Brain and Language 201 (2020) 104722

Contents lists available at ScienceDirect

Brain and Language

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

Cortical mechanisms of talker normalization in fluent sentences

Sophia Uddin*, Katherine S. Reis, 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:
Talker normalization
Working memory
Environmental sounds
Fluent speech
Source analysis
Acoustic
Attention

ABSTRACT

Adjusting to the vocal characteristics of a new talker is important for speech recognition. Previous research has indicated that adjusting to talker differences is an active cognitive process that depends on attention and working memory (WM). These studies have not examined how talker variability affects perception and neural responses in fluent speech. Here we use source analysis from high-density EEG to show that perceiving fluent speech in which the talker changes recruits early involvement of parietal and temporal cortical areas, suggesting functional involvement of WM and attention in talker normalization. We extend these findings to acoustic source change in general by examining understanding environmental sounds in spoken sentence context. Though there may be differences in cortical recruitment to processing demands for non-speech sounds versus a changing talker, the underlying mechanisms are similar, supporting the view that shared cognitive-general mechanisms assist both talker normalization and speech-to-nonspeech transitions.

1. Introduction

People seem to understand speech from different talkers with little difficulty. However, even among native speakers of a language, there is wide variability in the acoustic characteristics of various phonemes (Heald & Nusbaum, 2014b; Peterson & Barney, 1952). Even more acoustic-phonetic variability is introduced when accents and speech patterns specific to non-native speakers are taken into account. This variability in the relationship between acoustic patterns and phonetic categories introduces ambiguity into the recognition of a speaker's intended phoneme given that any particular acoustic pattern might map onto different phonetic categories (Nusbaum & Magnuson, 1997). Despite this acoustic-phonetic variability, listeners appear to quickly and easily understand utterances from different talkers, albeit with a small but reliable recognition performance reduction (Heald & Nusbaum, 2014b; Nusbaum & Magnuson, 1997). One explanation of this performance reduction is talker "normalization": the process by which listeners use talker vocal characteristics to resolve acoustic-phonetic ambiguities that are introduced when a talker change demands attention and extra processing (Nusbaum & Morin, 1992).

Though listeners can adjust to a new talker quickly, and are not usually aware of the fact that they are sensitive to the vocal characteristics of the new talker, there is evidence that this process may use working memory (WM), possibly to selectively direct perceptual attention towards acoustic cues needed to calibrate the speech for recognition. For example, when listeners must maintain a high WM load, they recognize spoken target syllables more slowly when the talker changes. When the talker does not change, however, the high WM load does not have the same effect (Nusbaum & Morin, 1992). These results strongly suggest that adjusting to a new talker draws on WM resources. Further, changes in talker result in slowed recognition and/or categorization for CV syllables, vowels, and whole words (Mullennix & Pisoni, 1990, Nusbaum & Morin, 1992; Kaganovich, Francis, & Melara, 2006; Wong, Nusbaum, & Small, 2004).

In order to examine the underlying neural mechanism for the effects of talker change, Wong et al. (2004) used fMRI to investigate contributions to talker normalization, both from traditional superior temporal language areas, and from a more distributed attention network. In this study, participants listened to lists of individual spoken words, and their task was to recognize a target word in the list. Each participant listened to lists in two conditions. In the "blocked by talker" condition, the words were all spoken by the same talker. In the "mixed talker" condition, the words were spoken by four different talkers. The results showed that listeners were slower to detect the target word in "mixed talker" conditions, which was expected from prior research (e.g., Nusbaum & Magnuson, 1997). Moreover, Wong et al. (2004) identified brain areas that were differentially active in blocked-talker and mixed-talker conditions. Two areas responded significantly differently to mixed and blocked talkers; these were middle/superior temporal areas, and the superior parietal lobule. The response in temporal cortex was

* Corresponding author at: 5480 S. Cornell Ave., Apt. 615, Chicago, IL 60615, United States.
E-mail address: sophiauddin@uchicago.edu (S. Uddin).

https://doi.org/10.1016/j.bandl.2019.104722
Received 13 June 2018; Received in revised form 4 November 2019; Accepted 13 November 2019
Available online 10 December 2019
0093-934X/

interpreted as resulting from increased recognition difficulty in the mixed-talker condition, and superior parietal activity was interpreted as reflecting a shift in perceptual attention that might address the increased recognition difficulty.

Unfortunately, the poor time resolution of fMRI makes it difficult to know if indeed the parietal activity was a consequence of, or a precursor to, the temporal cortex activity. However, EEG could provide information about the relative timing of activity in these cortical regions due to a talker change. For example, the N1 event-related potential (ERP) measured with EEG (usually peaking 50–150 ms after stimulus onset - (Näätänen & Picton, 1987) has been interpreted as reflecting a change in sensory attention (Picton & Hillyard, 1974), or the detection of a "trigger" that leads to reallocation of attention to something new (Lijffijt et al., 2009). It is generally thought that increased attention leads to an increase in N1 amplitude; this has been demonstrated for many types of auditory stimuli, including (but not limited to) music and fluent speech (Astheimer & Sanders, 2009; Heacock, Pigeon, Chermak, Musiek, & Weihing, 2019; Peng, Hu, & Chen, 2018; Picton & Hillyard, 1974; Snyder, Alain, & Picton, 2006; Zendel & Alain, 2014). To the extent that talker normalization involves attentional mechanisms, we might expect a change in talkers to increase N1 strength. In support of this idea, prior ERP research has suggested that talker changes increase N1 amplitude (Kaganovich et al., 2006). These findings support the idea that a talker change produces a shift in attention, likely directed towards relevant acoustic cues for the vocal characteristics of the new talker (Heald & Nusbaum, 2014b; Nusbaum & Magnuson, 1997; Nusbaum & Schwab, 1986).

Despite the strong link between the N1 and attention, however, there is a lack of research that has linked the N1 in a talker change context to empirically-informed brain areas. Due to the general involvement of the superior parietal lobule in perceptual attention tasks that do not necessarily involve speech (e.g., Yantis et al., 2002), Wong et al. (2004) interpret the involvement of this area in talker normalization as being related to the deployment of perceptual attention, as predicted by Nusbaum and Morin (1992) from behavioral research. We might therefore expect activity in this area to differ during the N1 based on whether the talker changes, as the N1 has been implicated in the attentional processes described above. This leads to specific hypotheses. First, the changes in parietal cortex reported by Wong et al. (2004) should be reflected in talker-variability-dependent changes in EEG. Second, if talker change is reflected in ERP responses, if such changes are found in the N1 component of the ERP, this would support the hypothesis that these are early sensory attention changes in processing. Third, the N1 changes should be more closely associated with superior parietal cortex than with superior temporal cortex, thus indicating that the parietal activity identified in prior research is the same as this early sensory attentional engagement in N1.

While the N1 has been extensively discussed in the context of auditory perceptual attention, there are other ERP components that may be informative for understanding how listeners accommodate to talker changes. Like the N1, the P2 is an ERP that has been observed to increase in amplitude with increased attention (Picton & Hillyard, 1974). The P2 is an ERP, usually peaking after the N1 at 150–250 ms after stimulus onset, that often occurs in response to auditory or visual stimuli (Lijffijt et al., 2009). If it is correct that the P2 increases in amplitude with attention, it might be expected to increase in amplitude with a change in talker. By this logic, the P2 might also co-occur with increased activity in the superior parietal lobule when the talker changes, as might reasonably be predicted for the N1 based on Wong et al. (2004) findings.

There is also evidence that the P2 reflects analysis of stimulus features (Luck & Hillyard, 1994; Näätänen & Winkler, 1999). Active theories of speech processing (e.g. Nusbaum & Morin, 1992) hold that when there is a talker change, there is greater possible ambiguity in terms of alternative interpretations of the speech signal. Active theories say that the alternative interpretations shift attention (predicting N1/parietal activity as described above). In turn, such attention shifts may trigger different processing of relevant auditory features and thus may result in increased P2 activity. Thus, another reason to expect P2 effects during a talker change is that listeners likely must analyze the features of the new talker's voice in order to select the most relevant acoustic cues for understanding this talker. This leads us to another prediction based on the temporal locations implicated by Wong et al. (2004). Unlike the superior parietal lobule, these temporal locations are active in speech processing and other complex auditory processing (Rauschecker & Scott, 2009). Given this information, it makes sense that they are active in talker normalization due to heightened analysis of the features of the speech signal. If the P2 does indeed reflect feature analysis, we might expect that source analysis involving the P2 would also reflect involvement of temporal speech areas, particularly when the talker is changing.

From this active perceptual framework, talker change can increase working memory load due to perceptual ambiguity (requiring the maintenance of multiple interpretations in WM) and shift attention, thus predicting interactions between measured working memory (WM) and the N1 and P2 potentials in a talker change paradigm. Given previous research showing an interaction between WM load and talker normalization – indicating that talker variability increases demand on WM (e.g., Nusbaum & Morin, 1992) – it makes sense to expect that the underlying neural processes would reflect this use of WM. Previous research has shown that measured WM capacity predicts N1 and P2 amplitudes in tasks requiring auditory selective attention (Giuliano, Karns, Neville, & Hillyard, 2014), such that higher WM capacity is associated with stronger attentional modulation of the N1 and P2. At one point, Engle (2002) proposed that WM is involved in the deployment of attention (also see Cowan, 2017).

In terms of recognizing speech when there is a change in talker, active theories (Nusbaum & Magnuson, 1997) hold that talker change increases the possibility of perceptual ambiguity. It has been suggested that maintenance of relevant information in WM interactions with attentional processes that affect early sensory processing of stimuli (Awh & Jonides, 2001). In terms of processing a talker switch, a talker change could lead to temporary ambiguity in phonetic interpretation of the utterance due to a change in the vocal characteristics of the talker. This ambiguity could increase the number of different potential meanings of the input, leading to a many-to-many mapping problem. Because these alternative interpretations of the input have to be stored in WM before one is eventually chosen, a change in source could increase the WM load. In accordance with Awh and Jonides, then, the maintenance of this information in WM could interact with attentional N1 and P2 processes.

An additional consideration is that previous research on the effects of talker change have focused on isolated discrete and silence-separated utterances using stimuli such as isolated vowels (e.g., Kaganovich et al., 2006), syllables (e.g., Morin & Nusbaum, 1989), or words (e.g., Mullennix & Pisoni, 1990). To what extent are the effects of talker variability due to the disruption that occurs between utterances versus the effects of a transition in fluent speech? The present study examined whether a fluent change in talker between a spoken sentence frame and final spoken word will affect speech processing in the same way as observed in a series of discrete and separate utterances. On the one hand, the change in vocal characteristics might be more disruptive since it is an ecological violation of speech (except in some circumstances) for one talker to fluently complete a different talker's sentence. On the other hand, message-level information from a meaningful sentence might override the need to normalize for talker differences. In order to establish that the talker normalization effects reflect a mechanism that operates during typical speech perception, it is important to replicate previous findings using fluent speech and in the context of language understanding. In the present experiment, participants were focused on understanding the meaning of the sentence as a whole, rather than a recognition or categorization task.

2. Experiment 1: Cortical sources and WM-dependence of talker normalization mechanisms

2.1. Introduction

Given that talker change increases demand on WM and requires changes in attention, we predict that a change in talker should increase N1 and P2 amplitudes. Indeed, an increase in N1 amplitude has been reported during talker changes (Kaganovich et al., 2006), likely reflecting an increase in the recruitment of attention to helpful features of the new talker's voice. To our knowledge, effects of talker change on the P2 have not been previously reported, although increased attention in other tasks has been found to increase P2 amplitude (Picton & Hillyard, 1974). If the effects of talker change are similar, we predict an increase in P2 amplitude in response to a talker change due to demands on attention and WM.

Given the role of superior parietal cortex in attention and in talker normalization as described in Wong et al. (2004), we predict that source analysis of the N1 and P2 scalp topographies will reveal increased superior parietal activity when the talker changes. If, as previously suggested (Luck & Hillyard, 1994; Näätänen & Winkler, 1999), the P2 also reflects stimulus feature analysis, then source analysis of P2 scalp topographies might reveal increased activity in auditory processing temporal areas (i.e., the areas found in the Wong et al. (2004) talker normalization study) when the talker is changing. This would be evidence that the parietal response to talker change is a fast sensory mechanism rather than a later attentional process during categorization or responding.

Finally, talker normalization could depend on individual differences in WM and thus the N1 and P2 responses may be moderated by WM capacity. Work by Giuliano et al. (2014) predicts that higher individual working memory capacity is associated with greater attentional modulation of the N1 and P2. In participants with greater WM capacity, there should be more ability to respond flexibly to the demands of talker change by deploying attention and sensory analysis. Therefore, we predict that talker change will elicit stronger N1 and P2 ERPs relative to the same talker condition in high-WM participants. On the other hand, for low-WM participants, the difference between same- and different-talker N1s and P2s might not be as pronounced. Research has suggested that low-WM participants have less flexible attentional control than high-WM participants (e.g., Awh & Vogel, 2008). In the context of talker change, then, high-WM participants should be better able to effectively use WM and deploy attention. However, if talker normalization is just reflective of demand on an active speech perception system (Heald & Nusbaum, 2014a) rather than mobilization of a specific process, then this system may operate constantly with responses dependent on the demands of perception. This means that the same processes should be part of normal speech perception even for a single talker, albeit to a lesser extent than when the talker changes. This could lead to a pattern in which both high- and low-WM participants have stronger ERPs when the talker changes, but the size of this difference does not change substantially based on WM. In this case, low-WM participants would have weaker N1 and P2 ERPs across the board, and high-WM participants would have stronger ones. If this account is correct, we expect main effects of talker change and WM capacity differences on the N1 and P2, without an interaction between these factors.

2.2. Methods

2.2.1. Participants

Participants were twenty-two (12 female, 10 male) adults from the University of Chicago and surrounding community. Their mean age was 19.59 years (SD: 1.0, range: 18–21). Twenty-one were right-handed and one was 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 (questionnaires available at https://osf.io/x8dau/). Participants received $30 cash for their participation. The target number of participants for recruitment was based on our previous studies with environmental sounds (Uddin, Heald, Van Hedger, & Klos, et al., 2018, Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018).

2.2.2. Working memory testing and analysis

Working memory was assessed by performance on an auditory *n*-back task. This task was administered in the lab prior to application of the electrodes for the EEG. The task involved actively monitoring a string of spoken letters of the alphabet, presented one at a time 3000 ms apart at 65–70 dB SPL. Participants pressed a button labeled "Target" if the current letter matched the letter presented *n* trials previously, and pressed a button labeled "Not Target" if the currently spoken letter did not match the aforementioned letter. The *n*-back task consisted of a 2-back task followed by a 3-back task. Both of these consisted of 30 practice trials (not analyzed) followed by 90 total trials (three runs of 30 letters). One third of the spoken letters were targets.

Performance on the auditory *n*-back was assessed using signal detection theory. Specifically, for each participant, we calculated the proportion of hits (correctly identified targets) and false alarms (incorrectly assigned "target" status to a non-target letter). These were then *z*-scored. Because proportions of 1 or 0 correspond to *z*-scores of ∞ or $-\infty$, respectively, we subtracted 0.5 from the total number of hits if there were a full 30 hits, and added 0.5 to the number of false alarms if there were 0 false alarms, so that real numbers would be obtained for the *z*-scores. (In theory we would have performed a similar adjustment if a participant had 0 hits or 60 false alarms, but this never happened). Adjustments like these are common practice in calculating d′ from *n*-back data (e.g., Van Hedger, Heald, Koch, & Nusbaum, 2015). After *z*-scoring the proportion of hits and false alarms, d′ was calculated for each participant as the *z*-scored hits minus the *z*-scored false alarms. Due to a ceiling effect in 2-back performance (7 participants had perfect scores, and another 7 missed only one trial), only 3-back d′ scores were used in further analysis.

2.2.3. Stimuli

The stimuli were spoken sentences with the last word of the sentence (the "target", always a noun) recorded separately to avoid co-articulation confounds. There were thirty-two targets; for each, there was a high-constraint (high cloze probability for match ending, median = 0.87, IQR = 0.25) and a low-constraint (low cloze probability for match ending, median = 0.16, IQR = 0.33) sentence stem. Cloze probability was determined based on written sentence completions from 66 Amazon Mechanical Turk participants. All sentence stems and endings were produced by an adult male speaker of Midwestern English (F0 = 131 ± 12 Hz, mean ± SE). A second set of ending words was produced by an adult female speaker of Midwestern English (F0 = 198 ± 3 Hz). In all cases, stimuli were digitized at 44.1 kHz with 16 bits of resolution and amplitude normalized to the same RMS level (~70 dB SPL).

For stimulus presentation, sentence stems and endings were spliced together in Matlab to form continuous sentences with no audible acoustic artifacts. In addition to half the sentences ending in a different talker, sentences were spliced together such that half the sentences contained a semantic mismatch comparable to mismatches used in an N400 study (e.g, Kutas & Hillyard, 1980). These mismatches were produced ahead of time by scrambling the targets and sentence stems. The resulting sentences were verified in a short written survey to ensure that they were not easily interpreted as sensical. Each word was presented an equal number of times in match and mismatch conditions.

Sentences were blocked by target type, such that there were four blocks of 32 sentences ending in the male talker (i.e., the same talker that said the rest of the sentence), and four blocks of 32 sentences ending in the female talker. Block types alternated across the

experiment, and the type of starting block (i.e. same or different talker) was counterbalanced across subjects. Subjects were told that a different talker would be saying some of the words, and warned before blocks in which the last word was said by the female talker. Within the blocks, stimuli were presented pseudo-randomly such that there was no particular pattern of matches vs. mismatches, or high vs. low constraint sentences, although within each block there was a 50–50 mix of matches vs. mismatches, as well as high vs. low constraint sentences.

Stimuli were presented at 65–70 dB over insert earphones (3 M E-A-RTone Gold) using Matlab 2015 (MathWorks, Inc., Natick, MA) with Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007).

2.2.4. Testing procedure

Participants were told what to expect from the EEG procedure, including the application of the saline electrode net, and the ~45-minute-long task of listening passively to sentences while minimizing eye blinks and movements. Each participant's head circumference was measured to fit the EGI electrode net. Participants were instructed to keep head motion and eye blinks confined to identified non-stimulus periods of time (several seconds between sentences). Participants were told to listen to the spoken sentences and think about whether they made sense. To encourage participants to pay attention, they were tested on recognition of the target words four times per block. Specifically, they heard a random sentence ending word matching the speaker saying the ending words in the current block. They 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, photographs of the electrode placement were taken by seating the participant in a geodesic dome containing eleven cameras (EGI, Eugene, OR). These photographs were used for determining the precise location of each of the 128 electrodes for each subject; the coordinates obtained through this process were used to increase precision for source analysis.

2.2.5. EEG setup

The 128 electrodes (embedded in an EGI saline Hydrocel Geodesic Sensor Net) were prepared by soaking in saline, and were applied to the participant. Impedance of each electrode was reduced to 50 kΩ or less by repositioning or rewetting with saline. EEG was continuously recorded and digitized at a sampling rate of 1000 Hz. Cz served as the online reference. The amplifier used was a 128-channel high-input impedance amplifier (400 MΩ, Net Amps™, EGI, Eugene, OR). Netstation 5 was used for data collection (EGI, Eugene, OR).

2.2.6. Data preprocessing

The online reference (Cz) was reincorporated into the montage. The EEG was re-referenced to the average of all electrodes and filtered with a 0.1–30 Hz bandpass (Tanner, Morgan-Short, & Luck, 2015) and a 60 Hz notch filter (to remove electrical noise) in BESA 6.0. The full EEG recordings were then segmented based on trial type; segments were defined as 100 ms before to 900 ms after the onset of the sentence-final word, i.e., the target. Trials with eye blinks, movement or muscle artifacts, or other contamination were removed from further analysis; exceptionally noisy channels were interpolated. For each trial, baseline correction was performed using the 100 ms preceding the onset of the sentence-final word. 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 all conditions.

Sensor locations unique to each participant were assigned using the net placement photographs taken in the geodesic dome for each participant.

2.2.7. Analyses

2.2.7.1. Topographic analyses. BESA 6.0 was used to generate participant-level averaged waveforms (as in Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018). BESA was used 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-Garcia, 2011).

We performed significance testing on the data using 5000 randomizations of the EEG topographies in RAGU to estimate baseline comparison data for analysis. This analysis is known as a TANCOVA, and it allowed us to compare observed topographic differences between conditions to the estimated topographic differences under the null hypothesis. In this way, we can assess if there are significant main effects of factors (e.g., talker) on the scalp topography of the elicited voltage. We included main effects of talker (same vs. different) and congruency (match vs. mismatch) as within-subjects factors, and WM (as measured by d' on a 3-back task) as a between-subjects factor. The output of the TANCOVA is a set of time windows, defined in milliseconds post-target onset, in which there are significant main effects of each of these factors—as well as their interactions—on the scalp topographies. The TANCOVA also provides average scalp topographies for the different factors at every time point. This allowed us to test whether a change in talker affects the patterns of neural activity during word understanding in sentence context, and if so, at what time points this happens. Similarly, it allowed us to test whether working memory (WM) interacted with ERPs, based on the prediction that WM could affect switching perception between different talkers.

We were also interested in the relationship of the talker-normalization-related sources identified in Wong et al. (2004) to ERP differences related to changes in talker, as well as WM. As the TANCOVA identifies time windows in which there are significant effects of talker and WM on scalp topography, as well as interactions between these factors, we further used these time windows for source analysis.

2.2.7.2. Source analysis. In order to relate our findings to the brain areas identified by Wong et al. (2004) for talker normalization, we performed source analysis using BESA 6.0. For source analysis, the lowpass filter was disabled to preserve as much information as possible for the process of transforming topographies into sources and to minimize possible distortions of the signal (cf., Widmann and Schröger, 2012). In general, we compared the variance explained by Wong et al. (2004) sources for same versus changing talker brain responses. If these sources better explain brain responses to a changing talker, this would provide additional evidence that these brain areas are active in talker normalization.

We performed source analysis in several different time windows that were previously identified in the TANCOVA as having main effects of talker, main effects of WM, or an interaction between these two factors. Because we expected talker change effects on the N1 and P2 ERPs, we limited our analysis to TANCOVA windows that occurred before 300 ms post-stimulus onset. Finally, we chose the P600 time window (628–675 ms, defined by a main effect of congruency in the TANCOVA characterized by P600 topographies; Table S.1, Fig. S.1) as a control window for source analysis. We chose this window because based on prior P600 source analysis (Shen, Fiori-Duharcourt, & Isel, 2016) we did not expect a significant contribution from the parietal and temporal sources identified by Wong et al. (2004). Also, we did not have any reason to expect differences in these sources based on talker change this late in the ERP.

Largely overlapping time windows for source analysis were combined into single windows, e.g. a window from 124 to 147 ms was combined with a window from 136 to 179 ms to make a larger window from 124 to 179 ms (Table 4). This was done in order to reduce the number of comparisons in further statistical testing, and still preserved a separation between N1 and P2 time periods as defined in the literature (e.g., Lijffijt et al., 2009; Näätänen & Picton, 1987). Performing

source analysis separately on these time windows allowed us to examine the time course for the activity of the sources identified in Wong et al. (2004), as due to time resolution, it was impossible to tell from fMRI whether temporal and parietal sources are active in talker normalization at the same time, or sequentially.

In each of these time windows, we fit three models to averaged condition-level data for each participant. The dipole models were, in Talairach coordinates: a "parietal model" including source dipoles at parietal locations identified by Wong et al. (2004) to be involved in talker normalization ([33 −67 45] and [−33 −67 45]), a "temporal model" including [56 −28 1] and [−56 −28 1], also identified by Wong et al. (2004) for involvement in speech perception, and a "both model" including all four coordinates as sources. For all models, the dipoles were held fixed at the aforementioned coordinates, but their orientations were allowed to vary. The subject-level averages used for source analysis were the "different talker" and "same talker" topographies, pooled across different constraint levels and congruency status. The models produce a residual variance (RV) i.e., the amount of variance in the scalp voltage maps that is left unexplained by the sources included in the model. Thus, a smaller RV indicates a better fit.

2.2.7.3. 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 as carried out by Potts and Tucker (2001), who used four adjacent electrodes in each ROI. This technique was also used in Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018); details can be seen in Table 1. The pooled ROI data were used for representing voltage traces in figures.

2.3. Results

Our first hypothesis was that, due to talker-normalization-related attentional reallocation, a change in talker would be associated with stronger N1 and P2 ERPs. If, on the other hand, talker normalization is not occurring in fluent sentences, or if for some reason attentional mechanisms are not involved, the N1 and P2 should not be significantly different between the same- and different-talker conditions. The TANCOVA revealed significant main effects of talker change on scalp topography 124–147 ms, and again 235–278 ms, after last word onset ($p < 0.05$, Table 2, Fig. 1 b,c, Table S.1). These time ranges correspond to N1 and P2 time ranges reported in the literature (e.g., Näätänen & Picton, 1987; Lijffijt et al., 2009). However, when the topographies were examined further, it was clear that these differences do not include a typical central negativity followed by central positivity that is characteristic of an N1-P2 complex, such as that elicited by environmental sounds in our previous work (Figs. S.3 and S.4 in Uddin, Heald, Van Hedger, & Nusbaum, et al., 2018). The voltage traces were also devoid of N1 and P2 peaks (Fig. 1d). Scalp topographies in both N1 and P2 windows were characterized by a left-lateralized, frontocentral positivity. In the same-talker condition in the N1 window, this positivity was modified by a posterior central negativity that was absent in the different-talker condition (Fig. 1a, left panel). In the P2 window, the positivity was more frontal in the different talker condition than in the same talker condition (Fig. 1a, right panel). Unfortunately, the lack of defined N1 and P2 peaks made it impossible to compare peak amplitude in the same versus different talker conditions. However, we were still able to perform source analysis using the scalp topographies in these time windows to test our hypotheses about parietal and temporal sources. We were also able to test if scalp topographies in these time windows interacted with WM.

Table 1
Electrodes included in ROIs. Numbers correspond to electrode numbers in the EGI Hydrocel 128-electrode Geodesic Sensor Net.

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

With regards to working memory, we hypothesized that if adjusting to a new talker in a fluent sentence involves WM-mediated attentional processes, we should find either main effects of WM on the scalp topography, or an interaction between factors of talker and WM, particularly during the N1 and P2 which are known to be modulated by attention (e.g., Picton & Hillyard, 1974). While the TANCOVA did not identify any time windows in which WM interacted significantly with talker, there were several windows with significant main effects of WM on scalp topography ($p < 0.05$, Table 3, Table S.1). As we predicted, these included windows during both the N1 (136–179 ms) and the P2 (220–258 ms).

Finally, we hypothesized that a change in talker would result in an increase in activity dependent on parietal sources during the N1 and P2 due to these sources' involvement in attention allocation (Wong et al., 2004; Yantis & Serences, 2003). We also hypothesized that there might be an increase in activity related to temporal sources during the P2, related to heightened auditory feature processing when the talker changes (Luck & Hillyard, 1994; Wong et al., 2004). If these hypotheses are correct, we should find that dipole models involving parietal regions better explains N1 and P2 topographies for the different-talker condition (as opposed to the same talker). We should also find that models based on temporal sources better explain P2 topographies for the different-talker condition. On the other hand, if these hypotheses are incorrect, we should not find significant differences between same- and different-talker responses with regards to the fit of models including these brain areas.

We used source analysis in BESA to calculate the percent of topographical variance explained by (1) parietal sources alone, (2) temporal sources alone, and (3) both parietal and temporal sources for both same- and different-talker ERPs. While the average residual variance was always lower for the different talker condition except in our control window from 628 to 675 ms (Table 5) – suggesting that the sources from Wong et al. (2004) better account for EEG measurements of neural activity in the different-talker condition than the-same talker condition – these differences were not always large. To statistically test the difference in model fit, we compared the RV for each model, in each time window, between same and different talker using one-sided, paired Wilcoxon rank-sum tests. As this led to twelve comparisons, a Bonferroni-corrected threshold of $p < 4.2E\text{-}3$ was used. In most of the time windows, including the control window as we expected, the difference between the fit of the models in the different and same talker conditions was not significantly different ($V \leq 174$, $p \geq 0.021$). However, in the 220–278 ms time window, the model including temporal sources was a significantly better fit for the different talker condition ($V = 191$, $p = 3.6E-3$, $d = 0.6$).

Though we did not find a significant difference between different and same talker with regards to the amount of variance explained by the parietal source model, we have reason to expect that differences between same and different talker might be mediated by working memory, potentially via attentional reallocation mechanisms that place a load on WM (Giuliano et al., 2014; Engle, 2002). Parietal sources have been implicated in attentional mechanisms (e.g., Yantis et al., 2002). Therefore, we expect the degree to which parietal sources are differentially active across talkers to depend on WM. In order to test this hypothesis, we calculated correlations between WM and the fit of the source models to the scalp topography elicited by a changing talker. Specifically, we correlated WM with the difference between model fit in

Table 2
Significance windows identified by TANOVA. There were no significant interactions of congruency (match/mismatch) and talker (same/different). There were, however, differences in topography based on talker, as well as short N400 and longer P600 congruency main effects. Windows that do not pass the 40 ms duration threshold, but that are located where we had an *a priori* expectation of a main effect, are italicized.

Factor	Window Start	Window End	Length	>=40 ms
Congruency main	18	25	7	FALSE
Congruency main	*433*	*443*	*10*	*FALSE*
Congruency main	589	621	32	FALSE
Congruency main	628	675	47	TRUE
Congruency main	686	818	132	TRUE
Talker main	*124*	*147*	*23*	*FALSE*
Talker main	235	278	43	TRUE
Talker main	314	394	80	TRUE
Talker main	437	498	61	TRUE
Talker main	520	640	120	TRUE
Talker main	713	729	16	FALSE
Congruency * Talker	NULL			

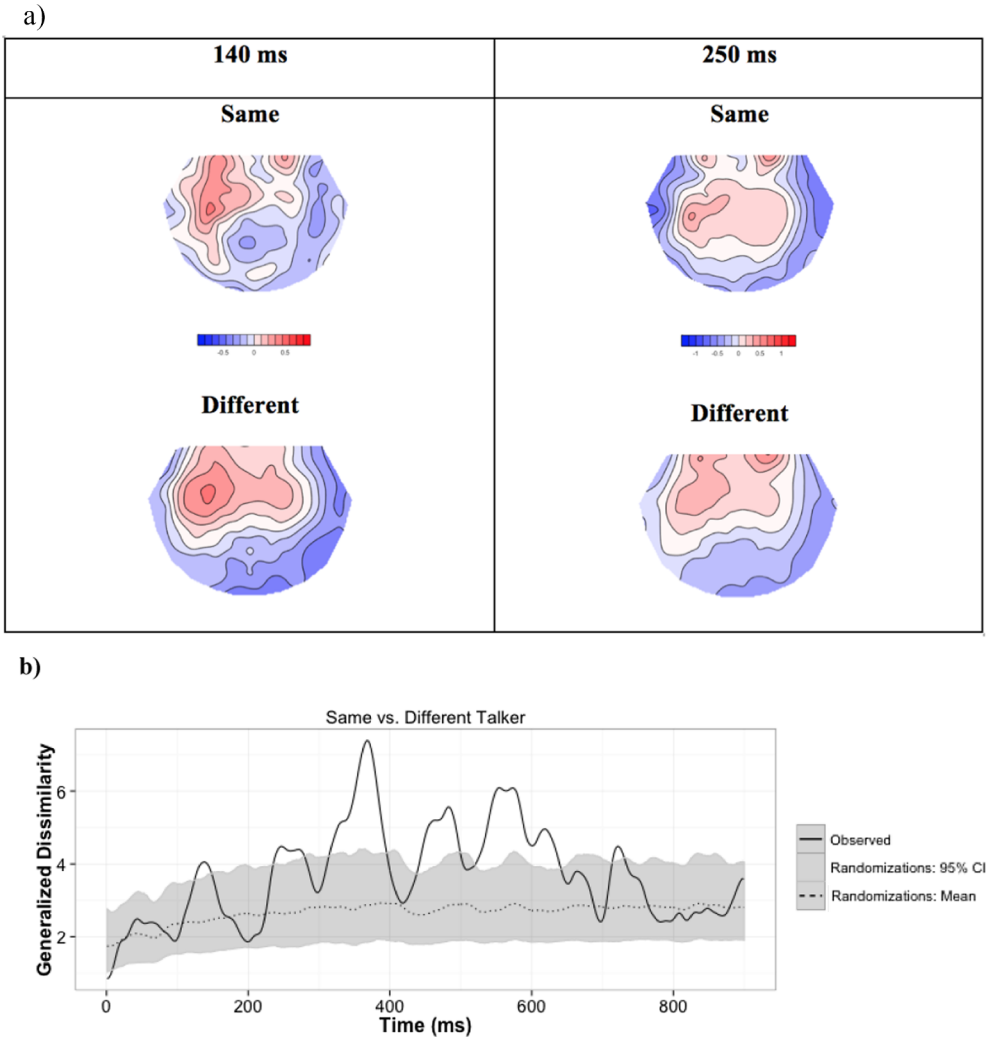

Fig. 1. N1 and P2 window topographies across talker. (a) Scalp topographies for sentence endings said by the same vs. a different talker at 140 and 250 ms post ending onset. Blue indicates negative potential; red indicates positive potential. (b) Time-varying generalized dissimilarity between raw same and different talker 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 widespread main effects of talker across the entire ERP. (d) Voltage traces for pooled same and different talker sentence endings in the nine examined ROIs. Note that negative is up and time = 0 ms corresponds to ending onset.

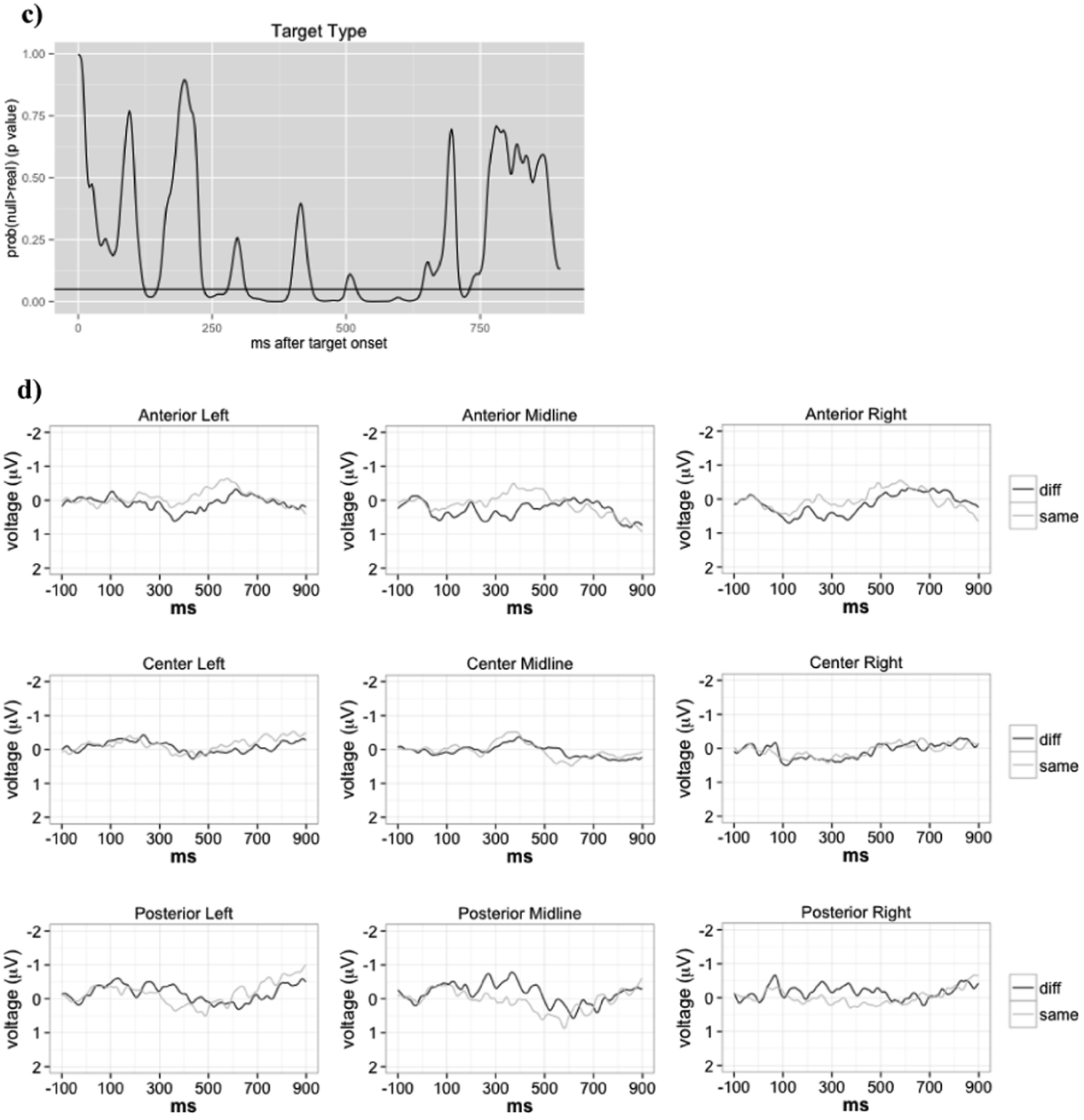

Fig. 1. (*continued*)

Table 3
TANCOVA windows with main effects and interactions of the between-subjects factor of WM.

Factor	Window Start	Window End	Length	> = 40 ms
d' main	3	29	26	FALSE
d' main	136	179	43	TRUE
d' main	220	258	38	FALSE
d' main	300	323	23	FALSE
d' main	690	700	10	FALSE
d' * Talker	NULL			

Table 4
Time windows examined in source analysis. Numbers are in milliseconds after target onset. The last window is a control in which there is not expected to be differential activity of parietal or temporal sources.

Start	Stop
124	179
220	278
628	675

the same- and different-talker conditions, that is: $RV_{same} - RV_{diff}$. In the N1 time window (124–179 ms), there was a significant correlation between WM and $RV_{same} - RV_{diff}$ for the parietal model (Fig. 2, $r = 0.51$, $p = 0.018$), which likely led to the marginally significant correlation between WM and $RV_{same} - RV_{diff}$ for the model including both temporal and parietal sources ($r = 0.41$, $p = 0.06$). The $RV_{same} - RV_{diff}$ of the temporal-only model was not significantly correlated with WM in this time window ($r = 0.34$, $p = 0.13$). This indicates that as WM increases, the parietal source model explains responses to a change in talker better

Table 5
Average residual variance (RV) for different vs. same talker in different time windows, with parietal, temporal, and "both" models tested separately. ***Bold italics*** denotes the window where the RV for different and same talker was significantly different after controlling for multiple comparisons. *Note that a lower RV indicates more variance explained by the model and thus indicates a better fit.*

Time win	Model	Different	95% CI			Same	95% CI		
124–179 ms	parietal	50.86	45.35	–	56.36	56.83	51.63	–	62.03
	temporal	40.36	34.97	–	45.74	47.00	40.29	–	53.72
	both	33.32	28.50	–	38.13	37.67	31.39	–	43.95
220–278 ms	parietal	55.35	48.67	–	62.03	62.29	56.41	–	68.17
	temporal	***42.84***	***36.69***	***–***	***48.98***	***51.24***	***45.48***	***–***	***57.01***
	both	35.08	28.23	–	41.93	42.17	35.81	–	48.52
628–675 ms	parietal	65.63	57.51	–	73.74	58.32	50.43	–	66.20
	temporal	54.76	44.99	–	64.54	45.83	36.91	–	54.75
	both	48.33	38.89	–	57.78	40.15	31.62	–	48.68

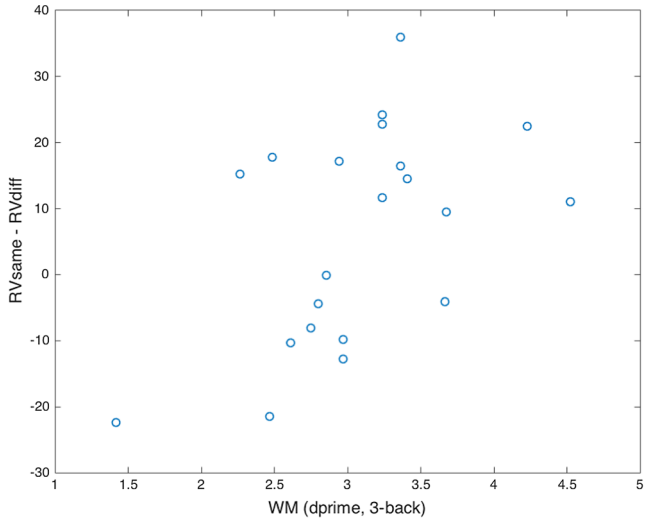

Fig. 2. Correlation between source analysis model fit and WM. For the N1 time window, correlation between WM as measured by d' on a 3-back task, and the difference in the goodness of fit of a parietal source model to scalp topography data between same and different talkers ($RV_{same} - RV_{diff}$).

than it does responses to the same talker.[1] It is worth noting that correlations between WM and $RV_{same} - RV_{diff}$ were not significant in other time windows ($r < 0.4$), supporting the idea that WM is differentially affecting the adjustment to a new talker via parietal sources specifically during the N1.

2.4. Discussion

Unfortunately, there were not distinct N1 and P2 peaks to test our hypotheses about talker change effects on peak amplitudes. This might be due to the lack of a silent gap between the end of the sentence stem and the last word; many studies examining N1 effects (e.g., Zhang, Peng, & Wang, 2013) include several hundred milliseconds of silence before the auditory stimulus of interest, although previous work with the present experiment's sentence stems followed by environmental sounds showed a clear N1-P2 complex which led us to believe that we might observe this in response to a changing talker (Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018). Despite this, our hypotheses about the underlying cortical sources active during a talker switch were largely supported. First, during the N1 window, we found a significant correlation between WM and the [same – different talker] difference in model fit for the model incorporating the parietal sources from Wong et al. (2004; Fig. 2). This indicates that for high-WM participants, the parietal sources better explain responses to a talker switch than responses to a consistent talker, whereas this pattern is reversed for low-WM participants. This finding fits nicely with findings that individuals' WM capacity affects their ability to deploy attention (e.g., Giuliano et al., 2014). These results suggest that when faced with a change in talker, high-WM participants are able to allocate attention to features of the new talker that will aid understanding. As the WM ~ $RV_{same} - RV_{diff}$ correlation is significant in the N1 time window (Fig. 2) but not the P2 time window, our results suggest that this WM-related attentional reallocation happens during the N1 and depends, at least in part, on the recruitment of the attention-related parietal sources identified by Wong et al. (2004). In contrast, low-WM participants likely have difficulty attending to aspects of the new talker's speech. These participants might be better able to attend to the familiar patterns of the same talker, which may be why the attention-related parietal sources better explain responses to the same talker for low-WM participants.

While we did not find a relationship between parietal sources and the P2 ERP, we did find support for our hypothesis that temporal sources from Wong et al. (2004) would better explain responses to the changing talker during the P2 time window (Table 5). As these temporal areas are associated with auditory processing of complex stimuli (Rauschecker & Scott, 2009) including speech (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), this finding agrees with previous work associating the P2 with auditory feature processing (e.g., Näätänen & Winkler, 1999). Interestingly, WM capacity did not correlate significantly with $RV_{same} - RV_{diff}$ for the temporal model in the P2 time window. This means that there was no systematic relationship between WM and the degree of heightened feature analysis taking place when the talker changes. Along with the results from the N1 time window, this points towards an account where, at least in processing a talker change in fluent speech, attentional mechanisms act earlier during the N1 in a WM-dependent manner via parietal cortical areas. After this, areas in temporal cortex mediate auditory feature analysis of the new talker's output during the P2.

3. Experiment 2: Do talker normalization mechanisms mediate auditory processing of environmental sounds in sentence context?

3.1. Introduction

As we have stated previously, the neural mechanisms behind understanding words and environmental sounds (ES) in the context of a spoken sentence appear to be remarkably similar (Uddin, Heald, Van Hedger, & Nusbaum, et al., 2018), arguing for domain-general mechanisms for auditory understanding, rather than specialized mechanisms for speech. Could it therefore be possible that the same

[1] Note that a high $RV_{same} - RV_{diff}$ indicates that the parietal model is better at explaining the different-talker topography than the same-talker topography, while an $RV_{same} - RV_{diff}$ of zero indicates that the parietal model is equally good at explaining the topography in the same and different talker conditions, and a negative $RV_{same} - RV_{diff}$ indicates that the parietal model is better at explaining the same-talker topography.

mechanisms involved in supporting a switch in talkers are also involved in allowing listeners to understand non-speech in speech context? It is no surprise that talker normalization is treated as solely a speech phenomenon in the literature (see Weatherholtz & Jaeger, 2016, for a review)—the notion of a "talker" is meaningless when dealing with non-human-produced environmental sounds such as train whistles or dogs barking. However, Laing et al. (2012) provide compelling evidence that the talker switch cost is not dependent on the language identity of the stimuli. Rather, it can be affected by the frequency range of tone sequences, as well as by speech stimuli.

Though environmental sounds are not speech, they do share some characteristics with words said by a novel talker. Like words, ES have easily-identified meanings. Previous work has shown that, both behaviorally and neurally, their meaning is processed in sentence context in a way similar to words (Uddin, Heald, Van Hedger, & Klos, et al., 2018, Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018). There are some general similarities between a change in talker and the presentation of environmental sounds in sentence context in that they are distinct in acoustic source compared to the rest of the sentence. Also, as with a change in talker, the spectral characteristics of environmental sounds differ acoustically from the previous speech, although such differences will be greater for environmental sounds. If the neural mechanisms that support speech recognition when there is a talker change are in fact more general than previously thought, WM, as well as the parietal and temporal sources identified by Wong et al. (2004) involved in selective attention and speech feature analysis, might play a similar role in adjusting to understanding non-speech in speech context. To test this, we carried out further analyses on previously reported ERP data comparing sentence understanding for speech and non-speech environmental sounds (Uddin, Heald, Van Hedger, & Nusbaum, et al., 2018).

We hypothesized that attention is recruited in order to process relevant acoustic features of the environmental sounds that are required for understanding given the shift in source. In sentence context, the environmental sounds are likely to have alternative interpretations, thereby loading WM and shifting attention to specific properties of the signal. In previous research, we have shown that N1 and P2 amplitudes are much greater for ES than for speech in sentence context (Uddin, Heald, Van Hedger, & Nusbaum, et al., 2018). In light of previous work implicating these ERPs in attentional processes (e.g., Picton & Hillyard, 1974), this finding supports the idea that attention recruitment is greater for ES in than words. However, the relationship of WM to these N1 and P2 effects has not been demonstrated. We hypothesize that WM-dependent attentional mechanisms are recruited to process environmental sounds (ES) in sentence context in a way similar to its recruitment for spoken words when the talker changes.

As in Experiment 1, we therefore predict that the TANCOVA will reveal either interactions between target type and WM, or main effects of WM on scalp topography, within N1 and P2 time windows. In Experiment 1, we found main effects of WM on scalp topography, with no talker-related interactions. As discussed in Section 2.1, this is expected if talker-change-related N1 and P2 effects occur to the same degree for both high-and low-WM participants, while baseline N1 and P2 activity in the same talker condition differs according to WM capacity. While it is possible that we could find this pattern for ES as well, it is worth examining the implications of increased processing difficulty for ES. Behavioral data suggest that ES are harder to process in sentence context than spoken words, although both are easily understood and similarly affected by factors such as constraint (Uddin, Heald, Van Hedger, & Klos, et al., 2018). ES are also farther from words said by a consistent talker, both psychologically and acoustically, than words said by a new talker. It is therefore possible that ES will place a larger load on WM than a changing talker, either through increased demands for recruitment of attention, or through increased ambiguity leading to the maintenance of more information in WM. This might lead to a pattern in which high WM participants are better able to modify attention to understand the ES, while low WM participants are limited in this ability by their WM capacity. We might therefore find interactions between WM and target type. Finally, if our hypothesis is not correct, and recruitment of attention for ES processing is not WM-dependent, we would expect no main effects or interactions involving WM in these time windows.

Perhaps the most interesting question is whether cortical sources implicated in talker normalization (Wong et al., 2004) are also active during comprehension of ES in spoken sentence context. We hypothesized that mechanisms for such adaptation are more cognitive-general than previously thought, and will be involved in adaptation to ES as well as a changing talker. In experiment 1, we demonstrated a WM-dependent contribution of parietal sources to changes in the N1 during talker normalization. We also found a contribution of temporal sources to talker-normalization-related changes in the P2. As discussed in Section 2.4, these results suggest heightened attention reallocation during the N1 period, and heightened auditory feature processing during the P2. We expect to find the same pattern for ES in sentence context. If our hypothesis is not correct, we would not expect to find significant differences in parietal and temporal contributions to ES and speech topographies during the N1 and P2 time windows. Instead, these dipole models should explain both types of responses equally well.

3.2. Methods

3.2.1. Participants

As this experiment consisted of new analyses involving previously published data (Uddin, Heald, Van Hedger, & Nusbaum, et al., 2018), the participants were the same as described therein. Specifically, they consisted of 23 (8 female, 13 male, 1 agender, 1 genderfluid) adults from the University of Chicago and surrounding community, with a mean age of 22.1 years (SD: 3.7, range: 18–29). Fifteen were right-handed and eight were left-handed[2]. 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 received either 3 course credits (n = 11) or $30 cash (n = 12) for their participation in the study.

3.2.2. Working memory testing and analysis

Working memory was assessed by performance on an auditory *n*-back task, which was administered prior to application of the electrodes for the EEG, as in Experiment 1. The *n*-back task and the determination of d′ from the data proceeded identically to Experiment 1. As in Experiment 1, only 3-back d′ scores were used in further analysis.

3.2.3. Stimuli

The stimuli were identical to those in Experiment 1, with the difference that instead of half the sentences ending in a changing talker, half of them ended in an environmental sound that was matched in meaning to the ending word of the sentence (Uddin, Heald, Van Hedger, & Klos, et al., 2018, Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018). Like the spoken words, the environmental sounds were digitized at 44.1 kHz with 16 bits of resolution and amplitude normalized to the same RMS level (~70 dB SPL). The stimuli can be found at https://osf.io/asw48/.

As in Experiment 1, sentence stems and endings were spliced

[2] This experiment had a higher proportion of left-handers than the previous one, which is a potential limitation of our comparison between the two experiments. However, our experimental questions concerned general attentional network activity rather than specific lateralized language processes, rendering handedness less important with respect to our hypotheses. As a further check, we analyzed right-handed participants alone, and found that they exhibited the same patterns as the whole group, suggesting that there are no important handedness-related differences affecting the parameters of interest (Fig. S.2).

together in Matlab to form continuous sentences with no audible acoustic artifacts. The block structure and conditions of this experiment were the same as in Experiment 1, except that blocks of 32 sentences with a different talker saying the sentence-final word were replaced by blocks in which the sentence ended in an environmental sound. Stimuli were presented at 65–70 dB over insert earphones (3M E-A-RTone Gold) using Matlab 2015 with Psychtoolbox 3.

3.2.4. Testing procedure and EEG setup

The testing procedure and EEG setup were identical to Experiment 1, with the difference that, due to mobility difficulties, one of the participants could not sit in the EGI geodesic dome to photograph the precise location of the electrodes.

3.2.5. Data preprocessing

EEG preprocessing proceeded the same way as in Experiment 1; this is also described in Uddin, Heald, Van Hedger, and Nusbaum, et al. (2018). As in experiment 1, participants were removed from further analysis if they had 50% or more artifact-contaminated trials in any one condition. One participant was removed for this reason; they 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 to estimate electrode locations (Electrical Geodesics Inc., Eugene, OR).

3.2.6. Analyses

3.2.6.1. Topographic analyses. Averaged waveforms were generated as in Experiment 1. For significance testing, a TANCOVA using a null hypothesis distribution generated from 5000 randomizations was performed in RAGU as in Experiment 1. The TANCOVA included main within-subjects factors of congruency (match or mismatch) and target type (environmental sound or spoken word), and WM (d′ on the 3-back task) as a continuous between-subjects factor.

As in Experiment 1, we performed further source analysis using time windows identified by the TANCOVA as having main effects of target type, WM, or interactions between the two.

3.2.6.2. Source analysis. Source analysis proceeded as in Experiment 1, using the same cortical source dipoles. The only difference was that instead of comparing responses to same vs. different talker, we compared responses to speech vs. environmental sounds. If these sources from Wong et al. (2004) better explain brain responses to environmental sounds, this would provide evidence that these brain areas are active not only in talker normalization, but in more general processing of acoustic source change.

For this experiment, the time windows identified by the TANCOVA for source analysis were 122–137 ms, 180–219 ms, and 273–293 ms. It should be noted that unlike Experiment 1, these windows were marked by interactions between WM and target type, rather than main effects of each.[3]

3.3. Results

We hypothesized that, like a change in talker, ES at the end of a sentence would recruit attention in a WM-dependent manner. As such, we expected to find either interactions between target type and WM, or main effects of WM on scalp topography, within time windows (i.e., N1 and P2) previously implicated in the deployment of attention. While the TANCOVA did not identify any significant main effects of WM on scalp topography, it did identify time windows in the vicinity of the N1 and P2 where there were significant interactions between WM and target type (ES or speech; $p < 0.05$, Table 6, Table S.2).

Table 6
TANCOVA windows showing interactions between WM and target type for Experiment 2. There were no main effects of WM.

Factor	Window Start	Window End	Length
d' * target type	122	137	15
d' * target type	180	219	39
d' * target type	273	293	20

In order to test our hypothesis that the cortical sources identified by Wong et al. (2004) would also be involved in processing ES in sentence context, we submitted these WM-by-target type interaction windows to source analysis. If our hypothesis is correct, we should find a pattern similar to Experiment 1, in which ES scalp topographies are better described by the aforementioned sources, particularly in N1 and P2 time windows. Except for the difference in the source data and time windows, this analysis (including statistical testing) was identical to that in experiment 1. We used a Bonferroni-corrected threshold of $p \leq 5.6E-3$ for nine tests. We found that the model including both parietal and temporal sources explained the ES data significantly better than the word data in both the 122–137 and 180–219 ms time windows ($V = 207$, $p = 3.7E\text{-}3$ and $V = 205$, $p = 4.6E-3$, respectively; Table 7), and approached significance in the 273–293 time window ($V = 201$, $p = 7.0E\text{-}3$; Table 7). This comparison did not reach significance for the other models (i.e., temporal-only and parietal-only). Unlike experiment 1, the fit of the models did not correlate significantly with WM; that is, participants with high WM did not exhibit a different pattern than participants with low WM when it came to the difference in model fit between speech and ES.

3.4. Discussion

For this experiment, we hypothesized that the same mechanisms underlying talker normalization would also be active in facilitating the transition between speech and meaningful nonspeech. First, we had hypothesized that ES in sentence context would elicit responses in the N1 and P2 time windows that differ based on WM, due to heightened attentional reallocation to relevant features of the ES. In agreement with this hypothesis, the TANCOVA revealed significant scalp topography effects due to interactions between WM and target type. Though this is in contrast with the main effects of WM found in experiment 1, it was not entirely unexpected. As suggested in Section 3.1, the increased difficulty and ambiguity of ES (compared to speech) might place a higher load on WM than a simple change in talker. As a result, individual differences in WM capacity might become more important in the period of attention deployment during the N1 and P2. Our results suggest that this is indeed what happens for ES in sentence context.

Second, we hypothesized that the cortical sources implicated in talker normalization would also mediate ES processing in sentence context. This hypothesis was also supported; dipole models including both parietal and temporal sources explained ES topographies significantly better than speech topographies in both N1 and P2 time windows. This difference was marginally significant in a third, late-P2 time window (273–293 ms). These results suggest that sources previously implicated in talker normalization are active in processing a switch from speech to ES. This finding agrees with previous work suggesting that talker normalization processes might actually be related to more general auditory processing (Laing, Liu, Lotto, & Holt, 2012). It thus appears that attentional processes depending on parietal areas, and auditory feature analysis depending on temporal sources, are common

[3] There were no windows with main effects of WM on scalp topography. Main effects of target type extended from approximately 100 to 700 ms, and thus a) did not provide discrete N1 and P2 time windows, and b) overlapped with the WM interactions.

Table 7
Average RV for ES vs. spoken words in different time windows, with parietal, temporal, and "both" models tested separately. Bold italics denotes the windows where the RV for ES and words was significantly different after controlling for multiple comparisons. *Note that a lower RV indicates more variance explained by the model and thus indicates a better fit.*

Time win	Model	ES	95% CI			Word	95% CI		
122–137 ms	parietal	49.78	42.81	–	56.74	63.35	54.90	–	71.79
	temporal	38.50	32.40	–	44.60	51.96	43.27	–	60.66
	both	***32.08***	***26.17***	***–***	***37.99***	***44.98***	***36.53***	***–***	***53.43***
180–219 ms	parietal	50.99	43.79	–	58.19	63.30	55.26	–	71.34
	temporal	37.77	31.23	–	44.32	51.28	41.82	–	60.74
	both	***31.56***	***24.72***	***–***	***38.40***	***44.49***	***35.56***	***–***	***53.42***
273–293 ms	parietal	52.33	45.12	–	59.55	64.35	55.21	–	73.48
	temporal	39.25	32.11	–	46.38	53.73	43.13	–	64.33
	both	31.20	24.72	–	37.68	46.80	36.76	–	56.83

features between ES processing in sentence context and processing of a changing talker.

Unlike experiment 1, there was no evidence that different sources were active in the N1 and P2 time windows in the current experiment. Whereas in experiment 1, parietal sources were implicated during the N1, and temporal sources were implicated during the P2, both sources were implicated at both time points in the current experiment. There are two reasons this might be the case. The first reason follows from the fact that ES are likely more difficult to process in sentence context than a talker change, as they are more different from speech, as well as less commonly encountered in sentence context in everyday life. Due to increased difficulty, perhaps different processing stages for ES take longer, and therefore overlap temporally, leading to both sources being active at both time points. Another possibility is that temporal smearing occurred due to averaging. There is a much closer temporal correspondence between two different talkers saying the same words, than there is between the words and matched environmental sounds. For example, it takes much longer to recognize the sound of a toilet flushing than it does to say the word "toilet". In this case, the time course of parietal and temporal source activation could be the same for ES and talker switches, but the temporal resolution is not good enough to detect this for the ES stimuli. Future experiments could use ES that are closely matched in length to their corresponding words in order to circumvent this issue.

4. General discussion

Talker normalization is a process, thought to require WM-dependent deployment of perceptual attention, that allows listeners to understand different talkers despite wide variability in acoustic characteristics of different talkers' speech (Nusbaum & Morin, 1992). Previous research has found N1 ERP amplitude increases in response to a change in talker, suggesting that attention is reallocated to features of the new talker's speech in order to aid talker normalization (Kaganovich et al., 2006). Moreover, fMRI studies have implicated parietal and temporal regions in processing changes in talker (Wong et al., 2004), suggesting that talker normalization involves both deployment of attention and acoustic feature analysis of the new talker's speech.

However, the time course of activity in the temporal and parietal talker normalization sources has not yet been identified. Moreover, talker normalization effects on the N1 ERP have not been examined in an EEG context with source analysis. Although attentional effects, as well as effects related to auditory feature analysis, have been connected to the P2 ERP, there remain few studies examining the P2 in talker normalization. Finally, the role of WM in these neural mechanisms of talker normalization has yet to be elucidated.

We used high-density EEG to examine scalp topographies in response to a change in talker. We hypothesized that we would find amplitude increases in the N1 and P2 ERPs when the talker switched, consistent with a reallocation of attention to new acoustic features. Unfortunately, while scalp topographies differed significantly between same- and different-talker responses during N1 and P2 time windows (Fig. 1b,c), no clear N1 and P2 peaks were obtained, either on an average or an individual level (Fig. 1a,d), with the exception of one participant (individual traces can be found at https://osf.io/x8dau). The weakness and variability of these potentials made it impossible to perform simple amplitude or latency analyses. Though we had reason to expect that an acoustic source change might elicit a strong N1-P2 complex (as it does for sentences terminating in environmental sounds—Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018), it is possible that the lack of a silent interval before the words of interest is responsible. For example, the talker normalization experiments by Zhang et al. (2013) employed a jittered silent interval of 300–500 ms before the last word. Further, the presence of background noise reduces the amplitude of the N1 and P2 in response to speech (Koerner & Zhang, 2015), suggesting that acoustic continuity (as was the case in our stimuli) may ablate these ERPs. Finally, the relative similarity of the words said by the two different talkers may have reduced the N1 and P2 to be undetectable when compared to the environmental sounds experiment, in which the sentence endings are conceptually quite distinct from the preceding speech. However, the existence of significant differences in scalp topography between the same- and different-talker conditions provides evidence that talker normalization processes do affect neural processing in the N1 and P2 time windows. Moreover, we were still able to perform source analyses on topographies during N1 and P2 time windows to assess the relationship of these topographies to the talker normalization sources in Wong et al. (2004). We also assessed the effects of WM on the scalp topographies in the same- and different-talker conditions.

With respect to WM, we hypothesized that individual WM capacity would predict differences in processing between the two talkers. As higher individual working memory capacity has been associated with greater attentional modulation of the N1 and P2 ERPs (Giuliano et al., 2014), we expected an interaction between talker and WM capacity. Given work highlighting active cognitive processes in speech perception (Heald & Nusbaum, 2014a), regardless of changes in talker, we also reasoned that we might find main effects of WM rather than interactions. This latter hypothesis was supported by main effects of WM on scalp topography in both N1 and P2 time windows (Table 3). This finding supports the idea that similar attentional mechanisms are recruited to process a consistent talker and a changing talker, but that these mechanisms are recruited to a greater extent for high-WM participants.

Finally, from one theoretic perspective, WM may not affect the deployment of attention, but rather attention acts as a "gatekeeper", focusing on the aspects of the input that are going to be retained in WM (cf. Awh, Vogel, & Oh, 2006). In this view, WM might be related to talker normalization and attention via a gating mechanism, in which attention focuses on aspects of the new talker's speech that need to be manipulated later in WM in order to produce understanding. Interestingly, the present results appear to refute this idea; if the deployment of attention itself is not influenced by WM, we would likely not see main effects of WM on ERPs (namely the N1 and P2) that are known to be modulated by attention. Thus, it seems most likely that WM is mediating attentional mechanisms which, in turn, are influencing the N1 and P2.

With regards to the source analyses, we expected to find that parietal sources (previously implicated in attention; Wong et al., 2004) were more active in response to a changing talker, due to attentional mechanisms involved in focusing on the idiosyncratic characteristics of the new talker. As both the N1 and P2 have been found to be modulated by attention, we hypothesized that these differences in parietal activation would be found in both N1 and P2 time windows. We did not find significant differences in the fit of the parietal source dipole models to same- vs. different-talker scalp topographies in either the N1 or P2

time window (Table 5). However, in the N1 time window, we found a significant correlation between individual WM capacity and the difference in how well the parietal model fit responses to the two different talkers (Fig. 2). Specifically, we found that for high-WM participants, the parietal sources explained responses to the changing talker better than they explained responses to the consistent talker. The effect was reversed for low-WM participants, such that the parietal sources better explained responses to the consistent talker. This effect was confined to the N1 window. These results suggest that WM modulates the reallocation of attention that is necessary in talker normalization. While high-WM participants were presumably able to recruit parietal areas to focus attention on relevant characteristics of the new talker, low-WM participants might not have been able to identify the right talker characteristics to aid processing. Instead, their attention may have been more effectively captured by the familiar speech patterns of the same talker who produced the sentence stems. While previous studies have examined the role of WM in talker normalization primarily by placing participants under a WM load (e.g., Nusbaum & Morin, 1992), the present study shows a neural effect of individual WM capacity on talker normalization processes. Such differences may be too small to pick up in a behavioral response time paradigm, but they support the interpretation that talker normalization requires WM-mediated attention reallocation. Finally, we did not find any such effects for the parietal sources during the P2 time window. These results suggest that, at least during fluent speech, WM-mediated attention effects in talker normalization are confined to the N1 time window, i.e. earlier in the processing stream.

As the temporal sources identified by Wong et al. (2004) are related to speech processing and, more generally, processing of complex auditory stimuli (e.g., Rauschecker & Scott, 2009), we hypothesized that these sources might be more active for a new talker during the P2 time window. This is because the P2 has previously been implicated in the analysis of acoustic features (Näätänen & Winkler, 1999). This is indeed what we found; during the P2 time window, Wong et al.'s (2004) temporal sources better explained scalp topographies in response to a changing talker (as opposed to a consistent talker; Table 5). This suggests that a change in talker elicits heightened auditory processing, likely due to acoustic feature analysis of the new talker's speech. Taken together with the results from the N1 time window, these results suggest a time course in which parietal sources modulate attention in a WM-dependent manner during the N1. After this, temporal areas are recruited for a more fine-grained analysis of the acoustic input during the P2.

In our second experiment, we hypothesized that neural mechanisms underlying talker normalization would also mediate the processing of environmental sounds in spoken sentence context. Such a finding would be in line with previous work suggesting that talker normalization (Laing et al., 2012), and language understanding in general (Uddin, Heald, Van Hedger, & Klos, et al., 2018, Uddin, Heald, Van Hedger, & Nusbaum, et al. 2018), rely on cognitive-general mechanisms that also facilitate other types of auditory processing. However, it is possible that, due to the specialization of language, the adjustment to acoustic change is dealt with differently for speech than for other auditory stimuli. Indeed, talker normalization is only rarely presented as an instance of a more general acoustic change adaptation phenomenon (a notable exception being Laing et al., 2012). If this is the case, we might expect that understanding an ES in sentence context might draw on different resources, such as cortical areas specialized for environmental sounds (cf. Leech & Saygin, 2011), rather than WM-mediated attentional reallocation.

The first analysis to test this hypothesis was a TANCOVA analysis of scalp topographies including a between-subjects factor of WM. We expected to find main effects of WM, or interactions between target type (ES vs. speech) and WM. This is because processing an ES in sentence context should require recruitment of attention, which, as in talker normalization, should rely on WM. There is precedent for this idea; the use of WM in the deployment of attention is thought to be a domain-general resource, not a speech-specific one (Engle, 2002; Giuliano et al., 2014). Therefore, it is reasonable to expect WM-mediated attentional differences in ES processing as well as in talker normalization. This hypothesis was supported by significant interactions between WM and target type in time windows encompassing the N1 and P2 ERPs (Table 6). Such interactions show that differences in ES and word processing in sentence context depend on WM capacity.

There is one important difference between the WM results in experiments 1 and 2. In experiment 1, there were main effects of WM on scalp topography in the N1 and P2 time windows (Table 3). In experiment 2, however, there were interactions between WM and target type (Table 6). This likely arises from the fact that while words are fluent and highly practiced, ES are more difficult to understand, and likely involve a higher level of ambiguity with regards to their identity (cf. Uddin, Heald, Van Hedger, & Klos, et al., 2018). This difference in difficulty might lead to a greater load being placed on WM during ES processing, as greater ambiguity likely leads to the maintenance of more possible interpretations in WM (cf. Nusbaum & Schwab, 1986). This would lead to a large WM load difference between ES and spoken words. Such a large WM load could accentuate WM-capacity-based processing differences between ES and words. This could lead to an interaction between target type and WM capacity, rather than the main effect that is found in the talker normalization experiment. Importantly, according to these explanations, there are not qualitatively different processes happening for ES and words. Instead, WM load likely increases in a graded fashion, such that it is lowest in processing a consistent talker, in the middle for processing a changing talker, and highest for processing ES.

For experiment 2, we also hypothesized that ES processing in sentence context would activate the same cortical areas as talker normalization, as it should involve both heightened attention reallocation and acoustic feature analysis in conjunction with WM to resolve ambiguity of interpretation. This hypothesis was supported by the source analysis; we found significant differences between ES and words with regards to the fit of dipole models including both parietal and temporal sources. Namely, models including both temporal and parietal sources better explained the variance for scalp topographies elicited by ES (when compared to topographies elicited by speech). These differences extended through both N1 and P2 time windows (Table 7). It therefore appears that, like talker normalization, a switch from language to ES involves attention reallocation and acoustic feature analysis mediated by the parietal and temporal cortical areas identified by Wong et al. (2004).

Whereas in experiment 1, effects related to parietal sources were largely confined to the N1 time window, and those related to temporal sources were largely confined to the P2, both sources were implicated in both time periods for experiment 2. As discussed in Section 3.4, this is likely due to temporal smearing as a result of increased stimulus length variability in experiment 2. Future work might examine ES and words that are matched in length; perhaps such experiments will reveal the parietal-then-temporal time course that we observed for a change in talker.

5. Conclusions

In summary, our analyses extend work by Wong et al. (2004) by indicating that in processing a change in talker, parietal sources are active in a WM-dependent manner in the N1 time window, and auditory-processing-related temporal sources are recruited later in the P2 time window. These results suggest a talker normalization processing stream that first involves WM recruitment to reallocate attention towards relevant features of the new talker's speech. These attentional processes involve parietal cortical areas, and happen during the N1 ERP. These effects then appear to be followed by auditory processing mechanisms that are not WM-dependent, which are mediated by

temporal cortical areas, and occur during the P2. Further work should be done to determine whether this processing stream is unique to talker normalization in fluent speech, or whether it also underlies talker normalization in vowel, syllable, or single word stimuli.

Furthermore, our results from a new analysis of previously collected ES data provide a novel link between the talker switch cost and processing of meaningful auditory non-speech stimuli: they suggest that mechanisms utilized in processing a switch in talkers may be utilized in understanding non-speech in the context of speech. These results suggest that the same cognitive-general mechanisms underlie the ability to quickly process both ES and words in sentence context, in turn supporting the idea that talker normalization and speech-to-nonspeech adaptation are two sides of the same coin.

Acknowledgements

The authors thank Leslie Kay, Dan Margoliash, and Nicho Hatsopoulos for helpful feedback on this work. They also thank Joel Snyder, Tahra Eissa, and Geoff Brookshire for data analysis advice, Peter Hu for assistance with data collection, and Willow Uddin-Riccio for assistance preparing the manuscript. This research is based on work supported by the Multidisciplinary University Research Initiatives (MURI) Program of the Office of Naval Research through grant, DOD/ONR N00014-13-1-0205 and by the University of Chicago MSTP Training Grant T32GM007281.

Note on Open Data and Materials

Stimuli, data and code for stimulus presentation and analysis are available on OSF at https://osf.io/x8dau/.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2019.104722.

References

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology, 80*(1), 23–34. https://doi.org/10.1016/j.biopsycho.2008.01.015.

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201–208. https://doi.org/10.1016/j.neuroscience.2005.08.023.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5*(3), 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X.

Awh, E., & Vogel, E. K. (2008). The bouncer in the brain. *Nature Neuroscience, 11*(1), 5–6. https://doi.org/10.1038/nn0108-5.

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

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

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160.

Giuliano, R. J., Karns, C. M., Neville, H. J., & Hillyard, S. A. (2014). Early auditory evoked potential is modulated by selective attention and related to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience, 26*(12), 2682–2690. https://doi.org/10.1162/jocn_a_00684.

Heacock, R. M., Pigeon, A., Chermak, G., Musiek, F., & Weihing, J. (2019). Enhancement of the auditory late response (N1–P2) by presentation of stimuli from an unexpected location. *Journal of American Academy of Audiology, 30*(6), 451–458 https://doi.org/info:doi/10.3766/jaaa.17047.

Heald, S. L. M., & Nusbaum, H. C. (2014a). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*. https://doi.org/10.3389/fnsys.2014.00035.

Heald, S. L. M., & Nusbaum, H. C. (2014b). Talker variability in audio-visual speech perception. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00698.

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research, 1114*(1), 161–172. https://doi.org/10.1016/j.brainres.2006.07.049.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16. https://doi.org/10.1068/v070821.

Koenig, T., Kottlow, M., Stein, M., & Melie-Garcia, 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. https://doi.org/10.1155/2011/938925.

Koerner, T. K., & Zhang, Y. (2015). Effects of background noise on inter-trial phase coherence and auditory N1–P2 responses to speech stimuli. *Hearing Research, 328*, 113–119. https://doi.org/10.1016/j.heares.2015.08.002.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. https://doi.org/10.1126/science.7350657.

Laing, E. J. C., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00203.

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. https://doi.org/10.1016/j.bandl.2010.11.001.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., ... Swann, A. C. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology, 46*(5), 1059. https://doi.org/10.1111/j.1469-8986.2009.00845.x.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*(3), 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x.

Morin, T. M., & Nusbaum, H. C. (1989). Cues to perceptual normalization of talker differences. S100 S100 *The Journal of the Acoustical Society of America, 86*(S1), https://doi.org/10.1121/1.2027256.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379–390. https://doi.org/10.3758/bf03210878.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125*(6), 826–859. https://doi.org/10.1037//0033-2909.125.6.826.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x.

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. *Talker Variability in Speech Processing,* 109–132.

Nusbaum, H. C., & Morin, T. M. (1992). *Paying attention to differences among talkers. Speech Perception, Production and Linguistic Structure*. IOS Press.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. *Pattern Recognition by Humans and Machines: Speech Perception: Vol. 1*, (pp. 113–157). San Diego: Academic Press.

Peng, C., Hu, C., & Chen, Y. (2018). The temporal dynamic relationship between attention and crowding: Electrophysiological evidence from an event-related potential study. *Frontiers in Neuroscience, 12*. https://doi.org/10.3389/fnins.2018.00844.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175–184.

Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191–200. https://doi.org/10.1016/0013-4694(74)90156-4.

Potts, G. F., & Tucker, D. M. (2001). Frontal evaluation and posterior representation in target detection. *Cognitive Brain Research, 11*, 147–156. https://doi.org/10.1016/S0926-6410(00)00075-6.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718–724. https://doi.org/10.1038/nn.2331.

Shen, W., Fiori-Duharcourt, N., & Isel, F. (2016). Functional significance of the semantic P600: Evidence from the event-related brain potential source localization. *Neuroreport, 27*(7), 548–558. https://doi.org/10.1097/WNR.0000000000000583.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18*(1), 1–13. https://doi.org/10.1162/089892906775250021.

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. https://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. https://doi.org/10.1016/j.cognition.2017.12.009.

Uddin, S., Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2018). Hearing sounds as words: Neural responses to environmental sounds in the context of fluent speech. *Brain and Language, 179*, 51–61. https://doi.org/10.1016/j.bandl.2018.02.004.

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition, 140*, 95–110. https://doi.org/10.1016/j.cognition.2015.03.012.

Widmann, A., & Schröger, E. (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. *Frontiers in psychology, 3*, 233. https://doi.org/10.3389/fpsyg.2012.00233.

Weatherholtz, K., & Jaeger, T. F. (2016). Speech perception and generalization across talkers and accents. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.95.

Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience, 16*(7), 1173–1184. https://doi.org/10.

1162/0898929041920522.
Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience, 5*(10), 995–1002. https://doi.org/10.1038/nn921.
Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology, 13*(2), 187–193. https://doi.org/10.1016/S0959-4388(03)00033-3.
Zendel, B. R., & Alain, C. (2014). Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiology of Aging, 35*(1), 55–63. https://doi.org/10.1016/j.neurobiolaging.2013.06.022.
Zhang, C., Peng, G., & Wang, W. S.-Y. (2013). Achieving constancy in spoken word identification: Time course of talker normalization. *Brain and Language, 126*(2), 193–202. https://doi.org/10.1016/j.bandl.2013.05.010.