



Published in final edited form as:

Biol Psychol. 2009 January ; 80(1): 23–34. doi:10.1016/j.biopsycho.2008.01.015.

Listeners modulate temporally selective attention during natural speech processing

Lori B. Astheimer and Lisa D. Sanders

Department of Psychology and Neuroscience and Behavior Program, Tobin Hall, University of Massachusetts, Amherst, MA, USA 01003

Abstract

Spatially selective attention allows for the preferential processing of relevant stimuli when more information than can be processed in detail is presented simultaneously at distinct locations. Temporally selective attention may serve a similar function during speech perception by allowing listeners to allocate attentional resources to time windows that contain highly relevant acoustic information. To test this hypothesis, event-related potentials were compared in response to attention probes presented in six conditions during a narrative: concurrently with word onsets, beginning 50 and 100 ms before and after word onsets, and at random control intervals. Times for probe presentation were selected such that the acoustic environments of the narrative were matched for all conditions. Linguistic attention probes presented at and immediately following word onsets elicited larger amplitude N1s than control probes over medial and anterior regions. These results indicate that native speakers selectively process sounds presented at specific times during normal speech perception.

Keywords

speech perception; selective attention; temporal orienting; ERP; auditory; N1; Nd

1. Introduction

The complexity and rapidly changing nature of speech signals present significant challenges to auditory sensory and neuroperceptual systems. Considering the amount of redundant information and irrelevant variability in speech, one such challenge is determining which of the overwhelming number of acoustic changes need to be processed in detail. Selective attention, the preferential processing of information selected on the basis of a simple feature, has proven to be extremely important in other complex perceptual tasks. The current study was designed to test the hypothesis that listeners employ temporally selective attention during speech perception to enhance processing of information presented at specific times.

1.1 Selective attention

The preponderance of research on attention has focused on spatial selection. There is ample evidence that both exogenous and endogenous cues can be used to direct selective attention to

Corresponding Author: Lori B. Astheimer, Department of Psychology, Tobin Hall, University of Massachusetts, Amherst, MA 01003, phone: 413-545-4761, fax: 413-545-0996, email: lastheim@nsm.umass.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

specific regions in space and that doing so results in improved behavioral responses to information presented at those locations (for reviews see: Cave and Bichot, 1999; Driver, 2001; Jonides and Irwin, 1981; Scharf, 1998). Spatially selective attention has also been shown to affect early neuroperceptual processing of both visual and auditory information. Event-related potentials (ERPs) elicited by images and sounds at attended compared to unattended locations have typically been shown to differ by around 80 ms after onset (Hillyard and Anllo-Vento, 1998; Luck et al., 1994; Picton and Hillyard, 1974; Woldorff et al., 1987). More specifically, sounds presented at attended locations, including speech, elicit larger auditory onset components including the first negative (N1) and second positive (P2) peaks (Hansen et al., 1983; Hillyard, 1981; Hillyard et al., 1973b; Hink and Hillyard, 1976; Schwent and Hillyard, 1975). Additional processing, distinct from the typical auditory onset response, is also evident for sounds at attended locations as indexed with a processing negativity (PN) or negative difference (Nd) that partially overlaps with the N1 and P2 time windows (Alho et al., 1987; Näätänen, 1982; Schröger and Eimer, 1993).

1.2 Temporally selective attention

Evidence from both the visual and auditory modalities indicates that observers can also allocate attention on the basis of simple features other than location (e.g., color, orientation, pitch) (Cave, 1999; Chen et al., 2007; Maunsell and Treue, 2006; Mondor and Lacey, 2001; Serences et al., 2005). Recently, attention directed to the time at which stimuli are presented has been shown to affect behavioral responses and neurophysiological indices as well (for reviews see: Correa et al., 2006a; Nobre et al., 2007). Initial studies of temporal orienting employed a modified version of the Posner spatial cuing paradigm (Posner, 1980) in which a cue presented at the onset of each trial provides information about where a target is most likely to occur. In analogous studies of temporally selective attention, the cue provides information about which of two possible *times* a target is most likely to occur. Valid temporal cues result in faster detection of stimuli presented at the earlier time (Coull and Nobre, 1998; Griffin et al., 2002; Miniussi et al., 1999). Similar studies that included more than two possible target times (Griffin et al., 2001) or catch trials on which no target occurred and responses had to be withheld (Correa et al., 2004; Correa et al., 2006b) have shown that temporally selective attention affects processing of stimuli presented at the later times as well.

To determine if temporally selective attention, like spatially selective attention, affects early perceptual processing, several studies have reported ERPs recorded while participants completed a temporal cuing task. Initial studies indicated that temporally selective attention affects processing indexed by the P300 (Griffin et al., 2001, 2002; Miniussi et al., 1999), indicative of response selection and preparation rather than early perception. However, in one of these experiments that required participants to make a difficult peripheral discrimination (Griffin et al., 2002 exp 1) and a more recent study that employed a perceptually demanding task (Correa et al., 2006a) temporal orienting was shown to modulate visual evoked potentials as early as 120 ms after onset.

Studies of temporally selective attention in the auditory modality have employed sustained attention paradigms in which listeners are asked to attend to the same time after cue onset for an entire block of trials and are never required to make a response to stimuli that occur at an unattended time (Lange and Röder, 2006; Lange et al., 2003; Sanders and Astheimer, in press). In these paradigms, like temporal cuing experiments, a slow negativity (CNV) developed prior to the attended time. Importantly, these studies also demonstrated larger-amplitude auditory evoked potentials (N1) in response to sounds presented at the attended time. These auditory temporally selective attention effects are remarkably similar in timing, distribution, and amplitude to those reported in auditory spatially selective attention studies

(Hillyard, 1981; Hillyard et al., 1973a; Schwent and Hillyard, 1975) although no direct comparison has been made.

1.3. Speech perception

Spatially selective attention has been shown to be most critical when more information than can be processed in detail is presented simultaneously at distinct locations, especially when distractors and targets share similar features (Awh et al., 2003; Duncan and Humphreys, 1989). If the same were true of temporally selective attention, it would be most critical for perception when more information than can be processed in detail is presented rapidly at a single location. Speech is a highly relevant example of a perceptually challenging stimulus containing rapidly changing information that may encourage listeners to use temporally selective attention. However, almost nothing is known about how, or even if, people use temporally selective attention to process speech.

If selective attention does play a role in efficient speech perception, it would be beneficial for it to be directed to times at which unpredictable information is presented. That is, when listeners are able to predict the acoustic information they will hear in the immediate future (e.g., the end of a highly constrained word), they may allocate the minimal amount of resources needed to confirm their predictions. In contrast, when listeners can predict that something important is coming up (e.g., the onset of the subject of a sentence) but not precisely what that information will be, they may need to process the acoustic information in more detail.

A similar argument has been made concerning event perception in general (Zacks et al., 2007). According to this theory, perceptual systems include event models that are constantly used to make predictions about upcoming stimuli based on previous experience. When a mismatch between these expectations and incoming sensory information occurs, event boundaries are perceived. The mismatch also serves as a cue for allocating cognitive resources over time, focusing them when prediction errors occur. The differential allocation of cognitive resources over time described by Zacks et al. (2007) is, in fact, the definition of temporally selective attention. This theory of event perception applies equally well to speech perception to the extent that word boundaries are processed like other event boundaries.

Several lines of evidence suggest that word onsets have a special status in speech perception. Behavioral studies indicate that auditory word recognition relies more heavily on word onsets than other segments within words (Connine et al., 1993; Marslen-Wilson and Zwitserlood, 1989). Further, ERP studies have shown that word onsets elicit larger amplitude N1s than acoustically similar word-medial syllable onsets when presented in continuous speech (Sanders and Neville, 2003; Sanders et al., 2002b). This word-onset negativity is not specific to one type of stimulus; it is evident when listeners process several types of continuous speech including normal English, Jabberwocky sentences in which all of the open-class words have been replaced with nonwords, and synthesized streams of nonsense syllables. These findings suggest that the larger N1 elicited by word onsets reflects a general processing difference rather than the actual process of speech segmentation. One putative processing difference is temporally selective attention directed to time windows that contain word onsets. This hypothesis is supported, in part, by similarities in the latency, amplitude, and distribution of the word-onset negativity and temporally selective attention ERP effects.

1.4. Hypotheses

Whether selected on the basis of location or time, attended auditory stimuli typically elicit a larger negativity peaking around 100 ms after sound onset (N1) that is largest at anterior and medial electrode sites (Hillyard, 1981; Hillyard et al., 1973a; Lange et al., 2003; Sanders and Astheimer, in press). If listeners selectively attend to the beginnings of words, the auditory

potentials evoked by acoustic onsets in this time range should be larger than those elicited by identical sounds in the middle and towards the end of words. However, not all portions of speech are likely to include abrupt acoustic onsets similar to the word-initial and word-medial syllable onsets employed in ERP studies of speech segmentation (Sanders and Neville, 2003; Sanders et al., 2002b). Therefore, an attention probe paradigm adapted from studies of spatially selective attention (Coch et al., 2005; Stevens et al., 2006; Teder-Sälejärvi and Näätänen, 1994) was used to index selective attention at multiple times during speech perception. In the spatially selective attention studies, attention probes presented from the same location as an attended narrative elicited larger auditory onset components than identical probes presented from the same location as an unattended narrative. To study temporally selective attention, onset components elicited by identical probes played at different *times* within a narrative were compared.

In the current study, attention probes were presented at and around word onsets as well as at random control times to examine the allocation of attention over time during speech comprehension. Separate experiments employed linguistic and nonlinguistic attention probes to examine the specificity of temporally selective attention during speech perception. Based on previous electrophysiological evidence and the relatively unpredictable nature of word onsets, it was hypothesized that listeners would direct attention to the time windows that contain word onsets resulting in larger amplitude N1s evoked by attention probes presented at these times.

2. Method

2.1 Participants

Forty participants (20 women) between the ages of 18 and 35 years (mean = 24 years 8 months) provided the data included in analyses. Twenty-two listened to the narrative with linguistic probes only, while 18 listened to it with nonlinguistic probes only. All participants reported being right handed, native English speakers, taking no psychoactive medications, and having no known neurological disorders. An additional eight participants completed the experiment; data from two were excluded due to technical problems and six for poor EEG quality. All participants provided written informed consent prior to the experiment and were compensated for their time at a rate of \$10/hr.

2.2 Stimuli

The narrative was adapted from Edward Abbey's reading of *Freedom and Wilderness* (1987). The 2.5 hour monologue was divided at natural pauses between sentences or phrase boundaries into 596 5–20 s segments. Each segment was saved in the left channel of a stereo WAVE file with a 44.1 kHz sampling rate. Linguistic attention probes were created by extracting a 50 ms excerpt of the narrator pronouncing the syllable "ba." Nonlinguistic probes were 50 ms 1000 Hz tones with 8 ms onset and offset ramps created in Sound Studio.

Two complete narratives were created by pasting either the linguistic probe or the nonlinguistic probe in the right channel of the sound files. Three-hundred attention probes were added to the narrative in each of six conditions: coincident with a word onset, 100 and 50 ms before and after a word onset, and at random control times (Fig. 1), for a total of 1800 probes. Word onsets were defined as the earliest indication of a new phoneme based on visual inspection of sound waveforms and listening to sentences using a gating procedure. To ensure the accuracy of word onset determinations, only words for which the onset times, as determined by three independent coders, fell within a 16 ms range were assigned attention probes. In order to keep the rate of probe presentation somewhat consistent throughout the experiment, over 650 additional repetitions of the probe sound were added, but ERP responses to these sounds were not

recorded. The specific word onsets that attention probes were associated with were selected such that the acoustic properties of the narrative were similar across temporal conditions resulting in probe positions that were never coincident with acoustic onsets in the narrative. Specifically, there were no significant differences (p 's > .20) in average intensity, peak intensity, average pitch, or pitch change in the segment of narrative 100 to 50 ms before probe onset, 50 ms before probe onset, during the 50 ms probe, or 50 to 100 ms after probe onset for the six conditions (Table 1). Attention probes were never presented within the first 2000 or last 500 ms of a sound file.

The narrative with attention probes was presented over two Dell speakers placed directly in front of participants and connected to a Dell computer using E-Prime software. Playback volume was adjusted such that peak intensity of the narrative was 65 dB SPL (A-weighted) measured at the location of participants. The linguistic probes were presented at the same level as the loudest parts of the narrative (65 dB-A) whereas the nonlinguistic probes were presented with higher intensity (70 dB-A). Small photographs of the southwestern United States, related to the general theme of the stories, were presented at the center of a black background on a computer monitor 152 cm in front of the participant. A new image was presented each time a new sound file began such that picture changes never occurred less than 2000 ms before or 500 ms after an attention probe. Images were presented at a maximum of 3.5 degrees of visual angle such that participants could view an entire photograph without making eye movements.

Procedure—All procedures were approved by university and department review boards at the University of Massachusetts, Amherst and were conducted in accordance with the ethical standards specified in the 1964 Declaration of Helsinki. Participants were fitted with a 128-channel HydroCel Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene OR) containing electrodes imbedded in small sponges soaked in a potassium chloride saline solution to reduce impedance at each electrode site below 50 k Ω at the beginning of the experiment. Additional saline was used to maintain impedances below 100 k Ω throughout the session. Continuous EEG was recorded at a bandwidth of .01–80 Hz while participants listened to the narrative and observed photographs on the monitor. Participants were informed that audible probes would be played throughout the narrative, but they did not need to respond to them in any way. Following the narrative, participants completed a brief comprehension questionnaire to ensure their attentiveness throughout the story. All participants included in analyses scored at least 80% correct on the comprehension questionnaire.

EEG was digitized (250 Hz) and segmented into epochs 200 ms before to 800 ms after attention probe onset, with the 200 ms pre-stimulus interval serving as a baseline. Trials that contained eye blinks, eye movements, or head movements, as determined by a maximum amplitude criteria defined for each individual by visual inspection of EEG were excluded from the final average. Only data from participants with at least 60 artifact-free trials in each condition were included in analyses.

Data for each probe condition and electrode site were averaged for every individual and rereferenced to the average mastoid measurements. To determine the distribution of auditory evoked potentials elicited by linguistic probes, peak latency and peak amplitude measurements were taken in the following time windows after stimulus onset, based on visual inspection of the waveforms: 50–80 ms (P1), 90–140 ms (N1), 150–220 ms (P2), and 250–650 ms (N2). For nonlinguistic probes, measurements were taken from 30–60 ms (P1), 70–120 ms (N1), 130–200 ms (P2), and 300–700 ms (N2) post-stimulus. Measurements were made at 60 electrode sites across the scalp and combined into 30 pairs of electrodes in a 5 (Medial-Lateral position, or ML) \times 6 (Anterior-Posterior position, or AP) grid (Fig. 2).

To examine attention effects more closely, subsequent measurements were taken at 60 electrode sites within a region of interest (ROI) where auditory onset components were most pronounced (Fig. 2). Peak amplitude and mean amplitude measurements during the same time windows described above were taken for all electrodes within the ROI. Again, average measurements from 30 pairs of electrodes were arranged in a 5 (ML) \times 6 (AP) grid. All waveform measurements were entered in 6 (Probe Time) \times 5 (ML) \times 6 (AP) repeated-measures ANOVAs (Greenhouse-Geisser adjusted). Post-hoc analyses were conducted for all significant ($p < .05$) main effects and interactions.

3. Results

3.1 ERP Waveforms

As shown in Fig. 3, both linguistic and nonlinguistic probes elicited a typical positive-negative-positive-negative series of peaks that were largest over medial anterior regions. There were, however, differences in the latency and maximum amplitude of waveforms between the two probe types.

Linguistic probes evoked a positive deflection (P1) that peaked at 63 ms and was largest at medial anterior electrodes [ML $F(4, 84) = 6.1, p < .005$; AP $F(5, 105) = 24.6, p < .001$]. The first negative peak (N1) had a latency of 112 ms and was largest at left medial electrodes across all anterior-posterior positions [ML $F(4, 84) = 5.2, p < .05$; AP $F(5, 105) = 1.3, p > .25$]. The second positive peak (P2) followed with a latency of 180 ms and a medial anterior scalp distribution [ML $F(4, 84) = 10.2, p < .001$; AP $F(5, 105) = 39.5, p < .001$]. A broad negativity (N2) followed, peaking around 300 ms after probe onset.

Nonlinguistic probes elicited peaks with significantly shorter latencies than linguistic probes [P1 $F(1, 38) = 127.8, p < .001$; N1 $F(1, 38) = 115.4, p < .001$; P2 $F(1, 38) = 95.8, p < .001$]. The P1 elicited by nonlinguistic probes peaked at 46 ms and was largest at anterior electrodes across all medial lateral positions [ML $F(4, 68) = 1.2, p > .25$; AP $F(5, 85) = 22.2, p < .001$]. This was followed by an N1 that peaked at 92 ms and had a medial anterior distribution [ML $F(4, 68) = 18.4, p < .001$; AP $F(5, 85) = 9.9, p < .005$]. This N1 had a significantly larger peak amplitude than the N1 elicited by linguistic probes, $F(1, 38) = 8.4, p < .01$. The P2 peaked at 158 ms and was largest at medial anterior electrodes [ML $F(4, 68) = 17.3, p < .001$; AP $F(5, 85) = 30.0, p < .001$]. Nonlinguistic probes also elicited a broadly distributed N2 that peaked around 300 ms after probe onset.

3.2 Attention Effects

For linguistic probes, there was no effect of probe time on mean amplitude of the P1 waveform, $F(5, 105) = 1.6, p > .15$. There was, however, an effect of probe time on N1 mean amplitude, $F(5, 105) = 3.5, p < .02$. Probes played at word onset elicited a significantly larger N1 than those played at a control time, $F(1, 21) = 10.5, p < .01$, (Fig. 4) or at 100 ms before word onset, $F(1, 21) = 76.3, p < .05$ (Fig. 5a). This effect tended to be most pronounced across central and anterior electrode sites [Probe Time \times ML \times AP $F(25, 525) = 2.2, p < .06$]. In addition, when mean amplitude measurements for probes played at -100 ms and -50 ms were averaged into a general “before word onset” category, N1 amplitude was larger for probes played at word onset than for probes played before, $F(1, 21) = 6.4, p < .02$. There was however, no difference in N1 amplitude between probes played at word onset and probes played after [$+50$ $F(1, 21) = .4, p > .5$; $+100$ $F(1, 21) = .5, p > .45$] (Fig. 5b). Average amplitude between 90 and 140 ms for all six linguistic probe conditions is plotted in Fig. 6b.

Probe time also had a significant effect on P2 mean amplitude, $F(5, 105) = 3.0, p < .03$. Additional comparisons revealed that probes played 100 ms before word onset elicited a

significantly larger P2 than probes played at word onset, 50 ms before, and 50 or 100 ms after [Word Onset $F(1, 21) = 13.3, p < .002$; $-50 F(1,21) = 8.0, p < .02$; $+50 F(1,21) = 7.8, p < .02$; $+100 F(1,21) = 4.5, p < .04$], an effect that appeared uniformly across all electrodes within the ROI. No other differences in P2 amplitude were evident. N2 mean amplitude was also affected by probe time, $F(5, 105) = 3.7, p < .02$. Probes played at word onsets elicited a larger N2 than those played at a random time, $F(1, 21) = 14.8, p < .002$ (Fig. 4), or at 50 ms after word onset, $F(1,21) = 12.2, p < .003$ (Fig. 5b). Probes played 50 ms before word onset also elicited a larger N2 than control probes, $F(1,21) = 4.8, p < .05$. The distribution of these effects did not vary across electrodes within the ROI.

Nonlinguistic probes did not show any of the early attention effects observed for linguistic probes, with no differences in mean P1 amplitude, $F(5, 85) = 2.4, p > .05$, N1 amplitude, $F(5, 85) = .9, p > .45$, or P2 amplitude, $F(5, 85) = 1.5, p > .2$, across probe times (Fig. 7). Average amplitude between 70 and 120 ms for all six nonlinguistic probe conditions is plotted in Fig. 6b. There was, however, an effect of probe time on N2 amplitude, $F(5, 85) = 3.0, p < .05$, with a smaller N2 elicited by probes played at control times compared to those played at word onset, $F(1, 17) = 13.1, p < .003$ (Fig. 7), before word onset, [$-100 F(1, 17) = 10.6, p < .006$; $-50 F(1, 17) = 6.7, p < .02$], or 50 ms after word onset, $F(1, 17) = 4.5, p < .05$.

4. Discussion

In the current study, ERPs elicited by attention probes varied according to both probe type and probe onset time, as summarized in Fig. 8. Linguistic attention probes played during the initial portions of words in a narrative elicited larger amplitude N1s than identical probes played during the final portions of words or at control times. This difference was not evident in response to nonlinguistic probes. For linguistic probes, there were also differences in P2 amplitude across probe times, although these results were likely influenced by interactions with the negative peaks that surround the P2. The only effect that was consistent across the two probe types was the modulation of the N2, which was smallest in amplitude for probes presented at random control times. Thus the N2 may be comparable to N400 responses reported in other studies. That is, this component may have been driven by words in the narrative since it was only evident for ERPs time-locked to probes that had a consistent temporal relationship with word onsets. The N1 effects indicate that listeners employ rapidly modulated temporally selective attention during natural speech processing. Specifically, attention is heightened at times that contain word-initial segments compared to other times that differ by as little as 100 ms. Temporally selective attention during speech perception affects early perceptual processing, as has been reported in other studies of auditory selective attention. Further, the early attention effects are similar to those reported in a broad range of studies of both temporally and spatially selective auditory attention.

4.1. Implications for Speech Perception

Temporally selective attention may be critical in allowing listeners to process the rapidly changing acoustic information that constitutes spoken language. Subtle differences in the timing, direction, or amount of change in specific formants have important implications for comprehension. Further, these subtle distinctions must be made extremely rapidly, as speech stimuli are presented at a rate of five syllables per second or faster (Miller et al., 1984). However, some of this information is predictable or redundant, and therefore does not need to be processed in as much detail as other sounds that provide new information. Skilled listeners may take advantage of differences in predictability to enhance processing of unpredictable, and therefore informative, segments of speech. In this manner, temporally selective attention could allow for efficient processing of an otherwise overwhelming amount of sensory information in speech.

The current study demonstrates that attention is directed to times that contain word-initial segments during narrative processing. This is an effective strategy for listeners because word-initial segments are relatively less predictable than word-medial segments (Aslin et al., 1999). Of course, this observation raises the critical question of what actually cues the allocation of attention to segments like word onsets. The model of event perception outlined by Zacks et al. (2007) proposes a mismatch between expectations and incoming information triggers both segmentation and the allocation of attention. The application of this model to speech processing and temporally selective attention would suggest that attention is captured exogenously by sounds that do not match listener's expectations. However, it is also plausible that an *inability* to predict what will happen next cues selective attention in an endogenous manner.

The demonstrated enhancement of attention at times containing word-initial segments broadens the implications of previous ERP findings on speech segmentation. Previous studies reported an increase in N1 amplitude in response to word onsets compared to acoustically matched word-medial syllable onsets (Sanders and Neville, 2003; Sanders et al., 2002b). However, the fact that the word-onset negativity has similar latency and distribution across conditions in which the process of segmentation depends on very different cues, and therefore, is likely a very different process, suggests it may be indexing a more general process than segmentation itself. Evidence of this enhanced negativity for unrelated probes in the current study supports that hypothesis and further provides evidence that the general process is temporally selective attention. Similarities in the polarity, latency, and distribution of the word-onset negativity, the effects of temporally selective attention on auditory evoked potentials, and the results of the current study suggest a relationship between the processes of segmentation and selective attention.

The nature of the relationship between temporally selective attention and segmentation remains ambiguous. Temporally selective attention could serve as a cue for segmentation such that events that are processed with attention are further categorized as event boundaries. Alternatively, segmentation could be one of the cues that directs temporally selective attention based on the assumption that event boundaries are likely to include important perceptual information. Finally, these two processes could be triggered independently by similar cues but be otherwise independent. Studies of speech segmentation indicate that listeners can flexibly use whatever segmentation cues are available including transitional probability, stress pattern, syntactic structure, and newly acquired lexical information (Aslin et al., 1999; Mattys et al., 2007; Sanders et al., 2002a; Sanders et al., 2002b). These same cues may also direct temporally selective attention to times that contain word-initial segments. Understanding what directs selective attention to word onsets is essential to determining whether temporally selective attention is a cause, consequence, or coincidence of speech segmentation.

The allocation of temporally selective attention during speech perception likely interacts with many other aspects of linguistic processing. To the extent that temporally selective attention is directed to information that a listener cannot predict, semantic context, syntactic structure, and phonological regularities will impact the segments that receive attentional resources. Further, the predictability of upcoming information in speech is not all-or-none. Listeners may be able to predict some information about a word (e.g., its syntactic role) but not the specific item (e.g., the lexical unit). The combination of syntactic predictability and lexical unpredictability might be particularly effective at directing listeners to attend to the speech stream. In future studies it will be important to explore the relationships between selective attention and higher level linguistic information as a potential factor in determining the latency of ERP and eye-tracking indices of the time course of language processing.

4.2. Implications for Temporally Selective Attention

In addition to providing insight into mechanisms of speech perception, the present findings also allow for further characterization of temporally selective attention as a perceptual tool. The observation of differential responses to probes presented only 100 ms apart relative to word onsets suggest that temporally selective attention can be modulated more precisely than previously reported. Studies involving temporal cuing and sustained attention paradigms have reported early ERP differences in response to stimuli presented at times separated by as little as 500 ms (Correa et al., 2006a; Lange and Röder, 2006; Lange et al., 2003; Sanders and Astheimer, in press). Although behavioral data have demonstrated that temporally selective attention can enhance target detection during rapid serial visual presentation (Correa et al., 2005) and target sounds presented in noise as little as 150 ms before or 100 ms after an expected time are rarely detected (Wright and Fitzgerald, 2004), enhancement of ERPs by temporally selective attention has not yet been demonstrated on such a small timescale. However, previous ERP studies involved explicitly cued or consciously sustained attention that required duration judgments to accurately allocate attention. Further, speech stimuli are far more complex than even the most challenging stimuli employed in previous studies. As suggested by Correa et al. (2006a), greater perceptual difficulty may contribute to earlier effects of attention. The complexity of speech may encourage listeners to use temporally selective attention and extensive experience with English may allow temporal modulation at a more precise level.

The attention effects reported in the current study resemble previous data on both temporally and spatially selective attention, highlighting potential similarities between the two mechanisms. First, studies of temporally selective auditory attention (Lange and Röder, 2006; Lange et al., 2003; Sanders and Astheimer, in press) and spatially selective auditory attention (Hillyard, 1981; Hillyard et al., 1973a; Schwent and Hillyard, 1975) all report that attended sounds elicit larger amplitude auditory onset components like the N1. This indicates that selective attention affects early perceptual processing regardless of whether sounds are selected on the basis of time or location.

Second, the results of this study demonstrate that listeners use temporally selective attention during speech perception without instruction; some features of speech, combined with previous linguistic experience result in an increase in attention at times that contain word onsets. Although it is not yet clear if attention is being drawn to word onsets by a physical cue in speech, directed by a mismatch between prediction and reality (Zacks et al., 2007), or cued by an inability to predict upcoming sounds, it is clear that, like spatially selective attention, temporally selective attention can be driven endogenously, exogenously, or by some combination of the two. The ephemeral nature of speech sounds may necessitate the use of temporally selective attention automatically, similar to the way complex scenes that overwhelm the visual system require the use of spatially selective attention (Pomplun, 2006; Rolls and Deco, 2006; VanRullen and Koch, 2003). Just as previous experience with the world causes endogenous attention shifts to relevant stimuli like people and houses during scene viewing, previous experience with a language may cause endogenous attention shifts to relevant times like those containing word onsets during speech perception.

Third, the present findings indicate that, like spatially selective attention, temporal selection can be constrained by target features. In the current study, the use of temporally selective attention was only apparent in responses to linguistic probes. It is unlikely that temporally selective attention was not utilized during the narrative containing nonlinguistic probes, since the listening task was identical in both cases. Instead, this difference suggests that multiple selection criteria were in used to allocate attention to the narrative. While this is inconsistent with previous evidence that temporally selective attention can be applied so generally that effects are observed across modalities (Lange and Röder, 2006), it is not surprising given that speech perception is a highly developed and specialized task. Since participants were instructed

to listen to the narrative for comprehension, they may have focused their attention on linguistic stimuli only, and therefore ignored all nonlinguistic probes regardless of their presentation time. This additional discrimination between linguistic and nonlinguistic sounds is analogous to studies of spatial attention involving conjunction searches that demonstrate that spatial attention can be guided by target features (Kim and Cave, 1995; Maunsell and Treue, 2006; Shih and Sperling, 1996). Thus, the current study adds to a growing body of evidence that spatially and temporally selective attention share many common properties, including neurophysiological correlates, endogenous and exogenous cues, and target specificity.

4.3. Conclusions

Temporally selective attention functions as a perceptual tool during speech perception, allowing listeners to enhance processing of informative stimuli such as word-initial segments. Although it is unclear what directs attention to times that contain word-initial segments, previous linguistic experience may allow listeners to make predictions about upcoming speech sounds and enhance attention to times when predictions cannot be formed and when predictions fail to match incoming sensory information. If predictions are driving temporally selective attention in speech processing, anything that interferes with the ability to make predictions, such as inexperience with a language, may affect subsequent levels of language processing. Understanding how temporally selective attention functions during speech for non-native speakers, children, and individuals with language and attention disorders may reveal why efficient perception can be difficult in some circumstances. Further, a greater understanding of temporally selective attention will give insight into how our perceptual systems process rapidly changing acoustic signals in order to derive vital information from spoken language.

Acknowledgements

This study was supported by an NIH NIDCD grant (RO3DC008684) to LDS as well as NIH training support (5T32 NS007490) for LBA. We thank Victoria Ameral, Molly Chilingirian, Ahren Fitzroy, Ralph McDonald, Preeti Putcha, Kathryn Sayles, and Michele Taylor for their work preparing sentence stimuli, William Bush, Molly Chilingirian, Nicolas Planet, and Yibei Shen for assistance collecting and analyzing data, and Charles Clifton and Matthew Collins for their feedback on earlier versions of this manuscript.

References

- Abbey, E. *Freedom and Wilderness*. North Word Audio Press; Minocqua, WI: 1987.
- Alho K, Donauer N, Paavilainen P, Reinikainen K, Sams M, Naatanen R. Stimulus selection during auditory spatial attention as expressed by event-related potentials. *Biological Psychology* 1987;24:153–162. [PubMed: 3607146]
- Aslin, RN.; Saffran, JR.; Newport, EL. Statistical learning in linguistic and nonlinguistic domains. In: MacWhinney, B., editor. *The emergence of language*. Lawrence Erlbaum Associates Publishers; Mahwah: 1999. p. 359-380.
- Awh E, Matsukura M, Serences JT. Top-down control over biased competition during covert spatial orienting. *Journal of Experimental Psychology: Human Perception & Performance* 2003;29:52–63. [PubMed: 12669747]
- Cave KR. The feature gate model of visual selection. *Psychological Research* 1999;62:182–194. [PubMed: 10490397]
- Cave KR, Bichot N. Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin and Review* 1999;6:204–223. [PubMed: 12199208]
- Chen Q, Zhang M, Zhou X. Interaction between location- and frequency-based inhibition of return in human auditory system. *Experimental Brain Research* 2007;176:630–640.
- Coch D, Sanders LD, Neville HJ. An event-related potential study of selective auditory attention in children and adults. *Journal of Cognitive Neuroscience* 2005;17:605–622. [PubMed: 15829081]
- Connine CM, Blasko DG, Titone D. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 1993;32:193–210.

- Correa A, Lupiáñez J, Milliken B, Tudela P. Endogenous temporal orienting of attention in detection and discrimination tasks. *Perception & Psychophysics* 2004;66:264–278. [PubMed: 15129748]
- Correa A, Lupiáñez J, Tudela P. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic Bulletin and Review* 2005;12:328–334. [PubMed: 16082814]
- Correa A, Lupiáñez J, Madrid E, Tudela P. Temporal attention enhances early visual processing: a review and new evidence from event-related potentials. *Brain Research* 2006a;1076:116–128. [PubMed: 16516173]
- Correa A, Lupiáñez J, Tudela P. The attentional mechanism of temporal orienting: determinants and attributes. *Experimental Brain Research* 2006b;169:58–68.
- Coull JT, Nobre AC. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience* 1998;18:7426–7435. [PubMed: 9736662]
- Driver J. A selective review of selective attention research from the past century. *British Journal of Psychology* 2001;92:53–78.
- Duncan J, Humphreys GW. Visual search and stimulus similarity. *Psychological Review* 1989;96:433–458. [PubMed: 2756067]
- Griffin IC, Miniussi C, Nobre AC. Orienting attention in time. *Frontiers in Bioscience* 2001;6:660–671.
- Griffin IC, Miniussi C, Nobre AC. Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time. *Neuropsychologia* 2002;40:2325–2340. [PubMed: 12417462]
- Hansen JC, Dickstein PW, Berka C, Hillyard SA. Event-related potentials during selective attention to speech sounds. *Biological Psychology* 1983;16:211–224. [PubMed: 6615954]
- Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical signs of selective attention in the human brain. *Science* 1973a;182:177–180. [PubMed: 4730062]
- Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical signs of selective attention in the human brain. *Science* 1973b;182:177–80. [PubMed: 4730062]
- Hillyard SA. Selective auditory attention and early event-related potentials: a rejoinder. *Canadian Journal of Psychology* 1981;35:159–174. [PubMed: 7248849]
- Hillyard SA, Anllo-Vento L. Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences USA* 1998;95:781–787.
- Hink RF, Hillyard S. Auditory evoked potentials during listening to dichotic speech messages. *Perception & Psychophysics* 1976;20:236–242.
- Jonides J, Irwin D. Capturing attention. *Cognition* 1981;10:145–150. [PubMed: 7198529]
- Kim MS, Cave KR. Spatial attention in visual search for features and feature conjunctions. *Psychological Science* 1995;6:376–380.
- Lange K, Rosler F, Röder B. Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: an event-related potential study. *Psychophysiology* 2003;40:806–817. [PubMed: 14696734]
- Lange K, Röder B. Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience* 2006;18:715–729. [PubMed: 16768372]
- Luck SJ, Hillyard S, Mouloua M, Woldorff M, Clark V, Hawkins H. Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception & Performance* 1994;20:887–904. [PubMed: 8083642]
- Marslen-Wilson W, Zwitserlood P. Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance* 1989;15:576–585.
- Mattys SL, Melhorn JF, White L. Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception & Performance* 2007;33:960–977. [PubMed: 17683240]
- Maunsell JH, Treue S. Feature-based attention in visual cortex. *Trends in Neuroscience* 2006;29:317–322.

- Miller JL, Grosjean F, Lomanto C. Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica* 1984;41:215–225. [PubMed: 6535162]
- Miniussi C, Wilding EL, Coull JT, Nobre AC. Orienting attention in time. Modulation of brain potentials. *Brain* 1999;122:1507–1518. [PubMed: 10430834]
- Mondor TA, Lacey TE. Facilitative and inhibitory effects of cuing sound duration, intensity, and timbre. *Perception & Psychophysics* 2001;63:726–736. [PubMed: 11436741]
- Näätänen R. Processing negativity: an evoked-potential reflection of selective attention. *Psychological Bulletin* 1982;92:605–640. [PubMed: 7156260]
- Nobre AC, Correa A, Coull J. The hazards of time. *Current Opinions in Neurobiology* 2007;17:465–470.
- Picton TW, Hillyard SA. Human auditory evoked potentials, II. Effects of attention. *Electroencephalography and Clinical Neurophysiology* 1974;36:191–199. [PubMed: 4129631]
- Pomplun M. Saccadic selectivity in complex visual search displays. *Vision Research* 2006;46:1886–1900. [PubMed: 16445960]
- Posner MI. Orienting of attention. *Quarterly Journal of Experimental Psychology* 1980;32:3–25. [PubMed: 7367577]
- Rolls ET, Deco G. Attention in natural scenes: Neurophysiological and computational bases. *Neural Networks* 2006;19:1383–1394. [PubMed: 17011749]
- Sanders LD, Neville HJ, Woldorff MG. Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research* 2002a;45:519–530.
- Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience* 2002b;5:700–703.
- Sanders LD, Neville HJ. An ERP study of continuous speech processing I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research* 2003;15:228–240. [PubMed: 12527097]
- Sanders LD, Astheimer LB. Temporally selective attention modulates early perceptual processing: event-related potential evidence. *Perception & Psychophysics*. in press
- Scharf, B. Auditory attention: the psychoacoustical approach. In: Pashler, H., editor. *Attention*. Psychology Press/Erlbaum; Hove, England: 1998. p. 75-177.
- Schröger E, Eimer M. Effects of transient spatial attention on auditory event-related potentials. *Neuroreport* 1993;4:588–590. [PubMed: 8513143]
- Schwent VL, Hillyard SA. Evoked potential correlates of selective attention with multi-channel auditory inputs. *Electroencephalography and Clinical Neurophysiology* 1975;38:131–138. [PubMed: 45943]
- Serences JT, Shomstein S, Leber AB, Golay X, Egeth HE, Yantis S. Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science* 2005;16:114–122. [PubMed: 15686577]
- Shih SI, Sperling G. Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance* 1996;22:758–779. [PubMed: 8666962]
- Stevens C, Sanders L, Neville H. Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Research* 2006;1111:143–152. [PubMed: 16904658]
- Teder-Sälejärvi W, Näätänen R. Event-related potentials demonstrate a narrow focus of auditory spatial attention. *Neuroreport* 1994;5:709–711. [PubMed: 8199343]
- VanRullen R, Koch C. Competition and selection during visual processing of natural scenes and objects. *Journal of Vision* 2003;3:75–85. [PubMed: 12678627]
- Woldorff M, Hansen JC, Hillyard SA. Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. *Electroencephalography and Clinical Neurophysiology Supplement* 1987;40:146–154. [PubMed: 3480115]
- Wright BA, Fitzgerald MB. The time course of attention in a simple auditory detection task. *Perception & Psychophysics* 2004;66:508–516. [PubMed: 15283074]
- Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR. Event perception: a mind-brain perspective. *Psychological Bulletin* 2007;133:273–293. [PubMed: 17338600]

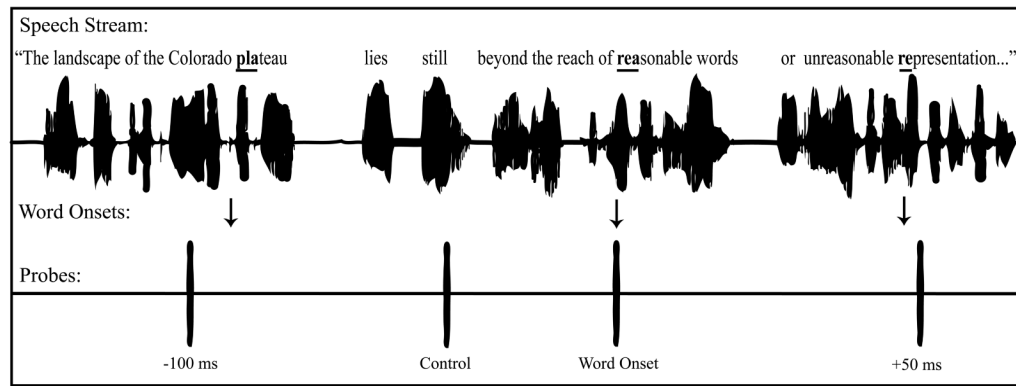


Fig. 1. Experimental paradigm. Linguistic and nonlinguistic probes (separate experiments) were presented at six times relative to word onsets in a narrative: concurrently with word onset, -100 ms, -50 ms, +50 ms, +100 ms, and at random control times.

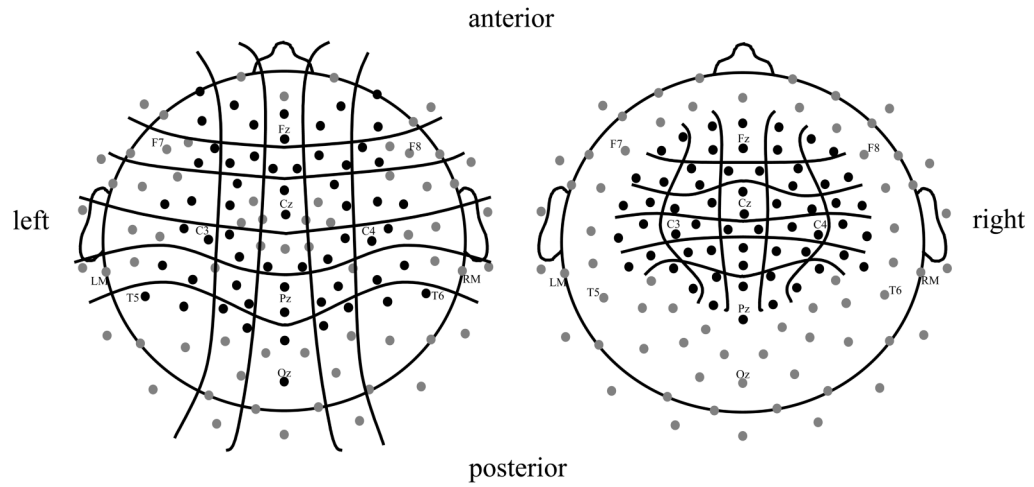


Fig. 2. Approximate scalp position of 128 electrodes. Data from the 60 electrodes shown in black on the left were included in initial analyses. Based on the distribution of auditory evoked potentials, subsequent measurements were taken from 60 electrodes in the region of interest (ROI) shown in black on the right. Measurements were combined into 30 pairs of electrodes in a 5 (Medial/Lateral position, or ML) \times 6 (Anterior/Posterior position, or AP) grid.

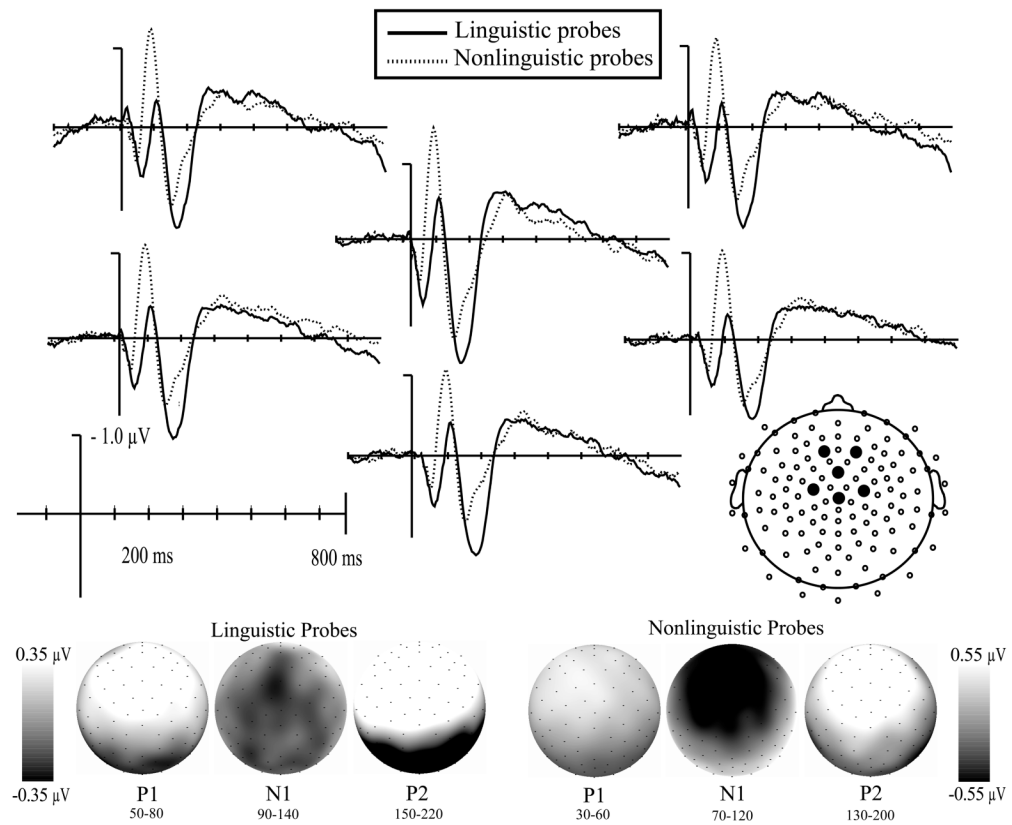


Fig. 3. Auditory evoked potentials time-locked to the onset of linguistic (solid line) and nonlinguistic (dotted line) probes. All probes elicited the typical positive-negative-positive oscillation largest at anterior-central and medial electrodes (topographic maps show mean amplitude). Linguistic probes elicited longer latency peaks and smaller amplitude N1s than nonlinguistic probes.

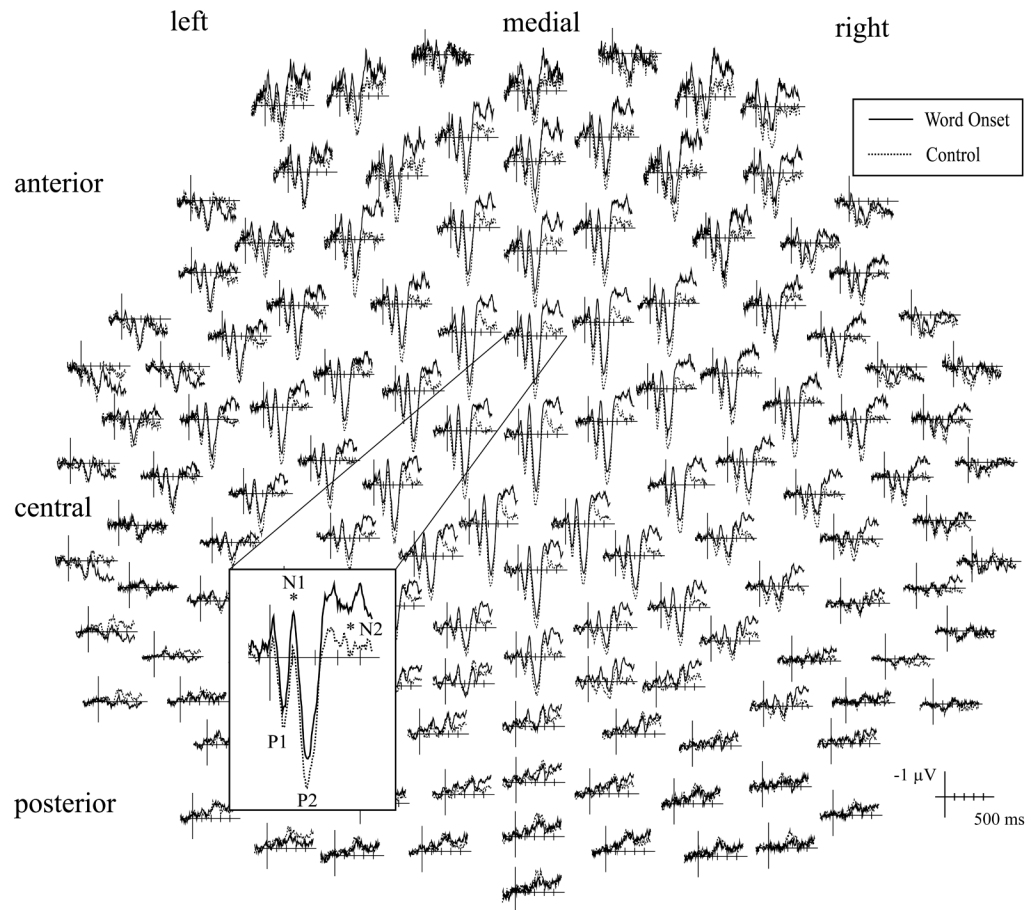


Fig. 4. Auditory evoked potentials elicited by linguistic attention probes presented at word onset (solid line) and random control times (dotted line) at all 128 electrode sites. Waveforms recorded at FCz are shown at a larger scale. Probes played at word onset elicited significantly larger N1 and N2 peaks at anterior-central and medial electrode sites compared to probes played at control times.

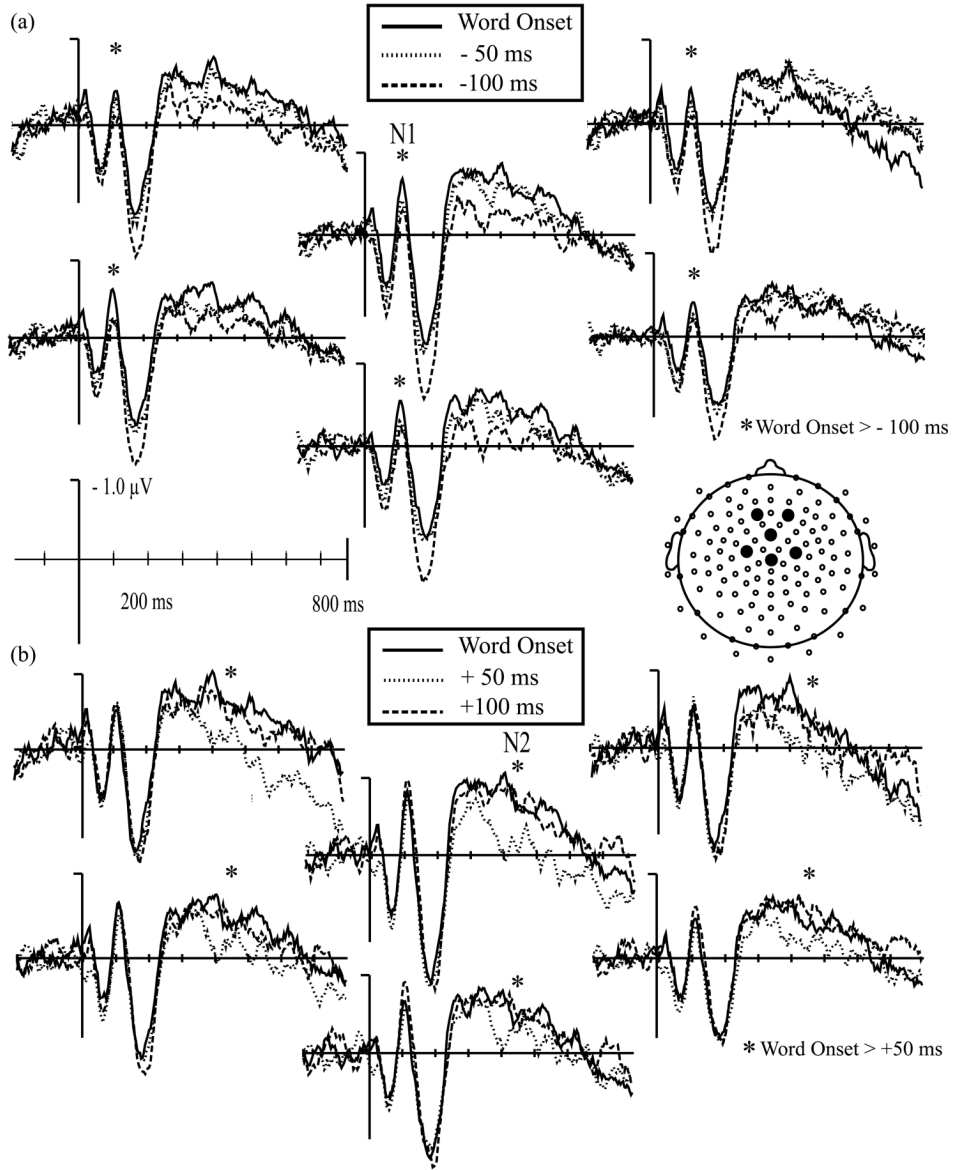


Fig. 5. (a) Auditory evoked potentials elicited by linguistic attention probes presented at word onset (solid line), 50 ms before (dotted line), and 100 ms before (dashed line) word onset. Probes played at word onset elicited a significantly larger N1 peak than probes played 100 ms before word onset. (b) Auditory evoked potentials elicited by linguistic attention probes presented at word onset (solid line), 50 ms after (dotted line), and 100 ms after (dashed line) word onset. Probes played at word onset elicited a significantly larger N2 peak than probes played 50 ms after word onset.

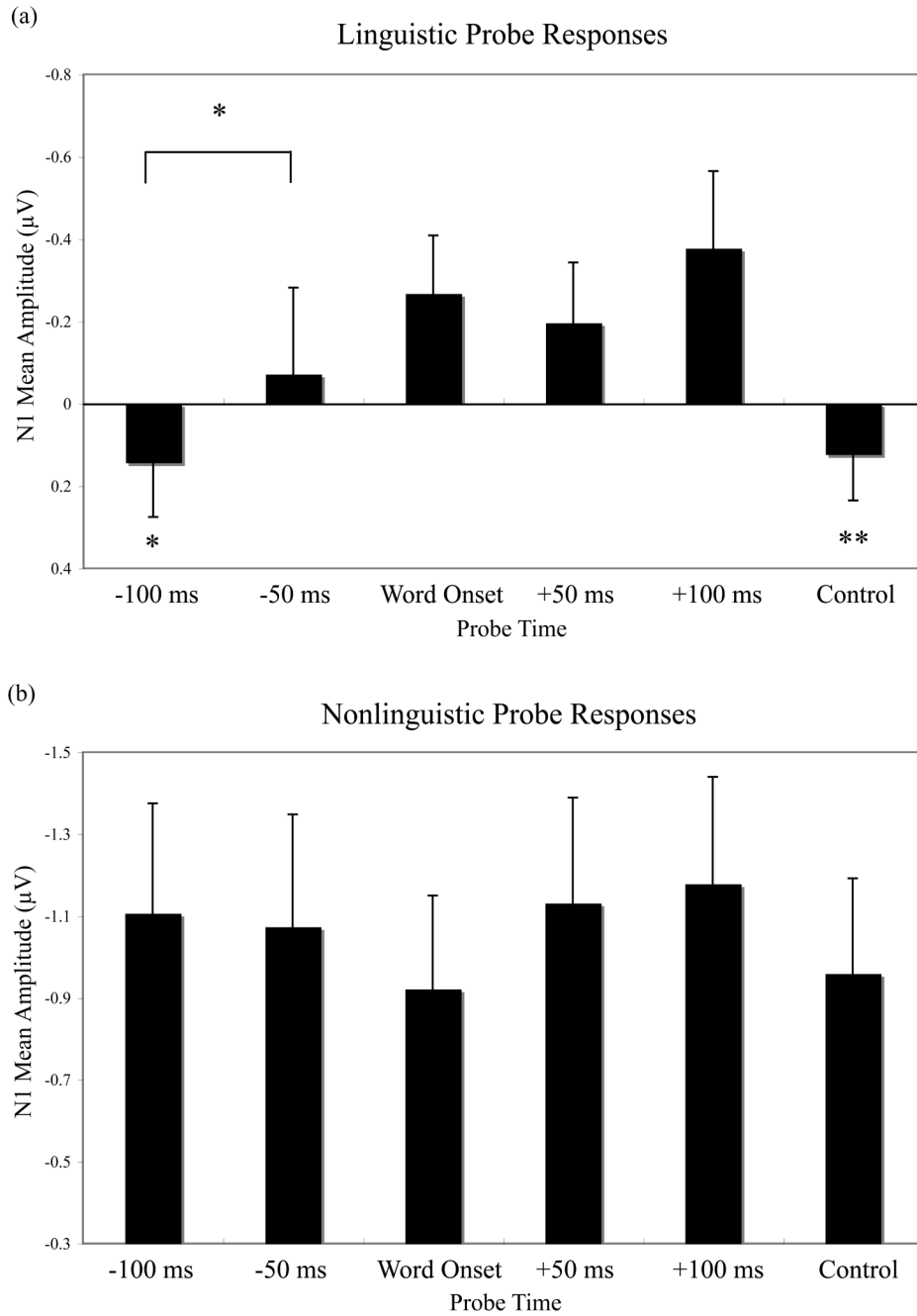


Fig. 6. Mean amplitude of auditory evoked potentials elicited by (a) linguistic probes from 90 to 140 ms across the 60 electrodes in the ROI and by (b) nonlinguistic probes from 70 to 120 ms at the same sites. N1 amplitude in response to linguistic probes presented at word onset was larger than for probes presented 100 ms before word onset or at control times. There were no differences in N1 amplitude for linguistic probes presented at and 50 or 100 ms after word onset. There were no significant effects of time condition on N1 amplitude for the nonlinguistic probes. * = significantly different from Word Onset at $p < .05$; ** = $p < .01$.

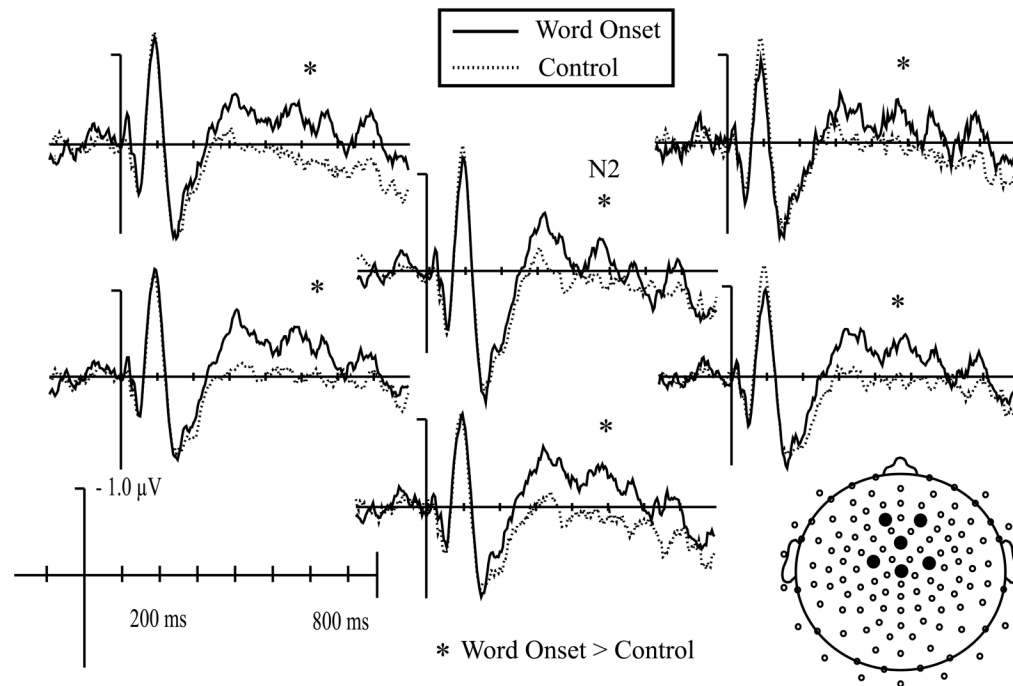


Fig. 7. Auditory evoked potentials elicited by nonlinguistic attention probes presented at word onset (solid line) and at control times (dotted line). Although no early attention effects were evident, probes played at word onset elicited a larger N2 than those played at a control times.

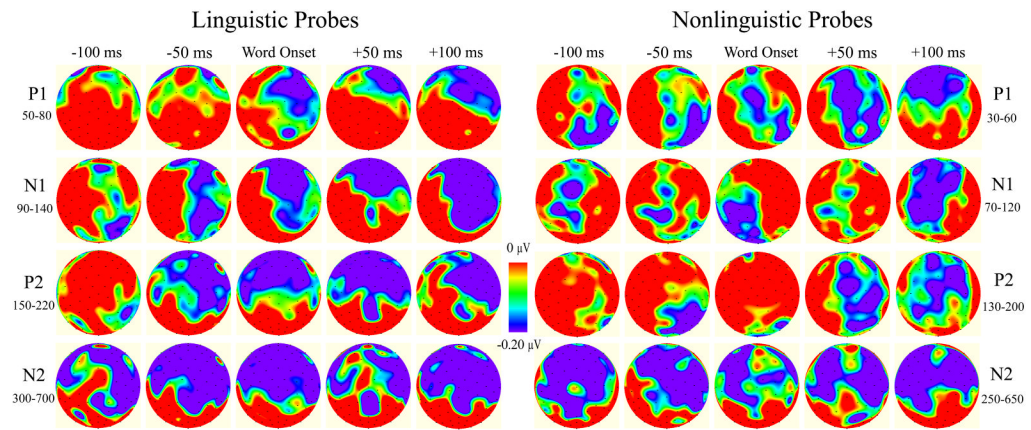


Fig. 8. Topographic maps of mean amplitude differences (probe time condition minus control) for linguistic and nonlinguistic probes, measured in P1, N1, P2, and N2 time ranges.

Table 1
Acoustic characteristics of the narrative measured during four time windows relative to probe onset.

Measure	Time Window	Probe Time		Word Onset	+50 ms	+100 ms	Control
		-100 ms	-50 ms				
Mean Intensity	-100 to -50 ms	69.3 (8.4)	70.5 (7.4)	69.3 (7.8)	67.4 (8.1)	67.7 (7.6)	69.1 (9.0)
	-50 to 0 ms	69.5 (7.7)	70.0 (7.8)	67.3 (8.1)	68.5 (7.5)	71.1 (7.2)	69.3 (9.1)
	0 to 50 ms	69.5 (7.3)	68.1 (8.3)	68.8 (7.3)	71.5 (7.0)	72.5 (6.3)	69.8 (8.5)
	50 to 100 ms	66.6 (8.6)	68.4 (7.0)	71.0 (7.9)	72.9 (6.1)	72.1 (6.8)	69.8 (8.7)
Maximum Intensity	-100 to -50 ms	70.9 (7.9)	72.0 (6.9)	71.1 (7.4)	69.4 (7.4)	69.7 (7.1)	70.7 (8.5)
	-50 to 0 ms	71.0 (7.1)	71.5 (7.2)	69.2 (7.3)	70.4 (7.0)	72.4 (6.9)	70.8 (8.6)
	0 to 50 ms	71.2 (6.6)	69.8 (7.6)	70.7 (6.7)	72.9 (6.7)	73.7 (5.9)	71.4 (7.8)
	50 to 100 ms	68.6 (7.8)	70.4 (6.4)	72.1 (7.7)	74.1 (5.6)	73.5 (5.9)	71.3 (8.1)
Mean Pitch	-100 to -50 ms	103.0 (11.6)	103.8 (13.0)	101.5 (14.4)	98.3 (11.9)	100.4 (13.2)	107.0 (12.8)
	-50 to 0 ms	101.1 (11.2)	102.2 (14.3)	98.5 (14.3)	101.7 (12.9)	107.9 (14.1)	106.8 (13.4)
	0 to 50 ms	98.5 (11.3)	98.4 (11.7)	101.8 (14.0)	107.6 (12.6)	110.5 (13.5)	105.9 (13.0)
	50 to 100 ms	95.6 (11.1)	101.8 (14.9)	107.5 (13.4)	111.3 (13.4)	110.5 (13.4)	104.9 (13.1)
Maximum Pitch	-100 to -50 ms	106.5 (12.4)	107.3 (13.3)	105.3 (14.7)	102.4 (12.5)	103.9 (13.5)	110.1 (12.9)
	-50 to 0 ms	104.2 (11.3)	105.3 (14.4)	102.7 (15.2)	104.6 (13.2)	111.4 (14.4)	110.0 (13.5)
	0 to 50 ms	101.7 (11.8)	104.4 (12.2)	104.7 (14.4)	111.0 (13.0)	113.6 (13.8)	109.4 (13.1)
	50 to 100 ms	99.6 (11.4)	104.7 (15.0)	111.2 (14.0)	114.3 (13.6)	113.4 (13.3)	108.1 (13.6)
Pitch Change	-100 to -50 ms	-3.6 (9.5)	-4.3 (7.8)	-4.2 (8.2)	-4.3 (8.7)	2.25 (8.5)	-2.4 (7.7)
	-50 to 0 ms	-2.5 (8.0)	-3.8 (7.7)	-3.7 (10.0)	1.6 (7.9)	2.9 (8.4)	-2.7 (8.0)
	0 to 50 ms	-3.8 (7.8)	-4.1 (8.6)	1.0 (7.9)	2.8 (8.9)	-0.8 (8.6)	-2.6 (8.6)
	50 to 100 ms	-4.4 (8.2)	.85 (8.3)	1.7 (8.8)	-1.4 (8.8)	-2.7 (7.4)	-1.6 (8.5)

Note: Average measurements for the 300 probes in each condition are shown with standard deviation in parentheses. Acoustic measurements of WAVE files were recorded using Praat software (version 4.0.41) on Apple computers. Intensity levels are reported in dB SPL and pitch measurements in Hz. There were no differences between the six probe conditions in any acoustic measure during the four measurement windows (p 's > .20).