*Psychophysiology, 40* (2003), 160–164. Blackwell Publishing Inc. Printed in the USA.


**BRIEF REPORT**

# Semantic integration in videos of real-world events: An electrophysiological investigation

TATIANA SITNIKOVA,[a] GINA KUPERBERG,[b,c] AND PHILLIP J. HOLCOMB[a]

[a]Department of Psychology, Tufts University, Medford, Massachusetts, USA
[b]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA
[c]Department of Psychiatry, Massachusetts General Hospital, Charlestown, Massachusetts, USA

**Abstract**

Event-related potentials (ERPs) discriminated between contextually appropriate and inappropriate objects appearing in video film clips of common activities. Incongruent objects elicited a larger negative-going deflection, which was similar to the N400 component described previously in association with words and static pictures and which has been argued to reflect the integration of semantic information into a mental representation of the preceding context. The onset of this potential occurred shortly after object presentation, indicating that semantic integration is a rapid online component of real-world perception. In addition, the anomalies in movies evoked a large late positive potential at posterior regions, suggesting that in event perception, semantic incongruity may trigger cognitive processes other than those mediating pure semantic integration.

**Descriptors:** Comprehension of real-world events, Viewing video film clips, Semantic anomalies, Event-related potentials (ERPs)

Comprehension of real-life scenes, for example, when observing another person do dishes, depends on the ability to rapidly integrate a continuous flow of visual information into "higher order" representations of meaning (Johnson-Laird, 1983). How and when this occurs in the brain, however, remains largely unknown. Here we report the results of an experiment that attempted to address these questions by recording event-related potentials (ERPs) while subjects viewed short video depictions of everyday events.

In the language domain, there is a wealth of evidence that ERPs are sensitive to online processes involved in the interpretation of sequentially presented words (see Kutas & Van Petten, 1994). In particular, one ERP component, the N400, appears to index semantic integration processes. The N400 has been described in association with words whose meaning does not fit with a preceding context in word pairs (e.g., Holcomb, 1993), sentences (e.g., Kutas & Hillyard, 1980) and larger texts (e.g., van Berkum, Hagoort, & Brown, 1999). Perhaps the most widely accepted account of the N400 argues that its amplitude is proportional to the "difficulty" or mental effort involved in integrating an item into the surrounding semantic context (e.g., Holcomb, 1993).

There have been several analogous studies of contextual integration of visually presented images. In these studies, an enhanced N400 has been elicited to critical picture stimuli that mismatched a single picture in priming paradigms (e.g., Barrett & Rugg, 1990; McPherson & Holcomb, 1999), successively presented pictures that conveyed stories (West and Holcomb, 2002), or a written sentence (e.g., Ganis, Kutas, & Sereno, 1996). Moreover, in some of these studies, a second and earlier negativity, the N300, has also been reported to overlap the more traditional N400 (e.g., McPherson & Holcomb, 1999). The N300/N400 elicited by pictures has been found to have a somewhat more frontal distribution than the N400 observed in most language studies (e.g., McPherson & Holcomb, 1999). One possibility is that the anterior portion of this complex may reflect image-specific semantic processing (West & Holcomb, 2002).

It could be argued that although humans frequently do process static pictures such as those presented in the above studies (e.g., in magazines and books), a much more common form of visual comprehension involves the viewing of dynamic images juxtaposed in a continuous flow. Therefore, an important outstanding issue is whether the comprehension processes engaged during static picture viewing are the same or similar to those employed during the viewing of dynamic images. One way to achieve more naturalistic processing is to use video film clips. Watching video clips evokes perceptual experiences that are remarkably similar to those elicited during the perception of

This research was supported by a grant HD25889 to PJH. We thank Sonya Jairaj, David R. Hughes, Kristi Kiyonaga, and Tanai Kamat for their assistance in preparing the materials and collecting the data.

Address reprint requests to: Tatiana Sitnikova, Department of Psychology, Tufts University, Medford, MA 02155, USA. E-mail: tatiana@neurocog.psy.tufts.edu.

events in the real world (e.g., Levin & Simons, 2000). And, although there have been no prior ERP studies using such stimuli, ERP studies of natural connected speech (e.g., Holcomb & Neville, 1991) and filmed hand and face movements of American Sign Language (ASL; Neville et al., 1997) suggest that it should be feasible to record ERPs when the critical signal extends over time, involves motion, and must be recognized within (and isolated from) a continuous sensory input.

In the current study, ERPs were recorded as participants viewed video clips of common activities in which a person manipulated an object that was either consistent or anomalous with the preceding context. For example, in one scenario, a man standing in front of a bathroom mirror applied shaving cream to his face and reached out for something. In the congruent condition, he grabbed a razor, and in the incongruent condition, he grabbed a rolling pin. After viewing each clip, participants indicated whether the depicted activity is common in real life.[1]

The primary aim of this study was to characterize the waveforms elicited by critical objects in the video scenarios. Specifically, we wanted to know if there are ERP components with similar properties to those elicited in previous picture and word studies (i.e., whether incongruent scenes elicit larger negativities than congruent scenes) and, if so, whether these differences are time-locked to the earliest possible point of anomaly detection (i.e., when an anomalous object first appears in the scene). This latter point is important because a recent neuroimaging study (Zacks et al., 2001) in which subjects viewed videos of real-world events has shown local brain activations to the boundaries between depictions of different action components. This finding suggests that the visual signal during event perception is parsed into simple action units. However, fMRI has a temporal resolution of seconds rather than milliseconds and therefore can give only a rough estimate of the timing of semantic processing.

## Methods

Sixteen right-handed native English-speaking volunteers (9 women, 7 men; mean age = 18.5 years) served as participants.

The stimuli were 80 pairs of color video film clips (for examples, visit http://neurocog.psy.tufts.edu/manuscripts.htm), each of which conveyed a simple plot involving a single character manipulating several real objects. All clips depicted typical real-life situations (e.g., shaving, cooking, etc). They were filmed using a digital video camera (Cannon model GL1), stored on digital video tape, and later, were transferred to a computer for editing and presentation. Clips were between 7 and 28 s in duration (mean = 16 s) and were presented without sound at a rate of 30 frames per second on a 17-inch computer monitor. All frames subtended approximately 4° of visual angle, and were centered on a black background.

[1]The anomaly detection task likely results in a somewhat unnatural focusing of attention on the anomalous events of interest. However, even with this liability, it has been shown to have a number of benefits (e.g., Holcomb & Neville, 1991; West and Holcomb, 2002). First, detection of anomalies necessarily requires that participants process the target items and their context at a deep semantic level (which is not guaranteed with many other tasks, such as word monitoring). Without deep semantic processing it would be difficult to interpret a null finding in a new area such as this. A second advantage is that this task allows trials to be sorted based upon the participants perceptions of anomalies. Again, in an initial study such as this, sorting ERPs based upon participants perceptions would seem to be a good idea.

All clips were structured in a similar way: In the beginning, one or more events were presented as a context (e.g., a character standing in the bathroom in front of the sink and mirror applied shaving cream to his face) and near the end of the clip the character manipulated a target object (e.g., stroked a razor across his face). We were careful to ensure that target objects (e.g., razor) did not appear in the clip until a "critical point" (e.g., until the character reached out for something and brought the razor into the scene). This critical point of the object's first discernable appearance (e.g., when an end of the razor's handle became visible) was determined by examining each clip, frame by frame, using a digital editing software (Ulead Media Studio Pro 6.0), and subsequently was used to time-lock ERP recording.

The two clips in a pair had the same lead-in context but had different endings. At the end of a congruent clip, the character used an object that was consistent with the context, whereas in the incongruent clip an unconventional object was used to perform the same action (e.g., the character stroked a rolling pin across his face in the shaving scenario). An object used in the incongruent condition in one pair was used in the congruent condition in another pair. The clips were arranged into two lists, each consisting of 40 congruent and 40 incongruent items. The assignment of clips and target items to lists was such that no clip context or target object was included twice in one list, and across lists all contexts and all target objects appeared in both the congruent and incongruent conditions. Half of the participants viewed list 1 and half viewed list 2.

Participants were instructed to decide whether each clip showed a scenario that one would witness in everyday life by pressing a "Yes" or "No" button at the "?" prompt that appeared 100 ms after the offset of the final frame of the clip. After the response, a fixation cross remained on the screen between the trials. Participants pressed a button to start presentation of each subsequent clip. Six additional clips were used in a practice block prior to the experimental run.

The electroencephalogram (bandpass, 0.01 to 40 Hz, 6 dB cutoffs; sampling rate, 200 Hz) was recorded at 57 scalp sites (for locations see Figure 1), the outer canthi of eyes (F9/F10), below each eye (IO1/IO2), the upper mastoid bones (T9/T10), and over the right mastoid (all referenced to the left mastoid). The ERPs (epoch length = 100 ms before critical-object appearance to 1,187 ms after object appearance) were averaged off-line after the trials with ocular artifacts (activity $> 60$ μV below eyes or at the eye canthi) were rejected: 9.5% of trials were rejected in the congruent condition and 7.4% in the incongruent condition. After averaging, the ERPs were rereferenced to a mean of the left and right mastoids.

Average ERPs were quantified by calculating the mean amplitudes (relative to the 100-ms baseline preceding object appearance) within three time windows (225–325 ms, 325–600 ms, and 600–900 ms after object appearance). These epochs roughly correspond to the time windows used in many previous studies to quantify the N300, N400, and late positive complex (LPC). Six analyses of variance (ANOVAs) for repeated measures were conducted to examine parasagittal columns of scalp electrodes along the anterior–posterior axis of the head. All analyses had a congruity factor (congruous/incongruous) and all but midline analyses had a hemisphere factor (left/right). The midline analysis had five levels of electrode site (FPz, Fz, Cz, Pz, Oz). The inner-medial analysis had three levels of electrode site (FC1/FC2, C1/C2, CP1/CP2). The outer-medial analysis had seven levels of electrode site (AF1/AF2, F1/F2, FC3/FC4, C3/

C4, CP3/CP4, P1/P2, PO1/PO2). The inner-lateral analysis had seven levels of electrode site (AF3/AF4, F5/F6, FC5/FC6, C5/C6, CP5/CP6, P5/P6, PO3/PO4). The outer-lateral analysis had nine levels of electrode site (FP1/FP2, AF7/AF8, F7/F8, FT7/FT8, T3/T4, TP7/TP8, T5/T6, PO7/PO8, O1/O2). The inferior analysis had three levels of electrode site (IO1/IO2, F9/F10, T9/T10). The Geisser–Greenhouse correction was applied to all repeated measures with more than one degree of freedom (Geisser & Greenhouse, 1959).

### Results

Figure 1 shows the obtained ERP waveforms and the corresponding voltage maps. ERPs to the critical objects were characterized by two potentials that started at around 200 ms after object appearance and continued until the end of the recording epoch: A negative-going wave was apparent at the more anterior sites, and a positive-going wave was evident at the more posterior sites. Early components of the visual ERPs (e.g., N1) could not be clearly seen, most likely due to the lack of discrete visual events separated by time (i.e., the early components were likely refractory due to the continuous stimulus presentation format; e.g., Davis, Mast, Yoshie, & Zerlin, 1966).

In the 225–325-ms epoch (N300), even though the ERPs at the frontal-central sites appeared to be more negative in the incongruent than congruent condition, this difference did not reach a conventional level of statistical significance (there was a

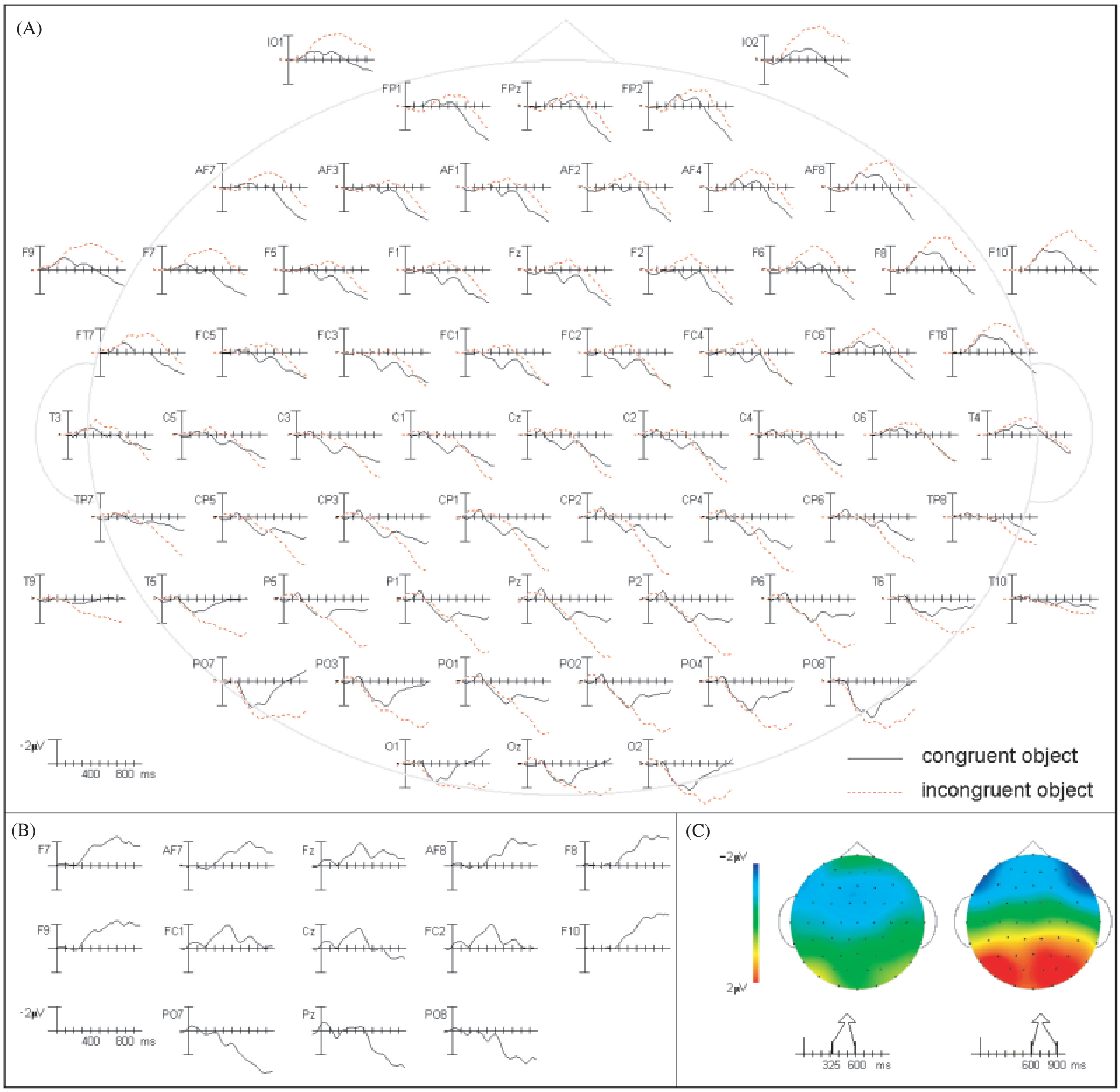


**Figure 1.** Average ERPs elicited by target objects in videos (A), difference waves obtained by subtracting incongruent from congruent condition at the selected electrode sites (B), and the corresponding voltage maps of the ERP differences in the N400 and LPC time windows (C).

**Table 1.** *Interactions between Congruity and Electrode Site Obtained in the ANOVAs Examining Parasagittal Columns of Scalp Electrodes*

| Analysis | Degrees of freedom | *F*-value |
|---|---|---|
| 325–600 ms (N400) | | |
| midline | 4,60 | 3.55* |
| inner-medial | 2,30 | 6.29* |
| outer-medial | 6,90 | 4.38* |
| inner-lateral | 6,90 | 4.51* |
| outer-lateral | 8,120 | 5.20* |
| inferior | 2,30 | 8.70* |
| 600–900 (LPC) | | |
| midline | 4,60 | 11.87** |
| inner-medial | 2,30 | 17.05** |
| outer-medial | 6,90 | 14.83** |
| inner-lateral | 6,90 | 19.44** |
| outer-lateral | 8,120 | 19.41** |
| inferior | 2,30 | 26.76** |

*$p < .05$; **$p < .01$.

trend towards significance of the main effect of congruity in the inner-medial analysis: $F(1,15) = 2.91$, $p = 1.00$).

In the 325–600-ms epoch (N400), incongruent objects elicited a larger negativity at the frontal-central electrode sites as indicated by the significant congruity by electrode site interactions obtained in analyses of all parasaggital electrode columns (see Table 1). Planned comparisons revealed that significant differences were present at frontal (Fz, F1/F2, F5/F6, F7), inferior-frontal (F9/F10, IO1/IO2), frontal-temporal (FT7/FT8), frontal-central (FC5/FC6, FC3/FC4, FC1/FC2), and central sites (Cz, C1/C2).

In the 600–900-ms epoch (LPC), the scalp areas where incongruent objects evoked more negative waveforms than congruent objects shifted to frontal-lateral sites. In addition, in this epoch, potentials at the posterior electrode sites were more positive in the incongruent condition than in the congruent condition. Significant interactions between congruity and electrode site were obtained in all analyses (see Table 1). Planned comparisons showed that the incongruent objects elicited an increased negativity at AF7/AF8, F5/F6, F7/F8, F9/F10, FT7/FT8, and IO1/IO2 sites and an increased positivity at central-parietal (CP5/CP6, CP3/CP4, CP2), temporal (T5/T6), temporal-parietal (TP7/TP8), parietal (Pz, P5/P6, P1/P2), parietal-occipital (PO7/PO8, PO3/PO4, PO1/PO2), occipital (Oz, O1/O2), and inferior-temporal sites (T9/T10).

## Discussion

The present data demonstrate a robust negative-going ERP elicited by objects in video depictions of everyday activities. The amplitude of this negativity was greater to contextually anomalous than to contextually appropriate objects. This difference started about 300 ms after the target object appeared in the video and at some sites continued until the end of the recording epoch. Overall, the morphological, functional, and temporal properties of this effect suggest that it is similar to the N400 previously reported in analogous paradigms using words and static pictures as stimuli (e.g., Kutas & Van Petten, 1994; West & Holcomb, 2002). Importantly, this N400 effect in videos started shortly after the critical objects first became visible, suggesting that there is a close temporal relationship between the processes of object identification and scene comprehension during viewing of videos.

There were also several differences in the results of this experiment and previous word and picture studies. First, in several previous reports, the N400 to pictures was preceded by an overlapping earlier negativity (the N300). In the current study, there was only a trend for a difference in the N300 window. One possible explanation for the absence of a significant N300 effect is that the timing of the appearance of target objects (and thus recognition) was somewhat more variable across the videos than across the static pictures used in previous studies. This likely resulted in a somewhat greater variation in time-locking to critical scenes, which might have in turn resulted in a somewhat smeared or shifted N300 (i.e., part of what is being identified as the N400 might actually be N300 activity).

There was also another difference in the time-course of ERPs between the current and prior experiments. At some sites, the duration of the present N400 effect was greater than 600 ms, whereas in most previous written word and picture studies, this component has been reported to last only for 200 to 400 ms. This prolonged time-course might also be explained by the variability in the timing of identification of different objects. Such variable timing could have resulted in the N400 effect itself being somewhat smeared across trials. In fact, a similar explanation has been proposed before for the lengthened N400 effects observed to spoken words (e.g., Holcomb & Neville, 1991), and ASL videos (Neville et al., 1997) that, like the present stimuli, had a variable identification point. To get a better estimate of the timing of semantic integration in videos, it might be useful to collect ERPs to critical items that have a clear point of appearance in the scene (e.g., at a scene change).

Another explanation for the extended ERP differences observed in movies could be exposure to additional information about the activity immediately after the critical scene. In the present videos, right after the critical object appeared in the scene, it was manipulated in a normal or bizarre way (e.g., the man with shaving cream on his face stroked the razor or the rolling pin across his face). Thus, in the incongruent movies, the enhanced N400 could be sustained due to processing of an anomalous action that unfolded over a few hundreds of milliseconds.

Another difference between the current results and those of previous studies using words is that the N400 effect for videos was more frontally distributed than the parietal-occipital N400 effect typically reported for words (e.g., Kutas & Van Petten, 1994). However, this anterior prominence for negativities is in keeping with a number of other studies that have used picture stimuli (e.g., McPherson & Holcomb, 1999; West & Holcomb, 2002). The present results are particularly similar to those of West and Holcomb's study in which the N300/N400 differences were evident only over more anterior areas but were not significant over more posterior electrode sites. Their study, like the current experiment, used complex scenes (rather than individual objects). It may be that scene complexity shifts the distribution of the effect towards more anterior sites.

A final difference between the findings of this study and previous word and picture studies is that, at posterior sites, critical scenes in anomalous videos elicited a late positivity that was dramatically larger than the positivity evoked by comparable scenes in congruent videos. There are at least two potential accounts for this posterior-positivity effect. One possibility is that the posterior ERP that was sensitive to the semantic manipulation in this study is similar in nature to the decision P3 (see Donchin & Coles, 1988). According to this view, detection of the anomalous scene might allow viewers to rapidly decide that this was an anomalous video, whereas no such decision was possible at the comparable point in congruent videos. This explanation seems plausible, especially considering that the task required participants to actively classify each video as congruent or anomalous. However, a recent follow-up study (Sitnikova, Kiyonaga, & Holcomb, 2002) casts doubt on this explanation. In this study, which was otherwise procedurally identical to the current study, participants did not actively classify the two types of scenarios, but instead answered occasional questions about content that had nothing to do with critical scenes in the videos. Without the classification requirement, there is no reason for participants to actively differentiate the videos and therefore there should not have been as large of a decision P3 effect for the anomalous videos. Nevertheless, a similar pattern of larger posterior positivities for anomalous scenes was found in the follow-up experiment.

Although the presence of an N400 effect strongly suggests that anomalous endings to videos caused problems in semantic integration, another possibility for the large posterior positivity effect is that it reflects participants' experiencing some different kind of processing difficulty, perhaps one that is not entirely semantic in nature. In the language comprehension literature, a late positivity, the P600, has been reported to a variety of syntactic processing difficulties (e.g., Osterhout & Holcomb, 1992), and also, under certain conditions, to semantic violations (e.g., Münte, Heinze, Matzke, Wieringa, & Johannes, 1998). Furthermore, a similar effect has been reported in at least one study of structural violations in music (Patel, Gibson, Ratner, Besson, & Holcomb, 1998), suggesting that the P600 is not necessarily specific to language. One interpretation of these findings is that the P600 reflects processes of reanalysis triggered by the detection of an error (e.g., Münte et al., 1998). Considering the manner in which our violations were constructed (by introducing a novel object in an otherwise congruent scene), it is possible that the anomalies forced participants to engage in a reanalysis of the scene that resulted in an enhanced late positivity. This possibility will be examined more systematically in future research.

## REFERENCES


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

Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology*, *21*, 105–113.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral & Brain Sciences*, *11*, 357–427.

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*, 89–106.

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

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*, 286–300.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

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

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.). *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press, Inc.

Levin, D. T., & Simons, D. J. (2000). Perceiving stability in a changing world: Combining shots and integrating views in motion pictures and the real world. *Media Psychology*, *2*, 357–380.

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

Münte, T. F., Heinze, H. -J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*, 217–226.

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating American Sign Language: Effects of sensory experience and age of acquisition. *Brain & Language*, *57*, 285–308.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory & Language*, *31*, 785–806.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.

Sitnikova, T., Kiyonaga, K., & Holcomb, P. J. (2002). Electrophysiological brain activity associated with comprehension of real-life events. Manuscript in preparation.

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*, 657–671.

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

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655.


(Received November 19, 2001; Accepted April 10, 2002)