Rhythm and Stress in Speech Cue Temporally Selective Attention

by

Yue Jiang

A Thesis

Presented to the Faculty of Mount Holyoke College

in partial fulfillment of the requirements

for the degree of Bachelor of Arts

with Honor

ACKNOWLEDGEMENTS

I thank my thesis advisor, Lisa Sanders, for her continuing patience and strict requirements in guiding me through this thesis project. Thank you for giving me this opportunity to obtain hands-on experience on one of the most important techniques in cognitive neuroscience research as well as managing a project through the end. I also thank Katherine Binder, Joseph Cohen, and Francine Deutsch, who are my thesis committee members, for their helpful feedback on my proposal and drafts.

I thank Dominique Simmons, for her time spent in the lab, running the experiment and scheduling participants when I was not able to be around every day. Thank you for your hard work. I would not be able to finish this project without your assistance.

I am grateful for the help from members of NeuroCognition and Perception (NCaP) Lab at the University of Massachusetts, Amherst. Thank you to Nicolas Planet, Lori Astheimer, Mara Breen, Benjamin Zobel, Patrick Taylor, and Ahren Fitzroy, for helping me fix problems encountered in experiment as well as giving me valuable advice in research.

My friends and family have been giving me support and comfort. Thank you for encouraging me when I am stressed and for sharing my excitement when the project progressed well.

I thank Ms. Reese, and Mr. and Mrs. Harap for generously providing financial support for my research.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Previous research shows that listeners direct temporally selective attention to the initial portions of words in continuous speech. This processing strategy is useful since speech signals change too rapidly for listeners to form representations of every detail, and the initial portions of words are particularly helpful in auditory lexical access. However, very little is known about the cues that direct temporal attention during speech processing. In the current study, event-related brain potentials (ERPs) elicited by word and syllable onsets in artificial languages were compared and lexical stress and the regularity of lexical duration were manipulated as potential attention cues. The results suggest that listeners can use stress or rhythmic regularity to direct attention to upcoming word onsets. The study provides information about how listeners preferentially process the most relevant segments in rapidly changing speech streams, an important step in extracting meaning from spoken language.

INTRODUCTION

Speech Processing is Challenging

In everyday conversations, listeners are able to extract meaning from rapidly changing speech signals. Although this process seems relatively easy to the listener, the complexity of speech makes receptive language processing quite complicated. For example, speech signals occur over time. Once a sound is spoken, listeners cannot refer back to it, as they can with printed text. For this reason, listeners have to know in advance which of the many sounds in speech are going to be important for understanding a speaker's meaning. Moreover, there are no reliable acoustic indications of when one word ends and the next begins that parallel the spaces between words in written English. Therefore, before listeners can determine what a word means, they have to use multiple segmentation cues to determine which sounds in speech streams go together such that lexical access is attempted on meaningful units rather than nonsense sound sequences.

Additionally, speech signals change very rapidly in time, at a rate of 5 syllables per second or faster (Miller, Grosjean, & Lomanto, 1984). In fact, the changes in speech are so rapid that it is unlikely listeners can process all portions of the speech stream in detail. Instead, listeners have to select the portions of speech that are most relevant to their task of comprehension, and direct attention to those segments (Astheimer & Sanders, 2009). For example, listeners tend to pay more attention to word onsets, which are less predictable and thus more important than the medial syllables in words.

Behavioral Studies of Speech Segmentation

A large body of literature indicates that listeners can use a wide variety of cues to segment speech. Previous research suggested that listeners can use lexical, syntactic, and stress-pattern cues to segment natural speech (Cutler & Foss, 1977; Frauenfelder, 1985; Sanders & Neville, 2000). For example, the ability to recognize one word before its offset indicates that whatever sounds come next must be the onset of another word. Lexical recognition serves as a segmentation cue even when no other information about where one word ends and the next begins is available. In a study by Sanders and Neville (2000), listeners performed better on a phoneme detection task with normal English sentences than with sentences in which all of the content words had been replaced by non-words. This result points to the importance of lexical recognition for speech segmentation. In the same study, participants performed better with sentences that maintained the syntactic structure of English, even without semantic content, than with sentences that were constructed entirely with non-words. The presence of familiar syntactic structure also aided listeners in segmenting the nonsense speech.

Second, listeners can use statistical information about how often sounds occur next to each other in speech to segment even unfamiliar or artificial languages (Saffran , Newport, & Aslin, 1996). That is, two syllables that are part of the same word are heard together more frequently than two syllables that cross a word boundary (i.e., the transitional probability between syllables in the same word is larger than that between syllables in different words). The transitional probability of syllable X followed by syllable Y is calculated with the function:

P (Y follows X) = frequency of XY / frequency of X

The lower transitional probability between syllables that occur in different words than those often occur within the same word is sufficient for statistical learning and allow listeners to segment speech in the absence of any other segmentation cues.

Third, listeners can use a familiar stress-pattern to segment speech (Sanders, Neville, & Woldorf, 2002). In English, and some other languages such as Dutch, lexical stress is most commonly placed on the initial syllables of words. That is, the stress pattern of "*te*lephone", with the first syllable pronounced louder, longer, and with greater pitch change than the second, is more common in English than words with a stress pattern such as "ba*na*na". As a stress-timed language, English has somewhat regular durations between strongly stressed syllables, although number of syllables between strong stresses may vary (Cutler & Mehler, 1993; Nakatani & Schaffer, 1978; Vroomen & de Gelder, 1995). Sanders and Neville (2000) found that words and non-words with the more typical English stress pattern (i.e., strong stress on the initial syllable) was also related to better ability to determine whether a target sound was the onset of a word or not. Native English speakers have been shown to use the fact that, strong stress typically indicates the beginning of a word, to segment both natural and artificial languages (Connine, et al., 1993; Marslen-Wilson & Welsh, 1978; Salasoo & Pisoni, 1985). Salasoo and Pisoni (1985) used a spoken word recognition task to investigate the relationship between the acoustic and semantic information that was available in sentences. They found that lexical stress on word-initial syllables was especially helpful when listeners were processing meaningful sentences. Further, Cutler and Butterfield (1992) analyzed a corpus of “slips of the ear” and conducted an experiment on the misperception of word boundaries. Segmentation errors most frequently took the form of extra boundaries inserted before stressed syllables that were actually in the middle of words and failures to segment speech before weak, word–initial syllables.

Although not specifically manipulated in previous studies, it is also possible that the rhythmic regularity of word boundaries aids listeners in segmenting speech. The ability to use each of the segmentation cues described above, lexical recognition, transitional probabilities, and lexical stress, might be greater when the word boundaries occur with at least some temporal regularity, as isochronous rhythms have been shown to increase performance for predictable auditory events (Jones & Boltz, 1989; Large &

Jones, 1999). Studies that used natural English sentences as stimuli likely included somewhat regular intervals between the strongly-stressed, word-initial syllables (Cutler & Mehler, 1993; Nakatani & Schaffer, 1978; Vroomen & de Gelder, 1995). Studies that used non-word stimuli designed to sound like English would have also contained somewhat regular intervals between boundaries (Sanders & Neville, 2000). Finally, the artificial language studies used to determine that listeners can use transitional probability to segment speech (Saffran , Newport, & Aslin, 1996) employed languages in which all of the words were equal in length (i.e., six three-syllable nonsense words). As such, as soon as listeners began consistently segmenting the continuous streams of speech into words based on statistical cues, they also gained the ability to use rhythmic regularity of onsets as an additional segmentation cue. Thus, rhythmic regularity may have contributed to listeners' ability to segment speech in all of the described studies.

Event-Related Potentials (ERP) Studies of Speech Segmentation

Although behavioral studies of speech segmentation have defined the cues listeners can use to segment speech, they do not differentiate between the cues that are used during online processing and those that affect how listeners remember or report what they heard. For example, lexical recognition might influence which words listeners report having heard at the end of an entire sentence, but may not play a role in how speech was initially segmented as it was first heard. The event-related potential (ERP) technique provides an online measure of speech segmentation. For example, acoustically similar word and syllable onsets in continuous speech elicit significantly different ERP waveforms. Specifically, word onsets elicit a larger first negative peak 90-120 ms after onset (N1) than syllable onsets that are not the beginning of a word (Sanders & Neville, 2003a; Sanders & Neville, 2003b). Although many acoustic features were matched between the initial and medial syllables of words, in natural speech it is possible that unmatched physical differences such as timbre and pitch existed between the two groups of syllables, and that these physical differences rather than segmentation are responsible for the ERP effects. However, similar effects have been reported for synthesized nonsense speech such that physically identical word onsets elicit a larger N1 after compared to before listeners learn to recognize the words in continuous speech (Sanders, Newport, & Neville, 2002). Since the only difference in the sounds before and after training was listeners' ability to segment the streams and recognize the words, the ERP effects must reflect that difference.

Temporally Selective Attention

The ERP technique enables researchers to measure the differential processing of physically identical stimuli under attended and unattended conditions, and has been used to study temporally and spatially selective attention in the auditory and visual domains (Driver & Frackowiak, 2001; Eimer, 1993; Hansen, Dickstein, Berka, & Hillyard, 1983; Heinze & Mangun, 1995; Hillyard, Woldorff, Mangun, & Hansen, 1987; Hink, Van Voorhis, & Hillyard, 1977; Neville & Lawson, 1987; Picton & Hillyard, 1974).

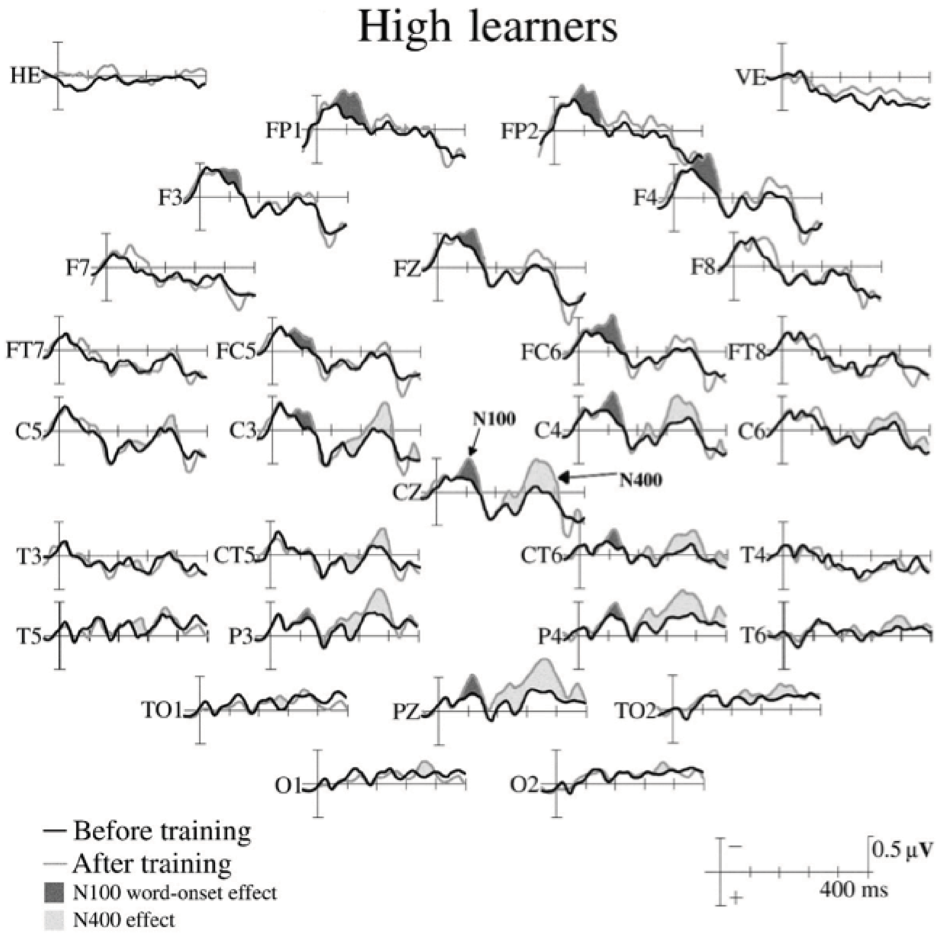
Although the ERP effect described in the study by Sanders and Astheimer (2000) and Sanders, Newport, and Neville (2002) indicates that speech must have been segmented, it may not index the process of segmentation directly. Both behavioral and ERP studies have shown that listeners can direct attention to specific times and that doing so improves performance for sounds presented at attended times and results in a similar ERP effect as that reported for segmentation (Abla, Katahira, & Okanoya, 2008; Lange, Rösler, & Röder, 2003; Sanders & Astheimer, 2008; Sanders & Neville, 2003; Sanders & Neville, 2003; Sanders, Newport, & Neville, 2002; Sanders, Ameral, & Sayles, 2009). Sounds presented at attended compared to unattended times elicit a larger amplitude negativity approximately 100 ms after onset (N1). The segmentation effect described above might reflect listeners' directing temporally selective attention to the initial portions of words presented in continuous speech. Since the initial segments of words have been shown to be particularly important for lexical recognition in English, directing attention to these segments might help listeners recognize words in rapidly changing speech streams.

The abilities to segment speech and to selectively attend to the initial portions of words have been demonstrated with natural language (Astheimer & Sanders, 2009). Astheimer and Sanders used ERPs to investigate the distribution of temporal attention during sentence processing. With twenty-two adult participants, they found that behaviorally irrelevant attention probes elicited larger auditory evoked potentials when presented during the initial portions of words rather than the medial or final portions of words or at random control times. These data show that listeners were selectively attending to the initial portions of words in continuous speech. However, it was not clear what characteristics of speech served as cues to help the segmentation process and the allocation of attention to the beginnings of words.

Attention Cues in Speech

Although it has been shown that listeners selectively attend to the initial portions of words in continuous speech, it is not known if the same cues used to segment speech are also important for cuing attention. That is, the same lexical recognition, transitional probability, lexical stress, and rhythmic regularity cues that help listeners to identify word boundaries may also serve as attention cues. In the study by Sanders, Newport, and Neville (2002), the researchers used the ERP technique to investigate the allocation of attention to the initial portions of words by presenting participants with continuous streams of six three-syllable nonsense words. The stimuli were constructed as described in a study by Saffran, Newport, and Aslin (1996). Eighteen right-handed, adult native English speakers first completed a behavioral pretest to detect any bias for the words in the artificial language or non-words constructed from the final portions of one word followed by the initial portions of another. Listeners were then exposed to the syllable stream for 14 minutes while having their ERPs recorded. They then were given another behavioral test to see how much they had learned just by listening to the syllable stream (based on the transitional probabilities). A twenty-minute training session followed, in which participants were explicitly taught the six words presented in isolation, and participants took a third behavioral test. ERPs were recorded again when participants listened to another 14-minute syllable stream, now with the ability to consistently segment the continuous speech into the familiar six three-syllable words. The experiment session ended with a fourth behavioral test. The results indicated that there were differences in the auditory evoked potentials for initial and medial syllables even before training. These differences likely reflected the acoustic differences in the sounds used as initial and medial syllables. Importantly, for the group of listeners who performed well on the behavioral test given immediately after training, there was also a syllable position by training interaction such that, as shown in Figure 1, initial syllables elicited a larger first negative peak (N1) after compared to before training. The same effect was not evident for medial syllables. Learning to recognize the words resulted in differential processing of initial syllables that was not evident as overall before-after effects for other syllables. This result suggests that listeners were directing attention to the initial portions of words in

Figure 1. Training effects on N1 and N400 amplitude for the group of participants who were most successful at learning an artificial language in Sanders, Newport, and Neville (2002). Figure is included with the permission of the authors.


High learners
HE
VE
FP1
FP2
F3
F4
F7
FZ
F8
FT7
FC5
FC6
FT8
C5
C3
N100
C4
C6
CZ
N400
T3
CT5
CT6
T4
T5
P3
P4
T6
TO1
PZ
TO2
O1
O2
— Before training
— After training
N100 word-onset effect
N400 effect
−
+
0.5 µV
400 ms

continuous speech after they were given enough information to consistently identify the word position of the syllables they heard.

The current study used a similar design but manipulated the rhythmic regularity of word onsets by employing two different artificial languages, both of which had lexical stress on the initial syllable of every word. Specifically, participants were asked to listen to a stream of synthesized speech constructed from six words repeated continuously in random order. The syllables that make up words occurred together more often in the stream (i.e., had higher transitional probability) than the syllables that cross word boundaries. If listeners are able to use the transitional probabilities or word-initial stress to segment speech and direct attention to the initial syllables, they should learn the words just by listening to the continuous stream. After listening to the stream, participants were explicitly taught to recognize the six words that make up the language. If lexical recognition is important for segmentation and directing attention to onsets, this training should affect both behavioral and ERP measures. The rhythmic regularity of lexical onsets and lexical stress was manipulated between subjects by giving some listeners Language A made of six three-syllable words and other listeners Language B made of two 2-syllable, two 3-syllable, and two 4-syllable words.

In a previous study (Sanders et al. 2002, Figure 1), it was shown that the initial portions of words elicit a larger amplitude N1 after compared to before training. In that experiment, training gave listeners the ability to use both lexical recognition and rhythmic regularity to direct attention to the initial portions of words. If adding stress to the first syllable of each word is sufficient for listeners to allocate attention in a similar manner even before training, training effects should disappear. If being able to predict when stress will occur even before training is necessary to allocate attention to the initial portions of words, presenting a language with stress on the word initial syllables but irregular word length should restore the previously reported effects, and word-initial, but not word-medial, syllables should again elicit a larger N1 after compared to before training. Therefore, it was hypothesized that participants who learned Language A would be able to segment the words even before the training session just by relying on the rhythmic stress, and thus the training effect would not be present in this group. On the other hand, participants who were taught Language B would not be able to anticipate when stressed syllables would occur in the stream until they had learned the words. Thus training would be necessary for this group to direct attention to the word onsets.

METHOD

Participants

This research project was conducted in the NeuroCognitive and Perception Lab at the University of Massachusetts, Amherst. The Institutional Review Board at the University of Massachusetts at Amherst approved all procedures. A total of 107 participants completed at least some portion of the experiment. However, data from 56 were excluded because of poor performance on the behavioral tests of word learning. The reason was that lexical knowledge was one of the independent variables in the current study, and it was important that participants were able to distinguish the target words from the non-word foils, so that we could compare the ERP when lexical access was and was not available. Data from another 15 were excluded because of artifacts such as bridging, eye blinks, and drift in the EEG recordings. Therefore, ERP data from 36 participants (half learned each language) were used for analyses. All participants were adults between 18 and 35 years old, were native English speakers and right-handed, and had normal hearing and normal or corrected to normal vision. Those who were included in analyses had not taken neurological or psychoactive medication in the past six months, and had not participated in any similar experiments with artificial languages. Participants were paid $10 for each hour of participation or received participation credits they could use in classes in the psychology courses at the University.

Material

Two artificial languages were synthesized using text-to-speech software (Acapela). In both languages, there were 6 target words made of 11 consonant-vowel (CV) syllables that are common and equally likely to be heard as the first syllable of a word (word-initial syllables) and at other positions in the word (word-medial syllables) in English: ba, da, pa, ta, bi, pi, ti, bu, du, pu, tu with durations ranging from 150-300 ms (Saffran , Newport, & Aslin, 1996; Sanders, Newport, & Neville, 2002). To mimic natural languages, some syllables occurred in more words than others, so that the transitional probabilities between syllables within words ranged from 0.17 to 1.0, while the transitional probabilities between syllables spanning word boundaries ranged from 0.09 to 0.17.

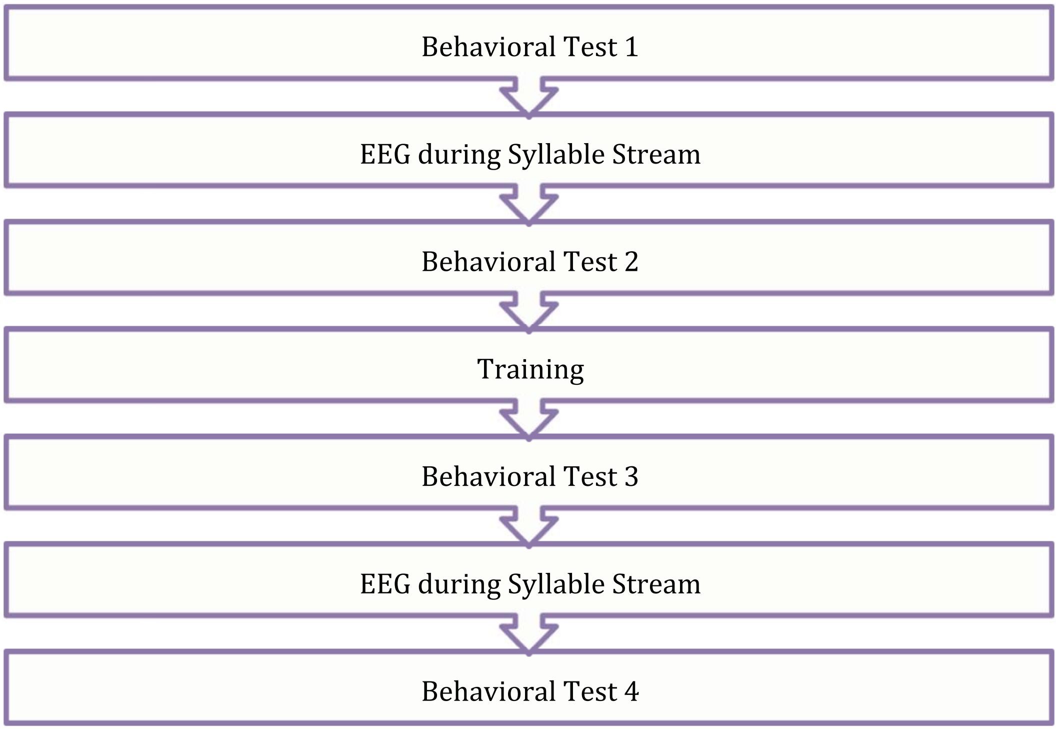
Language A was constructed from six three-syllable words (babupu, bupada, dutaba, patubi, pidabu, tutibu). Language B consisted of six words that varied in length from two to four syllables (bita, dabu, pababu, tuduti, bubapapu, pibutuda). For both groups, three-syllable non-word foils were formed from the first syllable of a three-syllable word followed by the last two of another three-syllable word (Language A: batibu, butaba, dutubi, padabu, pibupu, and tupada; Language B: paduti and tubabu). For Language B, two-syllable foils were created from the first syllable of a two-syllable word followed by the last syllable of another two-syllable word (bibu and data); four-syllable foils were formed from the first two syllables of a four-syllable word and the last two from another (bubatuda and pibupapu). The initial syllable of each word in both languages was stressed so that it was 5 db louder than the other syllables.

For each language, continuous sound files containing 200 repetitions of each word, arranged according to experimental objectives, were created by concatenating syllables in MATLAB. Each stream was divided into two 8-minute blocks consisting of the six target words occurring repeatedly and pseudo-randomly (a word did not occur two or more times consecutively). All sounds were stored as WAV files with a 44.1 kHz sampling rate.

Procedure

The experiment strictly followed the running protocols (Appendix A) and the procedure depicted in Figure 2. Participants signed the research consent form on their arrival (see Appendix B). Participants then completed an initial behavioral test designed to detect any preference for the words or non-words. The test also familiarized participants with other similar behavioral tests in the rest of the experiment session. In this first test, participants pressed one of the four buttons on a button box each time they heard a syllable string to indicate how much they liked the sequence. Each of the words and foils were separately presented three times in random order, and participants were thus asked to indicate their preference for 36 times. The button “1” indicated “dislike this word a lot”, “2” indicated “dislike this

Figure 2. Experiment Procedure.


Behavioral Test 1
EEG during Syllable Stream
Behavioral Test 2
Training
Behavioral Test 3
EEG during Syllable Stream
Behavioral Test 4

word a little", "3" indicated "like this word a little", and "4" indicated "like this word a lot". Performance was assessed by measuring the d' value. The d' is computed with the function

$d' = z(H) - z(F)$

H, hit rate = P (participant pressed 3 or 4 | the word appeared in training)

F, false alarm rate = P (participant presses 3 or 4 | the word did not appear in training)

Therefore, a higher d' score indicates higher hit rate and lower false alarm rate, indicating the listener prefers words in the artificial language and non-words.

After the first behavioral test, they listened to the stream of the language they were assigned to while having their EEG signals recorded. During each block, participants were asked to stay relaxed and alert while focusing on a fixation point, a white cross on the monitor screen in front of them. Between the two blocks, participants were given a break while the researchers checked the ERP facilities.

At the end of the ERP session, they took the second behavioral test to estimate to what extent mere exposure to the language could familiarize them with the target syllable sequences. In this test, they would press one of the four buttons on the button box to indicate how familiar they were with each syllable sequence, with "1" indicating "this word definitely appeared in the stream", "2" indicating "this word may have appeared in the stream", "3" indicating "this word may not have appeared in the stream", and "4" indicating "this word definitely did not appear in the stream". In this and the later tests, a higher d' value indicates listener's sensitivity to the difference between words and non-words.

A training session followed to let participants learn the target sequences. A PowerPoint slide with instructions and the six target syllable sequences were brought up on the monitor in front of participants. They would click on the speaker icon next to each word to listen to the words. They were instructed to use as much time as they needed and whatever strategies they wanted to use to learn these words, so that when they heard a word, they could recognize it. They were also told that a behavioral test similar to the previous ones would follow the training. Participants informed the researcher that they were ready to take the test, and a third behavioral test was presented. This time, participants were asked to indicate their confidence that a syllable sequence was one of the words that appeared in the training session by pressing buttons. "1" indicated "definitely not a word in training", "2" indicated "maybe not a word in training", "3" indicated "maybe a word in training", and "4" indicated "definitely a word in training". To ensure that participants were familiar enough to the words, a minimum of d' = 1.80 was required for them to pass the test and continue to the next session, and a higher d' score indicated better memory of the words.

If a participant failed to reach the criterion on the behavioral test immediately after training, she was asked to continue the explicit training on the words and to repeat the post-training behavioral test again. If a participant failed the test four times in a row, she was told that the experimental session was over and compensated for the time already given to the experiment.

Those who passed the test listened to the continuous stream for 16 minutes again after the training session and had their EEG recorded. The stimuli used in this ERP session were the same as in the previous one, where participants were instructed to look at the fixation point on the monitor screen while listening closely to the syllable stream as a review to prepare for a behavioral test that follows.

A final behavioral test confirmed that the memory of the words was retained during the last ERP session. For this final behavioral test session, a minimum of d'=1.0 was required for the data from a participant to be kept and analyzed. The d' criterion was lower in this final test because it was expected that the memories of the words would fade to some extent. As long as the performance was well above chance in this test, the participant was considered to have had the words in mind during the 16-minute ERP session. Participants were paid or awarded 4 research credits that could be used for partial fulfillment of requirements in Psychology courses. Participants were fully debriefed at the end of the experiment to ensure that their efforts in the session constituted an educational experience about research.

Apparatus

Presentation of instructions, fixation point, and sounds, as well as collection of manual response data, were controlled by E-Prime software running on a PC with a dual-monitor system such that experimenters can observe the stimuli being presented to participants. The E-Prime computer sent event markers to another computer that recorded EEG data. An external timing device was in place between the two computers to ensure synchronization of the two systems.

ERP methods adhered closely to the current recommended guidelines (Luck, 2005; Picton, et al., 2000). The Nyquist Theorem states that all of the information in an analog signal such as the EEG can be captured digitally as long as the sampling rate is at least twice as great as the highest frequency in the signal. Because participants were adults, the lower end of the bandwidth did not need to be high to exclude very low frequency waves caused by motions, like those produced by infants. Therefore, a bandwidth of 0.01 – 100 Hz was used to capture both low and high frequencies that may relate with stimulus processing. However, a 100Hz cutoff does not suppress everything above 100Hz (Luck, 2005). Therefore, a sampling rate of 250 Hz was used to capture the early components in waveforms such as N1 and avoid too huge a file size (Luck, 2005). EEG was recorded with a vertex reference using a 128-channel net system (Electrical Geodesics Inc., Eugene OR). Four sizes were available to achieve a close fit for every participant. Scalp impedances at all electrode sites were maintained below 100 kΩs, which has been shown to be sufficient to reduce high-frequency noise when using high input-impedance amplifiers, and the impedances at the key electrodes were maintained below 50 kΩs (Ferree, Luu, Russell, & Tucker, 2001). Impedences were lowered by using a saline conducting solution with potassium salt and baby shampoo. When applying the saline, participants were given breaks between application of the solution and recording to eliminate any “bridges” created between electrodes. Temperature of the experiment room in which EEG was recorded was maintained at or under 22 °C to avoid skin potentials that contribute to low-frequency noise.

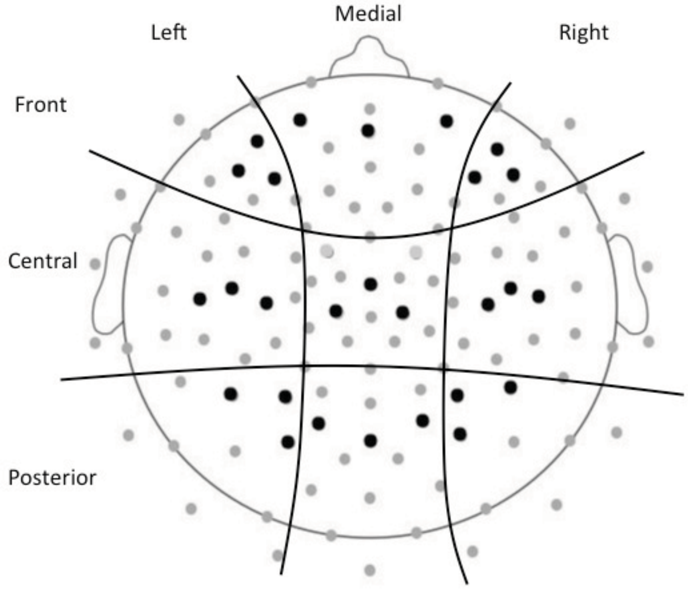
Data Processing

A 60 Hz filter was applied to the continuous EEG data to eliminate noise from electrical sources (e.g., the computer monitor). Then epochs, or time windows, were defined as 100 ms before to 500 ms after each stimulus onset. Artifact rejection criteria were applied to eliminate epochs that contained artifacts. In more detail, channel-specific maximum amplitude change criteria were set for each person based on observations of blinks, eye movements, and head movements made while participants listened to instructions. Initially, electrodes with amplitude changes that were greater than 70 μV within a trial were marked as bad channels, and trials with more than 5 bad channels were eliminated from analysis. The exclusion criteria were adjusted separately for each participant to include as many artifact-free trials as possible in the individual subject averages time-locked to the initial and medial syllables, before and after training. Only data from participants with at least 100 artifact-free trials in each of the 12 syllable × training × preceding word length interacting condition were included in analyses. The 600 ms EEG epochs were averaged for each condition and electrode site for every individual. Data were re-referenced to averaged-mastoid measurements to facilitate comparison with previous research. The 100 ms pre-stimulus interval was used as a baseline.

Average amplitude was measured 50–100 ms (P1), 110–160 ms (N1), and 200–495 ms after stimulus onset. Data from 27 electrodes in 9 regions (left-anterior, medial-anterior, right-anterior, left-central, medial-central, right-central, left-posterior, medial-posterior, and right-posterior) were selected from among the total 128 electrodes to investigate the scalp distribution of any differences in amplitude. The electrodes that provided data included in statistical analyses are shown in Figure 3.

Measurements from each of the 27 electrode sites were entered into omnibus 2 (syllable location) × 2 (before and after training) × 3 (lengths of the preceding word) × 3 (left, medial, and right regions) × 3 (anterior, central, and posterior regions) ×3 (electrode) repeated-measures ANOVAs

Figure 3. Electrodes that provided ERP data included in analysis.


Medial
Left
Right
Front
Central
Posterior

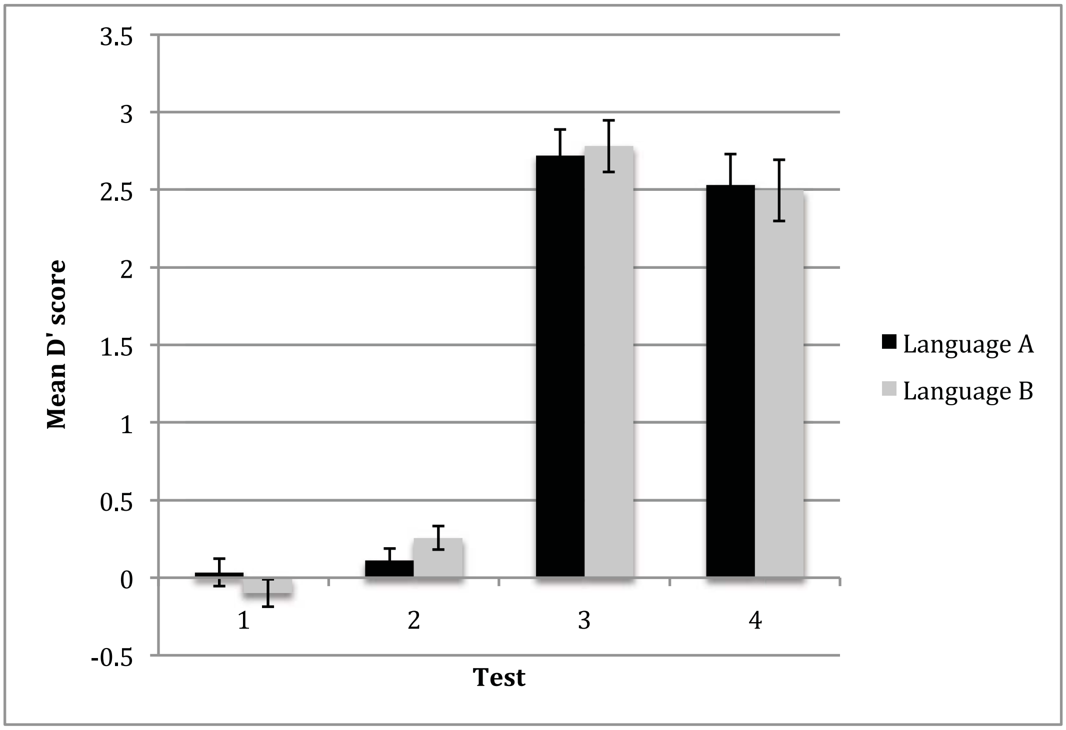
(Greenhouse-Geisser adjusted) in SPSS. Three-syllable words in Language A were randomly assigned to the conditions “followed 2-, 3-, and 4-syllable words” so that exactly the same analyses could be conducted on the two conditions, and the noise included in data in every condition could be similar for Languages A and B. Therefore, if there were any possible differences between groups, we would be able to compare data sets with the same sample size. Main effects of attention and interactions among attention and electrode location factors ($p < .1$) were explored by conducting ANOVAs at selected sets of electrodes as indicated by the specific interactions.

RESULTS

Behavioral Data

Behavioral data from the 36 participants who successfully finished the experiment and produced quality ERP data were entered into a 4 (test) × 2 (language) mixed ANOVA. Scores from the four tests significantly differed from each other ($F(3,102) = 251.18$, $p < .001$). Language type did not have a significant main effect on the scores ($F(1,34) = .005$, $p > .1$). It was predicted that the performance in the second test in both Languages A and B would improve if listeners could use transitional probabilities between syllables as a cue to learn the words. Post hoc tests indicated that across Language, the second behavioral test after exposure to the continuous stream of words did not improve compared to the initial test given before participants listened to the language ($p > .05$). This result indicates that participants did not learn to distinguish the words from non-words based on the transitional probability pattern or the presence of stress on the initial syllables just by listening to the syllable streams. As shown in Figure 4, training significantly raised the behavioral test scores above chance levels (chance = 0, one-sample $t(35) = 23.71$, $p < .001$). The scores on the final test remained well above chance performance (one-sample $t(35) = 18.29$, $p < .001$), and were not lower than those on the test immediately following explicit training (compared to test 3, $p > .1$).

Figure 4. Bar graph of performance on behavioral tests for participants who learned the rhythmically regular (Language A) and irregular (Language B) artificial languages.


3.5
3
2.5
2
1.5
1
0.5
0
-0.5
Mean D' score
1
2
3
4
Test
Language A
Language B

ERP data

Language A with Initial Stress and Regular Rhythm

The P1 component is the first positive peak occurring 50-100 ms after stimulus onset. It is usually seen as a sensory component. Training and syllable location did not have any effect on the P1 amplitude in participants who listened to the rhythmically regular Language A.

No significant training effect was found on the N1 component, the first negative peak that occurs 100-160 ms after the onset of either the initial or medial syllables. This lack of a training effect was predicted for Language A since both stress and rhythm cues were available to direct attention even before participants learned the words of the language.

A main effect of syllable was evident on mean amplitude 200-495 ms after stimulus onset over posterior regions (Syllable position × Anterior/Posterior interaction: F (2, 34) = .97, p = .389; Main effect of Syllable position at posterior sites only: F(1, 17) = 6.16, p = .024). However, the larger negativity evident in ERPs time-locked to the medial compared to initial syllables may actually reflect the acoustic differences between the final and initial, strongly-stressed, syllables. That is, 200-495 ms after a medial syllable, listeners were hearing a strongly stressed initial syllable. That would have occurred at a later time after the initial syllables. This result suggests that the transition from a softer final syllable to a louder initial syllable resulted in a larger posterior negativity. Training had no effect on this portion of the waveform.

Comparison between before and after training on P1, N1, and the later component elicited by the initial syllables is shown in Figures 6. Figure 7 shows the same comparison on ERPs elicited by the medial syllables. Waveform figures include data collected from electrodes indicated in Figure 5.

Language B with Initial Stresses and Irregular Rhythm

For P1 amplitude, there was a main effect of Syllable position at medial and central electrode sites (Syllable position × Laterality interaction: $F(2, 34) = .263$, $p = .646$; Main effect of Syllable position at medial sites only: $F(1, 17) = 4.51$, $p = .049$; Main effect of Syllable position at central sites only: $F(1, 17) = 7.47$, $p = .014$). The second syllables elicited larger P1s than initial syllables. This result was somewhat surprising since the louder initial syllables were expected to elicit larger auditory evoked potentials. No training effect was found.

No effects of training, syllable position, or interactions with the length of the preceding word were found on N1 amplitude, suggesting that participants did not allocate attention differently before and after training according to the syllable position or the length of preceding words.

A marginally significant effect of training was found on ERP amplitude 200-495 ms after stimulus onsets at central electrodes (Training ×

Figure 5. Waveform figures include data from the nine electrodes marked below.

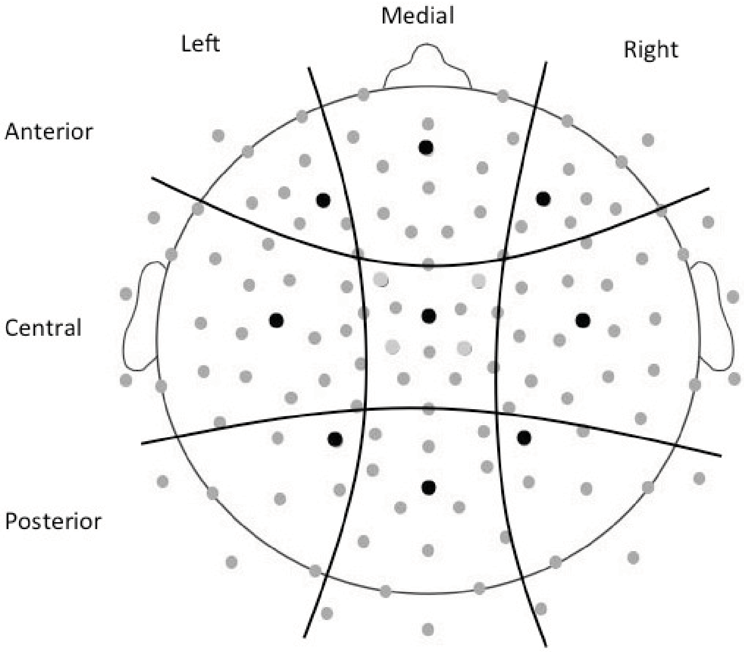

Medial
Left
Right
Anterior
Central
Posterior

Figure 6. ERP waveform elicited by word onsets before and after training in Language A with consistent word length.

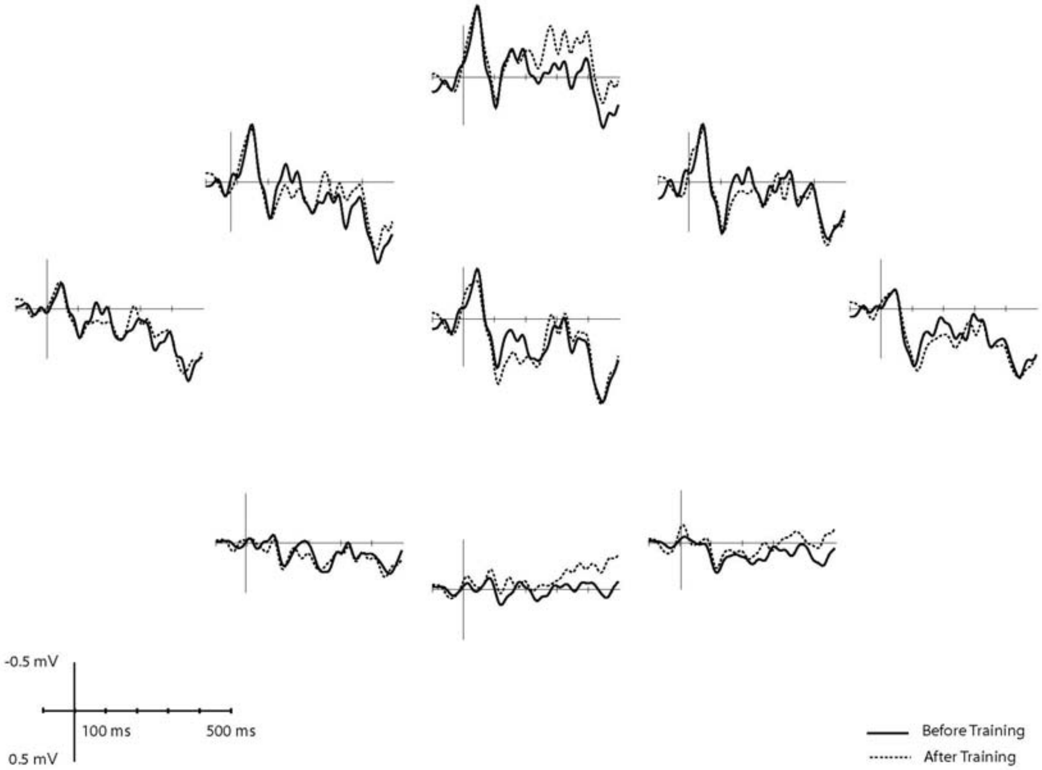
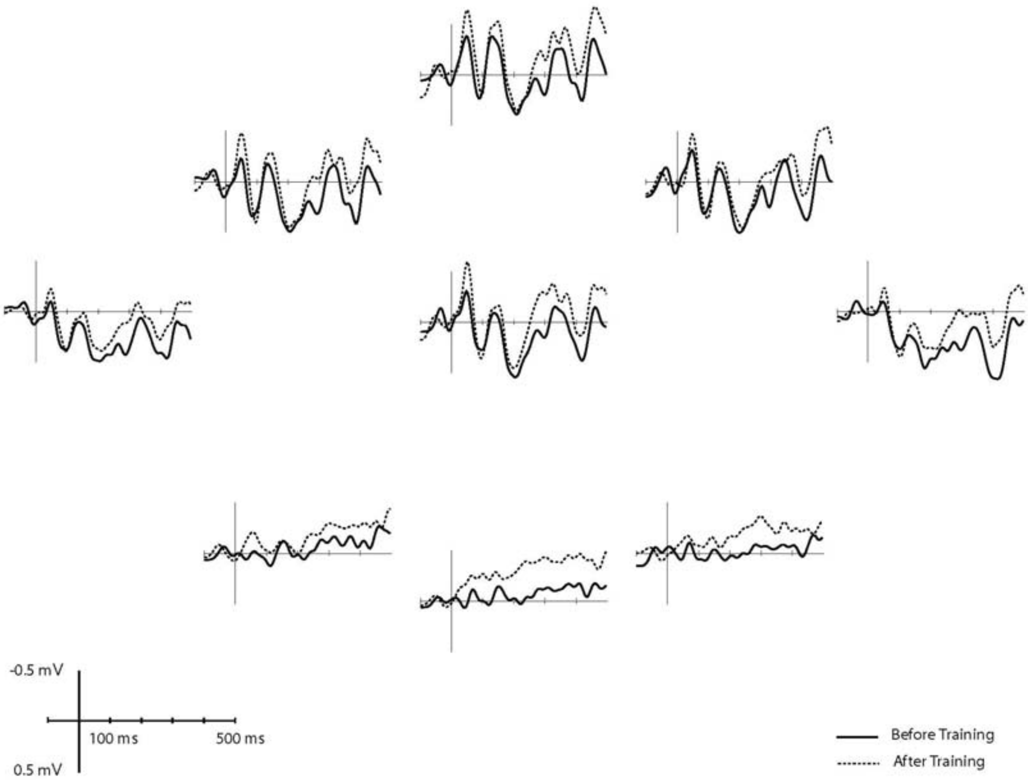

-0.5 mV
100 ms
500 ms
0.5 mV
Before Training
After Training

Figure 7. ERP waveform elicited by second syllables of the words before and after training in Language A with consistent word length.


-0.5 mV
100 ms
500 ms
0.5 mV
Before Training
After Training

Anterior/Posterior interaction: $F(2, 34) = 0145$, $p = .805$; Main effect of Training at central electrodes only: $F(1, 17) = 3.39$, $p = .083$). At these sites, the ERP amplitude was more negative before training. Once participants were familiar with the words and knew what to expect, a smaller response was observed. A marginally significant syllable by training interaction was observed over the posterior region (Syllable position × Training × Anterior/Posterior interaction: $F(2,34) = .15$, $p = .802$; Syllable position × Training interaction at posterior electrodes only: $F(1,17) = 3.36$, $p = .085$). However, further analyses at the posterior sites did not reveal a main effect of training for either the initial or second syllables.

ERP waveforms time-locked to the initial syllables of words before and after training are shown separately for words that followed 2-syllable words (Figure 8), 3-syllable words (Figure 9), and 4-syllable words (Figure 10).

Figure 8. ERP waveforms elicited by initial syllables after 2-syllable words before and after training in Language B with variable word length.

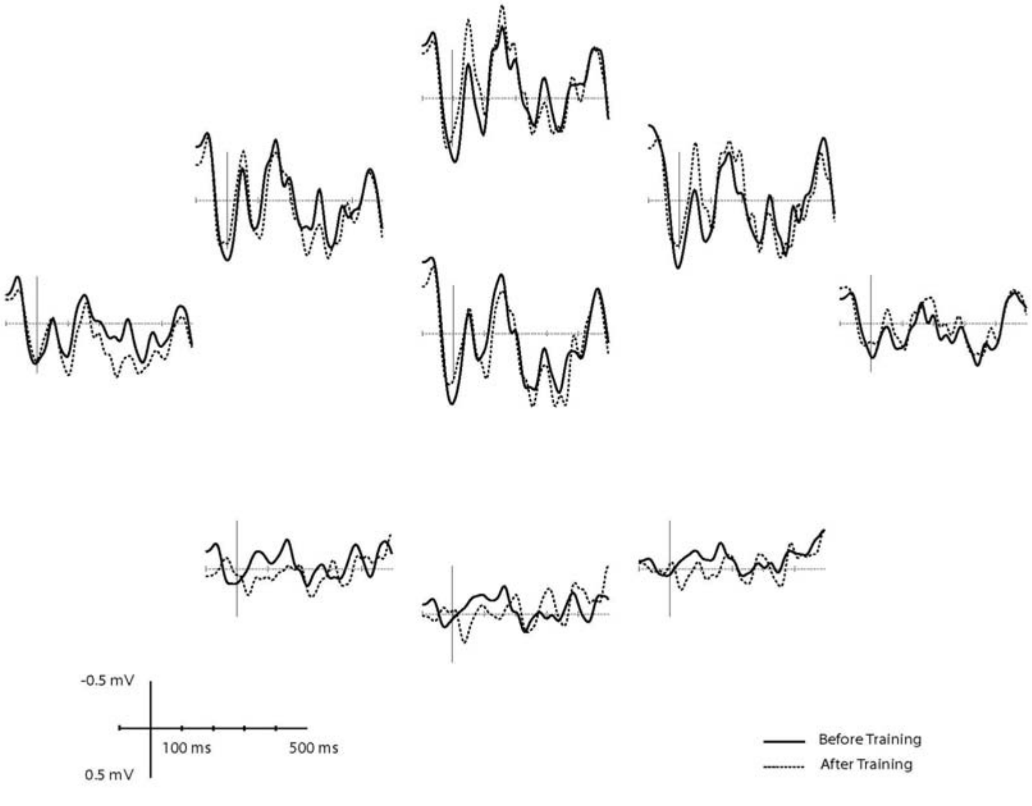

-0.5 mV
100 ms
500 ms
0.5 mV
Before Training
After Training

Figure 9. ERP waveforms elicited by initial syllables after 3-syllable words before and after training in Language B with variable word length.

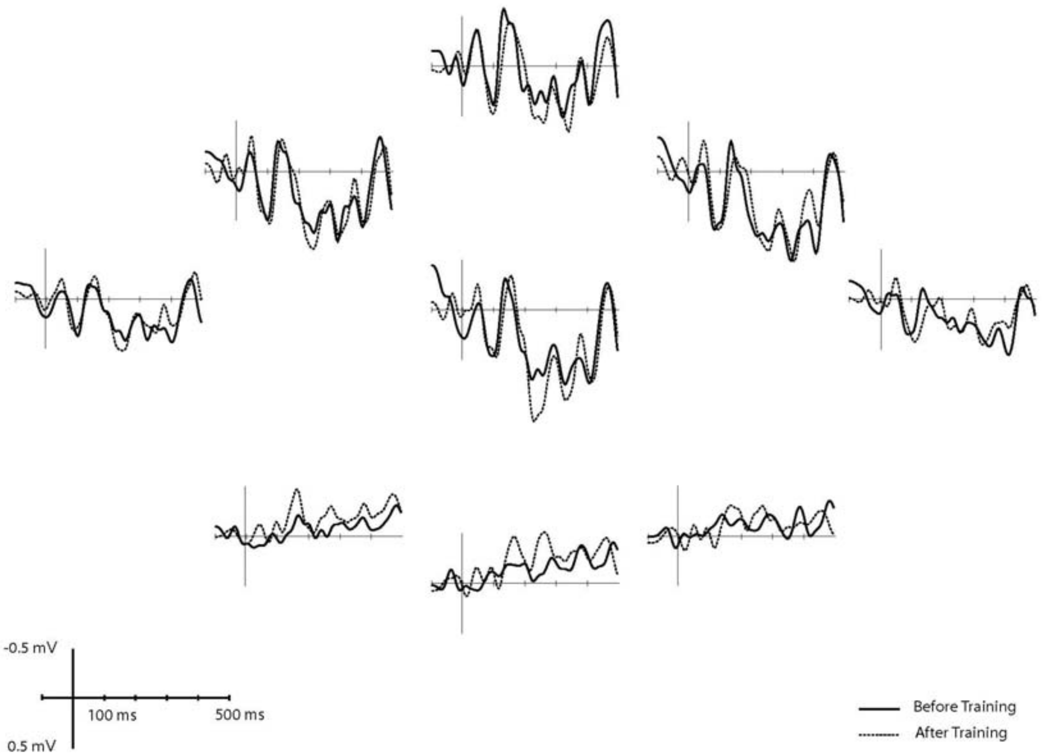
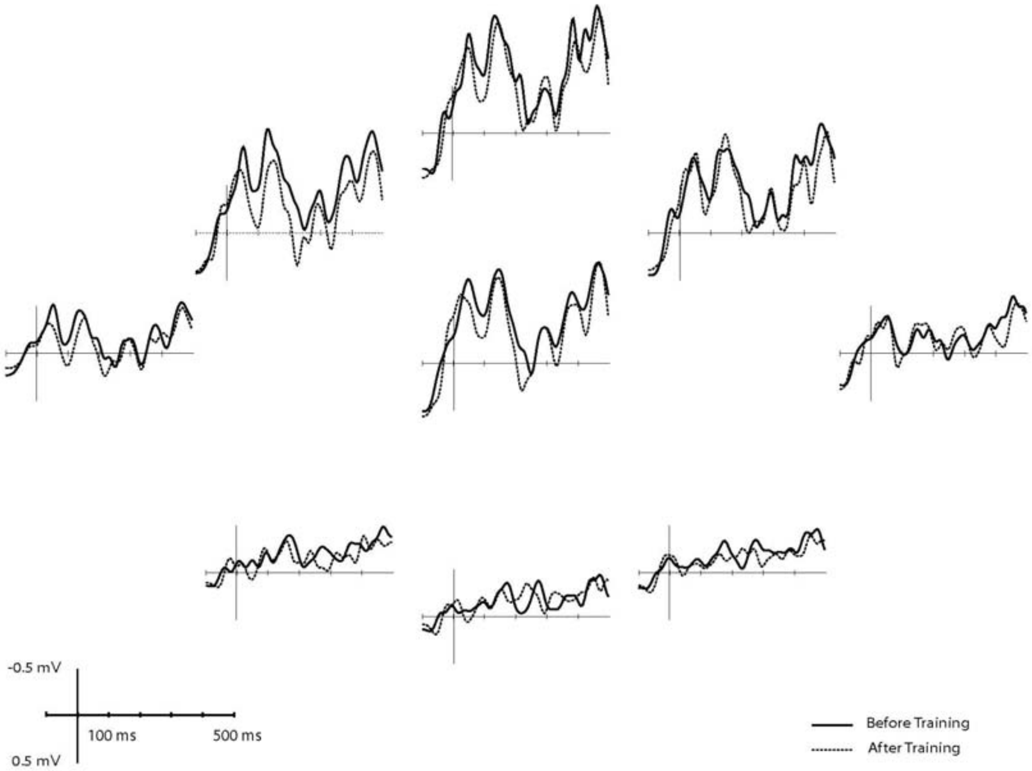

-0.5 mV
100 ms
500 ms
0.5 mV
Before Training
After Training

Figure 10. ERP waveforms elicited by initial syllables after 4-syllable words before and after training in Language B with variable word length.


-0.5 mV
100 ms
500 ms
0.5 mV
Before Training
After Training

DISCUSSION

In the present study, the behavioral test score did not improve from the first test to the second one after participants listened to one of the artificial languages for 16 minutes. Analyses on the ERP data from Language A revealed no significant overall effects of training or syllable position on P1 and N1 components, but syllable position had an effect on the ERP 200-495 ms after syllable onsets at posterior sites, with larger amplitude of ERPs following the second syllables compared to that of the ERPs after the initial syllables. Data from Language 2 showed a syllable position effect on P1 at the central electrodes. Surprisingly, the second syllables elicited a more positive P1 than the initial ones did. No training, syllable position, or length of preceding words was observed on N1. Training had a marginally significant effect on the ERPs 200-495 ms after syllable onsets.

Implications of Behavioral Data on Statistical Learning

Performance on the behavioral test after exposure to the continuous streams of speech was not above chance as was predicted. The result is surprising in light of the large literature indicating the importance of statistical learning as a mechanism in language processing and acquisition, i.e. listeners can learn words in an artificial language based only on the higher transitional probabilities for syllables within the same word compared to that for syllables that cross word boundaries (e.g. Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran, Newpot, Aslin, Tunick, & Barrueco, 1997). Further, in the current study, word learning was expected to be even stronger since participants had a familiar stress-pattern cue that could have contributed to consistent segmentation even before training. The lack of learning in the current study suggests that statistical learning may be a less important mechanism than previously reported. Statistical learning may be too fragile to support learning of a natural language, in which the transitional probabilities within and across words are more similar to each other than what is constructed with artificial languages.

Saffran, Aslin, and Newport (1996) had 8-month-old infants listen to an artificial language consisting of four 3-syllable nonsense words ("words") for two minutes. In addition, they constructed "non-words" and "part-words" as foils. "Non-words" were made of syllables that appeared in the familiarization stream but not in the order in which they appeared as words. "Part-words" were built with the final syllable of a word and the first two syllables of another word. The authors then assessed infants' preference by presenting a word, a nonword, or a part-word repeatedly and measuring the infants' listening time. They found a dishabituation effect in this study, i.e. infants attended to the unfamiliar foils for longer time than to the words that they heard. However, although the difference between the listening time for words and foils was statistically significant, the mean listening time for words and foils was very close to each other. While two minutes of listening might be enough for infants, older children and adults seem to need more exposure to the language before they can differentiate between words and non-words or part-words.

Saffran, Newport, & Aslin (1996) used the same language and asked adults to decide whether a syllable sequence was a word in the language after they listened to the speech stream for 21 minutes. Participants' performance in the discrimination task was above chance, which supported the hypothesis that distributional cues (transitional probability between syllables) were helpful in language learning. However, while results from the infant study showed that 8-month-olds might be more familiar with the part-words (constructed with two syllables from a word and one from another) than non-words (constructed with syllables that never followed each other in the language), the adult listeners in the non-word test did much better than those who took the part-word test.

Given that the results from the current study did not show any learning after the 16-minute exposure to the artificial language, it is possible that statistical learning does not always play an important role in artificial language learning. For example, when listeners are instructed to figure out where the words begin and end, as in the study by Saffran, Newport, & Aslin (1996), the instruction might help them more with learning than a simple instruction to "listen closely" as in the current study would. It is also possible that learning a language through the distribution cues can only be effective when the exposure to the words is much enough. In a study by Saffran et al.

(1997), when they doubled the time of listening to 42 minutes, or 600 repetitions for each word, participants' performance improved greatly. Sixteen minutes of exposure with 200 repetitions for each word during the first ERP session of both conditions in the current study could be insufficient for participants to improve their performance in the behavioral test. Furthermore, participants in the current study may not have been processing the artificial language as something important they should be trying to learn. During data collection, we noticed that fewer participants were able to learn the target words even after training toward the end of the semester than when the semester just started. People may have been less motivated to learn these words when there were many other things going on in their lives. This offers an alternative interpretation to why there were no early ERP effects on P1 and N1. Listeners could have been allocating attention to stressed syllables even before training, or it could be that they were not engaged in processing the language either before or after.

Interpretations of ERP Data

The results of analyses on the ERP data suggest that for participants who listened to Language A with approximately isochronous stresses, the sensory detection reflected in the P1 component was not affected by syllable position or training. Therefore, changes in any other components were unlikely to be due to acoustical differences in syllables. With the regular rhythm in Language A, listeners could adjust to the stress not show the difference between P1 elicited by initial syllables and that elicited by medial syllables.

There has been research suggesting that listeners form strong expectations about the rhythmic pattern, sometimes even when there is no physical indication of accents such as louder or longer sounds (Brochard, Abecasis, Potter, Ragot, & Drake, 2003). In this study, Brochard et al. used the ERP technique to look at subjective accenting, the phenomenon that when listening to identical sounds that are presented isochronously, listeners do not perceive the sounds as identical, but add accent to some of the sounds, and these accented sounds often form an organized rhythmic pattern. The authors made sounds of odd number (9th and 11th sounds, usually the down beat in music) louder or even number (8th and 10th sounds, usually the weak beat in music) softer as diviants, and compared the differences in the ERP elicited by these deviants. They found a larger ERP response to the soft sounds at odd-numbered positions, which indicated that listeners were more sensitive to things that deviate from their expectation, whereas the smaller response to the deviations at even-numbered positions could be explained by periodically deployed attention.

In another ERP study, Shaefer, Vlek, and Desain (2011) presented participants with sound sequences with stresses occurring on 2-, 3-, or 4-beat cycles to establish rhythm. The stresses later faded out and participants were instructed to imagine the accents. The researchers found an increase in amplitude on the early ERP components elicited by both the actual and imaginary stresses compared to other sounds in the rhythmic cycles, suggesting again that listeners can entrain to the external rhythmic patterns and can keep the rhythm even when the acoustic cues are absent.

The lack of syllable position effect on P1 in participants who listened to Language A may be related to this implicit rhythm processing. With the rhythmic regularity formed by the stresses at the first syllables of words with the same length, listeners could easily predict when a stress was to occur even before training, and thus did not need to attend to the transition from a softer sound to a louder one. On the other hand, the syllable position affected the P1 amplitude for participants who listened to Language B which consisted of words of different lengths and thus presented an irregular and unpredictable rhythmic pattern. In fact, the medial syllables in Language B elicited larger amplitude in P1 than the initial syllables did. It is possible that sensory processing, which was reflected by the P1 component, may have been affected by the predictability of the stresses. For both languages, syllable position did not interact with training, which may suggest that any effect of learning on the processing of continuous speech was not specific to the initial portions of words, but could have been an overall shift in how the streams were processed or even just how tired the participants were at the beginning compared to the end of the experiment.

As predicted, the amplitude of N1 did not change much after participants who were exposed to Language A went through the training session and learned to recognize the word, supporting the hypothesis that stress and rhythm are powerful cues to direct attention to word onsets even without lexical knowledge.

Training did not affect the N1 component in any regions in participants who listened to Language B, contradicting the prediction that there would be a training effect on N1 for a language with inconsistent word lengths, as mere stress might not be enough for listeners to segment continuous speech stream into words and then direct attention to word onsets. This result suggests that even without lexical knowledge and rhythm, listeners were able to use stress alone in deploying attention across time. Moreover, the N1 amplitude did not differ as a result of different syllable positions (initial or second) or different lengths of the preceding words (two, three, or four).

Training did not significantly change the ERP amplitude 200-495 ms after stimulus onsets in listeners of either Language A or B, indicating that the transition from a soft syllable (final syllable of the preceding word) to a louder one (initial syllable) that may affect processing did not change because of the training. Baseline correction in the preliminary data processing set the average amplitude of ERPs 100 ms before stimulus onset as 0μV and the fluctuations in amplitude immediately after stimulus onset reflect the comparisons between the potential at a moment and the baseline amplitude. Therefore, for the syllable position effects, the ERPs to the initial syllables were time-locked to the signal during presentation of final syllables, but the ERPs to medial syllables are related to the signal during presentation of initial syllables. The fact that no main effect of syllable position on the early auditory evoked potentials was observed when comparing initial to medial syllables means that the baseline correction effectively cancels out the early effects (which are higher-frequency components such as P1, N1, and P2). However, further away from the baseline, a slightly larger effect of syllable position was present on a lower-frequency component. As the ERP amplitude in this time window was less negative in Language B but not A, it is possible that training altered the way in which listeners process the transition from a soft sound to a louder one, as they might have been better at predicting when one word ends and the next begins after they knew the words better.

Conclusion

The results suggest that whenever stress or both stress and rhythm were available in speech, listeners could use these cues to allocate temporally selective attention without knowing the words. Compared with previous research on temporally selective attention in artificial language processing, results from the current study showed that adding salient prosodic cues into the artificial speech eliminated the training effect seen in the study using an artificial language without explicit stress or rhythm cues (Sanders, Newport, and Neville, 2002). In the study by Sanders et al. (2002), all the words in the artificial language they used were 3-syllable words, which made the stimulus somewhat rhythmically regular. However, listeners were not able to recognize the rhythmic pattern until they learned the words in the language.

Listeners also tend to direct attention to word initials in natural speech. Sanders & Neville (2000) used natural English speech to investigate the role of cues such as stress pattern in speech segmentation. They found that listeners were better and faster in detecting target phonemes when they occurred at word initials than when they were in the middle of words. When the target phonemes were not stressed, they were more easily detected when they were at the word initials. This result could be explained by that listeners allocate more attention to word initials than medial syllables. In another ERP study using natural English speech, Astheimer & Sanders (2009) again observed larger N1 amplitude at word onsets when the probes coincided with the initial syllable than when the probes occurred before the initial syllables, suggesting greater attention allocated to the initial portion of words.

Future Research

The current study answered the question what could cue temporally selective attention. Specifically, stress and rhythm, two important components of natural speech, are shown to be able to help listeners deploy attention to portions of speech. When the nearly isochronous rhythmic pattern was absent in Language B, stress was still a powerful cue to direct attention even without lexical knowledge. Given the convergent finding that listeners pay more attention to word initials than to medial portions, the results of the current study may have implications for natural speech processing. Future research can address this question on the distinction between the role of acoustic probing and predictability in natural speech using continuous streams with words of same or different length and manipulating stress positions within the words.

The results from the current study strongly support the hypothesis that stresses cue temporally selective attention. However, the effects of rhythm alone on attention remain ambiguous. An experiment that involves same artificial languages as used in the current study without stressing word onsets may shed light on the role that rhythm plays. Without the stress, listeners will be likely to present a training effect on the ERPs to the initial syllables no matter whether the words are of consistent or inconsistent length. However, if rhythm plays an important role in directing temporally selective attention to word onsets, then the training effect would be greater in the language with irregular rhythm than in the other one with regular rhythm, as listeners who listen to the rhythmically irregular language would rely more on the lexical knowledge to be able to recognize the words in the continuous stream than those who listen to the language with words of same length. Future research can also replace stresses with an irrelevant acoustic probe such as a click or beep like the one used in previous research (Astheimer & Sanders, 2009). In this way, we might be able to differentiate the roles of stress and the rhythm formed by the stressed syllables.

The ERP technique can map the experiment effects only on the scalp, but brain-imaging techniques such as functional magnetic resonance imaging (fMRI) will be helpful to localize the responses in the brain. However, given the relatively low temporal resolution of the fMRI technique, it may not be feasible to examine attention with narrow time windows that are often used in ERP studies. Nonetheless, it is possible to use brain imaging to study attention in speech processing. For example, in situations where the language becomes too difficult to process in detail or even maintain attention to the stimuli, such as when the speech stream is very fast, if rhythm does cue temporally selective attention, we may observe more activities over the entire cortex that can reflect attention when we add in isochronous rhythm that can help build a structural hierarchy and assist auditory signal processing by providing a scheme, whereas a language without an explicit rhythm pattern may be too difficult for listeners to pay attention to any aspect of the language.

Implications

The present study provides implications for how prosodic cues such as rhythm and stress can help with speech perception and learning. In natural speech, other cues including the semantic and syntactic context can also affect signal processing (e.g. Sanders & Neville, 2000). However, these cues often depend on the environment and can be difficult to control for the speakers to make themselves understood. On the other hand, speakers can vary the acoustic properties such as volume and timbre of their voices as well as rhythmic variability in their speech. Based on the present study, we may predict that, just as stress and rhythm can cue attention in speech perception, they may also be helpful in language learning, especially learning through listening to the language. Therefore, the audio material used for language learning, at least in English, should include a great proportion of rhythmic listening material such as poetry and songs.

Meanwhile, instructors can improve learning by exaggerating the parts they want to emphasize, which may be useful not only in foreign language instruction, but also in instructions of other subjects. Given the fact that the behavioral results of the present study did not strongly support statistical learning as an essential part of language learning, and that this result could have been due to the relatively shorter exposure time to the novel language, duration of exposure to the novel language may affect progression in learning. Therefore, language learners should be encouraged to listen to the language that they are learning as long as possible.

In addition, there are constraints only with which statistical learning can play an important role in language acquisition (Saffran, 2002). For example, grouping speech signals into hierarchical chunks has been found to facilitate speech processing as well as language acquisition (Saffran, 2002). Therefore, the rhythmic pattern in speech may assist listeners in learning the language. This theory, if true, can also be used to explain the lack of training effect in the experimental groups in the current study, especially in those participants who listened to Language A in the current study. In other words, it is possible that the rhythm pattern helped participants to distinguish between words and non-words, but the effect was not great enough to improve performances in the behavioral tests.

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.

Aslin, R.N., Saffran, J.R., Newport, E.L., (1999). Statistical learning in linguistic and nonlinguistic domains. In: MacWhinney, B. (Ed.), *The Emergence of Language*. Lawrence Erlbaum Associates Publishers, Mahwah, pp. 359–380.

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

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock " of our internal clock: direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, *14*(4), 362-366.

Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193-210.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception*1, *2. *Journal of Memory and Language*, *31*(2), 218-236. doi: 10.1016/0749-596X(92)90012-M.

Cutler, A., & Clifton, C. J. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhn, *Attention and performance* (pp. 183-196). Erlbaum: Hillsdale.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and speech*, (20), 1- 10.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, 21, 103-108.

Dimoska, A., McDonald, S., Pell, M., Tate, R., & James, C. (2010). Recognizing vocal expressions of emotion in patients with social skills deficits following traumatic brain injury. *Journal of the International Neuropsychological Society*, *16*(2), 369-382. doi:10.1017/S1355617709991445.

Driver, J., & Frackowiak, S. (2001). Neurobiological measures of human selective attention. *Neuropsychologia, 39*, 1257-1262.

Eimer, M. (1993). Spatial cueing, sensory gating and selective response preparation: An ERP study on visuo-spatial orienting. *Electroencephalography and Clinical Neurophysiology, 88*, 408-420.

Ferree, T., Luu, P., Russell, G., & Tucker, D. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology, 112*, 536-544.

Frauenfelder, U. H. (1985). Cross-linguistic approaches to lexical segmentation. *Linguistics, 23*, 669-687

Hansen, J., Dickstein, P., Berka, C., & Hillyard, S. (1983). Event-related potentials during selective attention to speech sounds. *Biological Psychology, 16*, 211-224.

Heinze, H., & Mangun, G. (1995). Electrophysiological signs of sustained and transient attention to spatial locations. *Neuropsychologia, 33*, 889-908.

Hillyard, S., Woldorff, M., Mangun, G., & Hansen, J. (1987). Mechanisms of early selective attention in auditory and visual modalities. In R. J. Ellingson, N. M. F. Murray & A. M. Halliday (Eds.), *The London Symposia (EEG Supplement 39)* (pp. 317-324): Elsevier Science.

Hink, R., Van Voorhis, S., & Hillyard, S. (1977). The division of attention and the human auditory evoked potential. *Neuropsychologia, 15*, 597-605.

Jones, M., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459-491.

Jones, M., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*, 313-319.

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.

Large, E., & Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119-159.

Luck, S. (2005). *An introduction to the Event-related Potential Technique*. Cambridge, MA: The MIT Press.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Miller, J.L., Grosjean, F., Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica 41*, 215–225.

Nakatani, L., & Schaffer, J. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234-245.

Neville, H., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task. III. Separate effects of auditory deprivation and acquisition of a visual language. *Brain Research, 405*, 284-294.

Picton, T., & Hillyard, S. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191-199.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R. J., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*, 127-152.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101-106.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172-196. doi: 10.1006/jmla.2001.2839.

Salasoo, A., & Pisoni, D. (1985). Interactions of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*, 210-231.

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

Sanders, L., Ameral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia, 47*, 1183-1186.

Sanders, L. D., Neville, H. J., & Woldorf, M. G. (2002). Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues. *Journal of speech, language, and hearing research*, *45*, 519-530.

Sanders, L. D., & Neville, H. J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research*, *43*, 1301-1321.

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

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

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature neuroscience*, *5*(7), 700-3. doi: 10.1038/nn873.

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological research*, *75*, 95-106.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 98-108.

Wright, B., & Fitzgerald, M. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics, 66*, 508-516.

Appendix A

Experiment Running Protocol

NABASTRESS

FALL 2010

SESSION NO.
SID:_______
Participant Name:______________________ Date:_____________ Time:________

Inclusion Criteria (ask on scheduling, confirm at ERP sessions)

- ☐ Birthdate __________________
- ☐ Age ______________________
- ☐ Sex______________________
- ☐ Handedness ________________
- ☐ Native Language ____________
- ☐ Net friendly________________
- ☐ No neurological issues _______
- ☐ No psychoactive meds _______

- ☐ Ethnicity: Hispanic or Latino___
- ☐ Racial Category
- ☐ American Indian / Alaska Native
- ☐ Asian
- ☐ Native Hawaiian / Pacific Islander
- ☐ Black / African American
- ☐ White
- ☐ Multiracial / Undisclosed

SET UP

☐ Prepare net solution
☐ Restart Computer.
☐ Reset Volume: Hit default first. Then adjust to **77%**
☐ Stimulus monitor on. G5 mode
☐ Clean towels and pipettes
☐ Speakers placed on the 2 lines on the top of the monitor. Volume: 4th dot from highest.
☐ Lighting: Normal
☐ Consent and information forms ☐ Check and receipt forms ready.
☐ **Restart Kings Canyon** ☐ **Restart Sequoia** ☐ Open impedance window
☐ Virus protection disabled ☐ Open NetStation ☐ Move to booth monitor
☐ Open E-Prime Run ☐ Open recording session: ☐ Monitor selector on Mac
☐ Open Experiment ☐ Allow zeros and gains to run
Set up person:____

PARTICIPANT INTERACTION

☐ Confirm inclusion criteria ☐ Information forms
☐ Procedure explanation
☐ Consent form
☐ Head circumference:___________
☐ Net size:___________

☐ Arm connected to amplifier
☐ Net connected to arm
☐ Arm connected to amplifier
☐ Net connected to arm
☐ Place net in solution.
☐ Procedure Explanation to Participant

IN THE COMPUTER ROOM

☐ Open NetStation
☐ Enter data with correct no. e,g S01, S02 for subject 1,2 etc.

☐ Open file Under NabaMaster folder.
☐ **Switch to PC**

BEHAVIOR PHASE I

<u>File</u> Naba*stress*BehavioralTest.ebs : Session 1, Test 1

Experiment instructions

In this experiment, you will be asked to listen to a series of sequences of sounds while we measure your brainwaves.

This experiment is broken up into five parts. First you are going to take a practice test. The purpose of this is to help you get used to the task and allow you ask any questions you might have. During the practice test you will hear nonsense words. Decide how much you like each word on a scale of one to four, with one meaning you don't like it at all and four meaning you like it a lot.

☐ End of behavioral test. **Switch to G5.**
☐ Net application
☐ Check Impedances (move to booth monitor)
  - Key electrodes < 50
  - All electrodes < 100.

☐ Open Dense waveform display.
☐Show brainwaves (blinks, eyemovements, tension, alpha, VERY IMPORTANT)
☐ Mention the fixation point to participant.
☐ Check volume once again on the PC. Make sure it on **77%**

☐ Open ERP file.
☐ **Switch to PC.**

ERP PHASE I

<u>File</u> Naba*stress*-ERP-BeforeTraining.ebs2: Session 1, Test 1

☐ Hit the SPACE bar after first instruction shows on screen.
☐ Hit SPACE bar again after 2nd screen to get to fixation screen.

Directions: The second phase of this experiment involves recording ERP data while you listen to a continuous stream of sound. Please pay attention to the things we discussed earlier, such as muscle tension and excessive blinking, that can interfere with data collection. You can relax in the purple chair while listening to this stream of sound but please try to stay alert because dozing and anything close to it also interferes with data collection. We will be in periodically to check the electrodes and make sure you are comfortable.

Application person:

☐ Mention that there will be a break between 2 streams of sound. This will be very short and only necessary if after checking up on subject over the intercom, he/she asks for the break. (max 5mins). Otherwise, go on to the next stream right away.
☐ Press SPACE bar to resume the second stream, unless there is need for a break.
☐ End of session. **Switch to G5**

Break:

RUN TIME NOTES:

BEHAVIORAL TEST #2
File Naba*stress*BehavioralTest.ebs: Session 2, Test 1
☐ **Switch to PC**
Directions: Now I will ask you to take another practice test similar to the one you took at the beginning of the experiment. This time, indicate how familiar each word is from the long stream you just listened to.
☐ End of session. **Switch to G5**

TRAINING
File Naba*sress*BehavioralTraining.ppt
☐ **Switch to PC**
☐ Have subject sit in red chair, placed in front of purple
Directions: Now, you will have the opportunity to memorize six different sound sequences. On the Powerpoint slide you will see six speaker buttons that will each play a special sequence; feel free to roll over these buttons as much as you like in your effort to learn their characteristics. Once you feel as though you know all six sound clips, let me know and we can begin the test.

Begin Train Time:______
End Train Time:_______

☐ End of training. **Switch to G5**
☐ esc out of .ppt page

BEHAVIORAL TEST #3
File Naba*stress*BehavioralTest.ebs: Session 3, Test 1 (if repeated, increment test #)
☐ **Switch to PC**

Directions: You will now hear sounds similar to the one you encountered during the practice tests. This testing process is the same as in the practice test, but now respond with your confidence that a word is one you learned. For example, if you are certain you learned a word, press 4.

Not only can it be difficult to memorize six sound sequences in the first place, it can be even trickier to tell the learned clips apart from the others. It is no problem if you do not reliably recognize the sequences on your initial attempt--you will have up to four more tries.

D' Value _____

☐ End of session. **Switch to G5**
☐ Have subject move back to purple chair.
☐ Check Impedances before running the next stream.
☐ Check volume at **77%** again, before running ERP.

ERP PHASE 2:
File Naba*stress*-ERP-AfterTraining.ebs2: Session 1, Test 1
☐ Switch to PC
Directions: The next phase of the experiment involves recording ERP data again. You will listen to a stream similar to the first, and the directions are the same. Please try to stay alert and focused on the stream.

Application person:
Break (if applicable):
RUN TIME NOTES:

☐ End of session. **Switch to G5**

FINAL BEHAVIORAL TEST:
File Naba*stress*BehavioralTest.ebs: Session 4, Test 1
☐ Switch to PC.
Directions: For this stage of the experiment we ask that you take one more behavioral test in order to assess how well you retained your knowledge of the six sequences.

☐ Switch to G5

AFTER EXPERIMENT
Participant procedure:
- ☐ Remove net, soak in warm water
- ☐ Pay and receipt ☐ Future experiments?
  Name, Signature, Address, SS #, Date, Amount
- ☐ Debriefing form.

Data checking procedure

NetStation file okay?

- ☐ correct name
- ☐ correct place
- ☐ looks okay

E-Prime file okay?

- ☐ correct name (.txt, .edat)
- ☐ correct place
- ☐ looks okay

- Save .edat file for each person.
- Quick analysis of data

Lab cleaning

- ☐ buckets emptied and rinsed
- ☐ towels in hamper
- ☐ enough towels for next run
- ☐ enough cash for next run
- ☐ files in right folders
- ☐ Go on SONA to award credit if applicable.

Appendix B

Consent form

Speech Segmentation: Naba*

NAME_________________________ DATE___________

You are invited to participate in a research study conducted by Dr. Sanders in the Psychology Department. You will be asked to complete a task on a computer while we measure your brain activity. We ask that you read this form and ask any questions you have before signing it. By signing this consent form, you indicate that you willingly agree to participate in this project.

DESCRIPTION

The purpose of this study is to learn more about the relationship between human brain activity and behavior. To investigate this, we will record brain waves while you complete a task.

PROCEDURES

You are being asked to participate in a single 2 hour session as a volunteer. If you agree to be in this study, we will ask that you do the following things:

1) We will ask you a few questions about your medical history. These questions will only be used to verify that you meet the criteria to participate in this study.
2) We will measure your brain activity while you complete a computer task. To record your brain activity, we use recording electrodes. These electrodes are small sensors that measure the normal, ongoing electrical activity in the brain. The electrodes are part of a net that will be placed on your head and removed after the experiment. To help the electrodes measure your brain activity, they will be wet with a saline (salty) solution.
3) You will sit in a comfortable chair in a dimly lit room and pay attention to different types of auditory stimuli such as beeps and words. You may also pay attention to visual stimuli such as pictures and words shown on the computer monitor.
4) At the end of this experiment, we may ask you if you are interested in participating in other experiments in this lab in the future. If you indicate that you are interested, we may contact you again. Signing this consent form, participating in the current study, and agreeing to be contacted about future studies does not obligate you in any way.

CONFIDENTIALITY

The data from this study and information you provide will be kept private.

1) All information will be identified by a number that will not be tied to your name.
2) This consent form, which indicates you did participate in at least some parts of this study, will be stored in a secure location.
3) In any report or presentation of the data, we will not include information that will make it possible to identify you.
4) All records of your participation will be destroyed after five years.

Initial______ Page 1 of 2

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 07/01/2010 Protocol #: 2009-0338
Valid Through: 07/19/2011
IRB Signature: Nancy C. Swett

POSSIBLE RISKS AND BENEFITS OF BEING IN THE STUDY

Risks: There are minimal risks associated with participating in this study. The brain waves are measured with standard equipment that is designed to make this procedure harmless.

1) Under rare circumstances (less than 1 out of 200 participants), people with very sensitive skin may have a reaction to the saline solution. A small red mark may be apparent at one or more electrode locations for a short period of time (less than three days). Some of the solution may remain in your hair until washed.
2) The computer task you will be asked to complete may be very difficult; incorrect responses reflect how difficult the task is rather than your own abilities.

Benefits: Our work is designed to answer some very basic questions about the relation between the brain and behavior. What we are doing should in no way be perceived as therapy, nor will it benefit you directly. The results of these brain wave tests may become useful for helping to treat certain types of patients who have neurological problems and for developing recommendations for education and rehabilitation.
Compensation: You will be compensated for your time with a personal check ($10 per hour, prorated for partial hours) OR experimental credit (1 credit per half hour, rounded to the nearest half hour). Your participation in the experiment is voluntary and you can withdraw at any time without penalty. You will still get the credits or payment for the time already spent in the study. If you are earning experimental credits through your participation, please understand that this is not the only way to do so. You may contact your instructor who will offer you an appropriate alternative activity.

CONTACTS AND QUESTIONS

Your participation is entirely voluntary. You many refuse to participate or withdraw at any time. If you decide to participate, you are free to withdraw your consent and discontinue participation at any time without penalty.

1) If you want more information about this research or have any research-related problems, please contact Dr. Lisa Sanders in the Psychology Department (Tobin 429) by phone, (413) 545-5962, or by email, lsanders@psych.umass.edu.
2) If you have any questions regarding your rights as a research participant, contact the Human Subjects Review Board (545-3428) or send an email to HumanSubjects@ora.umass.edu. You may also contact the Chair of the Psychology Department, Dr. Melinda Novak (Tobin 439) by phone, (413)-545-5958, or by email (mnovak@psych.umass.edu).

STATEMENT OF CONSENT

I acknowledge that my participation in this research project is voluntary. I have read the above information and the study has been explained to me to my satisfaction. I understand that I am free to withdraw from the study or refuse to participate at any time.

Print your name here ____

Signature _________________________________ Date ___________ Page 2 of 2

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 07/01/2010 **Protocol #:** 2009-0338
Valid Through: 07/19/2011
IRB Signature: Nancy C. Swett