Brain and Language 179 (2018) 51–61

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Hearing sounds as words: Neural responses to environmental sounds in the context of fluent speech

Sophia Uddin*, Shannon L.M. Heald, Stephen C. Van Hedger, Howard C. Nusbaum

Department of Psychology, The University of Chicago, 5848 S. University Ave., Chicago, IL 60637, United States

ARTICLE INFO

Keywords:
Environmental sounds
Language processing
Event-related potential
N400
Sentence understanding
Context

ABSTRACT

Environmental sounds (ES) can be understood easily when substituted for words in sentences, suggesting that linguistic context benefits may be mediated by processes more general than some language-specific theories assert. However, the underlying neural processing is not understood. EEG was recorded for spoken sentences ending in either a spoken word or a corresponding ES. Endings were either congruent or incongruent with the sentence frame, and thus were expected to produce N400 activity. However, if ES and word meanings are combined with language context by different mechanisms, different N400 responses would be expected. Incongruent endings (both words and ES) elicited frontocentral negativities corresponding to the N400 typically observed to incongruent spoken words. Moreover, sentential constraint had similar effects on N400 topographies to ES and words. Comparison of speech and ES responses suggests that understanding meaning in speech context may be mediated by similar neural mechanisms for these two types of stimuli.

1. Introduction

The question of whether speech understanding is mediated by a specialized neural system (e.g., Grodzinsky, 2000; Liberman & Mattingly, 1985) or more general neural mechanisms (Christiansen, Allen, & Seidenberg, 1998; Dick et al., 2001; Kleinschmidt & Jaeger, 2015; Leech, Holt, Devlin, & Dick, 2009) is a longstanding theoretical issue. This debate has often focused on the characteristics of language that set it apart from other kinds of information (e.g., Chomsky, 1986; Fodor, 1983), but more generally, it addresses age-old questions about the balance between specialization and modularity on the one hand, and distributed processes and domain-general mechanisms on the other hand.

Environmental sounds (ES) are auditory patterns that are meaningful, but not "linguistic": they lack internal phonological segments or higher-order linguistic structure, and in most cases are not produced by the human vocal tract. They are, however, easily recognized and categorized (Gygi, Kidd, & Watson, 2007; Warren & Verbrugge, 1984), and can be combined with each other or with words in order to form meaningful concepts (Ballas & Mullins, 1991). If speech perception is carried out by a separate dedicated processing mechanism, any similarity between understanding ES and spoken words would be due to chance and should not be systematic, whereas if perception and comprehension of spoken sentences is mediated by general auditory and cognitive processing, there should be substantial overlap between these processes.

An important characteristic of human language is the facilitative effect of context; it is well known that constraining contexts speed processes such as word recognition and sentence completion (Morris & Harris, 2002; Staub, Grant, Astheimer, & Cohen, 2015). Therefore, one way to address whether there are similarities between the processing of nonlinguistic stimuli and words is to ask whether we can understand a sentence that substitutes an ES for a spoken word. In doing so, we are asking whether the processes for understanding an item in light of its preceding context differ substantially between these types of stimuli. Readers easily understand "rebus" sentences in which a picture replaces a printed word (Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986), but it is possible that speech perception, as a more basic system than reading in human development (cf. Dehaene, 2011), might operate as a separate modular system (cf. Fodor, 1983). We have previously compared behavioral measures of perception of spoken sentences that end in either a word or an ES. In a gating paradigm (see Grosjean, 1980) constraining sentence frames (e.g., "he bought diapers for his __") reduced the duration of signal needed for recognition compared to general frames (e.g., "his back hurt from holding the __") similarly for words and ES. Further, response times for congruency judgments were similar for sentences ending in words and ES (Uddin, Heald, Van Hedger, Klos, & Nusbaum, 2018). While nonspeech stimuli can be understood, even substituted for words, in a spoken linguistic frame, similar patterns of behavior do not unequivocally indicate similar

* Corresponding author at: 5848 S. University Ave., Green 302, Chicago, IL 60637, United States.
E-mail address: sophiauddin@uchicago.edu (S. Uddin).

https://doi.org/10.1016/j.bandl.2018.02.004
Received 24 October 2017; Received in revised form 12 February 2018; Accepted 20 February 2018
Available online 06 March 2018
0093-934X/

underlying neural processing (cf. Reuter-Lorenz, 2002). Thus it is important to assess whether neural responses differ between ES and words when they are understood in spoken sentence contexts.

Of course, it is important to note that neural responses to ES and words might differ for reasons unrelated to their interaction with context. There are substantial acoustic differences between environmental and speech sounds (e.g., Lewicki, 2002). ES can be derived from a wide variety of sources, and can range from man-made sounds such as machinery, to nature-related sounds such as water rushing or an animal vocalizing. Therefore, ES have a much wider variety of sources and acoustic features than words pronounced by a single speaker (Lewicki, 2002). Moreover, fMRI studies show that ES recruit different cortical areas than speech (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Lewis, 2005; Vouloumanos, Kiehl, Werker, & Liddle, 2001), although more recent work suggests that cortical representations of environmental sounds and speech overlap substantially (Dick et al., 2001; Leech & Saygin, 2011; Leech et al., 2009). Therefore, while we expect differences in neural responses based on acoustic pattern and stimulus frequency differences between ES and spoken words, the important question is whether there are differences that reflect fundamentally different processes for understanding these stimuli in sentence context.

In processing an utterance's meaning, the N400 is a negative-going ERP in human EEG that can arise from a mismatch of a word with preceding context (Kutas & Hillyard, 1980a). Semantically incongruous words elicit larger negativities approximately 400 ms after word onset. Kutas and Hillyard postulated that this comes from neural processes related to sentence meaning repair, or reprocessing of the unexpected word in contextual integration. The N400's sensitivity to expectation violations can be used to investigate neural responses to ES in spoken sentence context.

Stimuli do not have to be linguistic to elicit an N400; they can also be pictures, environmental sounds, or other meaningful items (Kutas & Federmeier, 2011). While it is known that environmental sound probes presented after written word, picture, or spoken word primes elicit more negative N400s when probes and primes are incongruent (Cummings et al., 2006; Orgs, Lange, Dombrowski, & Heil, 2006; van Petten & Rheinfelder, 1995), this has never been tested in the context of a fluent speech sentence frame. This is an important distinction, because in previous experiments, primes and probes are single words or ES separated by a period of silence. Such isolated words or sounds stand alone as concepts. The present experiment, in contrast, presents either words or ES as the final item in a continuously presented sentence frame which is related in its meaning to a concluding final noun, if ending in speech. Before the final item is presented, it is not apparent whether it will be congruent or incongruent with the meaning of the antecedent sentence frame. Therefore, given a spoken sentence fragment, listeners continuously incorporate the meaning of each word with previous words in order to understand the sentence context before the final item arrives, at which point this final item is understood in this context. If the final item is a spoken word, listeners will certainly recognize and understand this word in the linguistic context established by the sentence frame. The question is what happens when the final item is not a word but is an environmental sound. Due to the continuous nature of the sentence stimuli in this paradigm, if there are processing costs, i.e. slower processing or obligatory extra processing steps, related to ES being more difficult to understand than words in context, such costs should be more apparent than in a paradigm where isolated primes and probes are presented several hundred milliseconds apart from each other. This is because (1) extra processing, or delays in processing, could manifest in the time between the presentation of primes and probes, and (2) a spoken sentence requires continuous attention and interpretation which will likely tax mechanisms for understanding beyond a prime/probe pair.

Though N400s to non-linguistic stimuli are routinely found using prime/probe type designs, it is possible that understanding ES in a fluent spoken sentence might draw on different neural mechanisms. While our behavioral work suggests against this possibility (Uddin et al., 2018), if this is true we might not expect any reliable N400 effects for ES in our experiment. On the other hand, ES in such a context may always be processed as incongruent because they are so different from speech. In this case, we would expect N400s to all ES regardless of the relationship between the meaning of the sound and the meaning of the preceding sentence frame.

If ES and spoken words are assigned to meaning, and that meaning is combined with preceding context in a similar fashion, we might expect a generally similar pattern of N400 results—higher-amplitude for sentence-frame incongruent final items—for both ES and word targets. However, it is also possible that ES are recognized and understood in these sentence frames similarly to words, but are more difficult to process in this context. While measured response times do not support this, as they are not substantially slower for ES (Uddin et al., 2018), increased processing cost could yield a delayed N400 without slowing response times (e.g., Ardal, Donald, Meuter, Muldrew, & Luce, 1990).

Finally, there is ample evidence that the constraint level of a sentence affects understanding of words in that sentence (e.g., Staub et al., 2015). These constraint effects extend to the N400: for words congruent with the preceding context, the N400 is larger in amplitude for low-constraint sentences (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas & Hillyard, 1984). However, for words incongruent with or highly unexpected in the preceding context, the N400 does not appear to depend on constraint level (Federmeier et al., 2007; Kutas & Federmeier, 2011). While obvious factors such as congruence with context may lead to similar neural responses for ES and words, it is possible that constraint information is a finer nuance of sentence processing that will only affect words. There is at least some precedence for this idea: Hendrickson, Walenski, Friend, & Love (2015) examined N400s to ES or spoken words presented after a picture. The ES or words were either matches (i.e., congruent with the preceding picture), near violations (incongruent but conceptually related), or far violations (incongruent and conceptually distant). For words, voltage in the 300–400 ms post-onset time window was graded based on degree of congruence with the picture, but for ES, near violations and matches were statistically indistinguishable. It is possible that in a fluent sentence paradigm, where ES must be rapidly understood in an unfolding spoken sentence, such differences between words and ES could be magnified. In our study, an interaction of sentence ending type (ES or word) with constraint would suggest that constraint differentially affects the interaction of words and ES with the context supplied by a full spoken sentence.

2. Methods

2.1. Participants

Participants were 23 (8 female, 13 male, 1 agender, 1 genderfluid) adults from the University of Chicago and surrounding community. Their mean age was 22.1 years (SD: 3.7, range: 18–29). Fifteen were right-handed and eight were left-handed. Participants completed questionnaires to ensure that they knew English to native proficiency, and that they were not taking medications that could interfere with cognitive or neurological function. Participants chose between 3 course credits (n = 11) or $30 cash (n = 12) for their participation. A power analysis of our behavioral data (Uddin et al., 2018) suggested that we needed a sample size of 16 participants to achieve a power level of 0.95; therefore, for this experiment, we rounded up to a goal of at least 20 participants.

2.2. Stimuli

Stimuli are available on Open Science Framework (https://osf.io/asw48/) and consisted of spoken sentence stems and acoustic endings ("targets") that were either spoken words or environmental sounds.

Sentence stems and separate ending words were recorded at 44.1 kHz by an adult male native speaker of Midwestern English. Stems and ending words were recorded separately to avoid co-articulation confounds. There were two levels of sentence frame constraint: half were "specific" (high cloze probability for match ending, median = 0.87, IQR = 0.25) and half were "general" (low cloze probability for match ending, median = 0.16, IQR = 0.33). Cloze probability was determined based on written sentence completions from 66 Amazon Mechanical Turk participants.

Each ending word was a noun that could also be represented by a matched environmental sound (e.g. "sheep" matched with the sound of a sheep vocalization; see Appendix Table A1). The environmental sounds were taken from online databases (e.g., soundbible.com), and if necessary were resampled to 44.1 kHz. They were amplitude normalized to the same RMS level (about 70 dB SPL, comfortable listening level) as the stems and spoken word targets. To ensure that the sounds were identifiable when heard alone, a small norming study of students in the department was conducted. Mean duration of spoken word targets was 0.502 s; mean duration of environmental sound targets was 0.838 s. Eight of the 32 environmental sounds involved repetition (e.g., the sound of a siren involves repeating pitch oscillations).

Sentence stems and target endings (nonspeech or speech) were spliced together in Matlab to form complete sentences. Waveforms were directly joined with zero silence between stem and target but no discernible clicks or acoustic artifacts were heard, and the speech flowed smoothly into the target sounds. Half the resulting sentences terminated in spoken word targets, and half terminated in matched environmental sound targets. In addition, mismatch (i.e. semantically incongruous) sentences were constructed by rearranging the targets and context sentence stems to mismatch in meaning. To generate these, the targets and stems were shuffled and the resulting sentences were verified in a short written survey to ensure that they were not easily construed to make sense. Thus, the "meaningful" i.e. congruent nature of the sentence depended on the last word of the sentence, which was replaced by an environmental sound for half the stimuli. This congruency (matched vs. mismatched) by target type (speech target vs. nonspeech target) by constraint (general vs. specific) design gave rise to eight types of sentence stimuli. Sentences were blocked by target type, such that there were four blocks of sentences ending in sounds, and four blocks of sentences ending in words. Block types alternated across the experiment, and the type of starting block (i.e. sound or word) was counterbalanced across subjects. Within each block, match and mismatch sentences were pseudo-randomly presented such that half were matches and half were mismatches. The sentences within each block were similarly divided and randomized between general and specific. The design was balanced such that each word and sound appeared in match and mismatch conditions an equal number of times. Stimuli were experienced at 65–70 dB over insert earphones (3M E-A-RTone Gold) and were presented using Matlab 2015 (MathWorks, Inc., Natick, MA) with Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007). The Matlab code used for stimulus randomization and presentation is available on Open Science Framework (https://osf.io/asw48/).

2.3. Testing procedure

The participants were informed about the EEG procedure, and head circumference was measured. Electrodes were applied, and participants were seated at a desk in front of a computer monitor and keyboard for the rest of the experiment. Participants were instructed to listen to the sentences and think about whether they made sense. They were instructed to keep eye blinks and other movements confined to the silent periods between the stimuli. To encourage participants to pay attention, they were tested on recognition of the target words or sounds four times per block. Specifically, they heard a random sentence target item (either an isolated sound or word, depending on the type of stimuli in the current block) and were asked, "Have you heard this item? If yes, was it in a meaningful or nonsense context?" In this case, "meaningful" refers to congruent/match and "nonsense" refers to incongruent/mismatch. They responded via button press with two buttons marked "yes" and "no" on the keyboard. After the experiment, the position of electrodes on participants' heads were imaged in an 11-camera geodesic dome (Geodesic Photogrammetry System, EGI, Eugene, OR) to determine the precise spatial location of all 128 electrodes (Russell, Jeffrey Eriksen, Poolman, Luu, & Tucker, 2005). One participant was unable to have the photos taken due to difficulties with mobility.

2.4. EEG setup

Saline Hydrocel Geodesic Sensor Nets with 128 electrodes (EGI, Eugene, OR) were used for the EEG recordings. After the net was applied, impedance was minimized (to 50 kΩ or less) by repositioning electrodes, or if necessary rewetting electrode sponges. Recordings were sampled at 1000 Hz and amplified with a 128-channel high-input impedance amplifier (400 MΩ, Net Amps™, Electrical Geodesics Inc., Eugene, OR). The software used for EEG data collection was Netstation 5 (Electrical Geodesics Inc., Eugene, OR).

2.5. Data preprocessing

Preprocessing was done in BESA 6.0. EEG recordings were filtered with 0.1–30 Hz bandpass (Tanner, Morgan-Short, & Luck, 2015), and a 60 Hz notch filter was applied to remove electrical noise. The recordings were then segmented based on trial type; trials were marked from 100 ms before to 900 ms after the onset of the sentence-terminal target sound or word. These trial segments were then examined for recording artifacts including eye blinks and movements; trials with eye blinks or other contaminating signals were removed, and exceptionally noisy channels were interpolated. The waveforms were baseline corrected using the 100 ms before target onset. Participants with 50% or more artifact-contaminated trials in any one condition were removed from further analysis. This procedure resulted in removal of one participant who lost over half the trials in the specific/mismatch/sounds condition.

Electrode coordinates from individuals' net placement images were used to assign individual sensor locations for each participant. For one participant who could not sit in the geodesic dome due to mobility difficulties, an average coordinate file provided by EGI was used (Electrical Geodesics Inc., Eugene, OR).

2.6. Analyses

2.6.1. Topographic analyses

We used BESA 6.0 to generate participant-level averaged waveforms and [mismatch – match] difference waves. We also used BESA to create ascii files of time-varying voltage at every electrode; these were used for topographic analysis in RAGU (Randomization Graphical User interface, Koenig, Kottlow, Stein, & Melie-García, 2011). Averaged waveforms and difference waves for each participant are available on Open Science Framework (https://osf.io/asw48/).

RAGU is an unbiased method for testing for statistically significant main effects or interactions between experimental factors using a scalp topographic map randomization procedure (a detailed description of this procedure is provided in the Supplement). RAGU has the advantage of using all 128 electrodes, i.e. the entire scalp topography, rather than requiring individual electrodes to be chosen for statistical testing. It relies on randomizations using only the collected dataset, and makes no assumptions about data distributions. We performed two analyses using 5000 randomizations of the data in RAGU.

The first analysis was intended to address our questions about differences between understanding words and ES in preceding sentence context. In this analysis, we analyzed [mismatch – match] difference topographies (as in, for example, Frishkoff & Tucker, 2001). This is because raw voltage between responses to ES and spoken words might

be different for many reasons unrelated to our manipulations in the present study, as discussed in the Introduction. Therefore, in this analysis, we examined the difference topographies for main effects of sentence constraint (specific vs. general) and target type (word vs. ES). This analysis identified time windows where there were significant main effects and interactions; it also output scalp topographies for the different conditions at each time point.

The second analysis included only factors of target type (word vs. ES) and congruency (match vs. mismatch), as it pooled together the two constraint levels. This analysis was conducted to (1) identify the time window in the vicinity of the N400 where there is a significant match vs. mismatch difference in topography, so that we could export data from this time window for N400 latency analysis, and (2) uncover potentially important differences between the responses to ES and words at other time points. Even though responses to ES versus spoken words might differ for many possible reasons unrelated to context effects as already discussed, we ran this exploratory analysis to see if any of these differences were of note. Note that in neither analysis did we pool together ES and words; sentence ending type was always a main factor being investigated as both a main effect and an interaction in our models.

Because our randomization analyses involve 5000 randomizations of the topographical maps at each time point, there are multiple statistical comparisons across time. To avoid false positives, we implemented a threshold of 40 ms for significance windows identified in our analyses (e.g., Guthrie & Buchwald, 1991). Time windows shorter than 40 ms showing significant main effects or interactions were not considered for further analysis.

2.6.2. Regions of interest (ROIs)

In order to represent most of the topography of the scalp in our analysis without arbitrarily choosing just a few electrodes, data were pooled into nine ROIs in a fashion similar to Potts and Tucker (2001), who used four adjacent electrodes in each ROI. Our ROIs and the component electrodes of each are listed in Table 1. The pooled ROI data were used for two purposes: (1) to represent voltage traces in figures, and (2) for statistical analysis of latency data described in the next section.

2.6.3. Latency analysis

As peak latency differences do not reliably reflect timing differences (Hansen & Hillyard, 1984; Luck, 1998), we used the time point dividing the area under the [mismatch – match] difference curve into two equal halves to estimate latency of the N400. To make sure we were looking at the N400, we limited this analysis to the N400 time window identified in our second randomization analysis (significant main effects of match vs. mismatch; 309–512 ms post target onset). For each subject, and for sounds and words separately, we pooled electrodes into the nine ROIs described above. These data were entered into a repeated measures ANOVA in R using the "ez" package (Lawrence, 2013). The dependent variable was latency in milliseconds post-target-onset; within-subject factors were ROI (9 levels, Table 1) and condition (speech vs. nonspeech).

Table 1
Electrodes included in our ROIs. Numbers correspond to electrode numbers in the EGI Hydrocel 128-electrode Geodesic Sensor Net. A spatial layout of this net is available on this study's Open Science Framework page https://osf.io/asw48/.

ROI	Electrodes	# electrodes
Anterior left	26, 27, 32, 33	4
Anterior midline	4, 11, 16, 19	4
Anterior right	1, 2, 122, 123	4
Center left	40, 45, 46, 50	4
Center midline	7, 55, 107, Cz	4
Center right	101, 102, 108, 109	4
Posterior left	58, 59, 64, 65	4
Posterior midline	71, 72, 75, 76	4
Posterior right	90, 91, 95, 96	4

3. Results

First, we assessed whether the present paradigm produces an N400. In the typical N400 time window 200–600 ms post-target-onset (e.g., Kutas & Federmeier, 2011), we found significant differences between match and mismatch conditions. Specifically, between 309 and 512 ms after target onset, we observed significant topographic ERP map dissimilarities between match and mismatch sentence endings ($p < 0.05$, Fig. 1, Table 2). The observed generalized dissimilarity between match and mismatch topographies in this time period exceeded the generalized dissimilarity obtained in at least 95% of the randomizations (Fig. 1a). The mean observed mismatch vs. match dissimilarity in this time window was 8.13; the mean dissimilarity expected due to random chance was 4.66 (95% CI: 2.91–7.37; Table 2). While there were other windows exhibiting significant main effects of congruency between (1) 254 and 286 ms and (2) between 612 and 639 ms, these windows did not pass our duration threshold ($\geq$40 ms) for further examination (Fig. 1b, Table S1).

The scalp topographies in the 309–512 ms time window indicate a stronger frontocentral, slightly right-lateralized negativity when the target is not congruent with the preceding sentence (Fig. S1a and b). By 430 ms, responses to match endings show a similar, albeit weaker, frontocentral negativity (Fig. S1a and b). This topography is similar to previous N400 studies with language presented in the auditory modality (Kutas & Federmeier, 2011; Kutas & Van Petten, 1994). It is important to note that while there was a significant main effect of congruency (in which a larger N400 occurs for both spoken word and ES mismatches as opposed to matches), there were no interactions between target type and congruency in this analysis ($p > 0.25$, Table 2, Table S1). Moreover, our analysis of [mismatch – match] differences (which will be discussed in more detail later) showed no main effect of ending type ($p > 0.25$, Table 2, Table S2). Taken together, these results indicate that there was no evidence for substantially different congruency-related N400 activity for ES and words.

Once we confirmed that our paradigm elicited N400 activity related to incongruency, we could ask our main questions about differences between speech and ES N400s. Most importantly, we wanted to assess if the N400's sensitivity to incongruency was similar for ES and words in the context of a fluent spoken sentence. Our [mismatch – match] difference topography analysis revealed that both ES and words had greater frontocentral negativities in mismatch conditions (Fig. 2a). In fact, there was no statistically significant difference between the [mismatch – match] topographies of words and ES in the vicinity of the N400 (Fig. 2b and c; $p > 0.25$; Table 2). The mean observed dissimilarity between words and ES [mismatch – match] topographies in the N400 time window identified above (309–512 ms) was 9.27. The mean observed dissimilarity that could be expected due to chance was 8.24 (95% CI: 5.44–12.58; Table 2). Thus, the congruency sensitivity of the N400 was not statistically distinguishable between words and ES. There were three very short windows after 600 ms where a significant main effect of target type was found (Fig. 2c, Table S2), however as the longest of these was 16 ms, none passed our 40 ms threshold for further consideration.

If understanding ES in a spoken sentence frame is more difficult than doing so for spoken words, the ES N400 could be delayed. We found no evidence that the N400 to ES was delayed; a 2 × 9 (target type [word, ES] × ROI [AL, AR, AM, CL, CR, CM, PL, PR, PM]) repeated measures ANOVA of [mismatch – match] difference wave latencies showed that there was no main effect of target type on latency [$F(1, 21) = 2.51$, $p = 0.13$, diff = 5.26 ms, $d = 0.56$, 95% CI (−1.25 to 11.78 ms); Table 3]. The mean latency for ES was 406.56 ms [401.51–411.61] and for words was 411.82 ms [407.62–416.03]. Therefore, even though the difference in latency trends towards

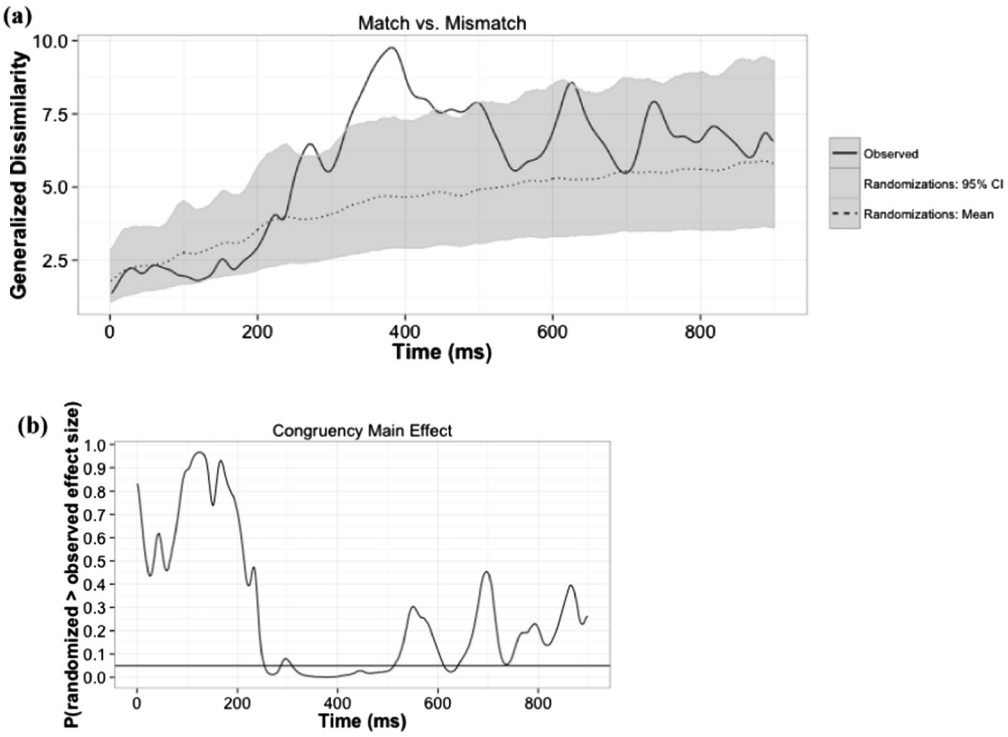

Fig. 1. (a) Time-varying generalized dissimilarity between raw match and mismatch topographies. To give a sense of the meaning of this effect size, the mean and 95% CI for the generalized dissimilarity expected due to random chance (estimated from randomizing the data) is also represented. (b) Time-varying p-value, i.e., proportion of randomizations leading to a larger effect size than observed. We can see that the largest window showing a reliable main effect of congruency is from approximately 300–500 ms post target onset.

Table 2
Summary of statistics for randomization analyses of topographical maps.

Window (ms)	Analysis	Factor	Observed generalized dissim. (GD)	Expected GD under null hypothesis	95% CI of GD under null hypothesis	p
309–512	Raw voltage, congruency × target type	Congruency (match vs. mismatch)	8.13	4.66	2.91–7.37	0.015*
309–512	Raw voltage, congruency × target type	Congruency * target type interaction	9.09	9.33	5.82–14.77	0.49
309–512	Mismatch – match difference, constraint × target type	Target type (words vs. ES)	9.27	8.24	5.44–12.58	0.28
359–421	Mismatch – match difference, constraint × target type	Constraint (general vs. specific)	12.60	8.09	5.30–12.53	0.025*
359–421	Mismatch – match difference, constraint × target type	Constraint * target type interaction	12.92	16.22	10.65–25.05	0.27

* $p < 0.05$.

significance, the trend is for N400s to ES to be earlier, not later, than N400s to words. There was also no evidence of a main effect of ROI [$F(8, 168) = 0.92$, $p > 0.25$] or an interaction of target type and ROI [$F(8, 168) = 0.99$, $p > 0.25$], indicating that N400 latencies were not statistically distinguishable between different regions of the scalp (Table 3).

Another important question we asked was whether sentence constraint affected responses to ES and words similarly. In order to address this question, we first asked if previous literature was replicated by finding constraint main effects on the N400. As outlined in the Introduction, based on previous literature we expect to find larger N400s for low constraint sentences when the ending is congruent, and roughly equal N400s for low and high constraint sentences when the ending is incongruent. By this logic, the [mismatch – match] difference wave for low constraint sentences should be smaller, as the larger match N400 would cancel out the large mismatch N400. For high constraint sentences, we should see a stronger N400 negativity in the [mismatch – match] difference wave, as the weak match N400 would do little to cancel out the mismatch N400. Our randomization analysis of the difference topographies showed a significant main effect of constraint between 359 and 421 ms after target onset (Fig. 3, Table 2, Table S2, $p < 0.05$). In this window, the mean observed dissimilarity between general and specific [mismatch – match] topographies was 12.60 (Table 2). The mean observed dissimilarity that could be expected due to chance was 8.09 (95% CI: 5.30–12.53; Table 2).

There were also main effects of constraint beginning at 259 and again at 445 ms, which did not reach our 40 ms threshold (Table S2); however, these windows exhibited the same topographic difference between general and specific as our longest window from 359 to 421 ms. Namely, [mismatch – match] topographies to endings after low constraint sentences exhibit a negativity that is shifted frontally relative to high constraint (Fig. S2a). If we focus on central midline and posterior regions, we can see that the N400 [mismatch – match] wave appears weaker in general conditions, as previous literature would predict (Fig. S2b). It is possible that previous studies have reported a weaker N400 in low constraint conditions because they focused on central regions rather than the entire scalp topography.

Once we demonstrated that constraint affects the N400, we asked whether it had similar effects on neural responses to ES and words in sentence context. If constraint affects ES and words differently, we would expect to see an interaction between target type and constraint in our randomization analysis. There was no evidence of such an interaction (Fig. 4b and c; $p > 0.25$); the mean generalized dissimilarity between 359 and 421 ms associated with the interaction was 12.92,

whereas the dissimilarity expected due to chance in this time period was 16.22 (95% CI: 10.65–25.06; Table 2). The topographies separated out for the four conditions (general/ES, specific/ES, general/word, and specific/word) all show the same pattern of a frontal shift for general conditions (Fig. 4a). Thus, constraint does not appear to differentially affect ES and words in the context of a spoken sentence.

As noted in the Introduction, there are reasons to expect a difference in the neural processing of speech and nonspeech sounds due to acoustic differences, frequency of experience, and differences in cortical activation patterns. The N1-P2 is a pair of event-related potentials (ERPs) that marks acoustic stimulus change (e.g., Hillyard & Picton, 1978); when it occurs after a change in an ongoing sound, it is called an acoustic change complex (ACC) (Kim, 2015). Our second randomization analysis revealed an ACC in response to ES following a spoken

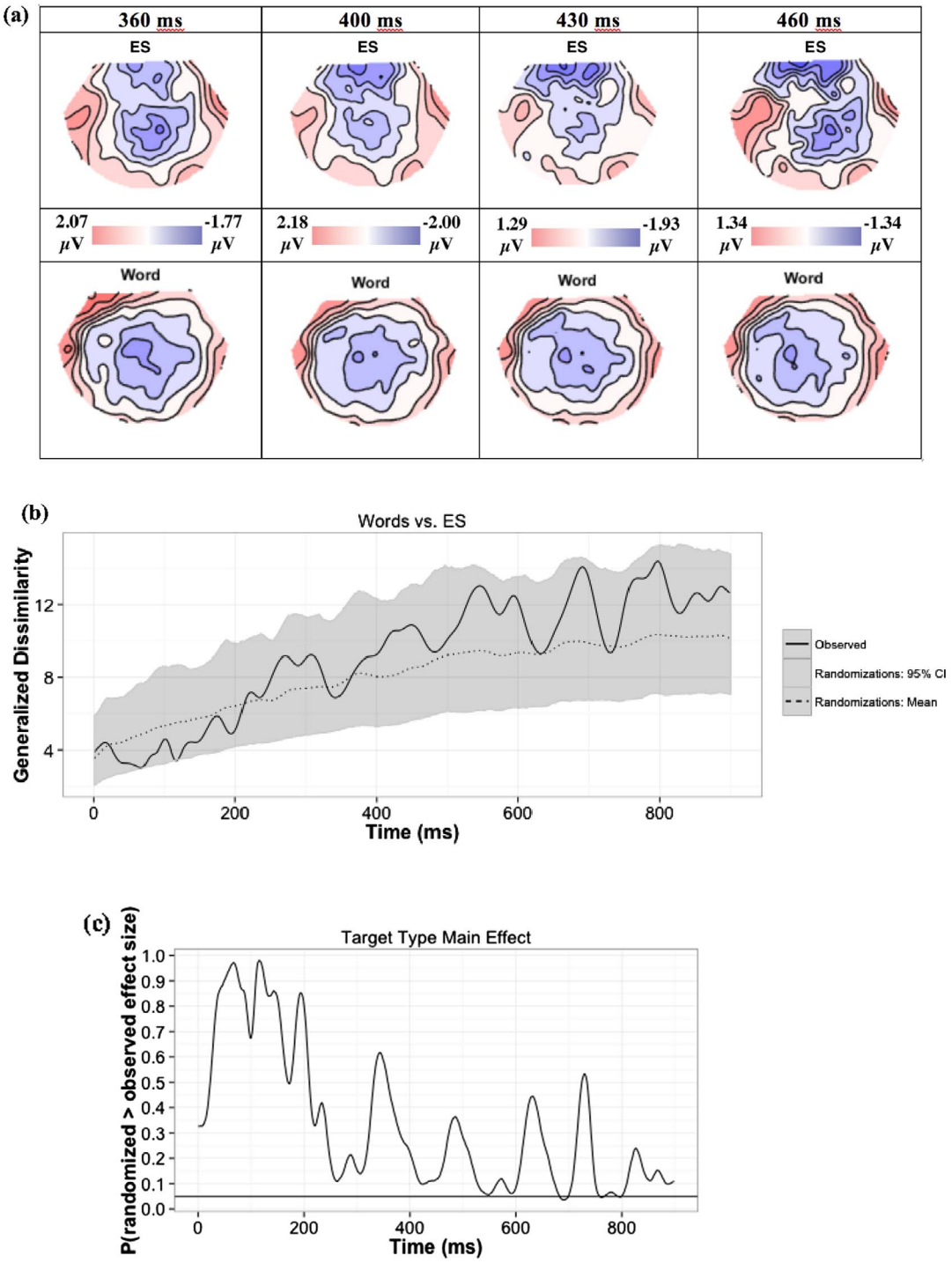

Fig. 2. (a) Scalp topographies for [mismatch – match] difference topographies for ES and words at 360, 400, 430, and 460 ms post target onset. Blue indicates negative potential; red indicates positive potential; colorbar shows correspondence of colors to microvolts. (b) Time-varying generalized dissimilarity between ES and word [mismatch – match] difference topographies. To give a sense of the meaning of this effect size, the mean and 95% CI for the generalized dissimilarity expected due to random chance (estimated from randomizing the data) is also represented. (c) Time-varying p-value, i.e., proportion of randomizations leading to a larger effect size than observed. We can see that there are no points in the vicinity of the N400 where the difference between word and ES topographies is statistically significant. (d) Voltage traces for ES and word [mismatch – match] difference waves in the nine examined ROIs. Note that negative is up and time = 0 ms corresponds to ending onset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

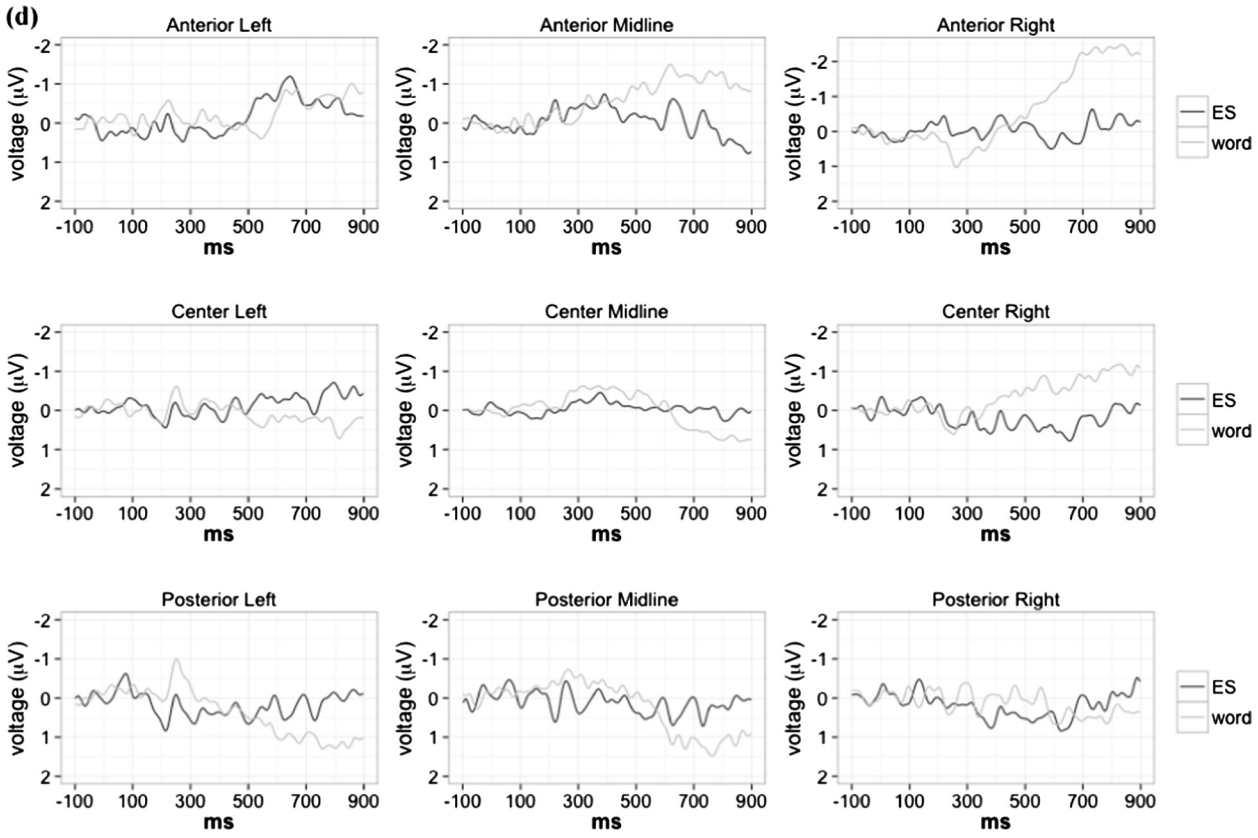

Fig. 2. (*continued*)

sentence, as shown clearly both from scalp topographical maps, and from voltage traces in ROIs (Fig. S3a and b). The presence of this ACC was statistically supported by significant main effects of Target Type on topographies in a long window (132–729 ms; $p < 0.05$) encompassing the ACC time frame in our second randomization analysis (Table S1). Though there were differences between raw voltage for ES and words in places other than the ACC, these were not a focus of the current study because (as already discussed in the Introduction) they could be due to many factors beyond the current study manipulations. Therefore, they will not be discussed further.

Finally, though there was no statistically significant interaction between congruency and target type, there is an apparent morphological difference between the shape of the N400 for ES compared to words (Fig. S4a and b, particularly anterior and center midline ROIs). Namely, it appears that the N400 to ES consists of two peaks instead of one. When we break this down by condition (Fig. S4a), it can be seen that the first peak appears to be more sensitive to congruency than the second one. This effect does not reach significance, but because of similarities with previous literature we will address this morphological difference in the discussion.

4. Discussion

The main question in the present study is whether environmental sounds and spoken words produce similar patterns of brain electrical responses when they are being understood in the same fluently spoken sentence context. Regardless of target type (word/ES), a significantly more negative frontocentral N400 occurred for mismatch conditions 309–512 ms after target onset (Table 2, Figs. 1 and S1). This finding replicates previous research showing stronger N400s to incongruent stimuli. The start of the significance window at 309 ms is characteristic of auditory N400s to fluent speech, which happen earlier than visual N400s (Kutas & Federmeier, 2011).

One crucial question was whether sentences ending in ES produce an N400 effect; one with properties similar to those for speech. The answer to this question is clear: [mismatch – match] difference topographies were not statistically different between ES and words. In both cases, difference topographies showed central negativities characteristic of the expected stronger N400 to mismatch stimuli (Fig. 2). This indicates that congruency with the preceding context affects the N400 to ES and words in the same way: for both, N400 responses are more negative in response to mismatches. To our knowledge, this is the first demonstration that ES can give rise to an N400 effect in a fluent speech context, and suggests some level of processing similarity with speech.

Incidentally, other studies have sometimes demonstrated lateral asymmetry effects in the difference topographies, such that the N400 congruency effect is more right-lateralized for ES (e.g., van Petten & Rheinfelder, 1995). We did not find such an effect, which would have shown up as a target type effect in the difference topography randomization analysis. Previous work shows that N400 lateralization can

Table 3
Summary of statistics for repeated measures ANOVA examining N400 latency.

Analysis	Factor	Latency difference	F	Cohen's d	95% CI	p
Latency repeated measures ANOVA	Target type	5.26 ms (ES earlier)	2.51	0.56	−1.25 to 11.78	0.13
Latency repeated measures ANOVA	ROI	–	0.92	–	–	> 0.25
Latency repeated measures ANOVA	Target type * ROI	–	0.99	–	–	> 0.25

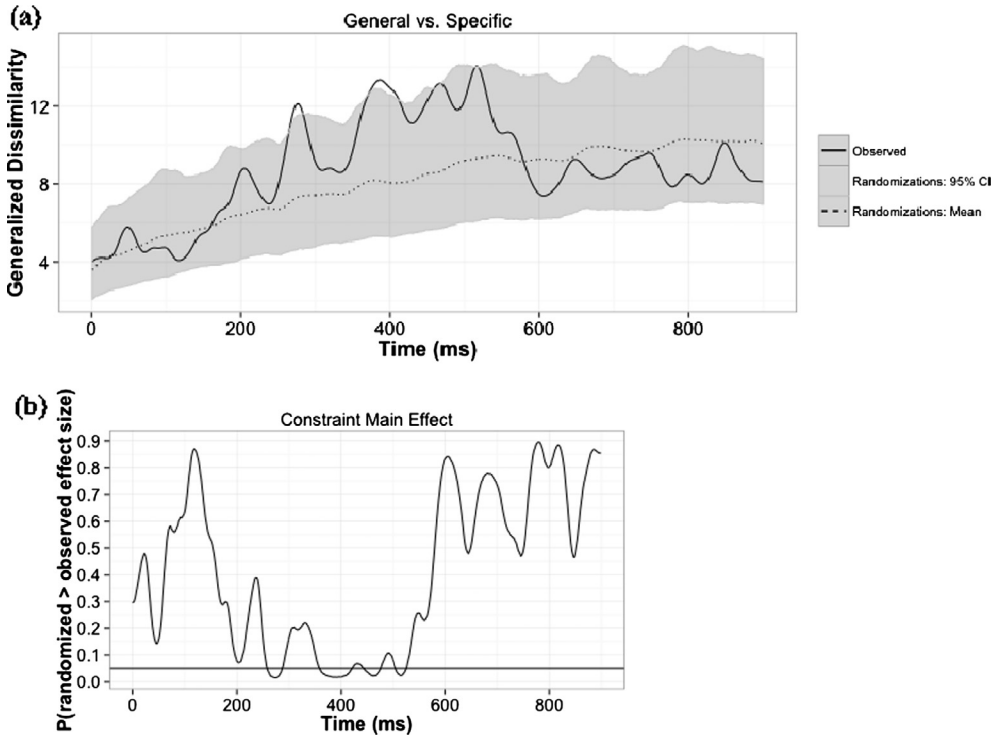

Fig. 3. (a) Time-varying generalized dissimilarity between general and specific [mismatch – match] difference topographies. To give a sense of the meaning of this effect size, the mean and 95% CI for the generalized dissimilarity expected due to random chance (estimated from randomizing the data) is also represented. (b) Time-varying p-value, i.e., proportion of randomizations leading to a larger effect size than observed. We can see that there is a main effect of constraint centered around 400 ms.

differ based on handedness (Fagard, Sirri, & Rämä, 2014; Kutas & Hillyard, 1980b). Therefore, a possible reason we failed to find lateral asymmetries between ES and words is the high proportion of left-handed participants in our study (over a third, whereas other studies often use all right-handed participants).

A further question we asked was whether environmental sounds produce N400s regardless of congruency with the sentence frame, simply by virtue of dissimilarity from speech. Our difference topography analysis also answers this question. If N400s to ES were always large, regardless of congruency with context, the [mismatch – match] difference for ES would approach zero, and a main effect of target type would be observed in our difference topography randomization analysis. Clearly this is not the case, as there is no main effect of target type on the [mismatch – match] differences. A second possibility was that ES are more difficult to understand than words in a sentence context, leading to a delayed N400. This hypothesis was also rejected; N400s to ES occurred 5.26 ms earlier than N400s to words. This agrees with previous findings that N400s are in fact slightly earlier in response to environmental sounds than to words (Cummings et al., 2006; Orgs et al., 2006), although it suggests that perhaps when the participant is focused on understanding the meaning of a full sentence, such differences are somewhat mitigated, as our 5.26 ms difference was quite small, and not statistically significant.

Finally, we asked whether the constraint level of the sentence would affect the N400 activity of ES and words differently. We did find main effects of constraint on scalp topography of [mismatch – match] differences. In particular, endings after general (low constraint) sentences elicited more frontally biased negativities, while endings after specific (high constraint) sentences elicited more central negativities. Importantly for our question, these main effects of constraint did not interact with target type; therefore, constraint appears to affect the N400 responses to ES and spoken words in a similar way. This finding goes against the prediction that constraint information might be too fine-grained or nuanced to affect ES to the same degree as words (cf. Hendrickson et al., 2015), although because we did not systematically vary near vs. far violations, the capability of this paradigm to thoroughly test this idea is limited. However, it is true that even when ES are being understood in a fluently spoken sentence—an artificial and unusual task—neural responses are affected by constraint the same way as spoken words. This suggests a surprising degree of seamless integration between these different stimulus types. Future experiments might more systematically vary the within-category nature of the semantic violations in order to test constraint effects on ES and spoken words in a more fine-grained way.

Interestingly, as outlined in the results section, previous literature predicts a weaker [mismatch – match] difference negativity for low constraint sentences, and a stronger negativity for high constraint ones. We found that this appears to be the case if we look at central and posterior regions of the scalp (Fig. S2b). However, looking at the entire scalp topography allows us to see that this appears to be a consequence of the negativity shifting frontally for low constraint conditions (Fig. 4a). Much of the previous literature on the effects of constraint on the N400 uses fewer electrodes and/or focuses on central scalp locations (e.g., DeLong, Urbach, & Kutas, 2005; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Therefore, it is possible that future work using higher density electrode arrays and topographical analyses could uncover interesting patterns involving cloze probability and/or constraint effects on the entire scalp topography.

We also noticed an apparent morphological difference between the N400 to ES and spoken words at central and frontal midline sites. In these regions, the N400 to ES appeared to consist of two peaks instead of one. Alternatively, it can be thought of as having a positive deflection in the middle. Though this morphological difference did not lead to any statistically significant effects, we found it noteworthy because such two-peaked N400s in response to ES have been reported before (Cummings et al., 2006; Hendrickson et al., 2015; Orgs et al., 2006). Hendrickson et al. theorize that this deflection could be P3b activity. The P3b is a centroparietal positivity with a latency of 300–450 ms, and is often observed to be larger in response to stimuli that are relatively more rare than others (Comerchero & Polich, 1999). Though it is typically associated with active participation in a task, it can be elicited by passive listening (Bennington & Polich, 1999). In some sense, ES are relatively more rare than words in our study due to the fact that the spoken sentences preceding every target consist entirely of words. However, because participants were instructed to pay attention to whether the sentence made sense or not, the ending targets were the

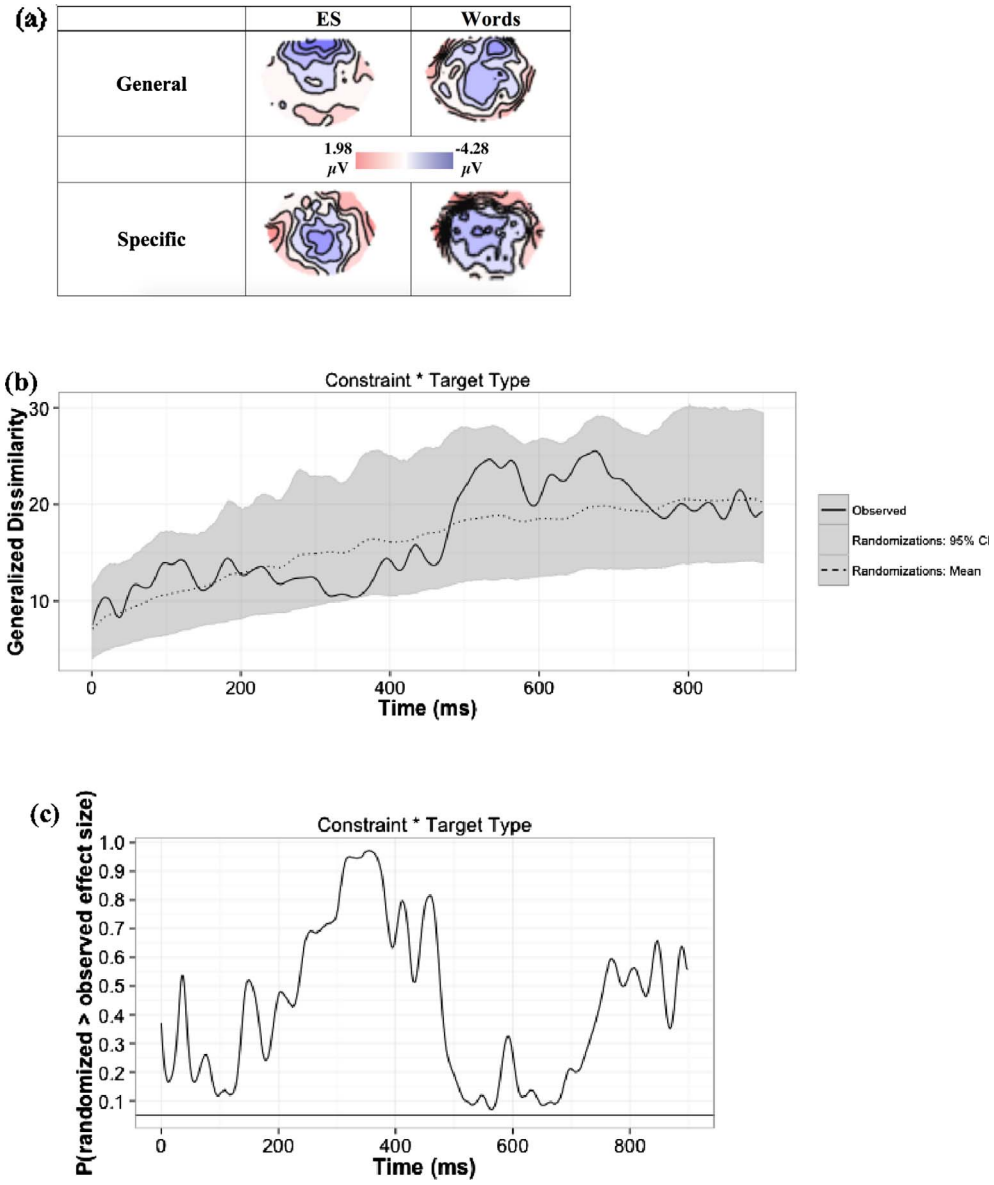

Fig. 4. (a) Scalp topographies for [mismatch – match] difference topographies for ES or word endings following low constraint (general) vs. high constraint (specific) sentences at 400 ms post target onset. Blue indicates negative potential; red indicates positive potential; colorbar shows correspondence of colors to microvolts. (b) Time-varying generalized dissimilarity associated with the interaction between constraint (general/specific) and target type (ES/word). To give a sense of the meaning of this effect size, the mean and 95% CI for the generalized dissimilarity expected due to random chance (estimated from randomizing the data) is also represented. (c) Time-varying p-value, i.e., proportion of randomizations leading to a larger interaction between constraint and target type than observed. Note that the p value always remains above the 0.05 threshold in this case, particularly in the vicinity of the constraint main effect, which is from 359 to 421 ms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

closest to task-relevant in our passive listening paradigm. Therefore, in a task-relevant sense, spoken words and ES occurred in equal proportions, making the P3b explanation unlikely. A more likely explanation is that favored by Orgs, Lange, Dombrowski, and Heil (2007) who found a remarkably similar double-peaked N400 to ES at central and frontal sites. Similarly to our voltage traces, the first of the two peaks appeared to be more sensitive to congruency than the second. Orgs et al. explained this as two separate N400 subprocesses. Though the two peaks in our study might be a length effect stemming from the longer ES than word stimuli (0.838 vs. 0.502 s, $p = 0.008$), this seems unlikely, as two-peaked N400s to ES have been found in cases where ES and spoken words were similar lengths (Hendrickson et al., 2015) and in cases where ES were all trimmed to 300 ms (Orgs et al., 2007). Moreover, word length has seldom been found to affect N400 latency, and has never been found to affect the number of peaks (Hauk & Pulvermüller, 2004). Further research is necessary to replicate and characterize this two-peaked N400, to assess whether it is characteristic of N400s to ES, and to uncover the functional significance of the two peaks.

5. Conclusions

To our knowledge, this is the first demonstration of an N400 in response to meaningful nonspeech in the context of fluent speech. This study suggests that even when embedded in fluent speech, environmental sounds can benefit from interpretation and context mechanisms typically used for understanding spoken language, along a similar timescale as spoken words. Our results suggest that neural mechanisms for integrating the meanings of words with context are flexible, and can adapt to accommodate environmental sounds, or at least such mechanisms are sufficiently general to accommodate processing of non-linguistic but meaningful acoustic patterns. Not only does congruency with context affect ES and words similarly in the context of a fluent spoken sentence—sentence constraint does as well. This work provides crucial evidence for the flexibility and adaptability of mechanisms, like linguistic ones, that at first glance appear to be quite specialized.

Statement of significance

This work compares electrophysiological responses to environmental sounds and words in spoken sentence context. Both types of stimuli elicited congruency-sensitive N400s, though nonspeech elicited an acoustic change complex and an N400 with slightly different morphology. This work enhances our understanding of specialization versus flexibility in the neural systems underlying language.

Acknowledgements

The authors thank Leslie Kay for advice on data analysis and revisions, Nina Bartram and Peter Hu for assistance with data collection, Tahra Eissa and Geoff Brookshire for advice on signal processing and data analysis, Thomas Koenig for answering questions about RAGU, and Willow Uddin-Riccio for assistance preparing the manuscript. This research is based on work supported by the National Science Foundation under Grant BCS-0116293 to the University of Chicago, and by the University of Chicago MSTP Training Grant T32GM007281.

Conflict of interest

None.

Appendix A

See Table A1.

Table A1
Paired environmental sounds and spoken words used in the current study.

Sound	Word
baby laughing	“baby”
camera shutter	“camera”
car engine revving	“car”
cashregister ch-ching	“cashregister”
cat meowing	“cat”
churchbells ringing	“churchbells”
clock ticking	“clock”
coin dropping onto hard surface	“coin”
cow mooing	“cow”
crow cawing	“crow”
dog barking	“dog”
creaky door closing	“door”
doorbell ringing	“doorbell”
drum set	“drums”
frog croaking	“frog”
guitar being strummed	“guitar”
gunshot	“gunshot”
helicopter	“helicopter”
car horn	“horn”
papers being ruffled	“paper”
phone ringing	“phone”
octave played on piano	“piano”
rooster crowing	“rooster”
saxophone notes	“saxophone”
servicebell ringing	“servicebell”
sheep bleating	“sheep”
ambulance/police siren	“siren”
sword being unsheathed	“sword”
toilet flushing	“toilet”
train whistle	“train”
water dripping	“water”
zipper	“zipper”

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2018.02.004.

References

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language, 39*(2), 187–205.

Ballas, J. A., & Mullins, T. (1991). Effects of context on the identification of everyday sounds. *Human Performance, 4*(3), 199–219. http://dx.doi.org/10.1207/s15327043hup0403_3.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309–312. http://dx.doi.org/10.1038/35002078.

Bennington, J. Y., & Polich, J. (1999). Comparison of P300 from passive and active tasks for auditory and visual stimuli. *International Journal of Psychophysiology, 34*(2), 171–177. http://dx.doi.org/10.1016/S0167-8760(99)00070-7.

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use.* Greenwood Publishing Group.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*(2–3), 221–268. http://dx.doi.org/10.1080/016909698386528.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*(1), 24–30. http://dx.doi.org/10.1016/S0168-5597(98)00033-1.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research, 1115*(1), 92–107. http://dx.doi.org/10.1016/j.brainres.2006.07.050.

Dehaene, S. (2011). The massive impact of literacy on the brain and its consequences for education. *Human Neuroplasticity and Education. Pontifical Academy of Sciences,* 19–32.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121. http://dx.doi.org/10.1038/nn1504.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review, 108*(4), 759–788.

Fagard, J., Sirri, L., & Rämä, P. (2014). Effect of handedness on the occurrence of semantic N400 priming effect in 18- and 24-month-old children. *Frontiers in Psychology, 5*. http://dx.doi.org/10.3389/fpsyg.2014.00355.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple

effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84. http://dx.doi.org/10.1016/j.brainres.2006.06.101.
Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology.* Cambridge: MIT Press.
Frishkoff, G. A., & Tucker, D. M. (2001). *Anatomy of the N400: Brain electrical activity in propositional semantics.* Institute of Cognitive and Decision Sciences, University of Oregon.
Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 23*(01), 1–21.
Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283. http://dx.doi.org/10.3758/BF03204386.
Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240–244. http://dx.doi.org/10.1111/j.1469-8986.1991.tb00417.x.
Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics, 69*(6), 839–855.
Hansen, J. C., & Hillyard, S. A. (1984). Effects of stimulation rate and attribute cuing on event-related potentials during selective auditory attention. *Psychophysiology, 21*(4), 394–405. http://dx.doi.org/10.1111/j.1469-8986.1984.tb00216.x.
Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology, 115*(5), 1090–1103. http://dx.doi.org/10.1016/j.clinph.2003.12.020.
Hendrickson, K., Walenski, M., Friend, M., & Love, T. (2015). The organization of words and environmental sounds in memory. *Neuropsychologia, 69*, 67–76. http://dx.doi.org/10.1016/j.neuropsychologia.2015.01.035.
Hillyard, S. A., & Picton, T. W. (1978). On and off components in the auditory evoked potential. *Perception & Psychophysics, 24*(5), 391–398.
Kim, J.-R. (2015). Acoustic change complex: Clinical implications. *Journal of Audiology & Otology, 19*(3), 120–124. http://dx.doi.org/10.7874/jao.2015.19.3.120.
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.
Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203. http://dx.doi.org/10.1037/a0038695.
Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience, 2011*, e938925. http://dx.doi.org/10.1155/2011/938925.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647. http://dx.doi.org/10.1146/annurev.psych.093008.131123.
Kutas, M., & Hillyard, S. A. (1980a). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science (New York, N.Y.), 207*(4427), 203–205.
Kutas, M., & Hillyard, S. A. (1980b). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99–116. http://dx.doi.org/10.1016/0301-0511(80)90046-0.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163.
Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of Psycholinguistics,* 83–143.
Lawrence, M. A. (2013). ez: Easy analysis and visualization of factorial experiments. R package version 4.2-2. Retrieved from <http://CRAN.R-project.org/package=ez>.
Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *The Journal of Neuroscience, 29*(16), 5234–5239. http://dx.doi.org/10.1523/JNEUROSCI.5758-08.2009.
Leech, R., & Saygin, A. P. (2011). Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex. *Brain and Language, 116*(2), 83–90. http://dx.doi.org/10.1016/j.bandl.2010.11.001.
Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience, 5*(4), 356–363. http://dx.doi.org/10.1038/nn831.
Lewis, J. W. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience, 25*(21), 5148–5158. http://dx.doi.org/10.1523/JNEUROSCI.0419-05.2005.
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36. http://dx.doi.org/10.1016/0010-0277(85)90021-6.
Luck, S. J. (1998). Sources of dual-task interference: Evidence from human electrophysiology. *Psychological Science, 9*(3), 223–227. http://dx.doi.org/10.1111/1467-9280.00043.
Morris, A. L., & Harris, C. L. (2002). Sentence context, word recognition, and repetition blindness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 962–982. http://dx.doi.org/10.1037//0278-7393.28.5.962.
Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition, 62*(3), 267–272. http://dx.doi.org/10.1016/j.bandc.2006.05.003.
Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology, 65*(2), 162–166. http://dx.doi.org/10.1016/j.ijpsycho.2007.03.003.
Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General, 115*(3), 281.
Potts, G. F., & Tucker, D. M. (2001). Frontal evaluation and posterior representation in target detection. *Cognitive Brain Research, 11*, 147–156.
Reuter-Lorenz, P. (2002). New visions of the aging mind and brain. *Trends in Cognitive Sciences, 6*(9), 394.
Russell, G. S., Jeffrey Eriksen, K., Poolman, P., Luu, P., & Tucker, D. M. (2005). Geodesic photogrammetry for localizing sensor positions in dense-array EEG. *Clinical Neurophysiology, 116*(5), 1130–1140. http://dx.doi.org/10.1016/j.clinph.2004.12.022.
Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1–17. http://dx.doi.org/10.1016/j.jml.2015.02.004.
Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*(8), 997–1009. http://dx.doi.org/10.1111/psyp.12437.
Uddin, S., Heald, S. L. M., Van Hedger, S. C., Klos, S., & Nusbaum, H. C. (2018). Understanding environmental sounds in sentence context. *Cognition, 172*, 134–143. http://dx.doi.org/10.1016/j.cognition.2017.12.009.
van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia, 33*(4), 485–508. http://dx.doi.org/10.1016/0028-3932(94)00133-A.
Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13*(7), 994–1005. http://dx.doi.org/10.1162/089892901753165890.
Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology. Human Perception and Performance, 10*(5), 704–712.