Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech

Lisa D. Sanders[1,3], Elissa L. Newport[2] and Helen J. Neville[1]

[1] Department of Psychology, University of Oregon, 1227 University of Oregon, Eugene, Oregon 97403-1227, USA

[2] Department of Brain and Cognitive Sciences, University of Rochester, Meliora 414, Rochester, New York 14627-0268, USA

[3] Present address: Department of Linguistics, University of Maryland, 3416 Marie Mount Hall, College Park, Maryland 20742-7505, USA

Correspondence should be addressed to L.D.S. (lsanders@wam.umd.edu)

Published online: 17 June 2002, doi:10.1038/nn873

Speech segmentation, determining where one word ends and the next begins in continuous speech, is necessary for auditory language processing. However, because there are few direct indices of this fast, automatic process, it has been difficult to study. We recorded event-related brain potentials (ERPs) while adult humans listened to six pronounceable nonwords presented as continuous speech and compared the responses to nonword onsets before and after participants learned the nonsense words. In subjects showing the greatest behavioral evidence of word learning, word onsets elicited a larger N100 after than before training. Thus N100 amplitude indexes speech segmentation even for recently learned words without any acoustic segmentation cues. The timing and distribution of these results suggest specific processes that may be central to speech segmentation.

To process natural speech, a listener must first break the continuous stream of sound into recognizable units. However, there are typically no reliable pauses between spoken words to indicate where one word ends and the next begins. Behavioral studies provide evidence that a wide range of segmentation cues contribute to adults' ability to segment continuous speech[1–3]. However, these behavioral studies are limited by their inability to establish the time course of speech segmentation and to distinguish between fast, online segmentation and slower linguistic processing that may influence performance on specific tasks. Further, behavioral studies cannot provide direct evidence about the brain systems involved in online segmentation. In addition, it is often difficult to use the same behavioral task with different groups of subjects. For example, evidence of speech segmentation is found in young infants[4,5], bilingual speakers[6,7], and monolingual adults using different tasks. Determining whether these groups are segmenting speech along the same time course and employing the same mechanisms requires designing tasks that can be accomplished by, and are equally engaging for, all groups.

The recording of ERPs provides an online measurement of speech segmentation suitable for listeners of all ages and backgrounds that also reflects the cortical organization of speech segmentation systems. We recently observed that initial syllables elicit a larger negativity around 100 ms (N100) than medial syllables presented in continuous speech[8]. This word-onset effect was found for initial and medial syllables matched on loudness, length and other acoustic characteristics. However, it remained possible that the larger N100s evoked by word onsets index small, uncontrolled physical differences in the different syllable types or in the syllables preceding them. To be certain that N100 word-onset effects index speech segmentation rather than acoustic characteristics that correlate with word boundaries, we measured ERPs while presenting the same physical stimuli before and after they were perceived as word onsets.

We recorded ERPs in response to six nonsense words presented as continuous speech before and after listeners learned the words as lexical items through training. We reasoned that by teaching listeners the lexical items of a nonsense language, they might begin segmenting continuous streams of those nonsense words. Thus, we could compare ERPs to the same stimuli when they were not segmented as lexical items (before training) and when they were segmented as lexical items (after training). Using this protocol, we showed that N100 amplitude indexes speech segmentation in the absence of acoustic segmentation cues.

Results

Behavior

The continuous streams of nonsense words in the present study have been used previously to show that mere exposure to distributional regularities is sufficient for listeners to learn to distinguish between nonsense words and part-word items on behavioral tests[5]. However, for the subjects in this experiment, performance on the tests given before (Mean percent correct (M) = 53.7%) and after 14 minutes of exposure (M = 53.5%) did not differ from each other or from chance. In order to induce segmentation more quickly (a previous study[5] used 21 minutes of exposure), we used a training protocol to teach the six nonsense words to the participants.

Performance on the behavioral test given immediately after training was well above chance (M = 79.5%), indicating that at

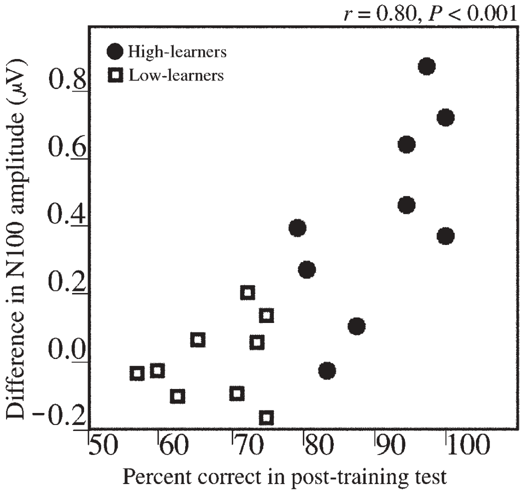

Fig. 1. Performance on behavioral tests after training (percent correct) plotted against difference in N100 amplitude before and after training (before training minus after training). Subjects' word learning as measured by the behavioral tests correlated with their N100 word-onset effects.

least some of the words had been learned. There were no differences in accuracy between this test and another test given after a second 14-minute exposure ($M = 79.2\%$), indicating that listeners neither learned new words nor forgot the ones they knew immediately after training. These scores were combined into a single post-training accuracy score ($M = 79.3\%$).

Event-related potentials

Across all subjects, training had no effect on N100 amplitude. However, the difference in N100 amplitude before and after training was highly correlated with individual performance on the post-training behavioral tests ($r = 0.80$, $P < 0.001$). Subjects who learned more of the words as measured by the behavioral test also showed larger N100 word-onset effects (**Fig. 1**).

To determine if the word-onset effect was significant in the group of subjects that showed the largest behavioral training effect, participants were divided into two groups based on a median split of post-training accuracy scores. The 9 subjects who showed the largest effect of training (before training $M = 55.1\%$, after training $M = 90.7\%$; $t_8 = 6.95$, $P < 0.001$) improved to a greater extent than the 9 subjects who showed the smallest effect of training (before training $M = 52.2\%$, after training $M = 67.9\%$; $t_8 = 4.55$, $P < 0.01$; group × training interaction, $F_{1,17} = 11.20$, $P < 0.01$).

Only high learners showed a significant effect of training on N100 amplitude (group × training × anterior/posterior interaction, $F_{5,80} = 2.91$, $P < 0.05$). For this group, the training × laterality × anterior/posterior interaction was significant ($F_{5,40} = 5.22$, $P < 0.01$). Over lateral electrodes, there were no main effects or interactions including training. However, over medial and midline electrodes, there was a main effect of training ($F_{1,8} = 6.29$, $P < 0.05$). Thus for high learners, word onsets elicited a larger N100 over medial and midline electrode sites after training (**Fig. 2**). For low learners, there was no effect of training on N100 amplitude.

To compare these results to our previous work[8], we did an additional comparison of the ERPs elicited by initial syllables and by medial and final syllables. Although this comparison lacks the advantage of contrasting the responses to physically identical stimuli, it was important to determine if training was specifically influencing the processing of word boundaries. Even before training, initial syllables elicited larger N100s than medial and final syllables across anterior electrode sites (position × anterior/posterior, $F_{5,80} = 7.94$, $P < 0.001$; at anterior sites only, position, $F_{1,16} = 13.74$, $P < 0.001$). This difference was likely due to the use of different syllables in these positions. However, it is also possible that the pre-training position effect indexed speech segmentation (based on learning from distributional regularities) that did not influence behavioral performance.

Importantly, high learners also showed a significant effect of training on N100 amplitudes elicted by syllables in different positions (group × position × training interaction, $F_{1,16} = 7.92$, $P < 0.01$). For the group of high learners (position × training interaction, $F_{1,18} = 15.64$, $P < 0.001$), the difference between N100 amplitude elicited by initial syllables and medial and final syllables was larger after than before training at anterior electrode sites (high learners after training, position × anterior/posterior interaction: $F_{5,40} = 8.548$, $P < 0.001$; at anterior sites only, position: $F_{1,8} = 21.40$, $P < 0.001$). No such interaction was found for the group of low learners.

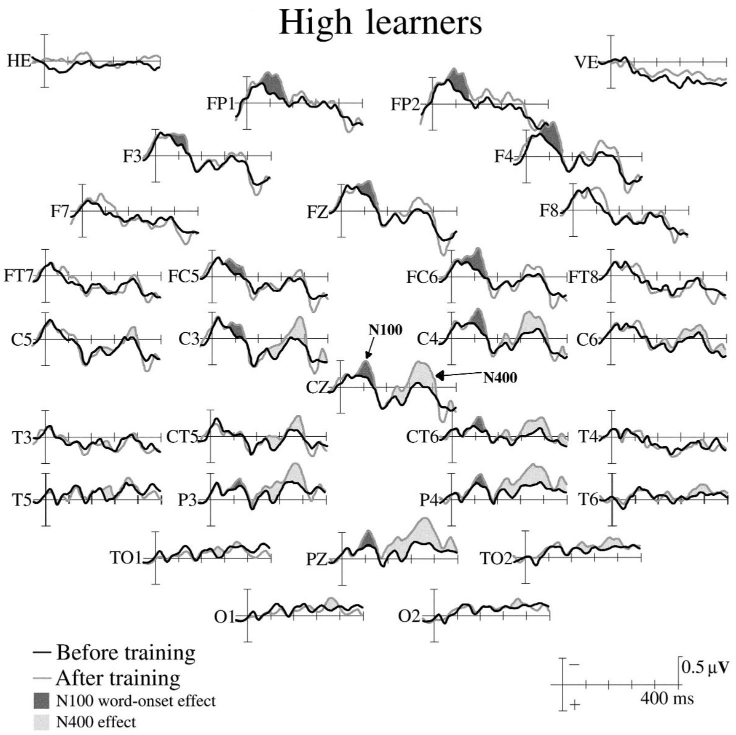

Fig. 2. ERPs averaged to word onsets before and after training for the subjects showing the largest behavioral learning effects (high learners). After training, word onsets elicited a larger N100 at midline and medial electrode sites. Words also elicited a larger N400 after training.

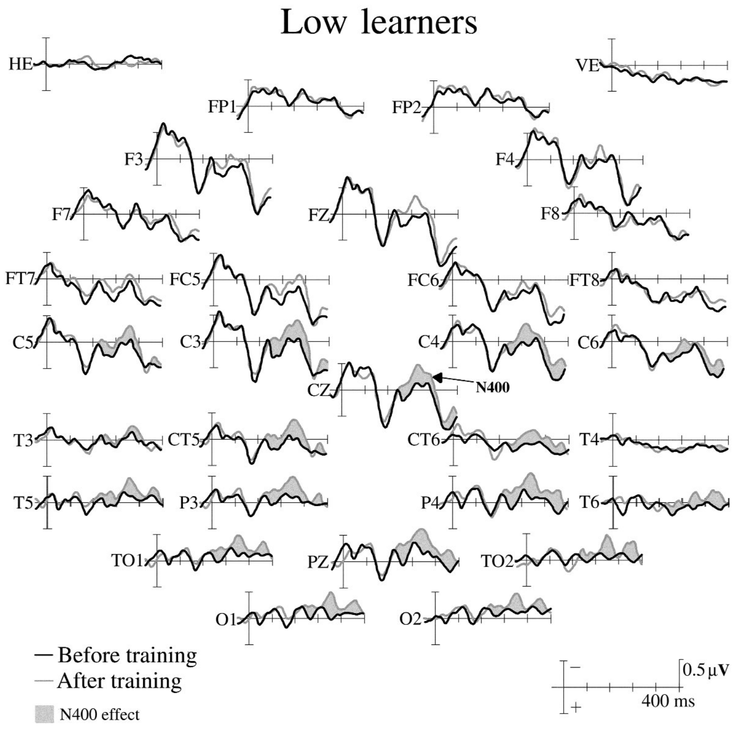

Fig 3. ERPs averaged to word onsets before and after training for the subjects showing the smallest behavioral learning effects (low learners). After training, words elicited a larger N400, similar to that found for the high learners.

For all subjects, the mean amplitude between 200 and 500 ms (N400) also showed effects of training (training × anterior/posterior interaction, $F_{5,85} = 4.90$, $P < 0.001$). The four most posterior rows of electrodes (reflecting the typical distribution of the N400) showed a main effect of training ($F_{1,17} = 6.57$, $P < 0.05$), indicating that the nonsense words elicited a greater negativity after training. There were no significant interactions with group, indicating that both high and low learners showed the N400 learning effect (**Fig. 3**).

The presence of N400 effects for both groups might influence the amplitude of earlier (N100) responses to medial and final syllables. However, the interaction of the word-level N400 and the syllable-level N100 would result in medial and final syllables eliciting larger N100s, whereas the opposite pattern was found. Furthermore, these two components have distinct distributions; the N100 was largest over anterior electrodes, and the N400 was largest over posterior electrodes.

Discussion

The N100 word-onset effect for nonsense words in the present study is remarkably similar to our observations in a study of processing real English[8]. For both real and nonsense words, word onsets elicit larger N100s across midline and medial electrode sites. The similarities in these findings are particularly striking considering the differences between the stimuli in the two studies. We observed the N100 word-onset effect in subjects listening to their native language, complete with semantic, lexical, syntactic, phonological and acoustic information. The N100 word-onset effect was also observed in subjects listening to nonsense sentences that contained only acoustic and phonological segmentation cues. In contrast, the present study used just 6 nonsense words learned during a 20-minute training session with no associated meanings and no acoustic segmentation cues, and again the N100 word-onset effect was observed. Behavioral studies show that a wide variety of cues can be used to segment speech; the N100 ERP response seems to index the perception of word onsets regardless of the type or number of segmentation cues available.

It is not clear whether the early ERP word-onset effect reflects differences in the way initial and medial sounds are processed or the process of segmentation itself. It is possible that linguistic onsets in continuous speech are processed like acoustic onsets and therefore elicit the same ERP components observed for acoustic onsets. However, mitigating evidence against this interpretation is that the distribution of word-onset effects (medial and midline) was distinct from the more lateral distribution of N100s in general ($F_{1,18} = 9.351$, $P < 0.001$). Alternatively, it is also possible that listeners direct greater attention to initial than to medial sounds. The effects of auditory attention to location and pitch are to increase N100 amplitude[9,10]; similar auditory attention effects may be elicited by word onsets in continuous speech.

One study of artificial language learning reports that a late negativity similar to the N400 found in the present study is sensitive to learning nonsense words[11]. After 50 hours of training on a 68-word written language, newly learned words elicit a larger negativity between 280 and 360 ms. During the same epoch, real English words elicit a larger negativity than pronounceable nonwords or consonant strings. These results are similar to findings from other studies in which words and orthographically legal nonwords elicit a larger N400 than consonant strings[12,13], and are consistent with an interpretation of the N400 as an index of lexical search. In the present study, listeners may not have conducted lexical searches at all during pre-training, before they were aware of repeating nonsense words in the continuous stream of syllables.

It is important to note that N400 effects, like earlier N100 differences, indicate that speech has been segmented. That is, before we can find any components that are time-locked to onsets in continuous speech, the speech must be processed as if it contains onsets. In the present study, if all syllables were processed in the same manner or if the syllables that were processed as onsets were distributed irregularly, N400s would not be time-locked to word onsets.

Interestingly, the group of subjects who showed the smallest behavioral word-learning effects and no early ERP word-onset effects (low learners) also had larger N400s after than before training. We observed a similar pattern of results in a study of late bilinguals listening to their non-native language[14]. In that study, native Japanese late learners of English did not show N100 word-onset effects when listening to English sentences; however, they did show larger N400s in response to words as compared to nonwords presented in continuous speech. There are several possible explanations for these findings. First, the high learners in the present study and the native speakers in the earlier study seem to have been segmenting speech differently, and in particular faster, than the respective groups of low learners and non-native speakers. That is, N100 amplitude may be indexing fast, online speech segmentation by the high learners and native speakers, whereas the later N400 effect may reflect slower or more variable segmentation. A related explanation is that non-native speakers and low learners were segmenting speech, but not using processes such as allocating greater attention to word onsets; if the N100 word onset effect reflects primarily these latter processes, the lack of N100 and presence of N400 in these groups would be consistent with this interpretation.

The results of the present study indicate that the N100 word-onset effect in continuous speech cannot be explained solely on the basis of acoustic differences in initial and medial sounds. Instead, the N100 effect indexes differences in the initial stages of processing of these two types of sounds. Differential processing of initial and medial sounds within a word indicates that speech has been segmented; therefore, the ERP word-onset effect can be used as an online measure of speech segmentation suitable for a wide variety of stimuli and for listeners of all ages and backgrounds. In addition, the timing and nature of this effect raise testable hypotheses concerning the specific mechanisms important in speech segmentation. Listeners who are more successful at segmenting speech (as measured by behavioral tests) show earlier segmentation effects. Word-onset effects occur very early, suggesting they involve either predictive or very fast, automatic processes. Furthermore, word-onset effects are similar even when the available segmentation cues are very different.

Methods

The procedure was approved by the University of Oregon Office of Human Subjects Compliance. Informed consent was obtained from all participants. Right-handed monolingual English speakers ($n = 18$) were first given 36 pairs of three-syllable nonsense words presented auditorally. On this pre-test, participants were asked to indicate which of the two items seemed more familiar. Each pair consisted of one of the six nonsense words that would later be presented in a continuous stream, and one part-word item constructed from the last syllable of one of the nonsense words followed by the first two syllables of another word.

Following the pre-test, participants were asked to listen carefully to a stream of sounds composed of the six trisyllabic nonsense words (babupu, bupada, dutaba, patubi, pidabu, and tutibu) as described[2]. The continuous stream was created from a pseudorandom list (the same item never occurred consecutively) of the 6 nonsense words repeated 200 times each. The list was then modified such that all spaces between the words were removed. From this list, a sound file was synthesized using a text-to-speech application. The resulting 14-minute speech stream contained no pauses or other acoustic indications of word onset (e.g., babupudutabatutibubabupubupadapidabu...).

ERPs were recorded from a 29-channel cap containing tin electrodes (Electro-Cap International, Eaton Ohio) during this portion of the experiment. Electro-oculogram was recorded from electrodes above and below and at the outer canthi of the eyes. Impedences at all scalp electrode sites were maintained below 3 kOhms. The EEG was amplified by Grass amplifiers with a bandpass of 0.01 to 100 Hz and sampled every 4 ms during the presentations of continuous speech. All electrodes were referenced to a single mastoid (right) online and later re-referenced to the average mastoid (left and right). Participants were asked to look at a fixation point presented on the computer monitor for the duration of the sound stream. They were also asked to remain relaxed, not move, and blink normally during this part of the experiment.

After 14 minutes of exposure to the stream of continuous nonsense words, a second behavioral test was given to determine if participants had learned some of the words by listening to the continuous stream. Because performance on this test did not differ from chance, a training procedure was implemented.

For the training portion of the experiment, participants were specifically instructed to learn the six trisyllabic nonsense words. During the first 10 minutes of training, the nonsense words were presented with a 500 ms ISI. As each word was heard from the speaker, the printed version of the word was presented on the computer monitor. During the second 10 minutes of training, the nonsense words were presented with a 100 ms ISI.

Immediately following training, subjects were given a third test. Participants were instructed to circle the number corresponding to one of the six words they had just learned. ERPs were then recorded for another 14-minute period while subjects listened to a continuous string of the nonsense words. Following this second ERP recording session, a fourth and final behavioral test was given.

Artifact rejection algorithms were used to reject trials during which blinks or eye movements occurred before ERPs were averaged. A comparison of the number of trials rejected before and after training revealed no significant differences. ERPs recorded before and after training were averaged to the onsets of each syllable (initial, medial and final). To test the hypothesis that the same ERP word onset effects described in previous studies would be found for recently learned nonsense words, we measured the peak amplitude between 70 and 130 ms (N100). We also hypothesized that learned items might elicit an N400, an ERP component typically elicited by lexical items, so we measured the mean amplitude between 200 and 500 ms to test this hypothesis.

These dependent variables were analyzed using a four-factor, repeated-measures ANOVA: training (before, after) × electrode hemisphere (left, right) × electrode laterality (lateral, medial) × electrode anterior/posterior position (six levels). Additional ANOVAs were conducted for specific electrode sites as was motivated by training and electrode site interactions, as well as with group (high learners, low learners) as a between-subjects factor.

Acknowledgments

We thank R. N. Aslin for help with the stimuli, Y. Yamada for help with data acquisition, P. Compton and R. Vukcevich for technical support, and D. Coch, S. Guion, D. Poeppel, M. Posner, M. Spezio and D. Tucker for comments on previous drafts. Supported by the National Institute on Deafness and Other Communication Disorders (NIH grants DC00128, DC00481 and DC00167) and by the National Institute of General Medical Sciences (Institutional NRSA 5-T32-GM07257).

Competing interests statement

The authors declare that they have no competing financial interests.

Received 17 April; Accepted 29 May 2002

1. Cutler, A. & Butterfield, S. Rhythmic cues to speech segmentation: evidence from juncture misperception. *J. Speech Lang. Hear. Res.* **31**, 218–236 (1992).
2. Saffran, J. R., Newport, E. L. & Aslin, R. N. Word segmentation: the role of distributional cues. *J. Mem. Lang.* **35**, 606–621 (1996).
3. Sanders, L. D. & Neville, H. J. Lexical, syntactic, and stress-pattern cues for speech segmentation. *J. Speech Lang. Hear. Res.* **43**, 1301–1321 (2000).
4. Jusczyk, P. W. Infants' detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* **29**, 1–23 (1995).
5. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
6. Cutler, A., Mehler, J., Norris, D. & Segui, J. Limits on bilingualism. *Nature* **340**, 229–230 (1989).
7. Sanders, L. D., Neville, H. J. & Woldorff, M. G. Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues. *J. Speech Lang. Hear. Res.* (in press).
8. Sanders, L.D. & Neville, H.J. An ERP study of continuous speech processing: segmentation, semantics, and syntax in native speakers. *Cogn. Brain Res.* (in press).
9. Hansen, J. C. & Hillyard, S. A. Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr. Clin. Neurophysiol.*, **49**, 277–290 (1980).
10. Hink, R. F., Hillyard, S. A. & Benson, P. J. Event-related brain potentials and selective attention to acoustic and phonetic cues. *Biolog. Psychol.* **6**, 1–16 (1978).
11. McCandliss, B. D., Posner, M. I. & Givón, T. Brain plasticity in learning visual words. *Cogn. Psychol.* **33**, 88–110 (1997).
12. Chwilla, D. J., Brown, C. M. & Hagoort, P. The N400 as a function of the level of processing. *Psychophysiology* **32**, 274–285 (1995).
13. Holcomb, P. J. & Neville, H. J. Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Processes* **5**, 281–312 (1990).
14. Sanders, L.D. & Neville, H.J. An ERP study of continuous speech processing: segmentation, semantics, and syntax in non-native speakers. *Cogn. Brain Res.* (in press).