Available online at www.sciencedirect.com
ScienceDirect
Neuroscience Letters 433 (2008) 93–97
Neuroscience Letters
www.elsevier.com/locate/neulet

N400 involvement in the processing of action sequences

Vincent M. Reid [a,*], Tricia Striano [b,c]

[a] Department of Psychology, Durham University, Science Laboratories, South Road, Durham DH1 3LE, UK
[b] Department of Psychology, Hunter College, New York, NY, USA
[c] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Received 18 September 2007; received in revised form 19 December 2007; accepted 30 December 2007

Abstract

Understanding others' sequences of action is a fundamental aspect of human movement and is the key to detecting goal directed behavior and intentional actions. Human action contains semantic information whereby logical chains of events are entirely dependent on the sequence in which they are performed. The sequential nature of action ensures that an observing individual anticipates the conclusion of a viewed action. This experiment reports on the results of 15 participants who viewed videos of an actress performing actions. Half the stimuli contained an anticipated conclusion to the action whereas half did not. Results from the passive viewing of stimuli depicting eating actions indicated an increased N400 response over frontal, central and parietal regions when viewing the unanticipated conclusions of the actions as compared with the amplitude for the anticipated condition. These results show that (1) neural systems exist to rapidly discern semantic information in actions, and (2) the N400 component, which predicts semantic information in language, also anticipates information within goal directed action.

Keywords: ERP; N400; Goals; Intentions; Social cognition

The ability to understand the actions of other humans is important for almost every aspect of social cognition from early development through to adulthood [1,21]. Humans are initially required to identify conspecifics from other components of the environment. Once this has been completed, the analysis of how and why the observed individual performs action must take place in order for the observer to understand the desires and the motivations of the observed [8,19,25]. It is only through ascertaining these characteristics and dispositions that the social intentions of the observed individual can be understood and an appropriate response can be prepared by the observer. Electrophysiological responses to action sequences have not been assessed, despite their fundamental importance to understanding social-cognitive functions in terms of neural correlates and their timecourse. The assessment of event-related potential (ERP) responses to action sequences is consequently the aim of the present study.

Mechanisms involved in the perception of action sequences may be similar to those associated with the processing of semantic understanding found within language research [18]. In research investigating the neural correlates of the processing of grammar and syntactic structure, sequence processing has been previously investigated. One particular Event-related potential (ERP) component, the N400, is elicited by words that do not fit within the preceding semantic context (e.g., [7]). Initially described over 20 years ago [15,16], the N400 has been instrumental in refining linguistic theory and has proved foundational for an understanding of the relationship between high-level cognitive processes and neurophysiological correlates to those cognitive processes [14].

The current study seeks to extend N400 research by utilizing current knowledge associated with its properties to assess the neural processing of human action. Our stimuli will feature no spoken language; however, we anticipate differing N400 responses for semantically anticipated and unanticipated action sequences. Recent research with adults has suggested that these distinctions exist when observing meaningful and meaningless hand postures [10] with the N400 also observed when sequences of words are followed by congruent or incongruent gestures [24]. These N400 responses are delayed in latency when compared to N400 components derived from language studies where single words are presented sequentially to form the sentence. The nature of the morphology of the N400 elicited by observing actions has yet to be determined although viewing pictures rather than words appears to induce a more anterior N400 effect [9,17].

* Corresponding author. Tel.: +44 191 334 3266; fax: +44 191 334 3241.
E-mail address: vincent.reid@durham.ac.uk (V.M. Reid).

0304-3940/$ – see front matter
doi:10.1016/j.neulet.2007.12.066

Research investigating the perception of action sequences has focused on how the adult human brain segments this action [25]. Logical chains of events are entirely dependent on the action sequence. For example, the process of retrieving a pen from a drawer requires small sequences of action to follow each other in order for the overall goal to be achieved. First, the drawer must be opened. Then the pen must be grasped and placed in a location so that the acting hand can close the drawer. Then the pen can be retrieved and used. Should this sequence be observed in a different order, then the goal of the action becomes obscured and the series of actions may be rendered meaningless. Zacks et al. [25] provided strong evidence that action is perceived and is processed in this way by the adult brain. Using functional magnetic resonance imaging, they showed that greater brain activity in frontal and parietal regions was observable when overall sequences of intentional action were completed when compared with brain activity derived from the observation of smaller, discrete units of action. How the human brain processes such information on the level of milliseconds and which mechanisms are utilized are issues that have yet to be investigated.

The current study investigates the neural systems involved in the processing of semantic information in action. We hypothesised that the close associations between language and action provide a basis from which to investigate action processing. Should actions be presented utilizing the same paradigms as seen in language research, we would expect to detect the existence of a similar N400-like effect for both language and action. We therefore conducted a study utilizing a paradigm designed to elicit the N400 component of the ERP. We predicted that unanticipated conclusions to actions would elicit a larger negative amplitude N400 than the negativity observed for anticipated actions.

15 adults (9 females, 6 males), with an average age of 24.8 years (range 20–30 years) participated in the study. Participants were recruited from the Leipzig area of Germany. All participants were from a middle class background and had normal or corrected-to-normal vision.

Three sets of two sequences of video clips featuring a female actress performing actions were created. These films were digitally encoded and processed into still frames at 15 frames per second to create video clips of 2500 ms. These sequences featured one condition in which the conclusion of the action was anticipated whereas the other trial featured an unanticipated conclusion. Presented stimuli included sequences of: (1) inserting food or an empty spoon into the mouth, (2) putting on both shoes before locomotion or only one shoe, and, (3) answering a telephone by putting it to the ear-mouth configuration or placing the telephone on the top of the head. These sequences were created with movie making techniques, where situation shots and close ups were utilized in order to enhance the information that was of importance. Example frames from these stimuli are depicted as Fig. 1. For example, the unanticipated condition featured the same video as those displayed in the anticipated condition except that during the collection of food, no food was collected by the spoon. The final sequence of the movie was also different, where the spoon was placed in the mouth. In this example, during the anticipated condition the spoon contained food, whereas for the unanticipated condition, the spoon was empty. When a spoon is put into a mouth, it generally contains food. Having no food on the spoon and still performing the action, we reasoned, would produce an N400 component in the resulting ERP.

Participants were briefed in what participation entailed. Participants were requested to reduce their eye movements, and to keep blinking and muscle movements such as swallowing to a minimum. A participation sheet was given to the participant prior to testing, explaining the procedure and all that it involves. At this stage participants were given the opportunity to ask questions and a voluntary consent form was presented. Research in this laboratory has been conducted under institutional protocols, in accordance with the 1964 Declaration of Helsinki.

Participants sat in a dimly lit sound-attenuated and electrically shielded cabin, at a viewing distance of 90 cm away from a 70 Hz 17-inch stimulus monitor. The stimuli were presented at 16 by 13 cm and were thus 8.26 degrees of visual angle. The experiment consisted of one block with 100 anticipated and 100 unanticipated trials of the stimuli described above, each presented with a 50% probability. Participants were instructed to watch the stimuli, but their attention was not drawn to any particular aspect of the films.

The two conditions were presented to the participant utilizing the software ERTS (BeriSoft Corporation, Germany) in a random order with the constraint that the same condition was not presented three times consecutively and that the number of

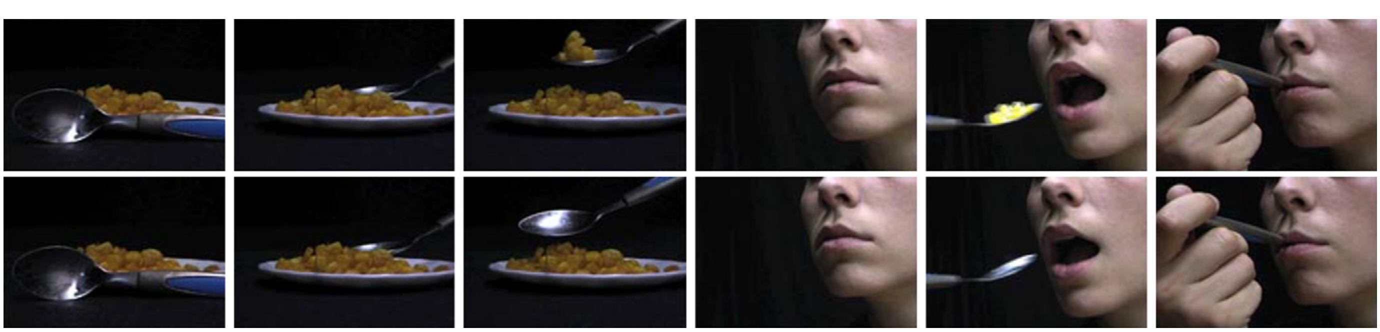

Fig. 1. Example frames from experimental stimuli. The top row depicts the anticipated action sequence. The bottom row depicts the unanticipated action sequence. Stimuli were short video clips depicting a plate of corn with a spoon. The spoon is picked up and either (a) collects corn or (b) does not collect corn. The spoon is removed. The final sequence initially depicts a closed mouth. The spoon with or without corn moves into shot and is then placed in the mouth.

presentations of each condition was balanced in every 20 trials presented. Each video was preceded by a fixation cross in the middle of the screen for 500 ms. The video stimulus was presented in the center of the screen for a total of 2500 ms. Between the presentations of the trials, the screen was blank for a random period of between 800 and 1000 ms.

EEG was recorded continuously with Ag–AgCl electrodes from 27 scalp locations in an extended 10–20 system, referenced to the vertex (Cz) and rereferenced to the linked mastoids after data collection. The ground electrode was positioned at electrode position FP1. Data was amplified via a Twente Medical Systems 32-channel REFA amplifier. Bipolar horizontal (outer canthi) and vertical (above and below right eye) EOGs were recorded to control artifacts caused by eye movements. The electrical potential was amplified with 0.3–20-Hz filter bandpass, digitized at a 250-Hz sampling rate, and stored on computer disk for the off-line analysis. EEG data was re-referenced offline to the linked mastoids.

The EEG recordings were segmented into epochs of waveform that comprised a 200 ms baseline featuring the last 200 ms of the clip within the sequence prior to the final section and 1000 ms of the final action. This consisted of either putting on both shoes before departing the visual scene vs. only one shoe; placing the telephone to the ear and mouth vs. the top of the head; or putting the food into the mouth or putting the empty spoon into the mouth (see Fig. 1). The baseline was thus the same action in both conditions. For the elimination of electrical artifacts caused by eye and body movements, EEG data was rejected off-line whenever the raw amplitude within a 200-ms gliding window exceeded 80 V for the eye electrodes and 50 V at any individual electrode. Data were also visually edited offline for artefacts.

We assessed the difference in the peak negative amplitude of the ERP in the anticipated and unanticipated action conditions. An ANOVA was performed in frontal, central and parietal regions as previous research into the properties of the N400 has suggested that the peak negativity can vary across these sites, dependent on the paradigm (i.e., video or still-frame) or the semantic content of the information [17,24]. Further, a visual analysis of the grand average and the individual averages indicated that this was the case with our data (see Fig. 2).

Visual inspection of the data suggested no differences between conditions for the sequences based around shoe and telephone use. Statistical analysis indicated no effects for these sequences ($p > 0.05$ in all cases). Accordingly, the following assessments of the ERP are based on data derived from the 'eating' sequences only. For statistical analysis a time win-

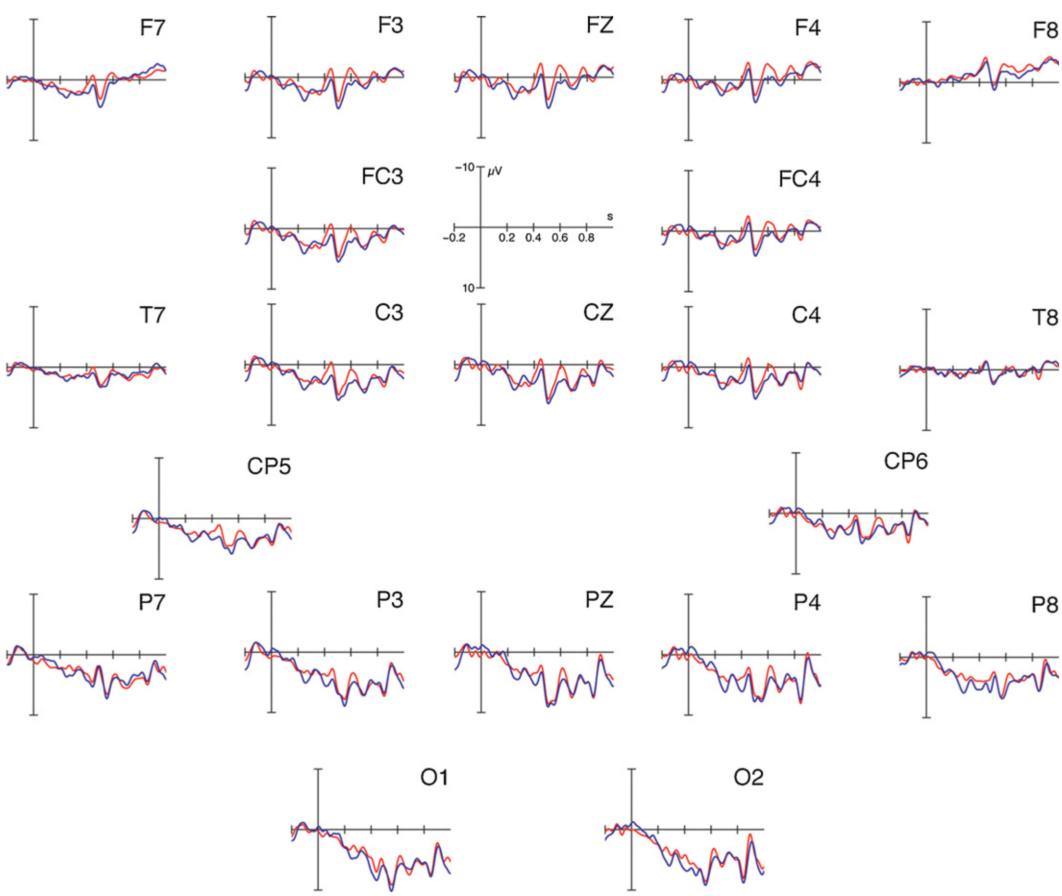

Fig. 2. ERPs recorded to anticipate final action (gray) and unanticipated final action (black). All electrodes in frontal, central and parietal locations recorded an enhanced negative response (peaking at 600 ms after the onset of the final shot of the film) to the unanticipated action compared with the anticipated action. This can be seen as the divergence between conditions of the upward curve of the ERP from 500 to 700 ms after the onset of the final action. Note: Positive is plotted down.

dow was chosen reflecting the onset of a negative amplitude deflection (500 ms) until the end of the deflection (700 ms) after stimulus onset. ERPs were evaluated statistically by computing the mean amplitude in the two conditions in the following regions of interest (ROIs): left anterior (F3), central anterior (FZ), right anterior (F4), left central (C3) centro-central (Cz), right central (C4), left posterior (P3), central posterior (Pz) and right posterior (P4). Variances of ERPs were analysed by a $2 \times 3 \times 3$ repeated measures ANOVA. Analysed factors were (1) condition (anticipated × unanticipated), (2) hemispheric lateralization (left, central, right) and (3) anterior-posterior topography (anterior, central, parietal).

The ANOVA indicated a main effect of condition $F(1,14) = 13.37$, $p = 0.003$, reflecting a larger peak amplitude negativity for the unanticipated action ($M = -0.795$, S.E. = 0.7) when compared with the anticipated action ($M = 0.827$, S.E. = 0.7). With negative plotted upwards, the N400 can be seen in channels from 500 ms as the upwards deflection of the ERP waveform. Two further effects were detected. An effect of laterality was found $F(2,28) = 5.41$, $p = 0.01$, representing a greater negativity at central (Mean = 0.134, S.E. = 0.739) and right hemisphere (Mean = −0.636, S.E. = 0.703) electrodes than left hemisphere electrodes (Mean = 0.551, S.E. = 0.651). A significant effect was also found for anterior–posterior relations $F(2,28) = 54.99$, $p < 0.001$, indicating a greater negativity at frontal electrodes (Mean = −2.123, S.E. = 0.689) than central electrodes (Mean = 0.38, S.E. = 0.7), which in turn were more negative than parietal electrodes (Mean = 2.21, S.E. = 0.732).

No other effects were found within the selected time window or within the wider context of the entire ERP. A parallel assessment of the peak latency of the N400 effect indicated no differences between the conditions.

In the present experiment, we assessed event-related potential (ERP) responses to the perception of videos depicting anticipated or unanticipated final actions. We found an effect for unanticipated actions for eating sequences, with enhanced negative amplitude for the unanticipated action when compared with the negative amplitude for the anticipated action. The ERP finding likely relates to the N400, which in previous studies related to sentence structure has been interpreted as indexing semantic processing of the information stream [14]. This result suggests that the adults processed the semantic information in the action sequence and built a prediction with respect to the termination of the action. When this was violated, an N400 response was produced.

The lack of an N400 component in presented sequences depicting the wearing of shoes and the answering of a telephone may be due to spatial smearing of the ERP between individual participants, with each participant identifying when in the action stream the observed action becomes either anticipated or unanticipated. Alternatively, given the repeated presentations of the sequences, low-level differences between conditions – such as a slight difference in hair arrangement of the actress – may have alerted the participants to the condition. The resulting observation of the conclusion of the action may have been anticipated in both conditions, resulting in no N400 component for these sequences.

An alternative explanation for the lack of significant results for the shoe and telephone sequences is that there is a fundamental difference between these types of action. Eating may well be regarded as a more primary act with less abstract goals, containing clearer, more concrete, intentions. In support of this notion, Fogassi et al. [6] demonstrated in the monkey cortex that relatively more mirror neurons coded the intention to eat than were recorded for the intention to place an object. Further, Iacoboni et al. [12] showed via fMRI with humans that there was higher activity in right frontal regions for coding the intention to eat than was seen in the intention to clean up a mess. This anatomical region may be similar to that which is generating the N400 effect found in the present study; however, in order to confirm this, a study similar to the one presented here would need to be performed with a larger montage of electrodes for the purposes of source analysis.

There is little work thus far that investigates ERPs as a result of the observation of moving images. Comparisons with other studies investigating the N400 component are therefore limited at this stage (see, for example, West and Holcomb [22]). Those studies that have used moving images report N400 components that are reduced in amplitude than when the paradigm utilizes static images [24], potentially due to spatial smearing of the ERP or the static image containing the key information of the moving images. It is also possible that the relative delay in the latency of the N400 found in the eating stimuli may be due to the use of moving images rather than static images (see Sitnikova et al. [20]). Even displaying a static image following a brief video delays the peak latency of the N400 [24]. Our data suggest that the N400 is a reaction to the first segment of the eating video when the spoon enters the frame. The following processing of the semantically unexpected event produces the N400 effect. What is fundamental is that in these data we have produced the expected effect despite the unknown factors associated with the application of the N400 paradigm to observing video and to this new aspect of semantic processing. Further research will determine whether delays in latency are due to the presentation format (video or static) or are related to the processing of action rather than language.

The result of an N400 component in the unanticipated eating condition suggests that past experience induces an expectation of a rational sequence of events. This raises the issue of when in development infants begin to process goal directed action by assessing their expectation of human actions. It is known that by 10–11 months, infants parse action sequences based on the goals of the observed individual [2]. Potentially the application of the paradigm reported in this article to young infants may provide evidence for the processing of anticipated action sequences. This may be due to the sensitivity of the ERP measure in assessing infant cognitive abilities when compared with the need to obtain an overt response in behavioral paradigms (see Ref. Striano and Reid [21]).

Research into the parameters of autism suggests that it is the very process of detecting and observing humans that may underlie many social-cognitive impairments observed in this atypical population [3,13]. The presented study may have potential ramifications for our understanding of why autism presents primarily in the social-cognitive domain, particularly if assessment of sequence processing within action is shown to be disrupted in this population. Potentially the failure to detect and extract important information within the action stream has a cascading effect on the development of further cognitive skills, such as language. This may be the case, with recent research that suggests that the identification of goal-directedness may also be critical to the development of language [4,5] and imitation [11], with potential implications for the understanding of autism [23].

In summary, we show that the conclusion of an observed action sequence of eating is anticipated, with the observer building predictions based on the prior components of action within the sequence. When the concluding action is unanticipated, a larger N400 ERP component is induced than when the concluding action is anticipated. This result suggests that neural systems exist to rapidly discern semantic information within human actions. Further, these systems are related to the N400 component of the Event-related potential.

Acknowledgements

This research was supported by the Transcoop award granted by the Alexander von Humboldt Foundation. We are grateful to those individuals who participated in the study.

References

[1] D.A. Baldwin, Interpersonal understanding fuels knowledge acquisition, Curr. Dir. Psych. Sci. 9 (2) (2000) 40–45.
[2] D.A. Baldwin, J.A. Baird, M.M. Saylor, A. Clark, Infants parse dynamic action, Ch. Dev. 72 (2001) 708–717.
[3] S. Baron-Cohen, Mindblindness: an Essay on Autism and Theory of Mind, MIT Press, Cambridge, MA, USA, 1995.
[4] M.C. Corballis, From hand to mouth: the gestural origins of language, in: M.H. Christiansen, S. Kirby (Eds.), Language Evolution, Oxford University Press, Oxford, 2003, pp. 201–218.
[5] M.C. Corballis, FOXP2 and the mirror system, Trends Cog. Sci. 8 (2) (2004) 95–96.
[6] L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, G. Rizzolatti, Parietal lobe: from action organization to intention understanding, Science 308 (2005) 662–667.
[7] A.D. Friederici, Towards a neural basis of auditory sentence processing, Trends Cog. Sci. 6 (2002) 78–84.
[8] C.D. Frith, U. Frith, Interacting minds: a biological basis, Science 286 (1999) 1692–1695.
[9] G. Ganis, M. Kutas, M. Sereno, The search for common sense: an electrophysiological study of the comprehension of words and pictures for reading, J. Cog. Neurosci. 8 (1996) 89–106.
[10] T.C. Gunter, P. Bach, Communicating hands: ERPs elicited by meaningful symbolic hand postures, Neurosci. Lett. 372 (2004) 52–56.
[11] M. Iacoboni, Understanding others: imitation, language, and empathy, in: S. Hurley, N. Chater (Eds.), Perspectives on Imitation: from Cognitive Neuroscience to Social Science, vol. 1, MIT Press, Cambridge, 2005, pp. 77–99.
[12] M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J.C. Mazziotta, G. Rizzolatti, Grasping the intentions of others with one's own mirror neuron system, PloS Biol. 3 (2005) E79.
[13] M.H. Johnson, Developmental Cognitive Neuroscience, 2nd ed., Blackwell, Oxford, 2005.
[14] M. Kutas, Views on how the electrical activity that the brain generates reflects the functions of different language structures, Psychophysiology 34 (1997) 383–398.
[15] M. Kutas, S.A. Hillyard, Event-related potentials to semantically inappropriate and surprisingly large words, Bio. Psych. 11 (1980) 99–116.
[16] M. Kutas, S.A. Hillyard, Reading senseless sentences: brain potentials reflect semantic incongruity, Science 207 (1980) 203–205.
[17] W.B. McPherson, P.J. Holcomb, An electrophysiological investigation of semantic priming with pictures of real objects, Psychophysiology 36 (1999) 53–65.
[18] G.A. Miller, P.N. Johnson-Laird, Language and Perception, Harvard University Press, Cambridge, MA, 1976, p. 1976.
[19] V.M. Reid, T. Striano, The directed attention model of infant social cognition, Euro. J. Dev. Psych. 4 (2007) 100–111.
[20] T. Sitnikova, G. Kuperberg, P.J. Holcomb, Semantic integration in videos of real world events: an electrophysiological investigation, Psychophysiology 40 (2003) 160–164.
[21] T. Striano, V.M. Reid, Social cognition in the first year, Trends Cog. Sci. 10 (10) (2006) 471–476.
[22] W.C. West, P.J. Holcomb, Event-related potentials during discourse-level semantic integration of complex pictures, Cog. Br. Res. 13 (2002) 363–375.
[23] J.H.G. Williams, A. Whiten, T. Suddendorf, D.I. Perrett, Imitation, mirror neurons and autism, Neuro. Biobeh. Rev. 25 (2001) 287–295.
[24] Y. Wu, S. Coulson, Meaningful gestures: electrophysiological indices of iconic gesture comprehension, Psychophysiology 42 (6) (2005) 654–667.
[25] J.M. Zacks, T.S. Braver, M.A. Sheridan, D.I. Donaldson, A.Z. Snyder, J.M. Ollinger, R.L. Buckner, M.E. Raichle, Human brain activity time-locked to perceptual event boundaries, Nat. Neuro. 4 (6) (2001) 651–655.