Prosodic Transfer and Phonological Learning in a Second Language Fluent Speech Segmentation Task

Tuuli Morrill Adams

Department of Linguistics, New York University, United States of America

tuuli.adams@nyu.edu

Abstract

Listeners use prosodic cues to facilitate lexical access when listening to fluent speech in their native language [1], [2]. This study investigates second language learners' ability to segment words from continuous speech, and the effect of native language prosodic structure on the perception of word boundaries in the second language. In an experiment conducted with natural language stimuli, English speakers learned words and then listened to fluent speech in a previously unfamiliar language (Finnish). After listening to fluent speech, they chose between pairs of correctly segmented real words and incorrectly segmented non-words, to identify possible words of Finnish. Results show that English speakers do exhibit a bias towards identifying words with first-syllable stress as real, likely an effect of a native language segmentation strategy [2]. However, the test group that learned words and then listened to fluent speech performed better than the group that did not listen to fluent speech. This suggests that learning words and hearing them in context aids in second language speech segmentation. More successful speech segmentation, in turn, promotes the learning of other phonological patterns, which makes learners more accurate at identifying possible words in the second language.

1. Introduction

In order to segment continuous speech, listeners use a combination of information about rhythmic structure, phrasal structure, and phonotactics to segment their native language [3], [4]. Infants have the ability to distinguish between native and non-native prosodic patterns, and rhythmic structure seems to play an important role in the development of their linguistic knowledge [5]. Evidence suggests that adult speakers have developed a metrically-based native language segmentation strategy, which is then the only strategy available to apply to a second language as well [6]. Using a native language segmentation strategy for another language can affect the ability to correctly segment words from fluent speech in a non-native language.

However, certain aspects of prosodic structure facilitate speech segmentation, even in cross-linguistic perception tasks. Phrasal structure, specifically the pauses found between intonational phrases, are identifiable even to non-native listeners as possible word boundary indicators [7]. Complications for predictions about cross-language segmentation behavior emerge when defining the similarities or differences between the prosodic structures of languages. It is possible that the rhythmic structure of language is determined along various dimensions of the phonetic implementation, such as the relationship between the lengths of different phonemes, as well as other aspects of their realization (quality, articulation) [8], [9]. Predictions for segmentation behavior based solely on the grouping of languages into metrical classes may miss some subtleties that emerge from a closer look at the phonetic information non-native listeners perceive.

In addition, research suggests that adults are able to use statistical tracking for transitional probabilities between syllables and apply this information to identify cohesive word-like units [10]. This indicates that some statistical learning mechanism, also hypothesized to play a large role in infants' first language acquisition [11], is available to adults learning a second language. However, many studies of statistical learning have, out of necessity, used very simplified artificial languages; it is not clear how exactly these abilities are used in the online processing of the complex natural speech signal.

The current study investigates adult second language learners' ability to segment words from fluent, natural language, and complements previous studies conducted on simple artificial languages or short natural stimuli. In an experiment, native English speakers learned words of an "unknown language," Finnish, and then listened to fluent Finnish speech. Afterwards, they were asked to identify possible words of Finnish, choosing between correctly and incorrectly segmented words that occurred in the fluent speech passage. If participants were segmenting words from the fluent speech while listening, they should be more likely to choose the correctly segmented word when identifying possible words of Finnish.

This paradigm represents a case of non-native segmentation of a language that is metrically similar to the native language; Finnish stress is always placed on the first syllable of the word, and the most frequent stress pattern in English is also word-initial [12]. If English listeners automatically transfer their native language segmentation strategy to the second language, they may assume that a stressed syllable signals the beginning of a word. However, other phonological and phonetic differences between the languages may not allow for a direct transfer of segmentation cues. For example, contrastive vowel length in Finnish allows unstressed syllables to contain long vowels, whereas in English, vowel length is a phonetic correlate of stress. By exposing participants to fluent speech in the second language, this experiment addresses the question of whether or not a similar native language metrical structure can benefit a non-native listener and enhance phonological learning.

The experiment also tests the effect of the learning environment on the adaptation of a native language segmentation strategy to a second language. If learning words of a language, followed by listening to fluent speech, facilitates speech segmentation and phonological learning, participants in the condition in which they complete both tasks should demonstrate some ability to identify words of Finnish. In a second experimental condition, participants learned words of Finnish, but did not listen to fluent speech. If the participants can learn some-

thing about non-native phonology by learning isolated words, and do not benefit additionally from listening to fluent speech, this group should perform as well at identifying Finnish words as the group that listened to fluent speech after learning words. A third group listened to fluent speech with no prior exposure to isolated words. If these English listeners automatically apply a metrically based segmentation strategy to the second language, they may be able to correctly identify stressed syllables as word-initial in the fluent speech; this could facilitate learning about possible phonological words in the language.

2. Experiment

An experiment was conducted using a between-subjects design, with three conditions and three groups of participants. The experiment consisted of three tasks: a Word-Learning task, a Fluent Speech Listening task, and a Possible Word Identification task, described in more detail in 2.1. Participants in the three conditions participated in different task combinations; the materials remained the same across groups.

2.1. Procedure

2.1.1. Word Learning Task

Participants learned 28 words of Finnish, known as the "foreign language," with a computer program written in E-Prime. Participants heard each word pronounced twice on presentation of a picture. Each picture was presented three times, twice with a female speaker's voice and once with a male voice, in random order. After the presentation of the words, participants had to reach criterion (82%) on a word recognition task, in which they heard a word and had to click on one of two pictures to correctly identify the word.

2.1.2. Fluent Speech Listening Task

In this task, participants listened to a Finnish narrative, divided into twelve section of approximately one minute each. To help participants stay motivated to pay attention, two words were played after each section, and participants were asked whether those words had occurred in the previous section. The narrative also contained the words from the Word Learning Task, and participants were told they would hear the words they had learned in the narrative. Other stimulus properties controlled for in the narrative are described below in 2.3.

2.1.3. Word Identification Task

Participants were presented with fifty-six pairs of real words and non-words. The real words had occurred in the Fluent Speech Listening narrative, and the non-words consisted of syllable sequences that were adjacent in the narrative, but spanned a word boundary. The participants were asked to identify which of the "words" was more likely to be a word of the language they were learning. The inter-stimulus interval was 1500 ms, and participants responded by pressing "1" or "2" on a button box.

2.2. Conditions

Participants were placed in one of three conditions. Group 1, the Full Training group, participated in all three tasks decribed in 2.1. The order of the tasks was as follows: the Word Learning task was followed immediately by the Fluent Speech Listening task, which was followed by the Word Identification task. Group 2, the Word Learning Only group, participated only in the Word Learning task, followed by the Word Identification task, with no exposure to fluent speech. Group 3, the Fluent Speech Only group, completed the Fluent Speech Listening task without having done the Word Learning task, and then completed the Word Identification task.

2.3. Materials

The stimuli were produced by native speakers of Finnish currently residing in New York City. They were recorded in a sound-proof booth with a Marantz PMD-680 Digital Audio Recorder.

2.3.1. Stimuli for the Word Learning Task

There were twenty-eight Finnish words in the Word Learning Task: 4 monosyllabic words, 19 bi-syllabic words, and 5 tri-syllabic words. Some of the words contained long vowels or geminate consonants, as is common in Finnish words.

The stimuli for the Word Learning Task were produced by two female speakers and one male speaker of Finnish. The speakers were instructed to read the words in isolation, as if telling somebody the name of an object. Each word was produced at least four times by each speaker, and the tokens were chosen for use in the experiment based on consistency in speech rate and intonation patterns. Minor manipulations of the pitch track were performed in Praat [13] on some tokens, in order to maintain consistent pitch patterns across words.

2.3.2. Stimuli for the Fluent Speech Listening Task

The narrative in the Fluent Speech Listening task consisted of a series of coherent, grammatical sentences which were read naturally as a "story." The sentences contained the 28 learned words, in addition to 28 new words which were matched in certain properties. Half of the learned words and half of the new words occurred as "high frequency" words, 10–12 times each throughout the narrative. Half of the learned and new words were "low frequency" words, occurring only three times in the narrative. The new words closely matched the learned words also in terms of syllabic structure; there were four monosyllabic words, twenty bi-syllabic words, and four tri-syllabic words.

Whenever a learned or new word occurred in the narrative, it was always followed by the same syllable, as in (1), where the word "melaa" is always followed by the syllable /li/. In this way, the transitional probabilities between the syllables within the words, and the syllables spanning the end of the word and the beginning of the following word were equal.

(1) Kun joka päivä **MELAA, LI**hakset pysyvät hyvässä kunnossa.

Half of the learned words and half of the new words occurred at phrase or sentence boundaries, and half occurred only phrase-medially. Most sentences in the narrative contained one or more learned or unlearned words.

The narrative was read by a female native speaker of Finnish, who had not produced stimuli for the Word Learning task. The speaker was instructed to read the narrative at a comfortable pace, and in as natural a style as possible. Each section of the narrative was recorded at least twice, and dis-fluencies and errors were eliminated from the final version of the narrative by splicing together the recordings. Splices occurred at phrase boundaries or pauses, so that there were no disruptions in intonation patterns. Two versions of the narrative were counterbalanced across participants, so that half of the participants heard the second part of the narrative first.

2.3.3. Stimuli for the Word Identification Task

Fifty-six pairs of words and non-words were constructed for the Word Identification task. The pairs each contained one correctly segmented learned or new word, paired with one incorrectly segmented non-word. Non-words were made up of the last syllable(s) of a learned or new word, joined with the first syllable(s) of the word that followed it in the narrative. From the example in (1), the stimulus pair would consist of the real word "melaa" and the incorrectly segmented "laali."

Table 1: *Sample stimuli from the Word Identification Task*

Pair with initial stress

Learned/New Word	Incorrectly Segmented*
UUni	NI-jo
MElaa	LAA-li
KUUma	MA-ilma

Pair with non-initial stress

Learned/New Word	Incorrectly Segmented
UUni	ni-JO
MElaa	laa-LI
KUUma	ma-ILma

*dashes represent the actual word boundaries and capital letters represent the syllable stressed in the stimulus

Because Finnish stress is always word-initial, the incorrectly segmented words were always presented with two stress patterns. The first was the stress pattern with which that syllable sequence appeared in the actual narrative; in this case, the stress would occur on a non-initial syllable in the non-word, since the first syllable in the non-word would be an unstressed final syllable of a real word (see Table 1, "Pair with non-initial stress"). The second pattern placed the stress on the initial syllable of the non-word, so that both the real word and the non-word in the pair would have initial stress. The non-words with initial and non-initial stress were counterbalanced, so that participants heard both types, but never the same non-word with two stress patterns. Example stimuli appear in Table 1.

The Word Identification stimuli were produced by the same female speaker who read the narrative for the Fluent Speech Listening task. The speaker was a native speaker of Finnish, a fluent second language speaker of English and Dutch, and an actress with training in voice quality manipulation. She was able to produce the non-words with stress placement on varying syllables without difficulty. The speaker produced all of the real and non-words in isolation, and produced the non-words with both initial stress and non-initial stress. Minor changes to the pitch contours were performed using Praat [13] in order to make the pitch ranges uniform across the stimuli.

2.4. Participants

Sixty adults from the New York University and New York City communities participated in the experiment, with twenty participants in each condition. All participants were native speakers of American English with no speech or hearing impediments.

3. Results

3.1. Overall Accuracy

Accuracy on the Word Identification task was measured for all participants. There was a main effect of condition, $F(1, 5) = 7.81$, $p < .05$. The average accuracy for new words (words that participants had *not* been exposed to in the Word Learning Task) in the Word Identification Task is significantly higher for the Full Training group (Group 1), 70.2%, than for the Word Learning Only group (Group 2), at 58.0%($p < .005$), and the Fluent Speech Listening Only group (Group 3), at 53.4%, ($p < .001$). The Full Training group was also better at identifying the learned words than the Word Learning Only group, at 89.5% vs. 81.3%. Both groups that completed the Word Learning Task identified learned words better than new words ($p < .001$).

3.2. Effect of Stress

All three groups were better at Word Identification for new words when the incorrectly segmented choice exhibited a non-initial stress pattern ($p < .05$). When the non-word had non-initial stress, the Full Training group was at 81.8% accuracy, the Word Learning Only group was at 67.5% accuracy, and the Fluent Speech Only group was at 58.9% accuracy in selecting the correctly stressed, real-word choice. However, only the Full Training group performed above chance when the incorrect word choice had the initial stress pattern ($p < .01$). In this case, the Full Training group was at 58.6% accuracy, the Word Learning Only group was at 48.6%, and the Fluent Speech Only group was at 50.4% accuracy (see Figure 1).

Participants were also more likely to choose correctly when the incorrect choice exhibited a mismatch between vowel length and stress, i.e., when the first syllable of the non-word had a long vowel, but was not stressed. Since vowel length is a phonetic cue of stress in English, the discrepancy between the location of the length cue and the pitch accent cue may have made listeners less likely to select these as possible words. This pattern was significant for the Word Learning Only group

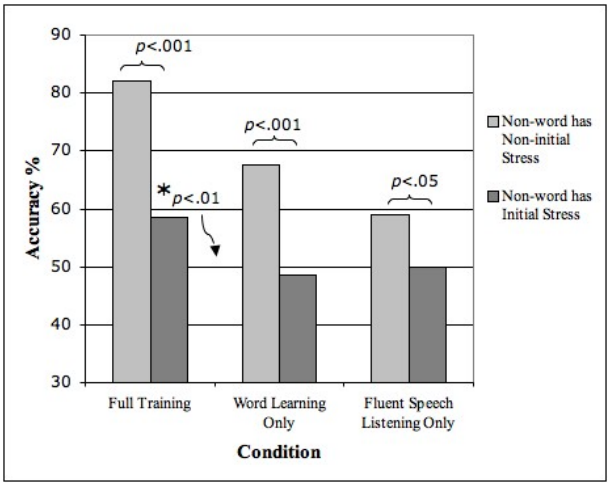

Figure 1: *Accuracy on Word Identification for new words only, indicating the stress pattern of the non-word in the pair.*

($p < .05$), approached significance for the Full Training group ($p = .064$), and patterned in the same direction, but was not significant for the Fluent Speech Only group ($p = .367$).

There were no effects of frequency of the word in the Fluent Speech Listening task (high vs. low frequency) and also no effects of phrasal position. This was true for both groups that were exposed to the fluent speech passage, the Full Training group and the Fluent Speech Only group.

4. Discussion

Native English speakers exhibit a bias towards identifying words with initial stress as real words. Since the bias exists for all of the groups, this result is likely due to prosodic transfer. As expected, the similarity in the metrical structures between Finnish and English allows native English listeners to benefit from transferring some aspect of their native language knowledge to phonological well-formedness judgments about words in a non-native language.

The Full Training group indicates that this result provides evidence for listeners making use of a segmentation strategy when listening to fluent speech. First, the result indicating that listening to fluent speech alone, without having learned isolated words first, does not provide as great a benefit in the Word Identification task, shows that it is the combination of learning words and then hearing fluent speech that facilitates learning. Word learning not followed by the Fluent Speech Listening task also fails to benefit the learners. Learning words, and then hearing them in the fluent speech context is facilitating word identification for the Full Training group. Crucially, they are better than the other groups at identifying the new words, which they had not learned in the Word Learning Task; this shows that they are learning something about the phonological structure of the non-native language that allows them to make generalizations about possible words.

There was no effect of word frequency, and no effect of word location with respect to phrase boundaries, for either of the groups that participated in the Fluent Speech Listening task. These groups were not more accurate at identifying words that had occurred at phrase boundaries than those that occurred phrase-medially, or words that had occurred twelve times in the narrative instead of only three. This is most likely due to the overall length of the narrative (over twelve minutes of continuous speech) in relation to the frequency of the words. These results indicate that listeners were not referring to explicit memories of words, even those that they may have perceived as cohesive word-like units when listening to the fluent speech. Therefore, the advantage the Full Training group had in word identification must stem from a more subtle combination of factors.

The phonological learning that takes place during the Fluent Speech listening task can be attributed to the adaptation of a native language segmentation strategy to the second language. Listeners who have learned a minimal amount of words in a second language (twenty-eight, in this case) are then able to hear these words in a fluent speech context. As a result, they can make more accurate predictions about word boundaries *following*, and possibly preceding, the learned word. Learning words in isolation may provide listeners with a strong indication that a metrically based segmentation strategy is transferable to the second language. When participants only heard fluent speech, they were not able to learn as much about the second language phonology; without exposure to definite word boundary cues, they may have been slower to attempt segmenting the second language or to apply their native segmentation strategy.

If listeners in the Full Training group were more accurately segmenting the second language, perceiving consecutive syllables as cohesive word-units, hearing fluent speech could allow them to track other cues and properties of phonological words. The Full Training group was the only one in which participants were above chance (at 58.6% accuracy) at identifying possible words when both stimulus choices exhibited initial stress. In these cases, participants could not rely on the stress pattern alone to judge phonological well-formedness, and must have been relying on information gained while listening to the fluent speech passage to aid in word identification.

5. Conclusion

The results of this study indicate that listening to fluent speech in a second language can be beneficial to adult learners. In cases where the prosodic structures of the native and non-native languages have certain similarities, it is possible that applying a native language segmentation strategy to the non-native language can yield some benefit. Even with a minimal amount of knowledge of a second language, listeners do attempt to segment continuous speech, and are able to learn phonological patterns in the process. It is worth noting that the environment of this experiment presented learners with a relatively small amount of exposure to fluent speech, and it is expected that listeners' knowledge of the second language would continue to increase with further exposure. This also indicates that second language segmentation abilities themselves could possibly improve and become more accurate with time.

6. Acknowledgements

I would like to thank Lisa Davidson, Maria Gouskova, Kevin Roon, Jason Shaw, and members of the NYU Phonetics and Experimental Phonology Laboratory for their comments on this paper.

7. References

[1] Norris, D., J. M. McQueen, A. Cutler, and S. Butterfield. 1997. "The possible-word constrain in the segmentation of continuous speech." Cognitive Psychology, 34: 191-243.

[2] Cutler, A. 1990. "Cognitive Models of Speech Processing," chapter in Exploiting Prosodic Probabilities in Speech Segmentation. MIT Press, Cambridge, Mass.

[3] Christophe, A., A. Gout, S. Peperkamp, and J. Morgan. 2003. "Discovering words in the continuous speech stream: the role of prosody." Journal of Phonetics, 31: 585-598.

[4] McQueen, J. 1998. "Segmentation of continuous speech using phonotactics." Journal of Memory and Language, 39(1): 21-46.

[5] Jusczyk, P., A. Cutler, and N. Redanz. 1993. "Preference for the predominant stress patterns of English words." Child Development, 64: 675-687.

[6] Cutler, A., J. Mehler, D. Norris, and J. Segui. 1986. "The syllable's differing role in the segmentation of French and English." Journal of Memory and Language, 25:385-400.

[7] Christophe, A., S. Peperkamp, C. Pallier, N. Block, and J. Mehler. 2004. "Phonological phrase boundaries constrain lexical access: I. adult data." Journal of Memory and Language, 51:523-547.

[8] White, L. and S. Mattys. 2007. "Calibrating rhythm: First and second language studies." Journal of Phonetics, 35:501-522.

[9] Ramus, F., M. Nespor, and J. Mehler. 1999. "Correlates of linguistic rhythm in the speech signal." Cognition, 73:265-292.

[10] Newport, E., and R. Aslin. 2004. "Learning at a distance 1. statistical learning of non-adjacent dependencies." Cognitive Psychology, 48(2):127-162.

[11] Saffran, J., E. Johnson, R. Aslin, and E. Newport. 1999. "Statistical learning of tone sequences by human infants." Cognition, 70(1):27-52.

[12] Cutler, A. and D. M. Carter. 1987. "The predominance of strong initial syllables in the English vocabulary." Computer Speech and Language, 2:133-142.

[13] Boersma, P. and D. Weenink. 2008. Praat: doing phonetics by computer (Version 5.0.04) [Computer program]. Retrieved from http://www.praat.org/.