Neuropsychologia 47 (2009) 1183–1186

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Note

Event-related potentials index segmentation of nonsense sounds

Lisa D. Sanders*, Victoria Ameral, Kathryn Sayles

Department of Psychology and Neuroscience and Behavior Program, University of Massachusetts, Amherst, MA 01003, United States

ARTICLE INFO

Article history:
Received 22 June 2008
Received in revised form 22 October 2008
Accepted 6 November 2008
Available online 13 November 2008

Keywords:
ERP
Speech segmentation
Statistical learning
Auditory
Selective attention
N1

ABSTRACT

To understand the world around us, continuous streams of information including speech must be segmented into units that can be mapped onto stored representations. Recent evidence has shown that event-related potentials (ERPs) can index the online segmentation of sound streams. In the current study, listeners were trained to recognize sequences of three nonsense sounds that could not easily be rehearsed. Beginning 40 ms after onset, sequence-initial sounds elicited a larger amplitude negativity after compared to before training. This difference was not evident for medial or final sounds in the sequences. Across studies, ERP segmentation effects are remarkably similar regardless of the available segmentation cues and nature of the continuous streams. These results indicate the preferential processing of sequence-initial information is not domain specific and instead implicate a more general cognitive mechanism such as temporally selective attention.

One of the challenges of perceptual processing is determining where one object or event ends and the next begins. For example, natural speech consists of multiple acoustic changes, with silence as common within as between words, forcing listeners to rely on other sources of information to segment the continuous streams of sound. Importantly, this process of segmentation must be carried out in a rapid, online manner. Event-related potentials (ERPs) have proven to be a critical tool for indexing online segmentation. For example, in participants who learned to discriminate between six three-syllable nonsense words and foils created from the same syllables in different sequences, word onsets in continuous streams elicited a larger negativity 70–130 ms (N1) and 200–500 ms (N400) after compared to before training (Sanders, Newport, & Neville, 2002). In this study, there were no acoustic cues in the synthesized speech that could have contributed to segmentation. Instead, both the N1 and N400 effects were dependent on listeners using newly learned lexical information acquired during a brief training session.

Synthesized speech is an extremely useful tool for determining the sources of information that can be used to segment continuous streams. However, it is also important to determine if similar mechanisms are employed during natural language processing. To address this question, ERPs elicited by acoustically similar word and syllable onsets in normal English, Jabberwocky in which the open-class words had been replaced with nonwords, and sentences in which all of the words had been replaced with nonwords have been compared (Sanders & Neville, 2003a). For all three types of sentences, word onsets elicited a larger amplitude negativity between 70 and 120 ms (N1). Evidence for the N1 differences for sentences composed entirely of unfamiliar nonsense words indicates lexical segmentation cues are not necessary to observe this early ERP index of speech segmentation. Instead, listeners were likely using multiple acoustic segmentation cues including allophonic variation, phonotactic constraints, and language-specific rhythmic properties (Brent & Cartwright, 1996; Cutler, Mehler, Norris, & Sequi, 1992; Jusczyk, 1999).

To make use of lexical and acoustic segmentation cues, listeners have to be familiar with a language and the language-specific associations between word boundaries and acoustic features. However, a great deal of evidence has shown that adults and infants can make use of transitional probabilities to segment speech with no previous knowledge of the language (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Listeners take advantage of the fact that syllables that are part of the same word are heard together more frequently than syllables that cross word boundaries to learn syllable sequences that are never presented in isolation and for which no other segmentation cues are available. Abla, Katahira and Okanoya (2008) demonstrated that statistical information alone can be used to segment sounds in the same fast, online manner reported for lexical and acoustic segmentation cues. Specifically, initial compared to medial and final tones in continuous streams elicited a larger negativity 80–160 ms (N1) and 300–500 ms (N400) with less exposure to the stream in the group of participants who showed behavioral evidence of larger amounts of statistical learning.

* Corresponding author at: Department of Psychology, University of Massachusetts, Tobin Hall, Amherst, MA 01003, United States. Tel.: +1 413 5455962.
E-mail address: lsanders@psych.umass.edu (L.D. Sanders).

0028-3932/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2008.11.005

Regardless of whether listeners are relying on newly learned lexical cues (Sanders et al., 2002), native-language acoustic cues (Sanders & Neville, 2003a), or transitional probabilities for tone sequences (Abla et al., 2008), sequence-initial segments elicit larger amplitude N1 and N400 components in listeners who have sufficient training or experience to demonstrate a high degree of familiarity with the sequences. Although these studies made very different segmentation cues available to listeners, the stimuli shared many important traits. Specifically, all of the studies employed very familiar sounds (speech or tones) that could easily be rehearsed. Therefore, it is still not clear whether or not the same neural indices of online segmentation generalize to novel stimuli that have to be processed without the benefit of representations of individual units built through a lifetime of experience. To test the hypothesis that novel sounds can be segmented in a similar manner to more familiar stimuli, the current study employed complex, nonlinguistic sounds that varied in length. Evidence that the ERP indices of segmenting novel sounds are similar to those reported in previous studies would implicate a flexible, domain-general cognitive process.

1. Method

Twenty right-handed adults (6 females) ages 19–27 years ($M=21$) contributed data. All were native English speakers and reported no neurological disorders or use of psychoactive medication. All participants provided written consent and were compensated $10/h for their time. Data from an additional five adults were collected but excluded from analysis because of artifacts in the EEG: low-frequency shifts ($N=2$), bridging between electrodes ($N=2$), and excessive blinks ($N=1$).

Eleven non-verbal sounds (e.g., glass breaking, elephant trumpeting, train whistle) were modified to mask the original source and discourage verbal labeling. Twelve listeners with no previous exposure were unable to recognize the modified sounds. The sounds (duration = 190–310 ms, $M=242$ ms) were combined to create six sequences of three sounds each (duration = 660–800 ms, $M=728$ ms) with some segments present in different positions in multiple sequences. For sounds A through K, the sequences were ABC, BDE, FGA, DHI, JEB, and HKB. Sequences were repeated 91 times each and compiled in random order (with the exception that no sequence could follow itself) to create a 6.6 min sound stream with higher transitional probabilities for sounds within sequences (0.31–1.0) than for sounds between sequences (0.1–0.2). The order of sequences was randomized three additional times for a total of four continuous streams saved as mono WAV files with an 11,025 Hz sampling rate.

To test learning of the sequences, six part sequences were created from the last two sounds of one sequence followed by the first sound of a different sequence as follows: BCF, DEJ, GAB, HIA, EBH, KBD. Each sequence was paired with a part sequence once to create a 36 item two-alternative forced-choice test. The two sequences in each test item were separated by 1 s of silence and test items by 5 s. Four separate randomizations of test items were created.

Participants were first asked to complete a behavioral test on which they indicated preference for the sequence or part sequence for each of the 36 test pairs. EEG was then recorded while two of the sound streams were presented. A second behavioral test followed for which participants were instructed to choose the sequence in each pair that sounded more familiar. Participants were then explicitly trained to recognize the six sequences by using a mouse to point a cursor at one of six icons on the computer monitor, which triggered the presentation of the corresponding sound sequence. Participants were instructed to memorize the sequences to recognize them on a subsequent behavioral test and were allowed to play each sequence as many times as wanted. A third behavioral test was given to assess learning and the training/testing procedure was repeated until the participant responded correctly on at least 32 of 36 test items (89% accuracy). Following training, EEG was again recorded while the two remaining streams of sequences were presented. As with the first presentation of the continuous streams, participants were asked simply to listen to the sounds. A final behavioral test was given at the end of the experiment. The order in which the four continuous streams and four versions of the behavioral test were presented was balanced across participants.

EEG was recorded from 128 electrodes using a 250 Hz sampling rate and a .01–100 Hz bandpass with impedance maintained below 50 kΩ at all locations. A 60 Hz filter was applied offline before EEG was segmented into 700 ms epochs beginning 100 ms before the onset of each sound. Trials were excluded if the voltage difference within a segment exceeded 100 μV at any electrode site including those used to monitor blinks and eye movements. A minimum of 100 artifact-free trials for each condition was required for data to be included in analysis. Averaged waveforms were re-referenced to the average mastoid measurements and baseline corrected using the 100 ms before sound onset.

Mean amplitude was measured in 50 ms bins staggered by 10 ms (e.g., 0–50, 10–60, 20–70) for the first 200 ms after sound onset. The first of five consecutive bins that showed statistically significant training effects was considered the onset of a segmentation index. Mean amplitude was also measured across the longer time-windows of 40–200 and 300–500 ms after sound onset for comparison with previous studies. Data from 108 electrodes collapsed into 12 groups of 9 were included in a repeated-measures ANOVA (Huynh–Feldt corrected): training (before, after) × sequence position (initial, medial, final) × anterior/posterior electrode position (4 levels) × left/medial/right electrode position (3 levels). Follow-up analyses were conducted at electrodes indicated by significant training × sequence position × electrode position interactions.

2. Results

Performance on behavioral tests is shown in Fig. 1. Accuracy was not above chance on the first test ($M=49\%$) but was on the second pre-training test ($M=55\%$, $t(19)=3.23$, $p<.01$). Participants spent 4–41 min training ($M=20$) and took the post-training test 1–5 times ($M=2.75$). Accuracy on the final post-training test was well above chance as required ($M=91\%$) and remained high ($M=86\%$) but significantly lower ($t(19)=2.33$, $p<.05$) on the test given at the end of the experiment.

The earliest of five consecutive 50 ms time windows for which there were significant training × sequence position × electrode

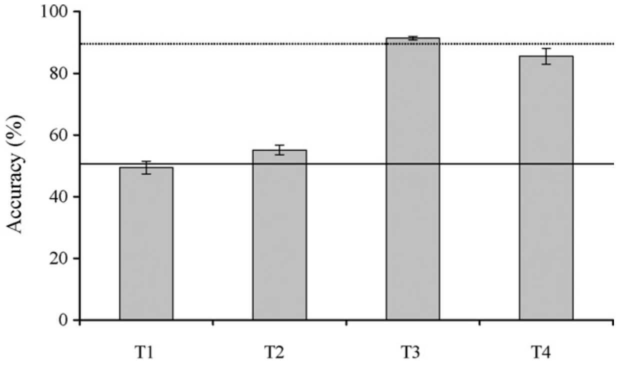

Fig. 1. Accuracy on four behavioral tests of sequence recognition: T1 = before any exposure to the sequences, T2 = after listening to two 6.6 min continuous streams, T3 = after mastering the sequences during training, and T4 = after listening to two additional 6.6 min streams. The solid line indicates chance performance (50%), the dotted line indicates training criteria (89%), and error bars indicate standard error.

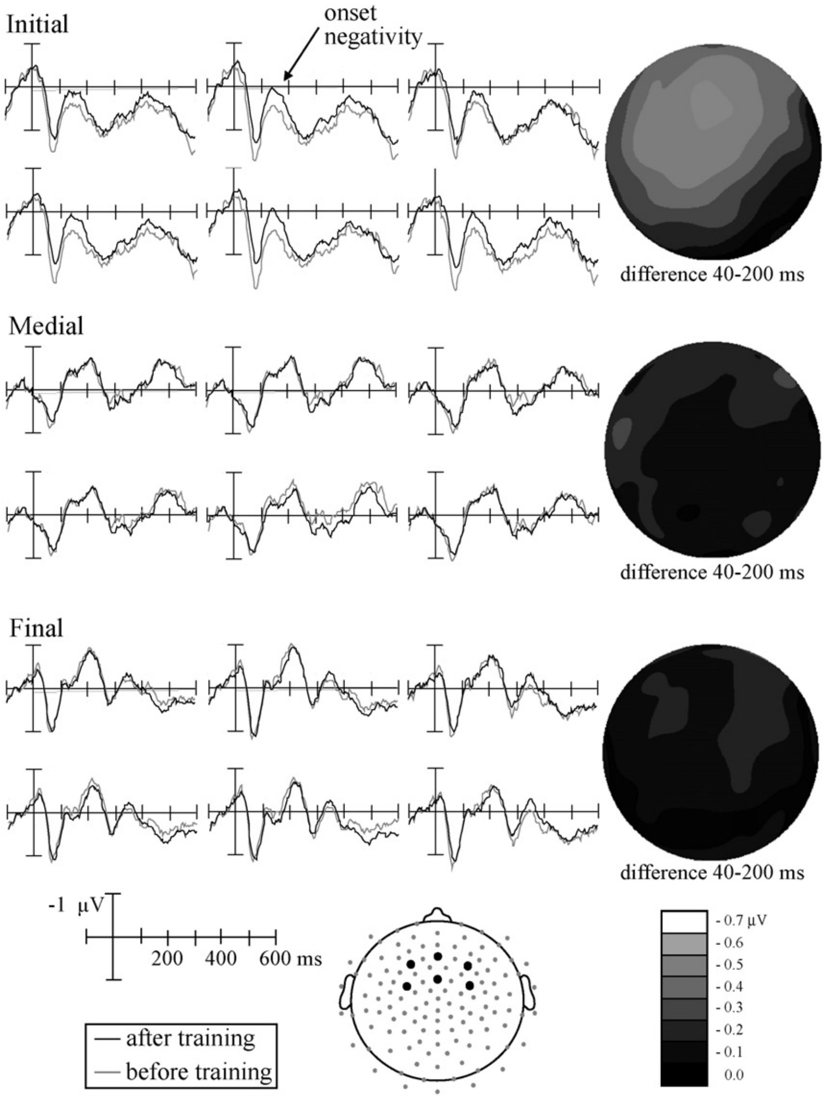

Fig. 2. ERPs time-locked to the onset of initial, medial, and final sounds in the sequences before and after training. Waveforms are shown for recordings made at the nine anterior and medial electrode positions indicated on the electrode map. Differences in mean amplitude (after minus before training) in the 40–200 ms time window across the entire scalp are shown in the topographic plots. Sequence onsets in continuous streams elicited a larger amplitude negativity between 40 and 200 ms after compared to before training.

position interactions on ERP mean amplitude was 40–90 ms (training × sequence position × left/medial/right: $F(4,76) = 3.84$, $p < .05$). A significant training × sequence position × left/medial/right interaction was observed on mean amplitude in the broader 40–200 ms window ($F(4,76) = 4.95$, $p < .01$). As shown in Fig. 2, training had a differential effect for the three sequence positions at medial electrodes ($F(2,38) = 3.27$, $p < .05$). Specifically, initial sounds elicited a larger negativity in this time window after compared to before training ($F(1,19) = 6.00$, $p < .05$). In contrast, no effect of training was observed for sequence-medial or sequence-final sounds. There was some indication that the negative difference for initial sounds was larger over the left than right hemisphere, but this interaction did not reach significance (training × sequence position × left/right electrode position, $F(2,38) = 3.20$, $p = .052$). No training or training by sequence position interactions were observed for mean amplitude 300–500 ms.

3. Discussion

When listening to continuous streams of sound, sequence-initial segments elicit a larger early negativity regardless of the available segmentation cues or nature of the auditory stimulus. The sequence onset negativity has been observed for natural, native language speech (Sanders & Neville, 2003a), nonsense sentences with acoustic features patterned from a native language (Sanders & Neville, 2003a), synthesized nonsense speech when no behavioral evidence of statistical learning was observed (Sanders et al., 2002), tone sequences when only statistical information was available as a segmentation cue (Abla et al., 2008), and, in the current study, for nonlinguistic, novel sounds that listeners learned to recognize but could not rehearse. The similarity of the sequence onset negativity across stimulus type and segmentation cues suggests it indexes a very general cognitive process.

A candidate general cognitive process that may be involved in segmentation across the wide range of conditions that have been explored is selective attention. Selective attention, the preferential processing of stimuli selected on the basis of a simple feature, has been shown to be critical for perception any time more information than can be processed in detail is present. For example, when too much information is presented simultaneously at distinct locations, spatial selection allows for preferential processing of the most relevant stimuli. A large body of evidence shows that when listeners are asked to attend to sounds from one location and ignore the sounds from another, sounds from attended locations elicit a larger anterior and medial negativity between 80 and 150 ms (e.g., Hillyard, Hink, Schwent, & Picton, 1973; Hink & Hillyard, 1976; Picton, Hillyard, Galambos, & Schift, 1971). However, perceptual systems are also overwhelmed by too much information presented rapidly in time. Under these conditions, temporal rather than spatial selection of relevant information becomes important. Recent studies indicate that listeners can use time as the simple feature to select stimuli for preferential processing such that sounds presented at attended compared to unattended times elicit a larger negativity over anterior and medial regions 80–150 ms after onset (Lange & Röder, 2006; Lange, Rösler, & Röder, 2003; Sanders & Astheimer, 2008).

The ERP indices of auditory spatially selective attention and auditory temporally selective attention are both remarkably similar to the sequence onset negativity. This similarity raises the hypothesis that the segmentation effects reported across a wide range of conditions arise from listeners selectively attending to the initial portions of sequences presented as continuous streams. Specifically, when listening to rapidly presented information, listeners may not be able to processes all of the acoustic changes in detail forcing them to select a subset for preferential processing. In continuous streams, the initial portions of sequences are likely to be particularly informative driving listeners to attend to the times initial information is presented and allocate fewer resources to the times subsequent information is presented. Recent evidence provides support for this hypothesis (Astheimer & Sanders, in press). Participants were asked to listen to a narrative that included attention probes with various temporal relationships to word onsets in continuous speech. Probes presented within the first 150 ms of words elicited a larger negativity between 80 and 150 ms than identical probes presented in the last 150 ms of words or at random control times. Since the acoustic environment of the probes was similar across conditions and the only relationship between probes and speech was defined by timing, the results indicate listeners employ temporally selective attention to preferentially process the initial compared to final segments of words in continuous speech.

In the current study, all listeners were trained to be highly accurate at recognizing the sequences. However, in previous studies that employed a set amount of training (Sanders et al., 2002) or relied on statistical learning (Abla et al., 2008), only those participants who showed behavioral evidence of learning a large proportion of the sequences also showed the N1 segmentation effects. Further, native Japanese speakers listening to normal English and English-sounding nonsense sentences did not show any evidence of segmentation within the first 300 ms after word onset (Sanders & Neville, 2003b). Studies employing a familiarization paradigm in which words were initially presented in isolation and then in continuous speech indicate that the earliest indication of segmentation for 10-month-old infants is 340 ms after onset (Kooijman, Hagoort, & Cutler, 2005) and in non-native speakers is 515 ms after onset (Snijders, Kooijman, Cultler, & Hagoort, 2007). Taken together these studies indicate that expertise, though not necessarily learning a language as a native speaker, is necessary for the rapid, online segmentation indexed by the early ERP effects.

The findings that only expert listeners show the N1 segmentation effects and that skilled listeners use temporally selective attention to process speech suggest a relationship between selective attention ability and receptive language processing. Recent evidence indicates there is a relationship between spatially selective attention and language processing in development (Stevens, Sanders, & Neville, 2006; Stevens, Fanning, Coch, Sanders, & Neville, 2008). Children with specific language impairment do not show the early differences in neurosensory processing of attended and unattended sounds evident in normally developing children. Further, computer-based training designed to improve receptive language skill affects both scores on standardized tests of language processing and ERP indices of spatially selective attention (Stevens et al., 2008). By employing nonlinguistic stimuli and measures of temporally as well as spatially selective attention, it will be possible to test the hypothesis that the mechanistic link between selective attention and speech perception is the allocation of temporally selective attention to the initial portions of sound sequences in continuous streams.

Acknowledgement

The research reported here was funded by a John Merck Scholars Fellowship to LDS, University of Massachusetts Commonwealth College Honors Research Grants to VEA and KES, and NIH R01DC00128 to Helen J. Neville. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDCD, NIH, the John Merck Foundation, or Commonwealth College. We thank Charles Clifton for many helpful comments on this manuscript.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*, 952–964.

Astheimer, L. & Sanders, L. (in press). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology*.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Cutler, A., Mehler, J., Norris, D., & Sequi, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

Hillyard, S., Hink, R., Schwent, V., & Picton, T. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177–180.

Hink, R., & Hillyard, S. (1976). Auditory evoked potentials during selective listening to dichotic speech messages. *Perception & Psychophysics*, *20*, 236–242.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*, 323–328.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*, 109–116.

Lange, K., & Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience*, *18*, 715–729.

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, *40*, 806–817.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin*, *92*, 605–640.

Picton, T., Hillyard, S., Galambos, R., & Schift, M. (1971). Human auditory attention: A central or peripheral process? *Science*, *173*, 351–353.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Sanders, L., & Astheimer, L. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Perception & Psychophysics*, *70*, 732–742.

Sanders, L., & Neville, H. (2003a). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, *15*, 228–240.

Sanders, L., & Neville, H. (2003b). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*, 214–227.

Sanders, L., Newport, E., & Neville, H. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*, 700–703.

Snijders, T., Kooijman, V., Cultler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain Research*, *1178*, 106–113.

Stevens, C., Fanning, J., Coch, D., Sanders, L., & Neville, H. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Research*, *1205*, 55–69.

Stevens, C., Sanders, L., & Neville, H. (2006). Neurophysiological evidence for selective auditory attention deficits in children with Specific Language Impairment. *Brain Research*, *1111*, 143–152.