect discrete, serially organized stages of mental activity underlying the comprehension of gestures. Rather, it may be the case that activating and integrating image-based representations may transpire in cascade.

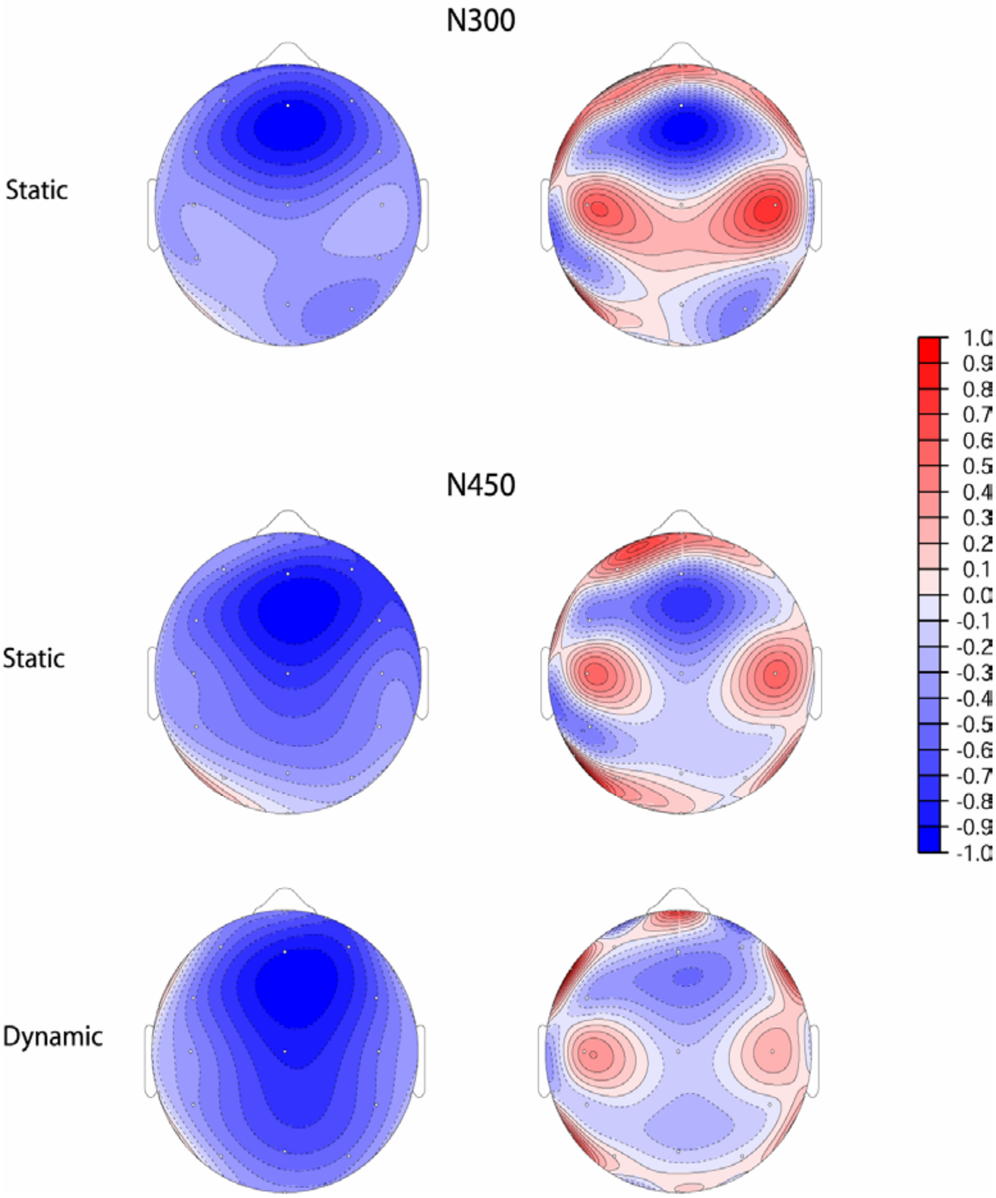

Figure 3. Normalized isovoltage maps (left) and current source density maps (right) computed from congruency effects elicited by dynamic and static gestures at 350 ms (top row) and 500 ms (bottom 2 rows) post-stimulus onset. (Activity elicited by dynamic gestures is not shown for the N300 time window because congruency effects do not start until 400 ms post-stimulus.)

Support for this idea can be found in Figure 3, which compares isovoltage and current source density maps of the N300 and N450 effects. Isovoltage maps were constructed by subtracting the mean amplitude of congruous from incongruous trials for data collected at each electrode site, and plotting interpolated values of the congruency effect (in microvolts) across the scalp. Current source density maps were computed from isovoltage maps, revealing an estimation of sources and sinks of radial current. For each map, normalization was implemented by dividing each data point by the maximum absolute value for that plot.

As shown in the isovoltage maps, both the N300 and N450 exhibit similar scalp topographies, though the N450 is more broadly distributed. Further, the current source density maps suggest that active dipoles within the N300 time window for static gestures are similar to those active during the N400 time window for both gesture types. These views of the data suggest that similar underlying neural generators are engaged both by static and dynamic gestures. Moreover, the N300 effect elicited by static gestures appears to engage similar generators to those mediating the N400 effect. Given these commonalities, it is possible that the N300 static gesture effect reflects the earlier activation of the same processes recruited during the comprehension of dynamic gestures.

Ultimately, conclusive resolution of the question of source localization would require a measurement technique with greater spatial resolution, such as MEG. Nevertheless, similarities in morphology (see Figure 2), distribution, and current source density between dynamic and static N450 effects suggest that both categories of stimuli likely engaged a highly co-extensive set of neural generators. Because static gesture stills were specifically extracted at points of maximal meaningfulness and discernability, it is possible that early congruency effects were evident with these trials, whereas variation in the uniqueness points of dynamic gestures may have prevented their detection within the N300 time window. Further, understanding static gestures may have required participants to draw more extensive inferences about their meaning, resulting in the more broadly distributed late congruency effect observed in response to static items but not in their dynamic counterparts.

A concern raised in a previous study using this paradigm touches upon the possibility that verbal information attained through lip-reading, rather than gestures themselves, was the source of the observed effects. One line of evidence against this proposal is the size of the speaker's mouth, which subtended less than ¼ of one degree of visual angle, minimizing the discernability of lip movement and other non-auditory information deriving from the physical production of speech. Moreover, as demonstrated by static items, even in the absence of speech based visual cues, the contextual congruency of gestural representations modulates brain activity.

In sum, the present study has confirmed and advanced existing research on the on-line comprehension of gestures. In keeping with previous findings (Wu & Coulson, 2004), ERPs time locked to the onset of contextually congruent and incongruent iconic co-speech gestures elicited a broadly distributed negative component – the N450 – with enhanced amplitude for incongruous as compared to congruous items. The gesture N450 was hypothesized to index the semantic integration of gestures with information made available in preceding cartoons in keeping with the N400 observed in picture priming paradigms. Additionally, contextually incongruent static gestures resulted in enhanced N300 relative to congruent items. The N300 was proposed to reflect the recruitment of visuo-semantic information in gesture comprehension. Overall, these findings are consistent with results from other studies measuring ERPs elicited by image-based media. Although the depictive power of iconic gestures does not derive from explicit visual features as is the case for photographs, line drawings, or video, gestures nevertheless engage visual semantic processes.

Method

Participants

27 volunteers were paid $24 or awarded course credit for participation. All were healthy, right handed, fluent English speakers without history of neurological impairment. Their mean laterality quotient, which is derived from the Edinburgh Inventory (Oldfield, 1971) and provides an index of handedness preference, was .735 (with maximal right handedness indicated by a score of 1). The data of three participants were excluded due to excessive artifacts, (greater than 40% of trials in critical bins).

From the remaining 24 individuals, the 16 best comprehenders were selected for the high comprehenders group, and the remaining 8 individuals were classified as low comprehenders. While accuracy rates in response to unrelated words were 94% for both participant types (s.d.=4%), high and low comprehenders accurately classified 86%

(s.d.=8%) and 68% (s.d.=11%) of related words, respectively. Only ERP responses from high comprehenders are reported due to the limited scope of this paper.

Materials

Stimuli were 160 cartoon clips paired with dynamic gesture video clips or static gesture freeze frames extracted from the gesture stream. Cartoon clips were constructed by digitizing televised cartoon shows (e.g. Tom and Jerry, Loony Tunes) into short, soundless segments. On average, cartoons lasted 3 seconds, and typically depicted one or two salient actions or events (e.g. Nibbles jabs Tom's foot and offers him a firecracker; a rock rolls towards the Roadrunner; Jerry rings a bell).

A naive individual was videotaped while describing these cartoon segments. He was told that his videotaped speech would be utilized in the construction of stimuli for a memory experiment, and instructed to describe each clip in as much detail as possible; however, no mention of gestures was made. Spontaneous gestures that were judged to represent elements within the corresponding cartoons were digitized into soundless segments of 48 frames each. Static gesture stills were created by extracting from each gesture stream a single frame in which key semiotic components of the gesture (e.g. hand shape, location, and orientation) were visible.

Primarily the head, arms, and upper torso of the speaker were visible in each gesture trial. Typically, the first frame coincided with the onset of the stroke phase of each gesture. In fewer than 9% of trials, the image sequence began in the preparation phase (e.g. the pre-stroke hold), primarily in cases where the stroke was executed very quickly. The presentation of each set of gesture frames lasted 2.3 seconds. On average, within each set of frames, gesture production extended for 2 seconds (s.d.=336). In 62% of trials, gesture production continued until the final frame. Gestures typically either re-enacted actions performed in the cartoon from a first person perspective (turning a doorknob, swinging a bat), or depicted salient features of an event (the path of a careening rock, the speed of falling apples) or salient features of static objects (the shape of a platter, the orientation of truck bed).

Congruous trials were those in which cartoon clips were paired with the original gesture produced while describing them. Incongruous trials involved mismatches. Eight lists were constructed, each containing 80 congruous and 80 incongruous trials (40 static and 40 dynamic of each type) and 80 related and 80 unrelated words. No cartoon, gesture, or word was repeated on any list, but across lists, each gesture appeared once as a congruous stimulus and once as an incongruous one. Words appeared once as a related item and once as an unrelated one following all four types of cartoon gesture pairs (congruent dynamic, congruent static, incongruent dynamic, incongruent static).

Procedure

Trials began with a fixation cross, presented in the center of a 17" color monitor. The cartoon and gesture clips were presented at a rate of 48 ms per frame with a 600 ms pause before the onset of the gesture (in order to allow participants time to establish central fixation). Although cartoons varied in length (mean=2949 ms, s.d.=900 ms), the duration of each gesture was exactly 2300 ms. One second after the offset of the gesture, a probe word either related or unrelated to the preceding context was presented (see Figure 1). A short pause (approximately 5 to 6 seconds) followed each trial as the next set of video frames was accessed by the presentation software. All video frames were centered on a black background and subtended approximately 10° visual angle horizontally and 7° vertically.

Participants were told that they would watch a series of cartoon segments, each followed first by video clips of a man describing either the immediately preceding cartoon, or a different one, and then by a probe word. They were asked to press YES or NO on a button box as soon as they felt confident that the probe word matched or did not match the preceding context. Response hand was counterbalanced across subjects. Four additional trials were used in practice block at the outset.

EEG Recording

The electroencephalogram (EEG) was recorded using tin electrodes at 29 standard International 10-20 sites (Nuwer et al., 1999), including midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz), medial (FP1, F3, FC3, C3, CP3 P3, O1, FP2, F4, FC4 C4, CP4, P4, O2), and lateral channels (F7, FT7, TP7, T5, F8, FT8, TP8, T6). Electrodes were also placed on the right mastoid for off-line re-referencing, below the right eye for monitoring blinks, and at the outer canthi for monitoring eye movements. All electrodes were referenced online to the left mastoid, and impedances maintained below 5 kΩ. EEG was amplified with an SA Instrumentation isolated bioelectric amplifier (band pass filtered, 0.01 to 40 Hz) and digitized on-line at 250 Hz.

EEG & Behavioral Data Analysis: Participants' mean accuracy was assessed with repeated measures ANOVA with the factors of Gesture Type (Static, Dynamic) and Congruency (Congruent, Incongruent). Artifact-free ERP averages time locked to the onset of gestures were constructed from 100 ms before stimulus onset to 920 ms after. Trials accurately categorized by participants were sorted and averaged. On average, critical bins contained 30 trials (±4). The mean artifact rejection rate was 22% percent (12% s.d.) for dynamic gestures and 23% (10% s.d.) for static ones.

Congruency effects were assessed by measuring the mean amplitude and peak latencies of ERPs time-locked to gesture onset from 300-400 ms (N300), 400-600 ms (N450), and 600-900 ms – in keeping with the intervals utilized in other paradigms involving complex visual stimuli (West & Holcomb 2002). Measurements were subjected to repeated measures ANOVA with the factors of Gesture Type, Congruency, and Electrode Site (29 levels). For all analyses, original degrees of freedom are reported; however, where appropriate, p-values were subjected to Geisser-Greenhouse correction (Geisser & Greenhouse, 1959).

To investigate the scalp distribution of ERP effects, we conducted 3 sorts of follow-up analyses: measurements of data collected from Midline sites (with 7 levels of anterior-posterior electrode location), measurements from Medial sites (with 2 levels of Hemisphere and 7 levels of anterior-posterior), and from Lateral sites (with 2 levels of Hemisphere and 4 levels of anterior-posterior).

References

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*, 201-212.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *developmental Psychobiology, 40*(3), 293-310.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Coulson, S., & Federmeier, K. D. (2003, in press). Words in context: ERPs and the lexical/postlexical distinction. *Journal of Psycholinguistic Research*.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(202-224).

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Eletrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40*, 730-747.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89-106.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95-112.

Gunter, T., C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters, 372*, 52-56.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to pictue stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*, 1339-1450.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic processing. *Brain and Language, 35*, 66-85.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24*, 259-276.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177-210). New York: Pergamon Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12).

Kutas, M., & Hillyard, S. A. (1980). Event-related potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*, 99-116.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.

Kutas, M., & Hillyard, S. A. (1988). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience, 1*, 38-49.

McNeill, D. (1992). *Hand and mind*. Chicago: Chicago University Press.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*, 53-65.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15-22.

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., Hinrichs, H., et al. (1999). IFCN standards for digital recording of clinical EEG. The international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement, 52*(11-14).

Oldfield, R. C. (1971). The assesment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia, 40*, 931-945.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*(3), 363-375.

Wu, Y. C., & Coulson, S. (2004). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *A supplement to the Journal of Cognitive Neuroscience*, 31.