BRAIN RESEARCH 1115 (2006) 92–107

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Auditory semantic networks for words and natural sounds

A. Cummings[a,b,c,*], R. Čeponienė[a], A. Koyama[a], A.P. Saygin[c,f], J. Townsend[a,d], F. Dick[c,e]

[a]Project in Cognitive and Neural Development, University of California, San Diego, USA
[b]San Diego State University/University of California, San Diego Joint Doctoral Program in Language and Communicative Disorders, USA
[c]Center for Research and Language, University of California, San Diego, USA
[d]Department of Neurosciences, University of California, San Diego, USA
[e]Birkbeck College, University of London, UK
[f]Department of Cognitive Science, University of California, San Diego, USA

ARTICLE INFO

Article history:
Accepted 13 July 2006
Available online 8 September 2006

Keywords:
ERP
ICA
N400
Word
Environmental sound
Semantic

ABSTRACT

Does lexical processing rely on a specialized semantic network in the brain, or does it draw on more general semantic resources? The primary goal of this study was to compare behavioral and electrophysiological responses evoked during the processing of words, environmental sounds, and non-meaningful sounds in semantically matching or mismatching visual contexts. A secondary goal was to characterize the dynamic relationship between the behavioral and neural activities related to semantic integration using a novel analysis technique, ERP imaging. In matching trials, meaningful-sound ERPs were characterized by an extended positivity (200–600 ms) that in mismatching trials partly overlapped with centro-parietal N400 and frontal N600 negativities. The mismatch word-N400 peaked later than the environmental sound-N400 and was only slightly more posterior in scalp distribution. Single-trial ERP imaging revealed that for meaningful stimuli, the match-positivity consisted of a sensory P2 (200 ms), a semantic positivity (PS, 300 ms), and a parietal response-related positivity (PR, 500–800 ms). The magnitudes (but not the timing) of the N400 and PS activities correlated with subjects' reaction times, whereas both the latency and magnitude of the PR was correlated with subjects' reaction times. These results suggest that largely overlapping neural networks process verbal and non-verbal semantic information. In addition, it appears that semantic integration operates across different time scales: earlier processes (indexed by the PS and N400) utilize the established meaningful, but not necessarily lexical, semantic representations, whereas later processes (indexed by the PR and N600) are involved in the explicit interpretation of stimulus semantics and possibly of the required response.

* *Corresponding author.* Center for Research in Language, 9500 Gilman Drive, UCSD Mail Code 0526, La Jolla, CA 92093-0526, USA.
E-mail address: acummings@crl.ucSD.edu (A. Cummings).

0006-8993/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2006.07.050

1. Introduction

Does our ability to derive meaning from words and sentences rely on language-specific semantic resources (Thierry et al., 2003), or do we use more domain-general sources of 'real-world' knowledge and memory (Cree and McRae, 2003)? One attractive method of contrasting meaningful linguistic and non-linguistic processing in the auditory domain has been to compare spoken language to environmental sounds, which have an iconic or indexical relationship with the source of the sound and thus, like nouns and verbs, can establish a reference to an object or event in the mind of the listener.

1.1. *Definition of environmental sounds*

Environmental sounds can be defined as sounds generated by real events – for example, a dog barking, or a drill boring through wood – that gain sense or meaning by their association with those events (Ballas and Howard, 1987). Like words, the processing of environmental sounds can be modulated by contextual cues (Ballas and Howard, 1987), item familiarity and frequency of occurrence (Ballas, 1993; Cycowicz and Friedman, 1998). Environmental sounds can prime semantically related words and vice versa (Van Petten and Rheinfelder, 1995) and may also prime other semantically related sounds (Stuart and Jones, 1995; but cf. Chiu and Schacter, 1995; Friedman et al., 2003, who showed priming from environmental sounds to language stimuli, but no priming in the reverse direction). Gygi (2001) and Shafiro and Gygi (2004) showed not only that spoken words and environmental sounds share many spectral and temporal characteristics, but that recognition of both classes of sounds breaks down in similar ways under acoustical degradation.

Environmental sounds also differ from speech in several fundamental ways. Individual environmental sounds are causally bound to the sound source or referent, unlike the arbitrary linkage between a spoken word's pronunciation and its referent. The 'lexicon' of environmental sounds is small, semantically stereotyped, and clumpy; these sounds are also not easily recombined into novel sound phrases (Ballas, 1993). There is wide individual variation in exposure to different sounds (Gygi, 2001), and correspondingly healthy adults show much variability in their ability to recognize and identify these sounds (Saygin et al., 2005). Finally, the human vocal tract is not capable of producing most environmental sounds (Aziz-Zadeh et al., 2004; Lewis et al., 2005; Pizzamiglio et al., 2005).

1.2. *Comparing environmental sounds to speech*

Despite these differences, comprehension of environmental sounds recruits many of the same cognitive mechanisms and/or neural resources as auditory language comprehension, when task and stimulus demands are closely matched (Saygin et al., 2003, 2005). Not only does spoken language and environmental sounds comprehension appear to develop similarly in typically developing school-age children (Dick et al., 2004, Cummings, Saygin, Bates, and Dick, submitted for publication), as well as in children with language impairment and peri-natal focal lesions (Borovsky et al., in preparation), but the severity of aphasic patients' language comprehension deficits predicts the severity of their environmental sounds comprehension deficits. Thus, behavioral, developmental, fMRI, and lesion data support a common semantic processor of auditory information within the brain (Saygin et al., 2003, 2005). However, the studies mentioned above either measured an outcome of semantic processing or an activation assessed over a large time scale. A possibility exists that during intermediate processing stages, lexical and non-lexical semantic information is processed by different mechanisms. Electrophysiological evidence is necessary to examine the rapid succession of these processing stages, and configurations of the associated neural networks, during word and environmental sound processing.

1.3. *The N400*

One particular event-related potential (ERP) component that can be used to assess the semantic processing of words and environmental sounds is the N400. The N400, a negative wave peaking at approximately 400 ms post-stimulus onset (Kutas and Hillyard, 1980a,b), is elicited by all visually or auditorily presented words. It is also an indicator of semantic integration of the incoming word with the foregoing content: the more explicit the expectation for the next word, the larger the N400 amplitude for words violating the expectation (Kutas and Hillyard, 1983; Kutas and van Petten, 1994; Halgren et al., 2002). The N400 can also be elicited by mismatching meaningful stimulus pairs: two words, two pictures, or a picture and a word (Koivisto and Revonsuo, 2001; Hamm et al., 2002; Ganis and Kutas, 2003; Perrin and Garcia-Larrea, 2003; Wang et al., 2004).

Both Van Petten and Rheinfelder (1995) and Plante et al. (2000) identified N400-related differences in meaningful verbal and non-verbal sound processing. Using a unimodal (auditory) priming experiment, in which either a spoken word preceded an environmental sound or vice versa, Van Petten and Rheinfelder (1995) found that the amplitude and latency of the N400 elicited by words preceded by environmental sounds were indistinguishable from the N400 elicited by a word–word pair. However, the scalp distributions of word versus environmental sound N400 were different. The sounds elicited a larger N400 over the frontal scalp, whereas the words elicited larger N400 responses at the parietal, temporal, and occipital electrode sites. The N400 was also somewhat larger over the right hemisphere for words and significantly larger over the left hemisphere for environmental sounds, suggesting hemispheric differences in the neural networks underlying the processing of words and environmental sounds.

Plante and colleagues (2000) tested healthy and learning-disabled adults using a cross-modal audiovisual paradigm. Here, verbal blocks consisted of visual–auditory word pairs: the first one printed on the screen and the second one spoken via an audio monitor (e.g., apple-orange or apple-dog). The non-verbal blocks consisted of picture-sound pairs: line drawings of objects, animals, or people, paired with either related or unrelated sounds (e.g., bird-birdsong or bird-