How transitional probabilities and the edge effect contribute to listeners' phonological bootstrapping success

Juwairia Sohail[a,b] and Elizabeth K. Johnson [a]

[a] Department of Psychology, University of Toronto Mississauga, Toronto, Canada

[b] Ontario Institute for Studies in Education, University of Toronto, Toronto, Canada

Manuscript Word Count: 5059 words

Corresponding Author:

Elizabeth K. Johnson

Department of Psychology

University of Toronto Mississauga

3359 Mississauga Road

Phone: (905) 569-4785

elizabeth.johnson@utoronto.ca

draft version

Abstract

Much of what we know about the development of listeners' word segmentation strategies originates from the artificial language-learning literature. However, many artificial speech streams designed to study word segmentation lack a salient cue found in all natural languages: utterance boundaries. In this study, participants listened to a speech-stream containing one of three sets of word boundary cues: transitional probabilities between syllables (TP Condition), silences marking utterance boundaries (UB Condition), or a combination of both cues (TP+UB Condition). Recognition of the trained words and rule words (words not in language, but conforming to its phonotactic structure) was tested. Participants performed equally well in the TP+UB and UB Conditions, scoring above chance on both trained and rule words. Performance in the TP condition, however, was at chance. Our results suggest that attention to UBs is a particularly effective strategy for finding words in speech, possibly providing a language-general solution to the word segmentation problem.

Abstract Word count: 152

Keywords: word segmentation, speech perception, prosody, spoken language, Edge Hypothesis, statistical learning

draft version

1. Introduction

Adults are thought to solve the word segmentation problem by relying on language-specific probabilistic cues to sound structure (e.g., Tyler & Cutler, 2009; Cutler, 2012). However, it is not yet clear how language learners acquire the language-specific knowledge needed to identify word boundaries in fluent speech. Research has suggested that both infants and adults begin the task of word segmentation by tracking transitional probabilities (TPs) between syllables (e.g., Saffran, Aslin, & Newport, 1996a; Saffran, Newport, Aslin, 1996b). Using TPs, listeners could segment an initial cohort of words (e.g., Thiessen, & Erickson, 2013; Thiessen & Saffran, 2003). From this cohort, listeners could then learn important language-specific segmentation cues, such as the placement of lexical stress or phonemes with respect to word boundaries (e.g., Sahni, Seidenberg, & Saffran, 2010; Swingley, 2005; Thiessen & Saffran, 2007). This syllable-distribution tracking explanation for the onset of word segmentation abilities is attractive in its simplicity and purported universal feasibility. However, tracking TPs is not the only possible universal cue to word boundaries. Utterance level prosody has been identified as another potentially important language-general word boundary cue (e.g., Christophe, Guasti, Nespor, Dupoux, & Van Ooyen, 1997; Daland & Pierhumbert, 2011; Golinkoff & Alioto, 1995; Endress & Hauser, 2010; Johnson & Seidl, 2008; Seidl & Johnson, 2006) that may provide listeners with an even more powerful universal segmentation cue than TPs between syllables (e.g., Endress & Mehler, 2010; Johnson, 2008; Shukla, Nespor, & Mehler, 2007). Here, we examine the relative efficiency of using TPs versus utterance level prosody to learn the sound structure of an artificial language and segment words from fluent speech. Our results support the hypothesis that the silences associated with

draft version

utterance boundaries provide language learners with a strong language-general cue to word boundaries, and that transitional probabilities between syllables may be a relatively more difficult segmentation cue to utilize (see Endress & Mehler, 2010; Gambell & Yang, 2005; Johnson & Tyler, 2010; Yang, 2004, for similar arguments regarding TPs).

Over the past 15 years, much support has been provided for a TP-based segmentation solution. After only two minutes of exposure, 5.5- to 8-month-old infants can segment words from an artificial language speech stream containing no pauses between words by tracking TPs between syllables (e.g. Saffran, Aslin, & Newport, 1996a; Thiessen & Saffran, 2003; Johnson & Tyler, 2010). Moreover, English-learning infants detect TPs between syllables in a well-controlled variant of a natural language (e.g., Jusczyk, Houston, & Newsome, 1999; Pelucchi, Hay, & Saffran, 2009). And corpus analyses suggest that TPs do indeed provide useful information about word boundaries in infant-directed speech (Swingley, 2005). Thus, all in all, there is compelling evidence in support of the notion that listeners may rely on TP cues to locate word boundaries in an unfamiliar language. However, tracking TPs in natural language has been argued to be a resource-intensive endeavour, and infants may have more trouble doing so with natural everyday-language input than with artificially-designed input (e.g. Gambell & Yang, 2005; Johnson, 2012; Johnson & Tyler, 2010; Mersad & Nazzi, 2012; Yang, 2004). Thus, although it is clear that infants are incredibly good at picking up statistical information from their environment, and TPs between syllables could in theory provide an excellent solution to the word segmentation problem, it seems prudent to at least entertain other supplemental (or perhaps even alternative) solutions to the word segmentation problem.

draft version

All languages of the world are structured in a prosodic hierarchy, clearly demarcating the boundaries between many linguistically relevant units in the speech signal (Selkirk, 1984). And all utterance and phrase boundaries necessarily align with word boundaries. Thus, attention to major prosodic boundaries, such as utterance and phrase boundaries, could provide listeners with a means to discover language-specific segmentation cues. Major prosodic boundaries are universally salient to all listeners (e.g., for evidence with infants see Christophe, Mehler, & Sebastián-Gallés, 2001; for evidence with foreign language listeners see Pilon, 1981), and a high proportion of the words heard by both infants and adults are flanked by at least one utterance boundary (henceforth referred to as UB-flanked words; Aslin, Woodward, LaMendola, & Bever, 1996; Johnson, Lahey, Ernestus, & Cutler, 2013; Johnson, Seidl, & Tyler, 2014; Van der Weijer, 1998). If listeners attended to the sounds occurring at the beginnings and ends of utterances, they could notice that phrases and utterances rarely begin or end in particular sounds, and by extension work out that words in the language probably rarely begin or end in particular sounds. That is, major prosodic boundaries could provide a universal strategy for bootstrapping language-specific segmentation cues such as position-specific allophones and phonotactics from fluent speech. Although there is ample evidence that prosodic boundaries influence listeners' segmentation of fluent speech (e.g., Johnson, Seidl, & Tyler, 2014; Seidl & Johnson, 2006; 2008; Marchetto & Bonatti, 2013; Ordin & Nespor, 2013; Peña, Bonatti, Nespor, & Mehler, 2002; Shukla, Nespor, & Mehler, 2007), there is little work focused on understanding how efficiently listeners can make use of this information to extract language-specific segmentation cues from the speech stream. Unlike past work pitting prosodic cues to word boundaries against TP cues to word

draft version

boundaries (e.g., Shukla, Nespor, & Mehler, 2007), this study jointly examines listeners' use of TPs between syllables and UBs with the goal of determining how efficiently listeners use each of these sources of information to learn a language-specific cue to word boundaries. More specifically, we examine infants' ability to use UBs and TPs to learn a phonotactic rule defining word boundaries. Our results fit well with past findings suggesting that 1) the silences universally associated with UB boundaries can be used to learn about word boundaries, and 2) TPs between syllables might not be the most powerful universal segmentation cue available to listeners. Both of these findings fit predictions of the Edge Hypothesis (Johnson et al., 2014; Seidl & Johnson, 2006), and dovetail nicely with related findings reported in the literature (e.g., Daland & Pierhumbert, 2011; Endress & Mehler, 2010; Shukla et al., 2007).

2. Experiment

In the current study, we examine listeners' use of TPs and UBs to find word boundaries in an artificial language. Adult listeners were briefly exposed to an artificial language containing nine CVCV nonsense words, all of which conformed to a simple phonologically-motivated phonotactic rule involving relative sonority. In this case, by phonotactic rule, we refer to a constraint on the positioning of a segment within a word (see Chambers, Onishi, & Fisher, 2003; Onishi, Chambers, & Fisher, 2002, for evidence that listeners learn similar rules from isolated words). One language contained all sonorant-initial words whereas the other contained all obstruent-initial words.

Participants were assigned to hear one of three variants of the language: the TP Condition contained only TP cues to word boundaries (i.e., as in many artificial

draft version

languages used to study word segmentation in the past, the language contained no silences or other prosodic cues marking word boundaries), the UB Condition contained only UB (but not TP) cues to word boundaries, and the TP+UB Condition contained both TP and UB cues to word boundaries. Note that in natural languages, UBs are marked by a host of acoustic cues besides silences (e.g., see Seidl & Johnson, 2008, for evidence that although silences are a particularly strong markers of UBs, even infants are sensitive to other prosodic cues marking UBs). In the current study UBs were artificially created, marked only by silent pauses. Thus, one could argue that UBs in natural language provide an even better cue to word boundaries than the UBs in our artificial language because the UBs in natural languages are redundantly marked my a multitude of acoustic cues.

After the exposure phase, listeners were subsequently tested on their recognition of trained words (words that occurred in the language) and rule words (words that never occurred in the language, but nonetheless conformed to its phonotactic structure) versus partwords in a 2AFC task. The purpose of including rule word test trials was to determine if listeners were simply memorizing the syllables occurring at utterance edges, or if they were learning the rules governing which consonant types could occur in different word positions. During the test phase, listeners were played pairs of words, and asked to indicate which of the two words belong to the language they had been exposed to. On trained word trials, listeners heard a word from the language versus a partword (i.e., a word that not only did not occur in the language, but also did not conform to its phonotactic rules). On rule word trials, listeners heard a rule word (i.e., a word that did not occur during the familiarization, but nonetheless conformed to the phonotactic rules of the language) versus a partword (see Peña, Bonatti, Nespor, & Mehler, 2002 for a

draft version

related use of trained versus rule words in an artificial language). The words of each language were constructed in such a manner that what formed a word in one language formed a partword in the other language (see Appendix).

Our predictions were as follows. If listeners can learn the phonotactic rules defining word boundaries in the artificial language, then they should perform above chance in both trained and rule word trials (if listeners only perform above chance in trained trials, then this would suggest they merely memorized specific word strings rather than learning a phonotactic pattern). Moreover, if attention to UBs enables listeners to learn the phonotactic rule of the language and segment words from speech, performance should be best for both trial types in the TP+UB Condition, because listeners are provided with two sources of information about word boundaries: TPs between syllables and UBs. Performance in the TP Condition should be better than chance, but not as strong as performance in the TP+UB. Performance in the UB Condition was difficult to predict because this is the first study to examine listeners' ability to bootstrap word boundaries from speech based on UBs alone.

2.1 Participants

Twenty-four native English speakers with normal hearing were tested in each of the three conditions (72 participants total; Mean Age = 20 years; 51 females, 21 males).

Participants were required to have learned English prior to the age of six and be currently using English as their primary language for communication. An additional 6 participants were excluded from the study for failing to meet the specified language criteria (3), failing to follow directions (2), and experimenter error (1). All participants were given

draft version

course credit or paid $5 for their participation in the study.

2.2 Stimuli

Eighteen CV syllables containing unreduced vowels were recorded in isolation by a female speaker, and then edited in PRAAT to ensure that the syllables were relatively uniform in amplitude (M=67 dB; SD=3.3), duration (M=0.31 s; SD=0.01), and pitch (M=188 Hz; SD= 1.8). The syllables were then combined into two different sets of 9 disyllabic words each, henceforth referred to as 'languages'. The words in each language were then concatenated in three different ways, resulting in a total of six training speech streams. In the first type of concatenation, used in the TP training condition, words were strung together in random order with no pauses in between words, resulting in higher TPs between syllables belonging to the same word (within word TP = 1.0, since no syllables occurred in more than one word) than between syllables belonging to different words (between word TP = .11, since there were 9 words in the language and every word was equally like to follow any other word). The second type of concatenation, used in the TP+UB training condition, differed from the training speech stream used in the TP training condition in only one way: a 650 ms pause was inserted every 3 to 5 words, resulting in a total of 215 UBs over the course of the training phase. For the purposes of the current study, the only cue to utterance boundaries was a brief silent pause in the speech stream (in natural speech utterance boundaries always involve silences, but are also marked by additional prosodic cues). The pauses were 650 ms in length because this was the shortest silent pause duration that gave the authors a clear impression of an utterance boundary (for simplicity sake, these utterance-boundary-length pauses will

draft version

henceforth be referred to as utterance boundaries, or UBs). These pauses were inserted every 3 to 5 words to create a speech stream that would have utterance boundaries with a frequency roughly reflective of that observed in natural speech (e.g., Johnson et al., 2013; 2014). Each word in the language was preceded and followed by a UB an approximately equal number of times. The third and final type of concatenation, used in the UB training condition, was very similar to the training speech stream used in the TP+UB training condition. The only difference between the UB and the TP+UB training conditions was that the words in the former were concatenated in a fixed rather than random order (see Curtin, Mintz, & Christiansen, 2005; Sahni, Seidenberg, & Saffran, 2010 for use of a similar procedure). Thus, the syllable transition cues between syllables in the speech stream were flat, or uninformative (TPs between and within words were all equal to 1.0). All training speech streams consisted of 95 tokens of each word, and the amplitude was ramped in and out over a 5 second window at the beginning and end of each soundfile to avoid providing an utterance boundary cue at the beginning and end of the training phase (following Perruchet, Tyler, Galland, & Peereman, 2004). Test items were constructed out of the same syllable recordings used in the training phase.

2.3 Procedure and Design

Participants were randomly assigned to one of the three training conditions: TP, UB, or TP+UB. All participants were tested in a quiet room with stimuli played over headphones at a comfortable listening volume. Participants were told that they would be listening to a "Martian Language." They were instructed to relax and listen, and that at the end of the training phase, they would be asked some questions about the language.

draft version

The test phase immediately followed training. After completing one practice trial, participants were presented with eighteen 2AFC trials and instructed to select which word occurred in the language they had just heard. On half of the trials, participants were presented with a partword (consisting of the last syllable of one word and the first syllable of another) versus a "trained" word (e.g., trained word "ripu" versus partword "pawi" for listeners exposed to Language A). On the other half of the trials, participants were presented with a partword versus a "rule" word (a word that was never heard during the training phase, but nonetheless conformed to the phonotactic structure of the language; i.e., rule word "raki" versus partword "tawa" for listeners exposed to Language A). Since the words of one language formed the partwords of the other language, the same test words could be used for all participants. In half of the trials the partwords occurred before the words; in the other half the words occurred first. Participants were instructed to respond on every trial, even if they felt they were guessing. Upon completing the test phase, participants were asked if they discovered any patterns in the language they listened to. The entire experiment took approximately 25 to 30 minutes to complete.

2.4. Results and Discussion

Mean percent correct on each of the two test trial types was calculated for each participant, with chance performance equal to 4.5 (or 50%) correct (see Figure 1). These means were then subjected to a 2 (Test Item Type: Trained Word versus Rule Word) by 3 (Condition: TP, UB, and TP+UB) by 2 (Language: Sonorant Initial versus Obstruent Initial) Mixed Design ANOVA, revealing main effects of Test Item Type, $F(1, 66) = 9.5$,

draft version

p = .003, η_p^2 = .13, Condition, $F(2, 66) = 12.6$, p < .001, η_p^2 = .28, and Language, F(1, 66)=23, p < .001, η_p^2 = .26. There was also a marginally significant interaction between Test Item Type and Language $F(1, 66) = 3.86$, p = .054, η_p^2 = .06, as well as between Condition and Language $F(2, 66) = 3.08$, p = .053, η_p^2 = .08. No other interactions reached statistical significance.

The main effect of Test Item Type was driven by higher performance on Trained Word than the Rule Word trials, t(71) = 3.08, p = .003. This suggests that listeners learned the phonotactic rule defining the language they were exposed to, but still perceived the trained words as more word-like (or familiar) than the rule words. The main effect of Condition was driven by poorer performance in the TP condition than in both the TP+UB condition, t(46) = 3.21, p = .0024, and the UB Condition t(46) = 3.76, p = .0005. Performance in the UB and TP+UB Conditions did not differ, t(46) = .427, p = .672.

Further analyses were carried out to determine whether performance on specific test item types was above chance in the three different conditions. Although overall performance collapsed across conditions was above chance on both Trained Word test trials [M=6.11; t(71) = 7.13, p < .001] and Rule Word [M=5.5; t(71) = 4.45, p < .0001)], learning success on the two trial types differed between conditions. Participants in both the UB and TP+UB conditions performed above chance on both Trained Word [UB: t(23)=8.3, p<.0001; TP+UB: t(23)=6.04, p<.0001] and Rule Word test items [UB: t(23)=5.08, p<.0001; TP+UB: t(23)=4.52, p<.0001]. However, participants in the TP condition failed to perform above chance with either test item type [Trained Words: t(23)=1.24, p=.11; Rule Words: t(23) = .419, p = .66].

draft version

The main effect of language was primarily driven by superior performance in the obstruent-initial language over the sonorant-initial language. This finding is in line with earlier studies reporting superior performance on an artificial language containing obstruent- as opposed to sonorant-initial words (Onnis, Monaghan, Richmond, & Chater, 2005). Interestingly, in our study, this effect was primarily driven by the TP Condition – the condition where no learning was observed [Sonorant-initial TP Condition: M=38%; Obstruent-initial TP Condition: M=66%; Sonorant-initial UB Condition: M=68%; Obstruent-initial UB Condition: M=75%; Sonorant-initial: TP+UB Condition: M=62%; Obstruent-initial: TP+UB Condition: M=77%]. We take this to indicate that all participants likely entered the lab with a pre-existing phonological bias to perceive obstruent-initial syllables as word-initial (much like the participants in Onnis et al., and possibly also the participants in Peña et al., 2002). However, only the participants assigned to the TP Condition failed to overcome this bias when it did not match the phonological structure of the artificial language they were exposed to. Participants in the two conditions involving UB cues, in contrast, were able to learn whether or not the pre-existing bias was appropriate for segmenting the artificial language they were presented with. Finally, the main effect of Test Item Type was primarily driven by the performance of participants who learned the Obstruent-initial language. On average, in the Obstruent-initial language, performance on the Trained words was 11% better than performance on the Rule words. In contrast, performance on the Trained words for those who were exposed the Sonorant-initial language was only 2% better than performance on the Rule words. This may have been because performance in the Sonorant-initial language was generally poor, and thus there was little room to show a difference in performance on

draft version

Trained versus Rule words. Although not central to the goals of the current study, our findings regarding listeners' unequal performance in the obstruent- and sonorant-initial language add to an interesting debate regarding listeners' ability to learn phonological patterns that either do or do not match the phonological patterns seen in a listeners' natural language (Buckley & Seidl, 2005; Endress & Mehler, 2010; Finley & Badecker, 2012; Saffran & Thiessen, 2003).

No participant reported explicitly noticing the phonotactic pattern of the language they were exposed to or noticed the repetition of the syllable sequence in the fixed-order UB training phase. Indeed, the majority of participants felt as if they were simply guessing in the test phase.

Our results demonstrate that listeners learned the phonotactic rule defining words equally well in the UB and the TP+UB training conditions. Listeners in the TP training failed to show evidence of segmenting the words from the speech stream or learning the phonotactic rule of the language. These results suggest that UBs may provide language learners with a more efficient segmentation strategy than attention to TPs. These results also support the notion that UBs may provide a strong universal cue for bootstrapping language-specific information about the sound patterns of words.

Listeners' failure to succeed in segmenting words from the TP Condition was admittedly surprising, given past studies reporting how good adults are at tracking TPs between syllables in an artificial language. However, there is growing evidence that TPs may be more difficult to track than initial work in the area suggested (e.g., Endress and Mehler, 2009; Johnson & Jusczyk, 2003; Johnson & Tyler, 2010; Gambell & Yang, 2005). Moreover, the language-learning task faced by participants in the current study

draft version

differed from past artificial language learning studies in many ways. For example, the artificial language used in Saffran, Newport, & Aslin (1996b) only contained 6 words whereas ours contained 9 (see Endress & Mehler, 2009; Brar, Tyler, & Johnson, 2013, for evidence that artificial languages become more difficult to segment with increasing numbers of words), and listeners in the Saffran et al. study were presented with three seven-minute training phases (with a clear utterance boundary present at the beginning and end of each phase). Thus, one could conjecture that our language may have been too hard to learn in the TP Only Condition. Of course, proposing these particular factors as a possible explanation for why participants performed so poorly in the TP Only Condition is merely speculative, as we did not systematically manipulate these factors in the current study (e.g., see Peña et al., 2002, for evidence that under some circumstances, learning can still occur with brief exposure to an artificial language with many words). Importantly, we do not see our results as evidence that listeners cannot segment an artificial language based on TP cues alone (as they have been shown to do so in numerous previous studies, as discussed in the Introduction). We simply conclude that, under the particular testing conditions employed in the present study, UBs provided a more efficient segmentation cue than TPs. The same language that is difficult to learn with only TP cues to word boundaries is readily learnable with the inclusion of utterance boundary length silences inserted every 3 to 5 words. Given this finding, and the fact that silences provide a universal cue to utterance boundaries in all natural languages, UBs may deserve greater attention as a language-general word segmentation cue than they currently receive in most models of adult and early speech perception.

draft version

4. Conclusions

Researchers have long pondered how infants first solve the word segmentation problem, and tracking TPs between syllables is one of the most widely-accepted explanations to date (e.g., Kuhl, 2004; Saffran, Werker, & Werner, 2006; Werker & Curtin, 2005). However, more recently, it has also been suggested that utterance boundaries may play an important role in solving the segmentation problem (e.g., Daland & Pierrehumbert, 2011; Endress & Hauser, 2010; Seidl & Johnson, 2006; 2008). The current study provides strong support for this view, demonstrating that adult listeners can rapidly acquire the phonological structure of an artificial language simply by attending to the edges of utterances.

The goal of this study was not to test whether or not listeners use TP cues to segment artificial languages (listeners' use of TP cues to segment words from speech has been demonstrated many times in the literature). Nor was the goal to pit TP cues to word boundaries against UB cues. Had this been our objective, we would have tested listeners on a variant of our language containing mismatched TP and UB cues to word boundaries (past studies pitting UB cues against TP cues have shown that UB cues provide important cues to listeners, constraining the way listeners compute TPs; Shukla et al., 2007). Rather, the goal of the current study was simply to examine how efficiently listeners can rapidly segment words from a difficult artificial language when provided with UB only, TP only, or UB and TP cues to word boundaries. In the training conditions involving UBs, we provided listeners with utterance lengths that roughly reflected the typical utterance lengths heard in natural language interactions (Johnson, Lahey, Ernestus, & Cutler; 2013; Johnson et al., 2014; Van de Weijer, 1998). And UBs were marked by

draft version

silences, a universal acoustic marker of utterance boundaries. The strength of the TP cues used in our study were roughly based on the TP cues used in past artificial language studies (e.g., Saffran et al., 1996), and were certainly much stronger than the TP cues provided in natural language input (e.g., Gambell & Yang, 2005; Johnson, 2012; Yang, 2004). Thus, compared to natural language, our artificial language provided listeners with UB cues similar in strength to those encountered by language learners in the real world (similar in terms of frequency of occurrence, but perhaps weaker in terms of acoustic realization since natural UBs are marked by a host of additional acoustic cues besides silences) and TP cues stronger than those encountered in real-world stimuli. Nonetheless, UBs still allowed listeners to more efficiently segment words from speech than TPs. One interpretation of this finding that we favor is that UBs provide language learners with better word boundary cues than TPs. However, other interpretations are possible. For example, UBs provide fairly deterministic cues to word boundaries, whereas TPs provide probabilistic cues to word boundaries. Thus, segmenting words from speech using TPs would be a more gradual process than segmenting words from speech using UBs. One could argue that this is the very reason why UBs are more efficient segmentation cues than TPs. But one could also argue that our results are an artifact of the artificial language methodology we have employed in the current study. In the real world, listeners are not provided with just 10 minutes of exposure to a language and languages contain more than 9 words composed of just 18 distinct syllables. Perhaps our short training phase or the fact that our language contained so few words (all of which landed along utterance boundaries many times within the span of ten minutes) may have increased the utility of UB cues relative to TP cues.

draft version

Another limitation of the current study is that we tested adults rather than infants. Thus, it is not yet clear whether our findings would extend to first language learners. Our intuition, however, is that these results will likely extend to children, since many past studies have shown that infant and adult language learners use similar cues to segment words from speech (e.g., Cutler, 2012; Saffran et al., 1996a; Shukla et al., 2011). Moreover, since utterance boundaries are universally salient to even newborn infants, and infant-directed speech consists of short utterances containing many edge-aligned words (e.g., Aslin et al., 1996; Johnson, under review; Johnson et al, 2013; 2014; Swingley, 2009; van der Weijer, 1998), the findings reported in this paper have potentially important implications for models of first language acquisition. Thus, in another line of work in our lab, we are currently examining *infants'* use of UBs to extract word boundary information from fluent speech. If infants are as attentive to UBs as the adults in the current study, then this would suggest that UBs play a key role in enabling infants to acquire the language-specific phonological cues defining word boundaries. It is worth noting, however, that even if our results did not extend to infants, they would still carry theoretical importance for adult L2 models of word segmentation. As a secondary question, we are also examining whether young infants, like adults, demonstrate a preference for obstruent-initial over sonorant-initial words.

Perhaps one of the most surprising aspects of the current study is how efficiently listeners utilized the UB cues to learn phonotactic cues to word boundaries. After only 10 minutes, listeners not only distinguished words that occurred in the language from words that did not occur in the language, they also distinguished words that followed the phonotactic rules of the language from those that did not. Indeed, when given equal

draft version

exposure to an artificial language containing either UB or TP cues, only those listeners who were presented with UB cues to word boundaries learned the phonological structure of the artificial language. When presented with a language containing only TP cues to word boundaries, listeners demonstrated no evidence of having learned either the words or the phonotactic patterns characterizing words in the language. Moreover, in the TP+UB condition, the presence of TP cues conferred no segmentation advantage over the presence of UB cues alone, suggesting that the participants in the TP+UB condition were relying primarily on UB cues to segment words from the artificial speech stream. Taking into consideration the fact that TPs in natural language are far more complex than those seen in artificial languages, the results of this study could be seen as evidence that a UB segmentation strategy may be very well suited for performing a first pass at segmenting words from natural language.

In conclusion, although attention to TPs offers a compelling explanation for how language learners first solve the word segmentation problem, it most likely does not provide the only solution. Tracking other statistics, such as the patterns of sounds that tend to occur along UBs, may provide listeners with another possibly more efficient language-general strategy for segmenting words from speech. Indeed, the importance of UBs in word segmentation is supported by several computational models of early word segmentation (e.g., Aslin et al., 1996; Brent & Cartwright, 1996; Monaghan & Christiansen, 2010). Thus, our findings add to a growing body of literature suggesting that listeners' attention to the ends of things (whether they be words, phrases, or even utterances) may enable infants to acquire more efficient language-specific word segmentation strategies (e.g., Endress & Mehler, 2010; Johnson et al., 2013; 2014;

draft version

Johnson & Seidl, 2008; Seidl and Johnson, 2006, 2008; Shukla, White, & Aslin, 2011). Future studies should examine whether attention to UBs still provides listeners with useful word boundary information when they are presented with more naturalistic language input, and whether listeners can use UB cues to learn other segmentation cues such as the positioning of lexical and phrasal stress.

draft version

5. Author note

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant awarded to EKJ, a SSHRC grant awarded to EKJ, and an NSERC USRA (Undergraduate Student Research Award) granted to JS. We thank Tania Zamuner for her theoretical input and comments on an earlier draft of this manuscript, and Jaspal Brar for assistance in testing participants. Preliminary results were presented at the International Conference on Infant Studies 2012 in Minneapolis, Minnesota.

draft version

6. Appendix

- insert Table 1 here -

draft version

7. References

Aslin, R., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Brar, J.K., Tyler, M.D., Johnson, E.K. (2013). What you see is what you hear: how visual prosody affects artificial language learning in adults and children. Proceedings of Meetings on Acoustics, 19, 060068.

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

Buckley, E., Seidl, A. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, 1, 289-316.

Chambers, K.E., Onishi, K.H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87, B69-B77.

Christoophe, A., Guasti, T., Nespor, M., Dupoux, M., & Van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, 12, 585-612.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2, 385–394.

Curtin, S., Mintz, T.H., & Christiansen, M.H. (2005). Stress Changes the Representational Landscape: Evidence from Word Segmentation. *Cognition*, 96, 233-262.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken*

draft version

words. Cambridge, MA: The MIT Press.

Daland, R. and J. B. Pierrehumbert (2011) Learnability of diphone-based segmentation. *Cognitive Science,* 35, 119-155.

Endress, A.D. & Hauser, M.D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, 61(2), 177-199.

Endress, A.D. & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Langauge*, 60, 351-367.

Endress, A.D. & Mehler, J. (2010). Perceptual Constraints in Phonotactic Learning. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 235-250.

Finley, S., & Badecker, W. (2012). Learning biases for vowel height harmony. *Journal of Cognitive Science,* 13, 287-327.

Gambell, T., & Yang, C. (2005). Mechanisms and constraints in word segmentation. Unpublished Manuscript, Yale University.

Golinkoff, R., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: implications for language acquisition. *Journal of Child Language*, 22, 703–726.

Johnson, E.K. (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. Williams (Eds.). *Statistical Learning and Language Acquisition.* Mouton de Gruyter.

Johnson, E.K. (under review). Building a proto-lexicon: an integrative approach to infant language development. *Annual Review of Linguistics*.

draft version

Johnson, E.K., Lahey, M., Ernestus, E., & Cutler, A. (2013). A multimodal corpus of speech to infants and adult listeners. *Journal of the Acoustical Society of America, 134, EL534-EL540.*

Johnson, E.K., & Seidl, A.H. (2008). Clause segmentation by 6-month-olds: a crosslinguistic perspective. *Infancy*, 13, 440-455.

Johnson, E.K. & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13, 339-345.

Johnson, E.K., Seidl, A.H., & Tyler, M. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. PLoS One, 9 (1). DOI: 10.1371/journal.pone.0083546.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science*, 3 (9), 323-328. doi:10.1016/S1364-6613(99)01363-7

Jusczyk, P. W., Houston, D. M., Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3), 159-207.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.

Marchetto, E., & Bonatti, L.L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, 67, 130-15.

Mersad, K., Nazzi, T. (2012). When mommy comes to the rescue of statistics: infants combine top down and bottom up cues to segment speech. *Language Learning and Development,* 8, 303-315.

Monaghan, P., Christiansen, M.H. (2010). Words in puddles of sound: modeling psycholinguistic effects in speech segmentation. *Journal of Child Language*, 37,

draft version

545-564.

Onishi, K.H., Chambers, K.E., Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-B23.

Onnis, L., Monaghan, P., Richmond, K., Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53, 225-237.

Ordin, M., & Nespor, M. (2013). Transition probabilities and different levels of prominence segmentation. *Language Learning*, 63, 800-834.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80, 674-685.

Peña, M., Bonatti, L.L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing, *Science*, 298, 604-607.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133(4), 573-583.

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research*, 10, 113-122.

Sahni, S.D., Seidenberg, & Saffran, J.R. (2010). Connecting cues: overlapping regularities support cue discovery in infancy. *Child Development*, 81, 727-736.

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

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

draft version

Saffran, J.R., & Thiessen, E.D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484-494.

Saffran, J.R., Werker, J., & Werner, L. (2006). The infant's auditory world: hearing, speech, and the beginnings of language. In R. Siegler and D. Kuhn (Eds.), Handbook of Child Development. New York: Wiley (p.58-108).

Seidl, A.H., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565-573.

Seidl, A. & Johnson, E.K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language*, 35, 1-24.

Selkirk, E. (1984). Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press.

Shukla, M., White, K. S. & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences,* 108, 6038-6043.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1–32.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50, 86-132.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, *364*, 3617-3632.

Thiessen, E. D., & Erickson, L. C. (2013). Discovering words in fluent speech: The contribution of two kinds of statistical information. *Frontiers in Psychology*, 3, 590.

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues

draft version

to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Thiessen, E.D., & Saffran, J.R. (2007). Learning to learn: Infants' acquisition of stress-based segmentation strategies for word segmentation. *Language, Learning and Development*, *3*, 73-100.

Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, 126, 367-376.

Van de Weijer, J. (1998). *Language input for word discovery*. Unpublished doctoral dissertation, Max Planck Series in Psycholinguistics 9.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234. doi: 10.1080/15475441.2005.9684216

Yang, C. D. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Science*, 8, 451- 456.

FIGURE CAPTION

Figure 1: Number correct broken down by Condition and Test Item Type (chance performance equals 4.5 correct out of 9; error bars indicate SE).

draft version