

available at www.sciencedirect.comwww.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}

^aProject in Cognitive and Neural Development, University of California, San Diego, USA

^bSan Diego State University/University of California, San Diego Joint Doctoral Program in Language and Communicative Disorders, USA

^cCenter for Research and Language, University of California, San Diego, USA

^dDepartment of Neurosciences, University of California, San Diego, USA

^eBirkbeck College, University of London, UK

^fDepartment 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.

© 2006 Elsevier B.V. All rights reserved.

* 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).

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-

barking). As in the first study, the N400 elicited by the spoken words was larger over the right hemisphere, whereas the N400 elicited by the environmental sounds was larger over the left hemisphere. The rather counterintuitive hemispheric predominance was attributed to “paradoxical lateralization”.¹ Thus, van Petten and Rhinefelder (1995) and Plante et al. (2000) concluded that the larger activations recorded on the right side of the in response to the words was due to predominantly left hemisphere involvement, and vice versa for the environmental sounds.²

1.4. Processing of nouns and verbs

Both the van Petten and Rhinefelder (1995) and Plante et al. (2000) studies used concrete animate and inanimate nouns to compare with environmental sounds. Whereas environmental sounds convey information about the object involved in the sound, they can also convey information about an event or action. Thus, it is possible that the semantic information they transmit might be more similar to that conveyed by a verb and thus may influence their electrophysiological signatures. Reports in the behavioral and neuroimaging literature regarding noun/verb differences suggest that this may be the case. For example, object naming (noun generation) and action naming (verb generation) are affected differently by word frequency (Szekely et al., 2005). ERP studies have indicated that nouns (associated with strong visual associations) and verbs (associated with motor associations) activate different cortical generators in both hemispheres (for a review, see Pulvermuller, 1999).

1.5. Goals of the present study

Here, we compared processing of environmental sounds with empirically matched nouns and verbs in an audiovisual cross-modal sound–picture match/mismatch paradigm. To examine the semantic processing of meaningful information (either lexical or not), we compared brain’s response to words and environmental sounds with the brain’s response to complex but ‘non-meaningful’ stimuli in the same experimental paradigm. Finally, we utilized a single-trial EEG analysis technique (here called ERP imaging) to examine which ERP components correlated with subjects’ behavior during conditions involving semantic processing.

¹ This is most often seen for motor potentials (cf. Boschert et al., 1983; Boschert and Deecke, 1986). For example, a unilateral foot movement produces larger potentials over the ipsilateral hemisphere as compared to the contralateral. This atypical result has been attributed to the fact that cortical representations of the foot are near the medial surface of the contra-lateral hemisphere but the neurons are oriented so that the current flow is greatest toward the opposite side of the head (Van Petten and Rhinefelder, 1995).

² In regards to cross-domain differences, it is worth noting that the visual primes in the Plante et al. (2000) study belonged to different input domains: printed words (lexical domain) vs. line drawings (non-lexical domain). Therefore, the observed N400 differences may have in part reflected the differences in integration across the different visual and auditory domains rather than differences in the processing of words vs. environmental sounds per se.

2. Results

2.1. Behavioral performance

2.1.1. Accuracy

Subjects responded more accurately in the environmental sound trials than in the word trials (stimulus type effect: $F(1,24)=11.343$, $p<0.003$; Table 1). There were no accuracy differences between the noun and verb conditions. A marginal Word Class \times Sound Type interaction ($p<0.06$) was observed, which was driven by the subjects in the Verb Word Class experiment being less accurate on word stimuli, but the subjects in the Noun Word Class condition being more accurate on environmental sound stimuli.

Subjects’ judgments of the non-meaningful sound trials were considered subjective. Nonetheless, the number of non-meaningful stimulus trials that subjects identified as matching and mismatching was examined to ensure that subjects did not have either a “match” or “mismatch” bias toward the non-meaningful sounds as a whole. On average, the subjects identified 71.7% of the experimenter-defined matching trials as matching and 78.5% of the experimenter-defined mismatching trials as mismatching. This indicated a fairly good agreement with the intended stimulus roles and showed that on these trials, the subjects were performing the task as expected.

2.1.2. Reaction time (RT)

The Sound Type effect was significant for reaction times ($F(2,21)=35.838$, $p<0.0001$; Table 1). However, it originated solely from the longer RTs in the non-meaningful sound trials, as compared with the meaningful sound trials. There was no overall RT difference between the word and environmental sound trials, or between the Noun and Verb Word Classes.

Because the main focus of this study was to compare word and environmental sound processing, the two meaningful sound types were examined without the non-meaningful sounds in the ANOVA model (Word Class \times Sound Type). A significant Sound Type \times Word Class interaction was observed ($F(1,24)=5.472$, $p<0.028$), which motivated independent

Table 1 – Accuracy and reaction time measures for all sound types recorded via button press response

Sound type	Accuracy (% correct)	RT (in milliseconds)
Nouns	96.57 (3.60)	760 (141)
Verbs	95.97 (3.82)	793 (151)
All words	95.37 (3.77)	773 (148)
Environmental sounds	97.12 (3.23)	789 (191)
Non-meaningful sounds	n.a.	934 (201)

n.a.—not applicable, standard deviation in parentheses.

Responses to the Nouns and Verbs are reported separately to show Word Class effects. Measures for Words, Environmental Sounds, and Non-Meaningful Sounds are pooled across the Noun and Verb Word Class experiments.

analyses of the noun and verb experiments. In the Noun experiment, the RTs to words were significantly faster than the RTs to environmental sounds ($F(1,11)=6.032, p<0.032$). In the Verb experiment, there was no effect of Sound Type ($p=0.511$).

2.2. ERP results

All sounds matching the pictures elicited ERPs characterized by an auditory N1–P2 complex, followed by a protracted positivity, maximal over the fronto-central electrodes. ERPs elicited by sounds mismatching the pictures were characterized by the N1–P2 complex followed by a negativity maximal over the centro-parietal areas (Fig. 1). Mismatch-minus-match ERP difference waveforms revealed two negativities: one maximal centro-parietally (300–400 ms, the N400), and another maximal frontally (500–700 ms, here called the N600). Whereas the N400 was clearly larger in amplitude in meaningful sound trials, the N600 was similar in amplitude for all stimulus types.

2.2.1. N400 peak latency

The responses to the non-meaningful sounds were small and inconsistent. Therefore, rather than forcing the selection of a peak, the non-meaningful sounds were not included in the latency analysis. In word vs. environmental sound ANOVA, a main effect of Sound Type was observed ($F(1,24)=49.066, p<0.0001$). The environmental sound N400 peaked significantly earlier ($M=331$ ms) than the word N400 ($M=401$ ms). The Word Class effect was not significant.

2.2.2. N400 onset latency

To rule out the possibility that the observed differences in word vs. environmental sound N400 latency were not caused by earlier recognition of the sounds, the onset latency of the word and environmental sound N400-s was assessed. It was measured at the electrode (Cz) for the most positive data value

between 150 and 400 ms, preceding the N400 peak. No differences in word vs. environmental sound onset latency were observed.

2.2.3. N600 peak latency

The N600 latency measures for words ($M=584$ ms; $SD=47$ ms), environmental sounds ($M=571$ ms; $SD=61$ ms), and non-meaningful sounds ($M=591$ ms; $SD=31$ ms) were all very similar. There were no significant main effects or interactions involving this measure.

2.2.4. N400 amplitude

We found a main effect of Sound Type on mean N400 amplitude ($F(2,21)=23.603, p<0.0001$; Table 2, Figs. 1 and 2). This was driven solely by the difference between the meaningful and non-meaningful sound trials. Post hoc contrasts revealed that the word-N400 (mean= -5.84 μ V) and environmental sound-N400 (mean= -5.96 μ V) amplitudes did not differ significantly from each other, but both were significantly larger than the non-meaningful sound-N400 (mean= -2.03 μ V, $p<0.0001$) amplitudes. When words and environmental sounds were analyzed in an ANOVA, no Word Class effect or Word Class \times Sound Type interaction was found.

2.2.5. N600 amplitude

There was no main effect of Sound Type on mean N600, with similar mean N600 amplitudes for words ($M=-2.03$ μ V), environmental sounds ($M=-2.25$ μ V), and non-meaningful sounds ($M=-2.09$ μ V). There were no significant effect of Word Class nor was there an interaction of Word Class \times Sound Type.

2.2.6. N400 scalp distribution

The scalp distribution of the N400 peak was first assessed using raw amplitude data to ensure that it was comparable with what is typically observed (Kutas and Hillyard, 1983). For

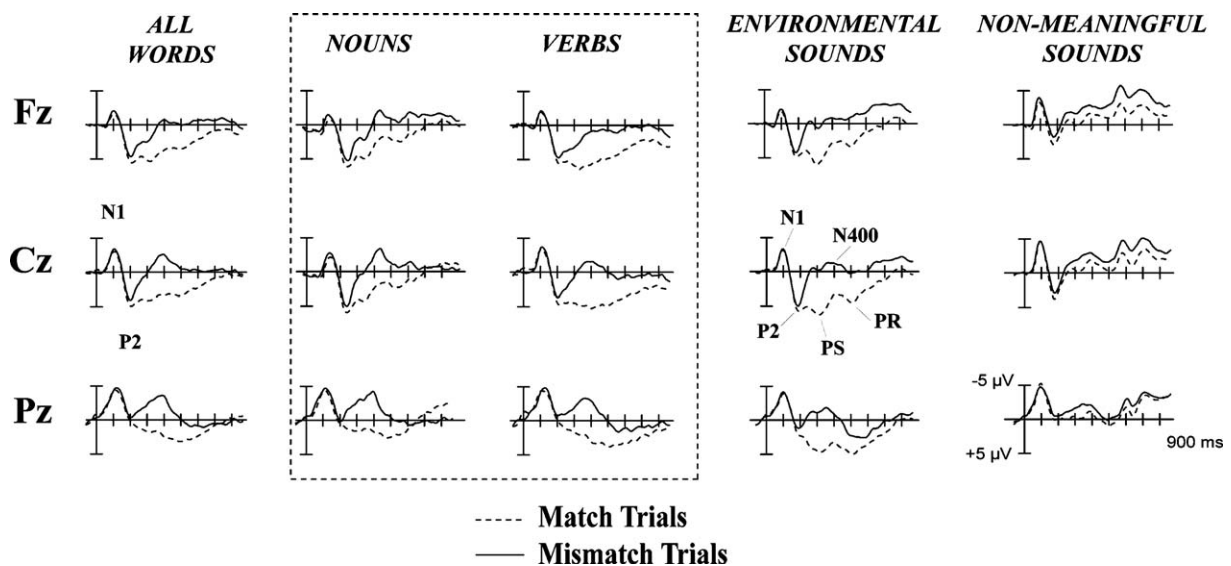


Fig. 1 – Matching and mismatching ERP responses to each stimulus type recorded at the midline electrodes. Responses to the Nouns and Verbs are shown separately. ERPs for Words, Environmental Sounds, and Non-Meaningful Sounds are pooled across Noun and Verb Word Class experiments. Early responses N1 and P2 are visible in all stimulus types, whereas the N400 is only visible in the meaningful stimulus responses.

Table 2 – Mean amplitude and latency of the N400 for all sound types recorded at the midline electrodes

Sound type	Amplitude (in mV)			Latency (in ms)		
	Fz	Cz	Pz	Fz	Cz	Pz
Nouns	-5.09 ^a (1.99)	-5.64 ^a (3.39)	-6.9 ^a (3.28)	394 (32)	384 (25)	398 (19)
Verbs	-5.56 ^a (3.03)	-8.02 ^a (3.82)	-7.11 ^a (3.37)	426 (42)	411 (42)	401 (49)
Words	-5.33 ^a (2.54)	-6.88 ^a (2.54)	-7.01 ^a (2.54)	411 (40)	398 (37)	399 (37)
Environmental sounds	-6.11 ^a (3.34)	-7.36 ^a (4.16)	-6.73 ^a (3.51)	323 (41)	332 (44)	330 (43)
Non-meaningful sounds	-2.22 ^a (2.08)	-2.31 ^a (2.20)	-2.19 ^a (2.05)	n.a.	n.a.	n.a.

Responses to the Nouns and Verbs are reported separately to show Word Class effects. Measures for Words, Environmental Sounds, and Non-Meaningful Sounds are pooled across the Noun and Verb Word Class experiments. Mean amplitude significance value is compared to pre-stimulus baseline measure.

^a $p=0.0001$; n.a.—not applicable, standard deviation in parentheses.

data pooled across all sound types, amplitude differences across six anterior–posterior levels were significant ($F(5,105)=20.242, p<0.0001$), with post hoc contrasts between the level pairs showing that this effect was driven primarily by the larger amplitudes at the centro-parietal (CP1/CP2; $M=-5.42 \mu V$) electrode sites compared with any other electrode pair ($p<0.004$). Additionally, the N400 was larger over the right ($M=-4.71 \mu V$) compared with left hemisphere sites ($M=-$

$-4.3 \mu V$; $F(1,21)=6.2, p<0.021$). Such right centro-parietal predominance is very consistent with those reported in the N400 literature (Kutas and Hillyard, 1980a,b, 1982; Kutas et al., 1988; Kutas and Iragui, 1998).

Among the three sound types, the mean amplitudes of both the word ($F(14,350)=10.006, p<0.0001$) and environmental sound ($F(14,350)=13.256, p<0.0001$) N400 differed by electrode site, whereas there were no electrode effects for the non-

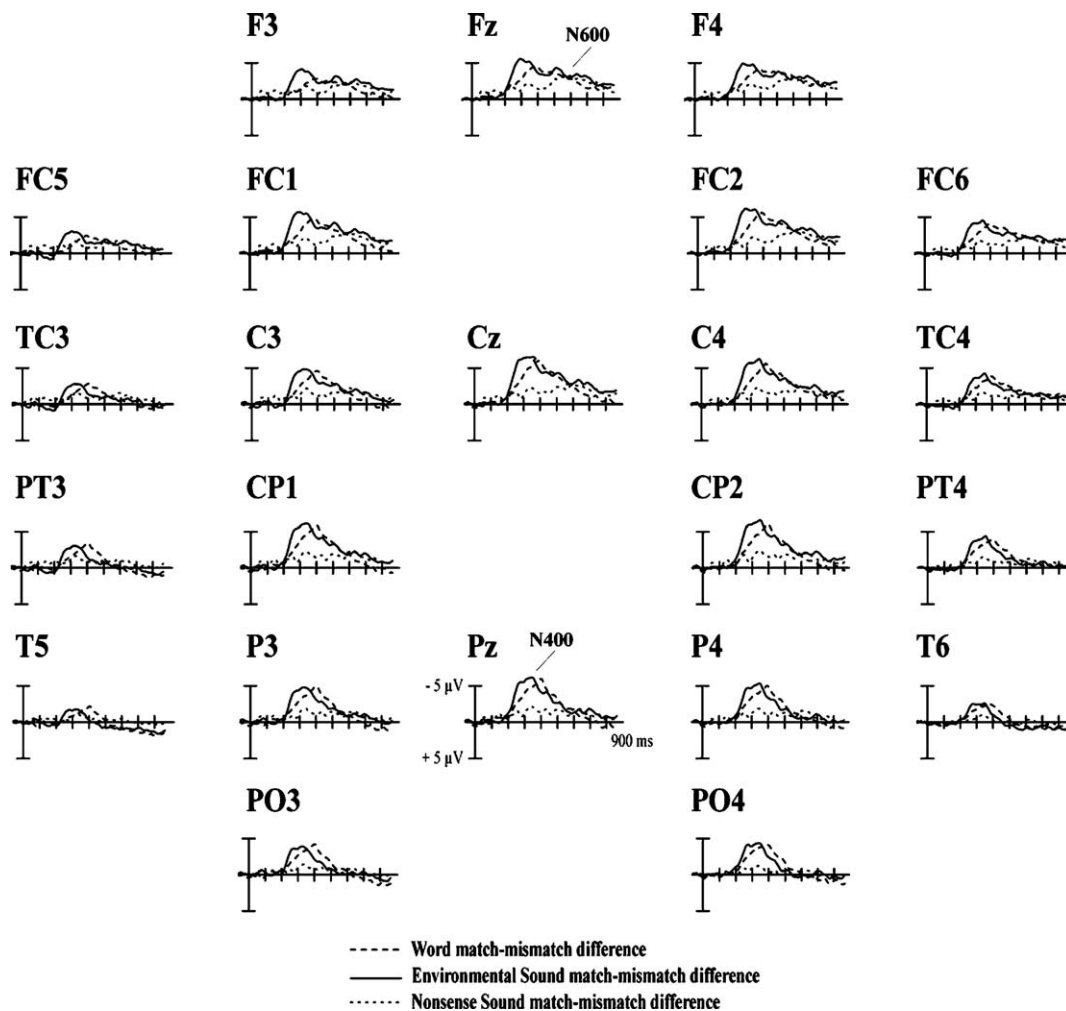


Fig. 2 – Mismatch-minus-Match ERP Difference Waves. ERPs to Words, Environmental Sounds, and Non-Meaningful Sounds are pooled across Noun and Verb Word Class experiments. The N600 is prevalent at the frontal electrode sites for all stimulus types, whereas the N400 effect in response to words and environmental sounds is most prevalent at centro-parietal sites.

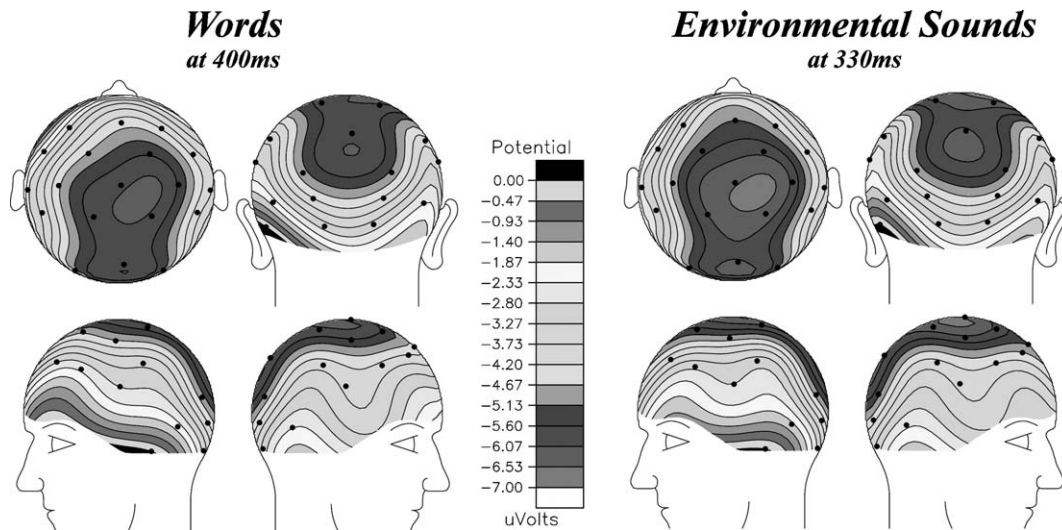


Fig. 3 – Scalp Density Voltage Plots. Plot shading represents the mean amplitudes of all words and environmental sounds at their peak latencies: 400 and 330, respectively.

meaningful sounds. Therefore, the non-meaningful sounds were not included in further Sound Type scalp distribution analyses conducted using z score normalized N400 amplitudes (see Methods). We found a significant Sound Type (Word vs. Environmental Sound) \times Electrode (15 levels) interaction ($F(14,154)=4.084$, $p<0.011$). This result motivated further anterior–posterior (6 levels) as well as left–right (2 levels) laterality analyses which yielded an interaction between Sound Type and Anterior–Posterior dimension ($F(5,125)=6.611$, $p<0.0001$; Fig. 3). Post hoc tests showed that the only difference between the two sound types occurred at the frontal electrodes (F3/F4), where environmental sounds elicited larger N400 deflections than the words ($p<0.003$). No laterality differences were found. This suggests that words and environmental sounds share fairly similar scalp distribution patterns, particularly in terms of laterality. There were no scalp distribution differences between the noun and verb N400.

2.2.7. N600 scalp distribution

As with the N400, we first assessed the distribution of the N600 with raw amplitude data. Amplitude differences across the six anterior–posterior levels were significant ($F(5,105)=24.999$, $p<0.0001$), with post hoc tests showing larger responses at the fronto-central (FC1/FC2) electrodes compared with other electrode sites ($F(1,21)=12.498$, $p<0.013$). There was no laterality effect. Normalized amplitudes were used to examine further the potential relationship between sound type and scalp distribution. There was no Sound Type (Word/Environmental Sound/Non-Meaningful Sound) \times Electrode interaction, suggesting all three sound types share similar scalp distributions for the N600.

2.2.8. Comparison of the N400 and N600

Scalp distribution analyses using normalized data at 6 anterior–posterior levels, 2 electrodes each, showed that for both word and environmental sounds, the N600 was distributed anteriorly when compared to the corresponding N400

($F(5,120)=16.45$, $p<0.0001$, and $F(5,120)=17.92$, $p<0.0001$, respectively). Furthermore, the Peak \times Stimulus Type \times Anteriority interaction was also significant in the Word vs. Nonsense sounds comparison ($F(5,120)=4.04$, $p<0.02$), with a similar trend in the Environmental Sounds vs. Nonsense sound comparison ($p<0.15$). This effect was driven by the fact that over the frontal scalp regions, the N600 did not differ over different stimulus types, but over parietal scalp regions, the N400 was larger for the meaningful than for the non-meaningful stimuli (Fig. 4).

2.3. Correlations between averaged N400 and RT

In order to examine whether there was a relationship between the N400 and behavioral performance, we ran correlation analyses between reaction time (RT) and the N400 mean amplitude, N400 peak latency, and N400 onset latency for electrode Cz. None of these correlations were significant.

2.4. Single-trial ERP analysis

Peak latencies of the averaged ERP peaks provide information about the timing of the respective neural processing stages. However, these latency measures are averaged across trials and lack information about the dynamic (trial by trial) relationships between the brain processes and behavior. In order to define which EEG phenomena are dynamically associated with behavioral performance during semantic processing, we performed single-trial ERP analysis (ERP imaging; Jung et al., 2001) on the word and environmental sound matching and mismatching trials.³ Fig. 5 demonstrates

³ Subjects' behavioral and ERP responses to the word and environmental stimuli were very similar, so the two were combined together for the ERP imaging analysis. The non-meaningful sounds were not included in the ERP imaging analysis because they did not invoke indices of semantic integration comparable to the meaningful sound stimuli.

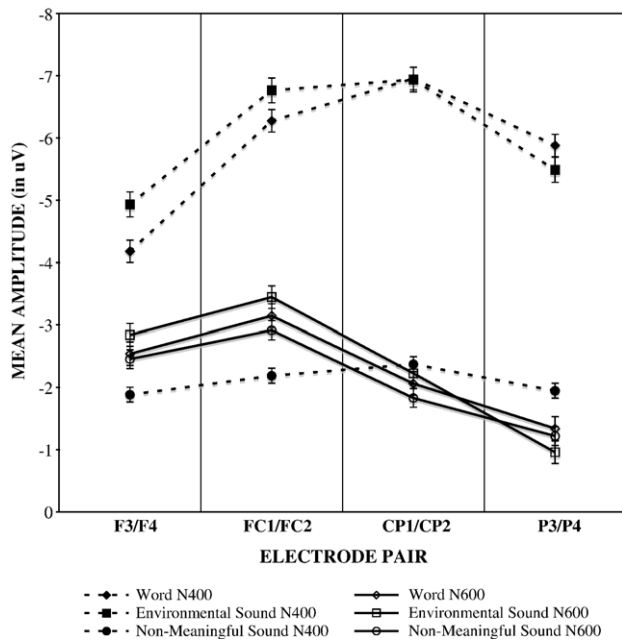


Fig. 4 – N400 and N600 Mean Amplitudes. Mean amplitudes for words, environmental sounds, and non-meaningful sounds are plotted by scalp anteriority. Amplitudes are the mean amplitude of each electrode pair (e.g., F3/F4). The N400 and N600 scalp distributions and their differential responsiveness to the meaningfulness aspects are clearly depicted here: the N400 magnitude was most prevalent at CP1/CP2 in response to both words and environmental sounds whereas the N600 magnitude was largest frontally with no sound type variation. Error bars show the standard error of the mean.

across-subjects single-trial color-coded ERPs (ERP images) at the Fz, and Pz electrodes, sorted by subjects' reaction times (top panel), N400 amplitude (middle panel), and sound length (bottom panel).

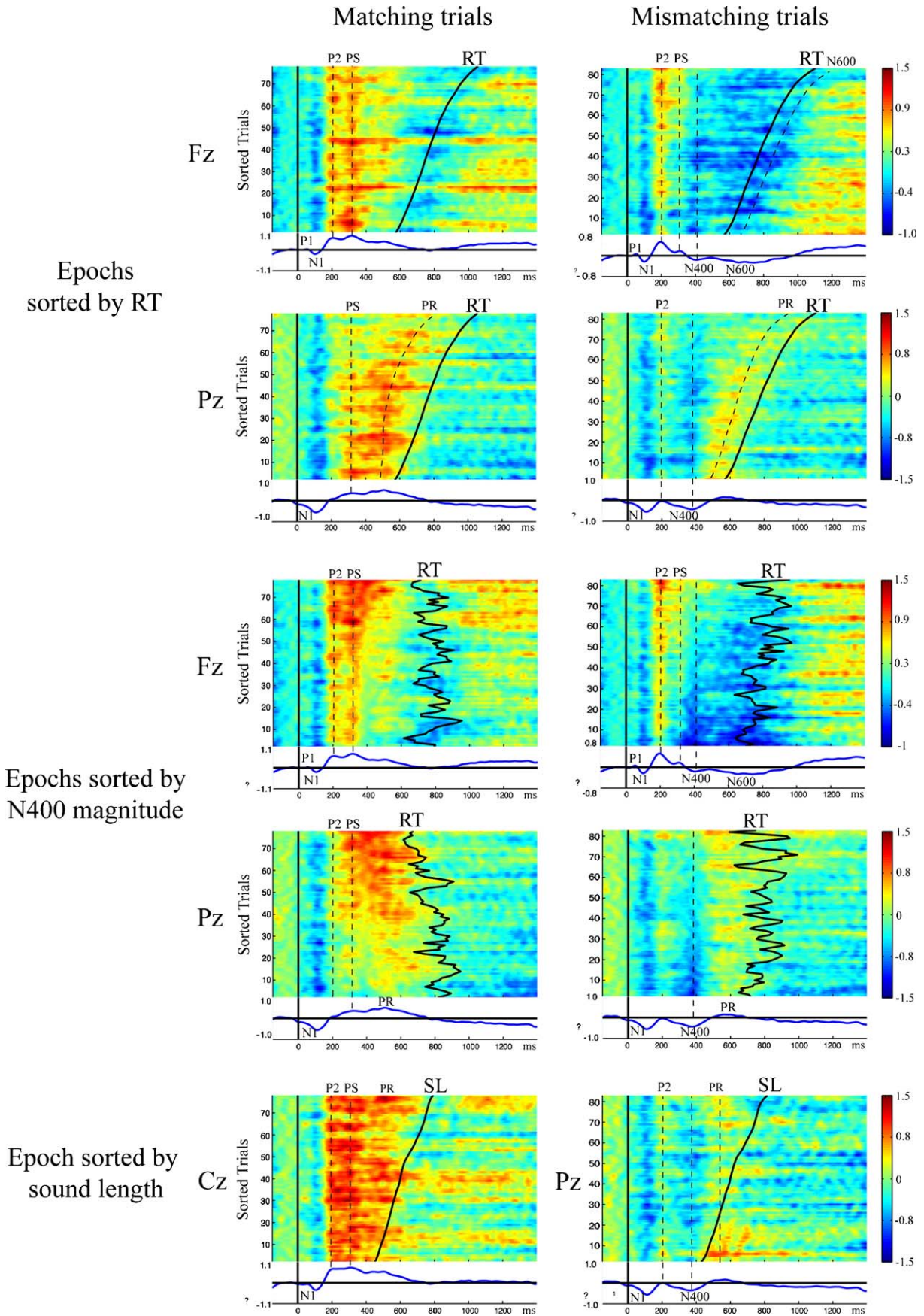
This three-dimensional (trials, time, amplitude) view into the evoked brain activity revealed at least three functionally distinct sets of activities that differed between frontal and parietal scalp regions. The first set comprised stimulus-onset-aligned activities, corresponding to the sensory ERP peaks P1 (50 ms), N1 (100 ms), and P2 (Ceponiene et al., 2005). In both

match and mismatch ERP images, these activities were most prominent and best expressed in the frontal channels, corresponding to the scalp distribution of the auditory sensory peaks. None of these were related to reaction times (top panel) or sound length (bottom panel) and will not be further discussed.

The second set comprised what we will somewhat loosely term semantic processing-related activities: the N400 and a positive peak we will refer to as the PS. The "S" denotes "semantic" because in matching trial ERPs, this peak differentiated meaningful (words and environmental sounds) from the non-meaningful stimuli (Fig. 1). In both ERP images and averaged ERPs, the PS appeared as the second peak of the extended positivity in the matching trials at ca. 320 ms and was best expressed over the frontal and central regions (Fig. 5, top panel, left column; see also Fig. 1). In the mismatching trials, the PS slightly preceded and largely overlapped with the subsequent N400 negativity at ca. 370 ms (Fig. 5, top panel, right column). Both the PS and N400 were aligned to stimulus onsets; their timing was not related to the behavioral response times (Pearson's product-moment correlations: match PS at Fz, $r=0.12$, $p=0.68$; mismatch N400 at Pz, $r=0.03$, $p=0.81$). However, the magnitude of these activities was linked with the reaction times (Fig. 5, middle panel): in the matching trials, there was a significant relationship between RTs and PS magnitude (i.e., the stronger the activity, the shorter the reaction time; Fz: $r=-.22$, $p<0.05$; Pz: $r=-.34$, $p<0.003$), whereas in the mismatching trials, a positive correlation was found between the N400 magnitude and RTs ($r=0.23$, $p<0.04$). Finally, neither the latency nor the magnitude of the N400 activity appeared to be associated with sound length (Fig. 5, bottom panel).

The third functional set was composed of response-related activities: frontally, the N600, which preceded and followed subject's behavioral responses and parietally, a positivity which we will call PR ("R" for "response"), which preceded the subjects' RTs by ca. 100 ms (Fig. 5, top panel). For both matching and mismatching trials, the parietal magnitude and latency of the PR was strongly correlated with reaction times (matching trials: latency, $r=0.42$, $p<0.0001$; amplitude: $r=-.31$, $p<0.005$; mismatching trials: latency, $r=0.43$, $p<0.0001$; amplitude: $r=-.34$, $p<0.002$). The frontal magnitude of the N600 showed a similar relationship in the matching trials ($r=0.24$, $p<0.03$), with a similar trend in the mismatching trials ($r=0.16$, $p<0.15$).

Fig. 5 – Group grand single-trial ERP images at the Fz and Pz electrodes. Matching (left column) and mismatching (right column) trials were sorted by subjects' reaction times (top panel), brain activity magnitude in the N400 latency range (350–425 ms; middle panel), and auditory stimulus length (SL; bottom panel). Only meaningful sound trials were included. *Top panel:* Three functionally distinct brain activity patterns were identified: (i) stimulus-onset aligned activities, corresponding to the sensory ERP peaks P1, N1, and P2 (most evident in frontal channels). (ii) Semantic processing-related activities, the PS and N400. The PS was most evident in matching trials over the frontal electrodes (top and bottom panels, respectively); in the mismatching trials, the PS largely overlapped with the subsequent N400 negativity. Both the PS and N400 were aligned to stimulus onsets; their timing did not influence behavioral response times. (iii) Response-related activities: frontally, the N600, which preceded and followed subject's behavioral responses; parietally, the PR, which preceded the subjects' response by ca. 100 ms. *Middle panel:* epochs sorted by the amount of negative activity (more negativity at the bottom) in the latency range of 350–425 ms. The magnitudes of the frontal PS and parietal PR were associated with reaction times. In the mismatching trials, a possible relationship could be seen between the reaction times and the magnitude of the N400 (both frontally and parietally), as well as with the magnitude of the N600 (frontally). *Bottom panel:* Duration of the PR activity appeared to be related to sound length. In contrast, neither the latency nor the magnitude of the N400 activity was appeared to be associated with sound length.



Although the duration of the compound positivity (P2+PS+PR) appeared to be related to the sound length (Fig. 5, bottom panel, left column), this was not due to the PR component. When separated from the larger positive complex by the N400 in the mismatching trials, it showed no relationship with the sound length (Fig. 5, bottom panel, right column).

In summary, ERP imaging revealed three main findings that would not have been revealed by conventional ERP peak-RT correlations. First, the slow positive deflection elicited by matching meaningful sounds is composed of at least three sub-components: the fronto-central sensory P2, fronto-central semantic PS, and the centro-parietal response-associated PR. Second, the timing of the PS and the N400 components of the ERPs are stimulus onset-locked but their magnitudes are related to the behavioral response times. Third, both the timing and magnitude of the PR component, and magnitude of the N600 component, appear to be tied to overt behavioral response.

3. Discussion

This study compared behavioral and electrophysiological responses associated with audiovisual semantic integration in nouns and verbs, environmental sounds, and non-meaningful auditory stimuli. The electrophysiological differences between meaningful verbal and non-verbal sounds were subtle and consisted of higher response accuracy and an earlier N400 latency to environmental sounds than words, as well as fine-grained N400 scalp distribution differences. No Word Class effects (nouns vs. verbs) were uncovered. In contrast, the non-meaningful stimuli elicited a negligible N400 and longer reaction times. Finally, our single-trial ERP imaging analyses revealed that the brain activity most closely paralleling the behavioral reaction times was a parietal positivity, the PR, following the N400 peak. Although magnitudes of the N400 and the underlying PS activities correlated with the RT behavior, their timing did not parallel subjects' response times.

3.1. Stimulus type dimension

We found relatively subtle N400 differences between words (either nouns and verbs) and environmental sounds. Whereas the N400 onset analysis showed no differences in the latencies of words and environmental sounds, the environmental sounds elicited an earlier and somewhat more anteriorly distributed N400 response than did the word stimuli. This suggests that although both sound types start the semantic integration stage at the same time, the environmental sound processing may proceed faster. One reason for this difference may be that the environmental sounds stimuli are much more variable on several acoustical parameters relative to the word stimuli. Thus, the listeners may be receiving more 'low-level' acoustical cues that disambiguate between competing environmental sound candidates, of which there are many fewer classes or types than in the case of nouns or verbs. As a consequence, the identification point of the environmental sounds may be

earlier as compared to the identification point of the words. This interpretation is consistent with behavioral results where semantically matched environmental sounds have been processed faster than their corresponding verbal labels in several prior studies in different subject populations (for a review, see Saygin et al., 2005). It is also possible that the latency differences may be due to the lexical (or not) nature of the stimuli: words may have to go through a lexical stage of processing before their semantic nature can be accessed whereas environmental sounds may directly activate the corresponding semantic representations, with a corresponding earlier N400 peak latency.

Because sound duration is known to affect auditory sensory ERPs (Kushnerenko et al., 2001), word, environmental sound, and non-meaningful sound stimulus sets were matched for mean and range of duration. For all stimulus types, sound durations ranged from 466 to 1154 ms. However, unlike the case of the auditory sensory ERPs, no evidence was found for a link between the N400 activity and sound length, as shown by N400 onset latencies (with no difference between the two sound types) and single-trial ERP imaging (Fig. 5, bottom panel).

It is also interesting to note that although the environmental sounds elicited a significantly earlier N400 than did the words, theoretically implying earlier semantic integration, the behavioral reaction times were not different for the two stimulus types. Strong clarifying evidence was provided by both the correlation analyses and our single-trial ERP image analysis, which showed that the timing of the N400 is not tied to the behavioral response time (Fig. 5). Thus, the N400 latency-RT discrepancy likely originates in the 'response stages' of processing. Whereas it may be easier to initially identify an environmental sound as indexed by the N400, the subsequent transformation of that identification into a response appears to take a relatively longer period of time for the environmental sounds than words. At least in part, this may be an experiential effect: an average person in the present-day society not only has more exposure to the verbal material than to meaningful natural sounds, but also in using words for communication. Therefore, word representations may have stronger and/or more widespread associations with the various response mechanisms than representations of the environmental sounds. Thus, translating the non-lexical meaningful auditory input into a behavioral response (i.e., match or mismatch) may take longer than translating lexical input.

Previous ERP studies (e.g., van Petten and Rhiensfelder, 1995; Plante et al., 2000) had found small laterality differences in the processing of speech and environmental sounds, with words evoking larger responses in the right hemisphere and environmental sounds eliciting larger responses in the left hemisphere. The present study did not find such laterality differences in the processing of words and environmental sounds. One possible reason why our results were not consistent with the earlier studies is due to different data analysis techniques. In contrast to van Petten and colleagues (1995, 2000) who used raw amplitudes in their laterality analyses, we used normalized mean amplitudes. It has been shown that when using non-normalized data, significant scalp distribution differences can be caused by mere differences in signal

strength rather than true distribution differences (McCarthy and Wood, 1985).⁴

In sum, the scalp distribution of the N400 in the present study does not appear to indicate substantial differences in the structure of neural networks processing verbal vs. non-verbal meaningful information. This is consistent with findings from studies with unilateral brain lesions populations that have shown a common processing breakdown of words and environmental sounds and common lesion locations (Saygin et al., 2003). Examination of other ERP components implicated in semantic processing, such as the PS peak noted in the present study may be a promising route for future research on this question.

3.2. Word Class dimension

Because previous studies (Dehaene, 1995; Pulvermuller, 1996, 1999; Szekely et al., 2005) have reported behavioral and electrophysiological differences in the processing of nouns and verbs, word class differences may also have been expected. However, neither behavioral nor electrophysiological (N400) differences were found. This null effect can possibly be attributed to the experimental paradigm. The task in the present study was a fairly simple picture/sound-matching paradigm, as compared to a more complex task of formulating and producing a verbal label (Szekely et al., 2005) or a lexical decision task (Pulvermuller, 1996). It is possible that word class differences may only be revealed when processing demands are increased or when more specific 'noun' or 'verb' tasks are associated with the experimental paradigm (Federmeier et al., 2000). Couching both environmental sounds and nouns/verbs within such tasks may serve to disambiguate the relative 'noun-ness' or 'verb-ness' of classes of environmental sounds.

3.3. Meaningfulness dimension

Large electrophysiological N400 differences were found between meaningful and non-meaningful stimuli. The meaningful stimuli (words and environmental sounds) elicited significantly larger N400 amplitudes than did the non-meaningful sounds. One explanation for this effect is that no pre-established semantic representations exist for the non-object pictures and non-meaningful sounds. Therefore, an expectation for the auditory stimulus could not be formed, and the semantic mismatch could not occur. However, the subjects were able to match the pictures and sounds based on their physical properties, and a small, though significant, as compared to baseline activation, N400 response was elicited in the non-meaningful trials (Table 2). These results may reflect the formation of rough, 'on-the-fly' semantic categories related to the non-meaningful sounds. The subjects underwent a brief practice session prior to beginning the experiment to acquaint themselves with the task and the differences between the "jagged" and "smooth" pictures and sounds.

⁴ However, our results were different from Van Petten et al. (1995) even when raw amplitudes were used. In an analysis of the word and environmental sound N400 scalp distributions with non-normalized data, again no Sound Type × Laterality interaction was observed ($p > 0.56$).

Therefore, it is likely that subjects formed intuitive semantic categories of "smooth" and "jagged" stimuli in order to perform the task. Violations of category membership are known to elicit a N400: stimuli that do not belong to a specific semantic category elicit larger N400 responses than stimuli that do "fit" into a category (Polich, 1985; Heinze et al., 1998; Federmeier and Kutas, 1999; Nunez-Pena and Honrubia-Serrano, 2005).

3.4. Dynamic links between brain and behavior

Although the sensitivity of the N400 component to semantic incongruity has been clearly demonstrated in previous studies, the dynamic links between brain processes generating the N400 and behavioral response have not been fully characterized. Knowing the nature of such relationships is important for understanding of the functional roles of brain processes under question. We utilized a single-trial ERP imaging technique to explore whether the timing and magnitude of the N400 activity and the underlying semantic positivity are associated with the behavioral responses on a trial-by-trial basis. The data shown in Fig. 5 confirmed the evidence that early sensory ERP peaks (P1, N1, P2) are "stimulus locked"; that is, they are generated in a strict orderly and timely fashion following onset of an external stimulus. This pattern suggests that these processes are concerned with the automatic processing of physical stimulus features, including feature analysis and synthesis stages.

Importantly, two later ERP phenomena, a positivity in matching meaningful stimulus trials (the PS, Figs. 1 and 5) and a negativity in mismatching trials (the N400, Figs. 2 and 5) that appear to be involved in semantic processing, were also stimulus locked. We suggest that the PS is related to semantic aspects of processing because (i) it was not elicited by non-meaningful stimuli (Fig. 1), (ii) its magnitude correlated with the response times, and (iii) it preceded the N400 component by less than 100 ms. Further, because the N400 reflects semantic integration, it is reasonable to expect it to follow the neural activity of semantic encoding (possibly the PS). Such a model closely corresponds to the intracranial activation patterns observed in N400-eliciting conditions (Halgren et al., 1994a,b, 2002; Halgren et al., in press).⁵

Finally, the fact that the timing of the PS and N400 activities are stimulus-locked suggests that they utilize pre-established, readily accessible, and obligatorily activated established semantic representations. It may be that the strength of activation of these representations influences behavioral performance, as suggested by the ERP image analyses.

This interpretation of the PS is consistent with Federmeier and Kutas (1999) who presented subjects with pairs of sentences in which the last word of the second sentence was an expected exemplar, a within-category violation, or a between-category violation. They reported a positivity akin to the PS found in our

⁵ These patterns consisted of a concurrent deep source and superficial inhibitory post-synaptic activity in response to semantic stimuli, bound to produce positive voltage at the scalp (a possible correlate of P2s), and a prolongation of the IPSP (inhibitory post-synaptic potential) source activity in the deeper cortical layer IV in response to semantic incongruity, a pattern that would correspond to a negativity at the scalp (possibly the N400).

data—which importantly also had a different scalp distribution than did their reported N400. The positivity was elicited by sentence-congruent end words and it was thought to it reflect the activation of semantic features. Specifically, items of the same semantic category (i.e., items sharing many semantic features) that were “expected” in a sentential context elicited a late positivity. Although the positivity of Federmeier and Kutas (1999) was elicited in reference to group-level semantic features, in our study it might refer to the single-item level.

Brandeis et al. (1995) also reported bilateral posterior positivities in the time range of our PS in a paradigm in which subjects silently read correct and incorrect versions of simple sentences with predictable color endings, and of more complex sentences with predictable composite word endings. Brandeis et al. (1995) interpreted their positivity as indicative of specific verbal processing of expected words at the end of sentences. It is not unlikely that the positivity reported by Brandeis and colleagues (1995) indexes the same type of cognitive mechanism as our PS, as it was also elicited only by semantically congruent (i.e., “correct”) sentences.

Although Federmeier and Kutas (1999) and Brandeis et al. (1995) did not specifically manipulate this pre-N400 positivity, a series of word recognition studies did (Rudell, 1991; Rudell et al., 1993; Rudell and Hua, 1995, 1996, 1997). Rudell and colleagues observed an occipital positivity that was evoked by visual presentations of words, pictures, and cartoons at approximately 200–250 ms post-stimulus onset. This positivity was interpreted as an index of stimulus recognition and was given the name “Recognition Potential” (RP). Rudell and Hua (1997) reported a low within-subjects correlation between RP latency and RT ($r=0.04$) indicating that the subjects who decreased their RT the most with training showed little tendency to also have the greatest decreases in RP latency. Thus, just as there was no correlation between our PS and subjects’ RTs, Rudell and Hua (1997) found no relationship between their RP and RT. Additionally, the elicitation conditions and timing of the RP are fairly similar to those of our PS. Modality differences (auditory in our study and visual in Rudell and colleagues’ studies) are likely to account for the scalp distribution differences. Therefore, it appears that the PS in the present study and the earlier reported RP index the same type of cognitive process, i.e., semantic stimulus recognition.

Finally, our ERP image data suggested a parieto-frontal network closely linked to behavioral performance, as reflected by the PR and N600 components (Figs. 2 and 5). Although the PR was maximal over the parietal scalp, it cannot be considered a P300-family response because it preceded, not followed, subjects’ RTs (Makeig et al., 1999). Further, the parietal scalp distribution, a long preceding time with respect to RTs, as well as an imperfect temporal relationship with the response times makes the possibility of the PR being a pre-motor response rather unlikely. This suggests that at least part of this activity is not related to response execution but is rather related with making a decision about stimulus match—one possibly informed by the processes indexed by the PS and N400 generators. In fact, an extensive literature search did not reveal another ERP component related to the PR. This is probably because previous studies did not use single-trial ERP imaging analyses so the PR could not have been teased apart from the larger P2–PS–PR complex.

In contrast to the N400, there were no differences over stimulus type in the N600 latency or amplitude, suggesting that N600 generation is not dependent on an established semantic representation of a stimulus. Rather, its frontal predominance, temporal proximity (could either precede a reaction time, follow it, or both; Fig. 5), and magnitude-RT relationship suggest that the N600 is related to stimulus-general processes, such as maintenance of task demands and response monitoring (Halgren et al., 1994b). This explanation is consistent with other ERP studies that have interpreted frontal negative slow waves as indicative of working memory or general, non-specific, cognitive processes such as attention (Itoh et al., 2005; Koelsch et al., 2003; King and Kutas, 1995).

4. Conclusions

The semantic integration of verbal and non-verbal meaningful information, as well as of verb and noun lexical categories, involves largely shared neural networks and processes (consistent with Saygin et al., 2005, 2003; Dick et al., submitted). The present study added temporal precision to previous work and revealed additional, subtler findings. The major difference between environmental sound and word processing might occur during the post-N400 stage of explicit cognitive processing, where the time to output is longer for environmental sounds than words, feasibly due to the experiential and encoding differences. Additionally, and in contrast to environmental sounds and words, the encoding of non-meaningful information does not involve the same types of neural activation. Thus, there appears to be differential activation of specialized semantic neural networks.

Additionally, a novel analysis tool, single-trial ERP image, provided important information about brain-behavior relationships. Using this tool, stimulus-locked, semantic-processing related, and behavioral response-related brain activity patterns were identified. A slow positive deflection elicited by expected meaningful stimuli has been reported in semantic tasks. Single-trial ERP image analysis allowed us to decompose this positivity into three functionally distinct subcomponents: the fronto-central sensory P2, the fronto-central semantic PS, and the centro-parietal response-associated PR. Based on their stimulus-locked timing and the RT-related magnitude, the PS and the overlapping incongruent-items’ N400 appear to activate pre-established, automatically accessible semantic representations. Finally, the PR had a strong relationship with subjects’ response times, possibly indexing decision-making processes.

5. Experimental procedure

5.1. Participants

Fourteen undergraduate subjects (7 male, mean age=22, range 19–35) completed the Verb Experiment and twelve undergraduate subjects (6 male, mean age=22.6, range 18–33) completed the Noun Experiment. All participants were right-handed native speakers of American English. All subjects signed informed consent in accordance with the UCSD Human Research Protections Program.

5.2. Study design

This study employed a picture-sound matching design to assess electrophysiological brain activity related to semantic integration as a function of auditory input type (Fig. 6). Subjects were presented with picture-sound pairs, with sounds beginning 600 ms after picture onset and both stimuli ending together. Colorful pictures of objects were followed by either a word or an environmental sound, whereas colorful pictures of non-meaningful patterns were followed by non-meaningful complex sounds (see below for details). Such a trial design allowed recording of clean visual ERPs to the pictures, left little time for conscious labeling of the pictures, and avoided a working memory load.

The study used a 2-within (Sound Type, Picture Context) × 1-between-subjects (Word Class) design. Subjects were randomly assigned to one of the two Word Class Experiments (Noun, Verb), which varied only in the grammatical category of the word paired with the pictures. All subjects heard three sound types (words, environmental sounds, and non-meaningful sounds), presented in semantically matching or mismatching Picture Contexts.

5.3. Stimuli

All auditory stimuli were digitized at 44.1 kHz with a 16-bit sampling rate. The average intensity of all auditory stimuli was normalized to 65 dB.

5.4. Environmental sounds

The environmental sounds used in this study were derived from the Saygin et al. (2005) study but were shortened to less than 1000 ms, while still maintaining their identifiability. The

54 sounds were drawn from several different semantic classes: animal cries ($n=15$, e.g., cow mooing), human non-verbal vocalizations ($n=3$, e.g., sneezing), machine noises ($n=8$, e.g., car honking), alarms/alerts ($n=7$, e.g., phone ringing), water sounds ($n=1$, toilet flushing), event sounds ($n=12$, e.g., bubbles bubbling), and music ($n=9$, e.g., piano playing). All non-verbal sounds ranged in duration from 400 to 870 ms (mean = 574 ms, SD = 104 ms).

5.5. Words

Words were chosen based on a norming study. During norming, two subject groups ($n=10$ in each) were presented with the environmental sounds and asked to provide a single word label for each sound (see Appendix A for stimuli list). One group was asked to label the sounds with nouns, the other with verbs. The noun label typically corresponded to the object that created the sound whereas the verb label typically corresponded to the object-related action. Subjects' responses were ranked on a 3 point scale: 2 = correctly identified, 1 = possible answer, but incorrect, and 0 = unknown or impossible answer. Subjects were more consistent in their noun responses, as compared to the verb responses, but the most frequent responses from both word classes were chosen to be the word labels that were the lexical correlates of the environmental sounds. Additionally, during the norming experiment, all subjects correctly identified the 54 environmental sounds chosen for the study.

In order to provide some variation in pitch and timbre to the word stimuli, they were pronounced by three North American speakers (one female and two male). The words were digitally recorded in a sound isolated room (Industrial Acoustics Company, Inc., Winchester, UK) using a Beyer Dynamic (Heilbronn, Germany) Soundstar Mk II Unidirectional

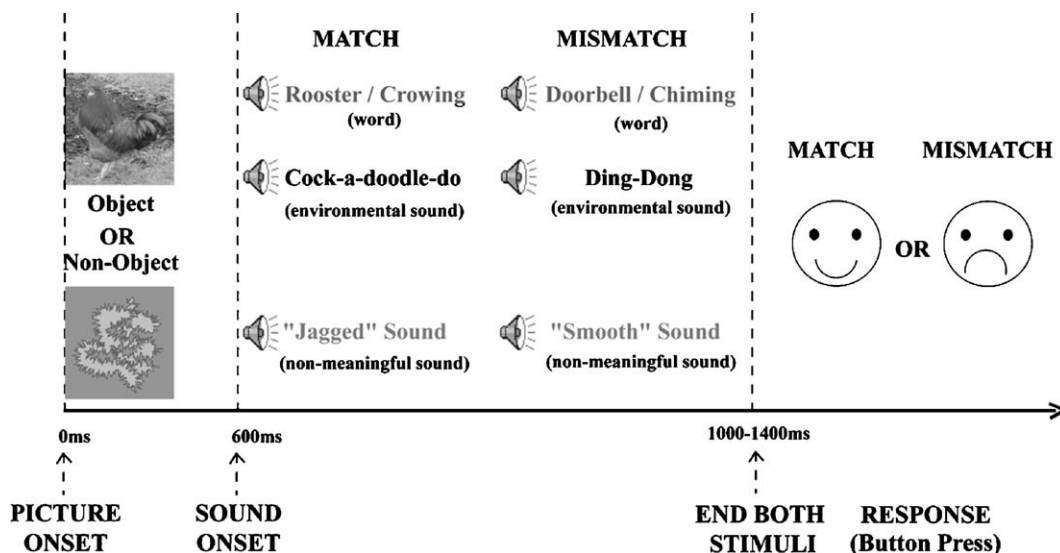


Fig. 6 – Pictorial representation of each trial presented in the experiment. Either an object or non-object picture was presented on a computer screen, then a word, environmental sound, or non-meaningful sound was presented via loudspeakers. The auditory stimulus either matched or mismatched the visual stimulus and subjects recorded their responses with a button press. The happy face represented a match trial response and the sad face represented a mismatch trial response. Half of the subjects had the match button on the right side of the button box, and the other half had it on the left side.

Dynamic microphone and Behringer (Willich, Germany) Euro-rack MX602A mixer. Noun stimuli ranged in duration from 262 to 940 ms (mean=466 ms, SD=136 ms) and verb stimuli ranged from 395 to 1154 ms (mean=567 ms, SD=158 ms).

5.6. Non-meaningful sounds

The 54 non-meaningful sounds were downloaded from a sound effect Web site (www.sounddog.com). They were matched in duration to the environmental sounds and consisted of computer-generated sounds that were not easily associated with any concrete semantic concept. In order to permit picture-sound matching, the non-meaningful sounds were chosen to portray either a 'smooth' sound, e.g., a harmonic tone, or a 'jagged' sound, e.g., cracking noise. Unlike the meaningful stimuli these were not extensively normed because we considered the non-meaningful sound and picture matchings "subjective". However, five raters came to a consensus about each sound fulfilling the 'smooth' or 'jagged' sound criteria.

5.7. Visual stimuli

Pictures paired with the meaningful auditory stimuli (environmental sounds and words) were full-color, digitized photos (480×480 pixels) of common action-related objects that could produce an environmental sound, and be described by a verb or a noun. Pictures paired with non-meaningful auditory stimuli were various colorful, non-object looking patterns chosen to represent one of two categories—either a smooth-looking pattern (e.g., wavy), or a jagged-looking pattern (e.g., sharp angles and broken lines), to go along with the 'smooth' and 'jagged' non-meaningful sounds. All pictures were presented in the middle of the computer monitor on a gray background and subtended 6.9° visual angle. The same set of pictures was used for both the matching and mismatching picture contexts. The only constraint on the semantic mismatching trials was that the mismatch had to be unambiguous (e.g., the picture of a basketball was NOT presented with the sound of hitting a golf ball).

5.8. Procedure

The stimuli were presented in blocks of 54 trials. Six stimulus blocks (324 trials altogether) were presented to every subject. Pictures and sounds were delivered using Presentation Version 9.70 software (www.neuro-bs.com, Neurobehavioral Systems, Albany, CA, USA). All pictures were presented with equal probabilities, in a random order without replacement. Pictures were presented on a computer screen situated 120 cm in front of the subject and sounds were played via two loudspeakers situated 30° to the right and left from the midline, as well as in front of the participant. Sounds were perceived to originate from the midline space. The subjects' task was to press one button (marked by a smiley face) if they thought the picture and auditory stimuli matched, and to press another button (marked by a sad face) if they thought that the stimuli mismatched. Subjects were instructed to hold the button box with both hands and use one thumb for each

button, i.e., both hands were active during the experiment. The sides of the match and mismatch buttons were counter-balanced across subjects.

5.9. Behavioral data analysis

Accuracy on a given trial was determined by a subject's button press response to the Picture-Sound pair. Whereas there were clear correct answers for the environmental sound and word trials, the non-meaningful sound and picture pairings were subjective. The non-meaningful stimulus trials were sorted into correct and incorrect based the subjects' subjective judgment, i.e., whether they had thought the sound had matched or mismatched the picture. By dividing up the data in this manner, we hoped to keep the all the "matching" brain processing response patterns in one group and all the "mismatching" brain processing responses in a different group. Nonetheless, in order to ensure that the subjects were effectively comparing the non-meaningful pictures and sounds, we looked for biases in their non-meaningful trial responses. Two biased subjects, who produced 0% "match" responses were excluded from analyses that included the non-meaningful sounds. The remainder of the subjects had a mean average of 74% of matching responses which ensured that they (i) did the task and (ii) their perception of these stimuli was similar to what was intended.

Reaction time was calculated from the onset of the auditory stimulus to the computer's registering of the button press. Because there was a wide variation in reaction times across and within subjects, median (as opposed to mean) reaction times were used in the data analyses.

Responses of subjects from both Word Class experiments were pooled together when word and environmental sounds stimuli were compared, whereas subjects' responses were divided into noun and verb bins for analyses of Word Class.

5.10. EEG recording and averaging

Continuous EEG was recorded using a 32-electrode cap (Electrocap, Inc., Eaton, OH, USA) with the following electrodes attached to the scalp, according to the International 10–20 system: FP1, FP2, F7, F8, FC6, FC5, F3, Fz, F4, TC3, TC4 (30% distance from T3–C3 and T4–C4, respectively), FC1, FC2, C3, Cz, C4, PT3, PT4 (halfway between P3–T3 and P4–T4, respectively), T5, T6, CP1, CP2, P3, Pz, P4, PO3, PO4, O1, O2, and right mastoid. Eye movements were monitored with two electrodes, one attached below the left eye and another at the corner of the right eye. During data acquisition, all channels were referenced to the left mastoid; offline, data were re-referenced to the averaged of the left- and right-mastoid tracings.

The EEG (0.01–100 Hz) was amplified 20,000× and digitized at 250 Hz for the off-line analyses. Prior to averaging, an independent-component analysis (Jung et al., 2000a,b) was used to correct for eye blinks and lateral movements. Sporadic artifacts (e.g., random, infrequent head or facial movements, but not the consistent artifacts such as blinks) were manually rejected from the data prior to ICA training in order to obtain

“clean” ICA components, including those accounting for blinks and lateral eye movements. These components were then removed from the data. This conserved up to 30% of the trials that would have been discarded had the traditional artifact rejection approach been used. Trials with remaining sporadic artifacts were rejected using voltage thresholds (Picton et al., 2000). Epochs containing 100 ms pre-stimulus and 900 ms post-stimulus time were baseline-corrected with respect to the pre-stimulus interval and averaged by stimulus type. A low-pass Gaussian filter was used to remove frequencies higher than 60 Hz. On average, the remaining individual data contained 102 word, 104 environmental sound, and 100 non-meaningful sound trials. ERP responses of the matching trials were subtracted from the ERP responses of the mismatching trials to create the difference waves for each sound type (e.g., Picture–Word Mismatch minus Picture–Word Match).

5.11. ERP measurements

Peak measurement was a multi-step process. Two negative peaks were seen in the grand average difference waveforms: one maximal centro-parietally and peaking at ca. 350 ms (the N400), and another maximal frontally and peaking at ca. 600 ms (the N600; Fig. 2). The N400 search window was determined by visual inspection of the grand-average waveforms. The grand-average N400 was measured from 200 to 500 ms at the 25 electrodes, where it was present in the grand-average waveforms: F3/F4, Fz, FC5/FC6, FC1/FC2, L41/R41, C3/C4, Cz, WL/WR, CP1/CP2, T5/T6, P3/P4, Pz, PO3/PO4, and O1/O2. Peak latencies were then averaged across the electrodes to determine the “center” latency of the N400. The “center” latency was then used to center a 50-ms window (25 ms on either side) to find the mean amplitudes for all electrodes and sound types. For the N400, the latency windows in the Verb Experiment class (words=375–425 ms, sounds=295–345 ms) were similar to those in the Noun Experiment (words=355–405, sounds=325–375 ms). The non-meaningful sounds did not elicit an obvious N400 peak; therefore, the non-meaningful sound difference wave was measured from the latency window decided by the environmental sound peak latency. The N600 search window was also determined by visual inspection of the grand-average waveforms. Because the N600 was a slow wave, the middle of the wave in the grand average (600 ms) was chosen as the “center” latency. As with the N400, a 50-ms window was centered around 600 ms for all electrodes and sound types (i.e., 575–625 ms for words, environmental sounds, and non-meaningful sounds).

Two-tailed independent-sample *t* tests were conducted to test the significance of the ERP peaks elicited by the three sound types on selected electrode sites, using appropriate alpha level corrections (Table 2). Within-group ANOVA analyses were performed using fifteen electrodes where there was a significant N400 effect: F3/F4, Fz, FC1/FC2, C3/C4, Cz, CP1/CP2, P3/P4, Pz, and PO3/PO4. When appropriate, Geiser–Greenhouse-corrected *p* values were used.

Scalp distribution analyses were completed to search for both laterality and anterior/posterior differences between the word and environmental sound stimuli. Mean amplitudes

from twelve electrodes: F3/F4, FC1/FC2, C3/C4, CP1/CP2, P3/P4, and PO3/PO4 were normalized using a *z* score technique, calculated separately for nouns, verbs, and environmental sounds (Picton et al., 2000). To analyze the relationship between components (N400 vs. N600) and sound type, an ANOVA analysis with Sound Type, Component, and Electrode (12 levels) was conducted on normalized N400 and N600 data.

5.12. Single-trial ERP imaging

It has been shown that traditional ERP trial averaging may mask the relationship between an ERP or electroencephalogram (EEG) components and a behavioral assay, such as reaction time (Jung et al., 2001) and thus may conceal aspects of the brain-behavior dynamics unique to semantic processing of words or environmental sounds. To assess possible relationships between electrophysiological brain phenomena involved in semantic processing and behavior, word and environmental sound (but not non-meaningful sound) single EEG trials were examined. The non-meaningful sounds were not included in the ERP imaging analysis because they did not require semantic integration as the meaningful sound stimuli did. Single-trial EEG epochs containing 0.15 s pre-stimulus to 1.4 s post-stimulus interval were extracted from sporadic-artifact pruned and ICA-cleaned continuous data. Because sound onset occurred during the time when the picture was still present and thus overlapped with visual ERP, these epochs were baseline corrected using 100-ms interval preceding the onset of a picture. Group ERP images were constructed by averaging, across subjects, single trials in order with their corresponding RTs. For example, the first single trial in a grand-average was obtained by averaging 24 epochs, one from each subject, with the shortest correct RT of each subject (two subjects could not be used for ERP imaging due to the stimulus coding differences). The second trial contained 24 epochs with the second-shortest RT values of each subject, etc. In this way, the reaction time–brain activity relationship for each trial was preserved for all the subjects. A fair signal quality for each epoch was also achieved, allowing us to measure the peaks of interest. For matching trials, at least 79 epochs were available in each subject, and for mismatching trials 84 epochs from each subject were used. In ERP image plots, the polarity of brain activity was color-coded (blue for negative and red for positive). The resulting 79 or 84 epochs were sorted by (1) subjects’ reaction times, (2) brain activity magnitude during the N400 latency range (280–380 ms), and (3) sound length (Fig. 5). For peak latency–RT correlations, the RT-sorted EEG epochs were measured. For peak amplitude–RT correlations, the magnitude-sorted EEG epochs were measured. The epochs were measured at the Fz and Cz electrodes, where the brain activity was the most representative of matching and mismatching trials, respectively. The N400 and N600 were measured as the most negative peaks from 300–500 and 500–800 ms windows, respectively; the PS and PR were measured as the most positive peaks from 280–380 and 450–800 ms, respectively. For all peaks, the mean amplitudes were integrated over 40 ms interval centered at the peak latency.

Acknowledgments

We are grateful to Mr. C. Williams and Ms. M. Evans for help with data collection and analysis and to Dr. M. Westerfield for her guidance and help with the ICA analyses. Our subjects were critical for the study. This study was supported by the NIH grant 5 P50 NINDS NS22343-18. The first author was supported by NIH training grants #DC00041 and #DC007361.

Appendix A

Noun label	Verb label	Noun label	Verb label
Alarm clock	Ringling	Frog	Croaking
Baby	Crying	Glass	Breaking
Baby	Laughing	Golf	Hitting
Ball	Bouncing	Helicopter	Whirling
Ball	Hitting	Horse	Galloping
Banjo	Strumming	Ice	Clinking
Bee	Buzzing	Jackhammer	Jackhammering
Bell	Dinging	Lion	Roaring
Bell	Ringling	Lips	Kissing
Bird	Chirping	Owl	Hooting
Bubbles	Bubbling	Phone	Ringling
Camera	taking a picture	Phone	Dialing
Car	starting a car	Piano	Playing the piano
Car	Honking	Rooster	Crowing
Cat	Meowing	Saw	Sawing
Chicken	Clucking	Saxophone	Playing the saxophone
Coins	Dropping	Seal	Barking
Cow	Mooring	Sheep	Baaing
Crow	Cawing	Siren	Siren
Cuckoo clock	Cuckooing	Soda	Pouring
Cymbals	Hitting	Trumpet	Playing the trumpet
Dog	Barking	Violin	Playing the violin
Doorbell	Ringling	Water	Dripping
Drill	Drilling	Whip	Whipping
Drum	Drumming	Whistle	Whistling
Duck	Quacking	Wind chimes	Chiming
Flute	Playing the flute	Zipper	Zipping

REFERENCES

- Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., Mazziotta, J., 2004. Left hemisphere motor facilitation in response to manual action sounds. *Eur. J. Neurosci.* 19, 2609–2612.
- Ballas, J.A., 1993. Common factors in the identification of an assortment of brief everyday sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 19 (2), 250–267.
- Ballas, J.A., Howard, J., 1987. Interpreting the language of environmental sounds. *Environment and Behavior* 19 (1), 91–114.
- Boschert, J., Deecke, L., 1986. Cerebral potentials preceding voluntary toe, knee and hip movements and their vectors in human precentral gyrus. *Brain Res.* 376, 175–179.
- Boschert, J., Brickett, P., Weinberg, H., Deecke, L., 1983. Movement-related potentials preceding toe plantarflexion and dorsiflexion. *Human Neurobiology* 2, 87–90.
- Brandeis, D., Lehmann, D., Michel, C., Mingrone, W., 1995. Mapping event-related brain potential microstates to sentence endings. *Brain Topogr.* 8 (2), 145–159.
- Ceponiene, R., Alku, P., Westerfield, M., Torki, M., Townsend, J., 2005. ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology* 42 (4), 391–406.
- Chiu, C.Y., Schacter, D.L., 1995. Auditory priming for nonverbal information: implicit and explicit memory for environmental sounds. *Conscious Cogn* 4 (4), 440–458.
- Cree, G.S., McRae, K., 2003. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.* 132, 163–201.
- Cycowicz, Y.M., Friedman, D., 1998. Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain Cogn.* 36 (1), 30–51.
- Dehaene, S., 1995. Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport* 6 (16), 2153–2157.
- Dick, F., Saygin, A.P., Paulsen, C., Trauner, D., Bates, E., 2004. The co-development of environmental sounds and language comprehension in school-age children. Poster presented at the meeting for Attention and Performance, Colorado.
- Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495.
- Federmeier, K.D., Segal, J.B., Lombrozo, T., Kutas, M., 2000. Brain responses to nouns, verbs and class-ambiguous words in context. *Brain* 123 (Pt. 12), 2552–2566.
- Friedman, D., Cycowicz, Y., Dziobek, I., 2003. Cross-form conceptual relations between sounds and words: effects on the novelty P3. *Cogn. Brain Res.* 18 (1), 58–64.
- Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res.* 16 (2), 123–144.
- Gygi, B., 2001. Factors in the identification of environmental sounds. Unpublished Doctoral Dissertation. Department of Psychology and Cognitive Science, Bloomington, Indiana University. PhD: 187.
- Halgren, E., Baudena, P., Heit, G., Clarke, J.M., Marinkovic, K., Chauvel, P., Clarke, M., 1994a. Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices. *J. Physiol. (Paris)* 88 (1), 51–80.
- Halgren, E., Baudena, P., Heit, G., Clarke, J.M., Marinkovic, K., Clarke, M., 1994b. Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes [corrected]. *J. Physiol. (Paris)* 88 (1), 1–50.
- Halgren, E., Dhond, R.P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J.D., Dale, A.M., 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage* 17 (3), 1101–1116.
- Halgren, E., Wang, C., Schomer, D., Knake, S., Marinkovic, K., Wu, J., Ulbert, I., in press. Processing stages underlying word recognition in anteroventral temporal lobe. *NeuroImage*.
- Hamm, J.P., Johnson, B.W., Kirk, I.J., 2002. Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clin. Neurophysiol.* 113 (8), 1339–1350.
- Heinze, H.J., Munte, T.F., Kutas, M., 1998. Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biol. Psychol.* 47 (2), 121–135.
- Itoh, K., Suwazono, S., Arao, H., Miyazaki, K., Nakada, T., 2005. Electrophysiological correlates of absolute pitch and relative pitch. *Cereb. Cortex* 15 (6), 760–769.

- Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Sejnowski, T.J., 2000a. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37 (2), 163–178.
- Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2000b. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111 (10), 1745–1758.
- Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2001. Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14 (3), 166–185.
- King, J., Kutas, M., 1995. Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* 7 (3), 376–395.
- Koelsch, S., Gunter, T., Schroger, E., Friederici, A., 2003. Processing tonal modulations: an ERP study. *J. Cogn. Neurosci.* 15 (8), 1149–1159.
- Koivisto, M., Revonsuo, A., 2001. Cognitive representations underlying the N400 priming effect. *Brain Res. Cogn. Brain Res.* 12 (3), 487–490.
- Kushnerenko, E., Ceponiene, R., Fellman, V., Huotilainen, M., Winkler, I., 2001. Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *NeuroReport* 12 (17), 3777–3781.
- Kutas, M., Hillyard, S.A., 1980a. Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol. Psychol.* 11 (2), 99–116.
- Kutas, M., Hillyard, S.A., 1980b. Reading between the lines: event-related brain potentials during natural sentence processing. *Brain Lang.* 11 (2), 354–373.
- Kutas, M., Hillyard, S.A., 1982. The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia* 20 (5), 579–590.
- Kutas, M., Hillyard, S.A., 1983. Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11 (5), 539–550.
- Kutas, M., Iragui, V., 1998. The N400 in a semantic categorization task across 6 decades. *Electroencephalogr Clin Neurophysiol* 108 (5), 456–471.
- Kutas, M., van Petten, C., 1994. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher, M.A. (Ed.), *Handbook of Psycholinguistics*. Academic Press, pp. 83–143.
- Kutas, M., Van Petten, C., Besson, M., 1988. Event-related potential asymmetries during the reading of sentences. *Electroencephalogr. Clin. Neurophysiol.* 69 (3), 218–233.
- Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Janik, J.J., DeYoe, E., 2005. Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* 25 (21), 5148–5158.
- Makeig, S., Westerfield, M., Jung, T.P., Covington, J., Townsend, J., Sejnowski, T.J., Courchesne, E., 1999. Functionally independent components of the late positive event-related potential during visual spatial attention. *J. Neurosci.* 19 (7), 2665–2680.
- McCarthy, G., Wood, C.C., 1985. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr. Clin. Neurophysiol.* 62, 203–208.
- Nunez-Pena, M.I., Honrubia-Serrano, M.L., 2005. N400 and category exemplar associative strength. *Int. J. Psychophysiol.* 56 (1), 45–54.
- Perrin, F., Garcia-Larrea, L., 2003. Modulation of the N400 potential during auditory phonological/semantic interaction. *Brain Res. Cogn. Brain Res.* 17 (1), 36–47.
- Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson Jr., R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37 (2), 127–152.
- Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., Di Russo, F., 2005. Separate neural systems for processing action- or non-action-related sounds. *NeuroImage* 24 (3), 852–861.
- Plante, E., van Petten, C., Senkfor, A.J., 2000. Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia* 38 (13), 1669–1684.
- Polich, J., 1985. Semantic categorization and event-related potentials. *Brain Lang.* 26 (2), 304–321.
- Pulvermuller, F., 1996. Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology* 33 (4), 317–333.
- Pulvermuller, F., 1999. Words in the brain's language. *Behav. Brain Sci.* 22 (2), 253–279 discussion 280–336.
- Rudell, A.P., 1991. The recognition potential contrasted with the P300. *Int. J. Neurosci.* 60 (1–2), 85–111.
- Rudell, A.P., Hua, J., 1995. Recognition potential latency and word image degradation. *Brain Lang.* 51 (2), 229–241.
- Rudell, A.P., Hua, J., 1996. The recognition potential and word priming. *Int. J. Neurosci.* 87 (3–4), 225–240.
- Rudell, A.P., Hua, J., 1997. The recognition potential, word difficulty, and individual reading ability: on using event-related potentials to study perception. *J. Exp. Psychol. Hum. Percept. Perform.* 23 (4), 1170–1195.
- Rudell, A.P., Cracco, R., Hassan, N., Eberle, L., 1993. Recognition potential: sensitivity to visual field stimulated. *Electroencephalogr. Clin. Neurophysiol.* 87 (4), 221–234.
- Saygin, A.P., Dick, F., Wilson, S.M., Dronkers, N.F., Bates, E., 2003. Neural resources for processing language and environmental sounds: evidence from aphasia. *Brain* 126 (Pt. 4), 928–945.
- Saygin, A.P., Dick, F., Bates, E., 2005. An on-line task for contrasting auditory processing in the verbal and domains and norms for younger and older adults. *Behav. Res. Methods* 37 (1), 99–110.
- Shafiro, V., Gygi, B., 2004. How to select stimuli for environmental sound research and where to find them. *Behav. Res. Methods Instrum. Comput. Special Issue: Web-Based Archive of Norms, Stimuli, and Data: Part 2* 36 (4), 590–598.
- Stuart, G.P., Jones, D.M., 1995. Priming the identification of environmental sounds. *Q. J. Exp. Psychol.* A 48 (3), 741–761.
- Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arevalo, A.L., Vargha, A., Bates, E., 2005. Timed action and object naming. *Cortex* 41 (1), 7–25.
- Thierry, G., Giraud, A.L., Price, C., 2003. Hemispheric dissociation in access to the human semantic system. *Neuron* (38), 499–506.
- Van Petten, C., Rheinfelder, H., 1995. Conceptual relationships between spoken words and environmental sounds: event-related brain potential measures. *Neuropsychologia* 33 (4), 485–508.
- Wang, Y., Cui, L., Wang, H., Tian, S., Zhang, X., 2004. The sequential processing of visual feature conjunction mismatches in the human brain. *Psychophysiology* 41 (1), 21–29.