Available online at www.sciencedirect.com
ScienceDirect
Biological Psychology 80 (2009) 23–34
BIOLOGICAL PSYCHOLOGY
www.elsevier.com/locate/biopsycho

Listeners modulate temporally selective attention during natural speech processing

Lori B. Astheimer *, Lisa D. Sanders

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

Received 28 September 2007; accepted 29 January 2008
Available online 10 March 2008

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 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., 1973; 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

* Corresponding author. Tel.: +1 413 545 4761; fax: +1 413 545 0996.
E-mail address: lastheim@nsm.umass.edu (L.B. Astheimer).

0301-0511/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.biopsycho.2008.01.015

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, 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., 1973; 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