2010

Statistical Bootstrapping of Speech Segmentation Cues

Nicolas O. Planet
University of Massachusetts Amherst

STATISTICAL BOOTSTRAPPING OF SPEECH SEGMENTATION CUES

A Thesis Presented

by

NICOLAS O. PLANET

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements of the degree of

MASTER OF SCIENCE

May 2010

Psychology

STATISTICAL BOOTSTRAPPING OF SPEECH SEGMENTATION CUES

A Thesis Presented

by

NICOLAS O. PLANET

Approved as to style and content by:

Lisa D. Sanders, Chair

Andrew L. Cohen, Member

Joe V. Pater, Member

Melinda A. Novak, Department Head
Psychology

ABSTRACT

STATISTICAL BOOTSTRAPPING OF SPEECH SEGMENTATION CUES

MAY 2010

NICOLAS O. PLANET

Directed by: Professor Lisa D. Sanders

Various infant studies suggest that statistical regularities in the speech stream (e.g. transitional probabilities) are one of the first speech segmentation cues available to infants. Statistical learning may serve as a mechanism for learning various language specific segmentation cues (e.g. stress segmentation by English speakers). To test this possibility we exposed adults to an artificial language in which all words had a novel acoustic cue on the final syllable. Subjects were presented with a continuous stream of synthesized speech in which the words were repeated in random order. Subjects were then given a new set of words to see if they had learned the acoustic cue and generalized it to new stimuli. Finally, subjects were exposed to a competition stream in which the transitional probability and novel acoustic cues conflicted to see which cue they preferred to use for segmentation. Results on the word-learning test suggest that subjects were able to segment the first exposure stream, however, on the cue transfer test they did not display any evidence of learning the relationship between word boundaries and the novel acoustic cue. Subjects were able to learn statistical words from the competition stream despite extra intervening syllables.

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Natural speech does not have the same explicit cues to word boundaries that are present in written language. The short breaks present in speech streams do not reliably identify word boundaries. This is most evident when we listen to a foreign language. Natural speech in a foreign language is often perceived as a rapid and continuous stream of sound. We must be able to use an assortment of cues in order to segment speech into a meaningful string of words. Segmentation cues can be divided into two general categories: lexical and acoustic. Lexical cues involve things like the uniqueness point of a word, lexical competition and top-down knowledge (e.g McClelland & Elman, 1986, however, see Norris, McQueen, and Cutler, 2000 for an argument against top down effects.) In the last few decades, researchers have identified and studied the contributions of several segmentation cues including stress, prosody, phonotactics, and transitional probabilities. The proposed study is designed to investigate how we learn the cues to segment speech. One hypothesis is that a universal cue can serve as a basis for a type of bootstrap learning that allows for the use of segmentation cues that are language specific. A likely candidate for this "foundation" cue is a type of statistical cue known as transitional probabilities.

There are some who argue or assume that the word segmentation problem can be avoided by the possibility of learning words in isolation and using this lexical knowledge to extract words from longer utterances (see Pinker, 1984). Brent and Cartwright (1996) point out that there are a few weaknesses to this position. One counterpoint to isolated word learning theories is that it there is no proposed method for discriminating single word and multiword utterances. Another counterpoint is that some words rarely or never occur in isolation, such as determiners. There is also considerable evidence suggesting that the speech input to young children predominantly consists of multiword utterances. For example, a series of studies by Aslin, Woodward, LeMendola and Bever (1996) were designed to analyze the speech used by mothers to their infants. As part of these studies mothers were given the task of teaching their 12-month-old infants 3 novel words using whatever method they preferred. Analysis of the speech input revealed that more than 10% of the mothers never used the novel words in isolation and for the majority of mothers the infant directed speech consisted mostly of multiword utterances. Additionally, there is evidence that infants have the ability to segment novel words from fluent speech. Jusczyk and Aslin (1995) presented 7.5 month-old infants with fluent natural speech productions of multiple sentences containing a novel target word. The testing phase of the experiment provided evidence that the novel words were recognized, suggesting that the infants were able to segment the unfamiliar words from the passages. If hearing words in isolation is not a prerequisite for learning words then there must be other important segmentation cues available to language learners. Researchers are looking towards statistical learning as a possible explanation.

All natural languages have statistical regularities in how phonemes are concatenated. The transitional probability between two syllables is one type of statistical cue that has been studied extensively. The idea that transitional probabilities can be used as a segmentation cue is based on the fact that syllables (or other sublexical units) that form a word co-occur more frequently than syllables that cross word boundaries. The exact transitional probability can be calculated for a sequence of syllables (XY) by dividing the frequency of XY by the frequency of X. A high transitional probability indicates that Y is strongly predicted by X. A low transitional probability indicates a weak relationship between X and Y and can be interpreted as a word boundary (Saffran, Newport, & Aslin, 1996). To make this concrete, consider Saffran et al.'s (1996) example using the word *baby*. The transitional probability of the syllable *bi* given *bay* can be computed using the following equation:

$$\frac{\text{Frequency of bay.bi}}{\text{Frequency of bay}}$$

The value given by the above equation will very likely be greater than the value resulting from the word external pair *bay-too* in the equation below.

$$\frac{\text{Frequency of bay\#too}}{\text{Frequency of bay}}$$

Initial studies of distributional cues in language were restricted to computer models. It was not long before researchers were able to create the proper experimental design to study human speech segmentation using statistical cues.

To investigate the possible role of transitional probabilities in speech perception, Saffran et al. (1996) created a speech segmentation paradigm sometimes referred to as the artificial language learning paradigm that has been adopted by many of the segmentation studies that followed. Six trisyllabic words (babupu, bupada, dutaba, patubu, pidabu, and tutibu) were created from 12 consonant-vowel syllables. Transitional probabilities within words ranged from .31 to 1. Transitional probabilities between words were much lower and ranged from .1 to .2. The words were presented in random order and produced by a speech synthesizer. After 21 minutes of exposure to the artificial language, adult subjects took a two-alternative forced-choice test. Test questions pitted words against non-words and part-words. Non-words were comprised of a sequence of syllables that never occurred during the speech stream. Part-words, on the other hand, were made by taking one of the six words and replacing the first or last syllable with a syllable from another word (e.g. babudu). During testing, subjects were able to distinguish the trisyllabic words from non-words and part-words. Additionally, the variability of transitional probabilities within words (.31 -1) allowed them to show that a higher average transitional probability within a word led to better learning.

The ability of humans to use statistical information to segment speech has important implications especially if this ability is available to infants during language acquisition. Saffran, Newport, and Aslin (1996) tested the ability of 8-month-old infants to use transitional probabilities. After only two minutes of familiarization with a speech stream, infants were able to differentiate words from non-words and part-words. To rule out the possibility that the effects seen in the previous study were based on hearing words more frequently than any other test item, a follow-up study by Aslin, Saffran, and Newport (1998) equated the frequency of words and part words. To do this, Aslin et al. (1998) created a four-word corpus. During the speech stream presentation, two of the words occurred twice as frequently as the other two. The part-words formed across the boundaries of the more common words occurred as frequently as the less common words. During the testing phase, infants were presented only with the less common words and the more frequent part-words. Even when part-words and words were heard the same amount of times, 8-month-old infants showed the ability to discriminate between them. This provided strong evidence that statistical cues were enough to segment speech and that this ability is available very early.

Statistical learning is not specific to linguistic material. Saffran, Johnson, Aslin, and Newport (1999) varied transitional probabilities between pure tones and tested the ability of adults and infants to segment the stream into “tone words.” After exposure to three 7–minute tone streams, adult subjects were able to differentiate tone words from non-words and-part words. They also found that tone words with higher average transitional probabilities were learned best. In a second experiment, Saffran and colleagues presented 8-month-old infants with a 3-minute tone stream and found that infants were able to discriminate tone words from part words. It has even been shown that primates are able to segment the same speech streams that were used by Saffran et al. (1996) (Hauser, Newport, & Aslin, 2001). Multiple studies (Turk-Browne, Isola, Scholl, & Treat, 2008; Kirkham, Slemmer, & Johnson, 2002) suggest that this mechanism may be used for extracting regularities in a variety of modalities.

The statistical learning studies discussed so far have focused on the learning of statistical patterns between adjacent syllables or sounds. There are instances in natural language in which regular patterns between nonadjacent syllables or phonemes are observed (e.g. infixation in Tagalog or consonant templates in Semitic languages). Newport and Aslin (2004) found that this type of statistical learning is more heavily constrained by natural language. In their first experiment, they found that subjects were unable to learn words with nonadjacent syllable dependencies. They performed multiple variations including lengthening the learning stream, making it an implicit learning task, simplifying the language, and changing the phonemes used but none of these changes resulted in the learning of nonadjacent dependencies between syllables. It turns out that statistical learning is constrained by the available patterns in natural speech. The nonadjacency patterns found in natural language tend to be between phonemic units and not syllables. Newport and Aslin (2004) created a new experiment with nonadjacent patterns between consonants and another experiment with patterns between nonadjacent vowels. The results showed successful statistical learning and further extended the power of statistical cues in language learning. This early, possibly innate ability to extract statistical regularities from language and other auditory information could provide an explanation of how listeners begin to build their collection of acoustic segmentation cues.

Not all speech segmentation cues are employed across all languages, as would be expected with transitional probabilities. For example, while phonotactic cues are present in all languages, the specific phoneme combinations that are legal or illegal differ from language to language. Surprisingly, the ability to use basic units of language (e.g. syllables or moras) as word boundary cues in speech segmentation is also language specific. A study by Mehler, Dommergues, Frauenfelder, and Segui (1981) found evidence that French speakers use syllable boundaries to segment speech. For example, participants were faster to respond to a target consonant-vowel (CV) sequence *pa* within a word such as *palace* which is clearly syllabified as [pa][lace]. Conversely, participants were faster to identify the consonant-vowel-consonant (CVC) sequence *pal* within a word such as *palmier* which in French has the syllable structure of [pal][mier]. A replication of this experiment by Cutler, Mehler, Norris, & Segui (1983) used English speakers and did not find the same syllable effects. Even with English words that had clear syllable boundaries there were no significant differences in reaction times to targets when they matched the complete syllable of a word and when they did not. An additional experiment by Cutler et al. (1983) tested French speakers using English stimuli. The results indicated that the French speakers continued to use the syllabification strategy with English words. Further studies with Korean (another proposed syllable timed language) have shown that the syllable segmentation strategy is not something specific to French but something that is likely applied to all languages in the syllable rhythm class (Kim, Davis, & Cutler, 2008). Similar to the French or Korean speaker's use of syllable structure to segment, Japanese speakers have been shown to take advantage of certain properties of their language to help segment speech. Japanese is believed to be a mora timed language and numerous studies suggest that this is evident in the way it is produced and processed (e.g. Kubozono, 1989; Katada, 1990; Otake, Hatano, Cutler & Mehler, 1993). Just like French and Korean speakers applying syllable segmentation strategies to native and nonnative stimuli, Japanese speakers will attempt to use moraic segmentation strategies to segment speech in unfamiliar languages (e.g English) that do not have regular moraic structure (Cutler & Otake, 1994).

The likely explanation for the results of the moraic and syllable segmentation studies lies in the differences between languages belonging to separate rhythm classes. French and Korean are believed to be syllable-timed languages and the majority of words have clear syllable boundaries. English, on the other hand has many words with ambisyllabic segments: phonemes that belong to two syllables. This makes most syllable boundaries in English ambiguous. However, as evidenced in the results, it is not the properties of the language alone but also the acquisition process of language learners that leads to the different segmentation strategies. One possibility is that through the interactions of syllable boundaries with other segmentation cues, the French or Korean child or infant learns that syllable boundaries are a consistent and efficient cue for segmentation. Alternatively, Cutler, Mehler, Norris and Segui (1986) propose that syllabification is just one of many strategies available to the language processing device. During language acquisition, learners develop specific segmentation strategy preferences based on the phonological properties of their language. So in the case of syllabification, French and Korean speakers develop a preference for the strategy over alternatives, while English speakers do not. It is important to note that the lack of segmentation by syllables for English speakers does not equate to a lack of classification of syllables.

English speakers may not use sound unit segmentation strategies but other unique prosodic properties of the English language may afford the use of segmentation strategies that are not employed by speakers of other languages. Prosodic properties of the speech input include things like rhythm, intonation and stress. Stress has been shown to be an important segmentation cue for English speakers. Stress may involve a relative change in pitch or duration of a phoneme or syllable that can be used to put emphasis on that unit. In English, stress occurs on the first syllable of about 83-90% of open-class words, which includes nouns, main verbs, adjectives, and adverbs (Cutler & Carter, 1987). Cutler and Norris (1988) found that, among native English speakers, identification of a monosyllabic word in a two-syllable cluster was facilitated by a strong-weak stress pattern. Strong syllables are identified by the presence of full vowels (e.g eye and pill) and weak syllables contain reduced vowels which are usually schwa (e.g second syllable in ion). In the Cutler and Norris (1988) study, subjects were quicker to identify *mint* in *mintef* (strong-weak) than in *mintayve* (strong-strong). They argue that this is a result of segmentation cued by the identification of strong syllables as word onsets. According to this theory the second strong syllable in *mintayve* triggers segmentation (*min-tayve*), which slows down the access of the word *mint* because it crosses a segmentation boundary. Even within the processing of natural English speech, which contains many lexical, semantic, and acoustic cues, there is a measurable effect of stress cues, suggesting that it plays an important role not only in language learning but normal speech processing (Sanders & Neville, 2000). More specifically, stress cues can account for 3% of the accuracy in identifying target phonemes a speech stream. The effect of stress cues increases when semantic cues are removed and further increases when syntactic cues are removed.

Additional studies have looked to other stress timed languages to assess whether stress segmentation was specific to English. Dutch and Finnish also have predominant initial syllable stress to an even greater extent than English such that it is nearly deterministic in Finnish. Vroomen, Tuomainen, and de Gelder (1998) found that in a task which involved the detection of CVCV words within CVCVCV segments, stress cues aided segmentation for Finnish speakers to a greater extent than disharmonious vowel sequences which can also be a word boundary cue in vowel harmony languages like Finnish. Another experiment in the Vroomen et al. (1998) study found that stress also improved segmentation for Dutch Speakers but not for French speakers, who arguably have no association between stress and word boundaries. The relation between the stressed syllable and a word boundary may be crucial for the use of stress segmentation strategies. In Spanish roughly 70-80% of multisyllabic words have penultimate stress (Harris, 1983); this means that in Spanish, stress is a reliable cue to word boundaries, comparable to English. However, as shown by Toro-Soto et al. (2007) stress did not improve Spanish speakers' ability to segment nonsense words from a continuous stream. In fact, penultimate stress on artificial words reduced performance to chance levels. This suggests that the placement of stress near word boundaries may be important for it's viability as a segmentation cue.

Infants can perceive stress patterns as early as about 2 months of age, however there is no evidence for an ability to use stress for segmentation at this age (Jusczyk & Thompson, 1978). Jusczyk, Cutler and Redanz (1993) established that the infant's listening preference for words with a strong-weak pattern develops sometime between 6 and 9 months of age. Jusczyk, Houston, and Newsome (1999) performed a series of experiments that provided support for stress-cued segmentation of fluent English speech by 7.5 month olds. For example, if the word *guitar* (weak-strong) was consistently followed by *is*, the infants showed a preference for the segmentation of *taris* over *guitar* By creating a speech stream with conflicting stress and statistical cues, Johnson and Jusczyk (2001) found evidence that 8-month-old infants rely more heavily on stress cues than transitional probabilities to segment an artificial language. More specifically after listening to a stream with statistically coherent iambic words, infants at this age preferred the trochaic part-words. The finding of infants at such an age using stress segmentation cues over other conflicting cues suggests that it may be the first cue used for speech segmentation. However, to further investigate this issue, Thiessen and Saffran (2003) studied developmental changes in the use of segmentation cues in infancy. First, they found a preference for stress over statistics with 9-month-old infants. They then repeated the same experiment with 7-month-old infants and found the reverse pattern. The 7-month-olds relied more heavily on statistics to segment speech. One interpretation of these findings is that at 7-months of age the infant is transitioning from statistics to stress as the primary speech segmentation cue.

The research discussed converges on a sort of learning process in which the ability to compute and analyze the statistical regularities in speech bootstraps the learning of other segmentation cues (e.g. stress). Bootstrapping is an important concept for infant learning and in particular language acquisition. Infant bootstrapping theories usually involve an innate mechanism that allows for independent learning with little to no teaching or feedback from the parent or caregiver. Bootstrap learning models have been proposed for the emergence of developmentally crucial abilities such as joint attention (Nagai, Hosoda, & Asada, 2003), social and emotional behavior (Yale, Messinger, Cobo-Lewis, & Delgado, 2003), and the perception of goal-directed action (Biro & Leslie, 2007). Language acquisition presents a special challenge for both innatist and empiricist theorists. Although interactive top-down models of speech segmentation may be compatible with adult segmentation they cannot explain how an infant with little to no lexical knowledge can begin to segment. Theories that posit inherent knowledge of parameters that select for language specific segmentation strategies would still require a preliminary analysis of the speech input (Cairns, Shillcock, Chater, & Levy, 1997). The universality and domain general aspects of statistical cues and statistical learning provide reason to believe that they are the foundation for speech segmentation. Additionally, the sub-lexical nature of transitional probabilities makes it a suitable statistical cue given the constraints on infant learning.

It has been shown that learning early in life is constrained by the limited and immature cognitive capacities of infants. For example, adults are able to hold more information in short-term memory than children (see Dempster, 1981, for a review). However, despite the cognitive limitations of infants they show a propensity for language learning that far exceeds that of adults. A theory termed the "Less Is More" hypothesis has been proposed by Newport (1988, 1990) to explain the superior language learning capabilities of infants and young children relative to adults. This theory suggests that the limited processing abilities and memory capacity of children benefits their ability to learn in tasks that require componential analysis. This theory developed from well known evidence of critical periods and Newport's own work with deaf learners of American Sign Language (ASL) and learners of English.

A set of studies (Newport & Supalla, 1990; Newport, 1990) has used three groups of congenital or pre-lingual deaf signers: native learners, early learners, and late learners. They investigated the relationship between age of acquisition and the comprehension and production of ASL motion verbs. The sign productions were examined at various levels of syntax and morphology. Participant's abilities were scored, error patterns were measured and behavior was qualitatively analyzed. Compared to native signers, late learners were more inconsistent with their use of ASL morphology, and they produced more ungrammatical forms. Further analysis showed that the effects could not be explained by related factors like years of experience and other possible influences such as input differences or differences in social or intellectual deprivation. Altogether, this provides strong evidence that the earlier the age of acquiring a first language, the more proficient the person will be.

Johnson and Newport (1989) researched the critical period hypothesis with studies of second language acquisition. They studied the effect of age of arrival on competency in English among native Korean and Chinese speakers. The results were consistent with the first language results, such that even when controlling for the amount of experience with English, the earlier the age of arrival the more proficient the speaker is.

Newport (1990) notes important characteristics of the errors made by non-native signers in the previous studies. One common type of error is "frozen" structures: unanalyzed multi-phoneme constructions that are frequent in ASL and used by the late-learner in contexts where some of the morphemes are ungrammatical. Additionally, the structures produced by late-learners are more variable and inconsistent. Together, this suggests that late-learners are learning the language by acquiring whole word constructions without fully analyzing the morphological structure. Kersten and Earles (2001) found that presenting adults with small bits of linguistic information, mimicking infant processing, resulted in better learning of meaning and morphology compared to when they were presented with the full complexity of the language.

Although this theory has been proposed for the acquisition of meaning-to-morpheme relations, it can be extended. The limited cognitive abilities of infants should not selectively affect morphology acquisition. Instead, we should find that the learning of other aspects of linguistic information is affected by processing limitations in similar ways. It follows that speech segmentation should also be influenced by these same constraints. Given the lack of lexical knowledge, the constraints on infant learning and evidence for the syllable as the smallest unit of segmentation (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Bertoncini & Mehler, 1981), transitional probabilities are very likely to be the easiest and most accessible statistical cue for infants to begin to segment speech. The reliance on transitional probabilities is likely to be an important difference between infant L1 acquisition and adult L2 learning. Finally, through statistical learning it seems that infants can learn new cues that have lower processing demands and these new cues become the preferred way to segment. The infant studies suggest that this process is occurring but there is little direct evidence that statistical bootstrapping of segmentation cues is possible. On the other hand, through isolated word learning and frozen structure learning, adults may be paying less attention to statistical cues in L2 learning and thus the statistical bootstrapping of various segmentation cues are less likely to occur. Despite this possibility we know that adults are still able to use statistical cues to segment speech so it reasonable to believe that adults are a suitable subject for testing the statistical bootstrapping hypothesis.

The proposed study is a modified artificial language learning paradigm designed to investigate the hypothesis that it is possible to use transitional probabilities to learn a novel acoustic speech segmentation cue. To test this, participants will be presented with a speech stream in which transitional probabilities consistently line up with a novel acoustic cue placed on the last syllable of each word. Listeners will then be tested to see if they are able to learn the association between word boundaries and the novel acoustic cue and apply it to a new set of words. It is important that this cue is not present in

English so that we can rule out any effect of language experience on the listener's ability to use it. Finally, statistics and the newly learned cue will be pitted against each other to see which cue listeners rely on more heavily.

CHAPTER 2

METHODS

The experiment was divided into two sessions. The first session began with the cue position preference test to assess any bias the participant may have to associate the novel cue with a word position. Following the test, participants listened to a random stream with no transitional probability cues to word boundaries, only the novel acoustic cue every third syllable. After listening to the random stream participants took the familiarized cue position preference test to assess if the rhythmic properties of the continuous speech stream changed listener's preferences. Two to seven days later participants returned for the second session. Session two began with a word learning stream in which transitional probabilities reliably identified word boundaries and the novel acoustic cue was on the last syllable of all words. The stream was followed with a test to assess word learning. Immediately following the word-learning test, participants were given all new words in the cue transfer test to see if participants have learned an association between the novel cue and word boundaries and apply it to new syllable stimuli. After the second test participants listened to another stream in which transitional probability cues and the novel acoustic cue were pitted against each other. The stream was followed with the competition test to see whether participants relied more heavily on the novel acoustic cue or transitional probabilities to segment the stream.

Participants

Thirty-six adult undergraduate students at the University of Massachusetts were recruited for the experiment. Four participants did not complete the full study. All participants were native English speakers. All participants provided written consent before the experiment. Participants received either class credit or payment for their participation.

Stimuli

Speech was synthesized using Acapela speech synthesis software. Syllables were recorded individually. All volume and pitch adjustments were made in PRAAT. All words in the artificial language were three-syllable nonsense words. The words were created from 8 consonants (d, p, b, m, r, t, l, k) and 5 vowel sounds (oh, ah, ee, oo, ay; IPA: [o], [a], [i], [u], [e]) resulting in 36 unique syllables used in the experiment. All syllables began with a consonant and ended in a vowel sound (CV). No word or syllable appeared in more than one block. A maximum of 5 different consonant sounds and 4 different vowel sounds were used in each block. In all blocks, the novel segmentation cue was a pitch manipulation on word final syllables in which the pitch fluctuated by +/- 10 Hz across the duration of one syllable in a sinusoidal pattern. The average pitch (160.6 Hz) of this modulated syllable was the same as the unchanged syllables. Syllables ranged in length from 193 ms to 363 ms with an average duration of 287 ms.

The first speech stream was composed entirely of words and the novel segmentation cue was on the last syllable 100% of the time. In this stream, the four words did not share any of the same syllables. Words were randomized with the one exception that no word was presented two times in a row. Transitional probabilities in the first stream were 1.0 within words and on average .33 between words. The four words in the stream were *dahpeebah, maymeedoh, raybeepah, rohdeepay*. Each word was repeated 150 times.

The word-learning test was composed of four words and eight part-words. Two different types of part words were created for this test. Four of the eight part-words were made of the last two syllables of a word followed by the first syllable of another, the other four were formed from the last syllable of a word followed by the first two syllables of another. This resulted in half of the part-words with the novel segmentation cue on the middle syllable and the other half with the cue on the first syllable. Words were presented with the novel segmentation cue on the last syllable. Words and part-words were repeated six times each during the test. Each word was paired with three of each part-word type, yielding a 36-item test. The cue transfer test was constructed in the exact same way only with new syllables that have not been used in any of the previously presented stimuli.

The second speech stream contained 4 multi-word sequences. These sequences were composed of 5 syllables. The secondary cue was always placed on the fifth syllable. Transitional probabilities between the first three syllables were high (1.0) and the probabilities between the last three were low (.2). Probabilities between multi-word sequences were the lowest (.07). A total of 12 different syllables were used to create all four sequences. This design made it possible for the stream be segmented in two ways. If the listener prefers to segment using transitional probabilities they will choose the first three syllables in each chunk over the last three when asked which grouping forms a word. If the listener relied more heavily on the novel segmentation cue, they should indicate the last 3 syllables form a word. The four sequences included *lahmahleerookoo, tookahrooreelee, kayrahkooleeree, mootahreekooroo.*

The competition test contained the four transitional probability words that made up the first three syllables of the five-syllable sequences and the four novel segmentation cue words that made up the last three syllables. Each of these words was repeated nine times during the test. Each transitional probability word was paired with each of the novel segmentation cue words to create the 36-item test. The novel segmentation cue appeared on the test items in the same way that they appeared during the speech stream such that transitional probability words have no pitch change on any syllable, and the novel segmentation cue words have the pitch drop on the last syllable.

Procedure

The experiment was divided into two separate sessions. Sessions were separated by a minimum of one day and a maximum of one week. The first session began with a 36-item two alternative forced choice cue position preference test. Participants were asked to pick the items that sounded like a more plausible word in a foreign language. After the test, participants listened to a continuous stream which contained syllables presented in pseudo-random order with the only stipulation being that every third syllable has the novel segmentation cue. Despite the randomness of the stream, participants were told that it contained words and that it was their task to learn the words for a second pretest to assess learning. The stream lasted for 14 minutes and was divided into two 7-minute sections. For the familiarized cue position preference test that followed the stream, participants were asked to indicate which item they believed was a word in the stream they had listened to. Participants were informed that some pairs were both nonwords. If neither item sounded right to them then they should indicate which item sounds better, similar to the criteria used on the first pretest. After this test the session was concluded and subjects were informed that there were indeed no words in the stream.

In the second session, participants were asked to listen to two streams of speech in a foreign language. After each stream, participants were tested to determine if they correctly segmented each language. They were told that the stream would contain words but they would be void of meaning and there was no grammatical structure. In the first block, participants listened to a stream of six three-syllable nonsense words repeated 150 times each in pseudo-random order. This stream lasted approximately 14 minutes and was divided into two 7-minute sections with a 3-minute break between each section. After listening to the speech stream, participants were given a 36-item two alternative forced choice word-learning test. For each pair of items, participants were asked to choose which of the two was a word in the language they heard. This test assessed the listener's ability to learn the nonsense words based on the transitional probabilities between syllables in the continuous stream.

After the word learning test participants were given a 36-item cue transfer test which had all new words and syllables. The task was to pick the items that made better words. The participants no longer had transitional probabilities to guide their decisions so the novel segmentation cue was the only indication of proper word structure in the language.

In the second speech stream, the novel segmentation cue and statistical regularities were pitted against each other. Participants listened to a stream of five-syllable items repeated 150 times each during the 28-minute stream. After participants listened to this stream, they took a 36-item test to determine if they learned the words in the third stream by using statistical regularities or the novel segmentation cue.

CHAPTER 3

RESULTS

Results from the cue position preference test indicated that subjects did not pick the last syllable cue (LS) words more than chance (t(31)=-1.69, p =.100). However, subjects were more likely to pick the first syllable cue (FS) words more than chance (t(31)=2.45, p=.020). The mean score for proportion of FS items picked was .57 or 13.7 out of 24. After exposure to a random syllable stream subjects were still no more likely to pick the LS words than chance on the familiarized cue position preference test (t(31)=.745, p = .462), although there was a significant increase of .07 for proportion of LS words picked (t(31)=-2.064, p=.047). Additionally, on this test subjects were not more likely to pick FS words than chance (t(31)=.56, p = .582). The change in proportion of FS words picked from before to after the listening stream was not significant (t(31)=1.15, p= .259).

When the subjects returned for the second session they were first exposed to a speech stream containing four words with the novel acoustic cue on the final syllable. On the word learning test following exposure, subjects were more likely than chance to pick words (t(31)= 2.76, p=.010). Mean proportion of words chosen was .62 or 14.9 out of 24. When comparing the familiarized cue position preference test with the word learning test result we see a significant increase of .09 in the proportion of LS words chosen (t(31)= -2.12, p=.042).

Directly after the word learning test participants were given a cue transfer test without hearing a new stream. The cue transfer test contained 4 new words and 8 new part-words composed of syllables that were not heard in the previous streams. The only cue to distinguish words from non-words was the placement of the novel acoustic segmentation cue on the last syllable. Subjects were no more likely to pick the words than chance ($t(31)=.62$, $p=.542$). Additionally, the proportion of words picked was not significantly different from the proportion of LS words picked in the familiarized cue position preference test ($t(31)=.28$, $p=.778$).

In the last section of the experiment, participants listened to a stream that could be segmented in at least two ways depending on whether they relied more heavily on statistical cues or the novel segmentation cue. Results on the competition test showed that participants had a tendency to prefer statistical words although this was only marginally significant ($M=.57$, $t(31)=1.99$, $p=.055$). The proportion of statistical words picked was not significantly different from the proportion of LS words chosen in the familiarized cue position preference test ($t(31)=-.89$, $p=.382$). Additionally the proportion of statistical words chosen in the competition test was not significantly different from the words chosen in the word learning test ($t(31)=1.16$, $p=.253$)

Because not all of the subjects provided evidence for statistical learning it was necessary to divide the participants into groups so that we could better assess the learning of a novel segmentation cue from statistical learning. Participants were divided into two groups. The criterion for the group division was a .1 increase in proportion of selected LS words from the average of the pretest scores to the word learning test score. The division resulted in a high-performer group ($n=17$) that showed better than chance performance ($M=.78$) on selecting words on the word learning test ($t(16)=6.05$, $p<.001$) and a low performer group that did not differ from chance ($M=.44$, $t(14) = -1.59$, $p = .135$).

In high-performers, the proportion of words chosen was significantly higher than the proportion of LS words chosen on the familiarized cue position preference test by .23 ($t(16)=5.384$, $p<.001$). Despite the division of groups, neither high-performers nor low-performers performed better than chance on the cue transfer test ($t(16)=1.03$, $p=.318$, and $t(14)=-.31$, $p=.765$). High performers were at better than chance for picking the statistical words on the competition test (proportion=.63, $t(16)=3.09$, $p=.007$) while low performers showed no preference ($t(14)=.074$, $p=.942$). The proportion of statistical words chosen on the competition test was not significantly different from the proportion of words chosen in the familiarized cue position preference test for high-performers ($t(16)=-.09$, $p=.109$). Finally, for high performers only, the proportion of words chosen in the word learning test was significantly higher than the proportion of statistical words chosen on the competition test ($t(16)=2.84$, $p=.012$).

Comparisons between the high and low-performer group showed significant differences only for performance on the word learning test ($F(1,31)= 30.79$, $p<.001$) and the competition test ($F(1,31)= 4.30$, $p=.047$) with high-performers selecting a higher proportion of statistical words on both tests.

CHAPTER 4

DISCUSSION

Together the results indicate that statistical learning was not an effective mechanism for learning this particular novel acoustic segmentation cue. The results of the cue position preference test given at the beginning of the experiment suggest that participants have a preferential bias for words with the novel segmentation cue on the first syllable. It is not definite why this cue is being identified with word initial position. It may have something to do with interference from the subject's native language. It has been previously shown that participants are influenced in artificial language learning paradigms by the properties of their native language. As early as 8 months of age there is a preference for artificial words which follow the general English stress pattern even when it conflicts with statistical cues (Johnson & Juscyzk, 2001). Additionally, Finn and Hudson Kam (2008) found that adult listeners were better at learning words in an artificial language when they were consistent with the phonotactics of their native language, suggesting that L1 experience was interfering with the ability to identify word boundaries in the continuous speech streams. This finding is especially interesting given that adult subjects, unlike infants in other studies, were aware that the task did not involve their L1 and therefore could have inhibited their L1 processing if possible. Additionally, Weber and Cutler (2006) have shown that L1 phonotactic interference effects can occur even with high proficiency in an L2. In this study, participants were given the task of detecting English words in nonsense sequences. Even though the task only involved detecting English Words, German listeners who were highly proficient in

English were better able to detect English words when the word boundaries were consistent with German-specific phonotactic cues.

It is possible that the biases observed on the first pretest are due to phonological effects from English. Despite the efforts to create a truly novel cue, the pitch changes involved in the cue may still be too perceptually similar to English stress. English stress is mainly perceived as changes in three acoustic parameters: intensity, duration and pitch. Listeners rely most heavily on pitch and duration cues and rely least on intensity (Fry, 1955). Altering just the pitch or duration of a syllable is enough for adult listeners to perceive the syllable as stressed (Streeter, 1978). The novel cue was designed so that its average pitch was no different from the rest of the syllables, which makes it quite different from the stress in English where stressed syllables tend to be higher in pitch relative to unstressed syllables. Nonetheless, because adults are highly sensitive to pitch changes as an indication of stress, listeners could have considered any kind of change in stress in the otherwise monotone stream as a stress-like prominence.

Although participants did show an initial bias for the novel segmentation cue in word initial position, this bias was not present on the familiarized cue position preference test. This suggests that listening to the randomized stream effectively eliminated existing biases for cue positions. If the listener initially made an association between the novel cue and English stress then they either unlearned the association or they discovered that it was not an effective cue in this particular language and were able to "switch it off" in a sense. There have been a few studies showing the ability to exert control over linguistic knowledge through inhibition or language-specific selection mechanisms. For example, Levy, McVeigh, Marful, and Anderson (2007) showed that during speech production

English speakers learning Spanish inhibited the phonology of corresponding English words when producing a Spanish word. However for highly proficient bilinguals there is evidence that is consistent with a language-specific selection mechanism rather than inhibition (Costa & Santesteban, 2004). It has also been shown that language-switching mechanisms are active with recently acquired linguistic knowledge gained in an artificial language learning setting. For example, Altmann, Kwan, and Goode (1995) trained subjects on two different artificial grammars and allowed them to later choose the grammar to be tested on. There was no evidence for the influence of the alternative grammar, which suggested that they had intentional control over application of linguistic knowledge even though both grammars consisted of the same items but with differing relationships between them.

When subjects returned for the second session they first listened to a stream of four words randomly organized. Performance on this test was better than chance with a mean percentage of 62%. Initially, the only cues subjects had to segment these streams were transitional probabilities, however it is possible that some subjects were able to use the transitional probabilities in conjunction with the novel segmentation cue as the relationship between the word boundaries and acoustic cue was learned. The performance on the word-learning test is comparable to adult performance in other studies using part-words such as Saffran et al. (1996) (M=65%), Perruchet and Desaulty (2008) (M= 60%), Toro, Sinnett, and Soto-Faraco (2005) (M=69%), and Toro-Soto, Rodriguez-Fornells, and Sebastien-Galles (2007) (M=70%). We might have expected higher performance than observed given that we used fewer words and higher within word transitional probabilities than most of the comparable artificial language learning studies mentioned. However, some of the vowel and consonant sounds used in this experiment have not been used in the majority of artificial language learning experiments and may be more difficult to learn.

The division of the subjects into two groups resulted in a high-performer group that showed significant learning of the words and another low performer group that did not show evidence of word learning. By focusing on the high performance group we should be able to better assess the ability to learn the novel cue from the statistics. However, the high performer group did not show any evidence for learning the novel cue when they were given a new test with a new set of words. One issue that may be raised concerning this assessment of cue transfer was that subjects did not actually apply the cue to the new set of words in the context of word segmentation. If the subjects successfully used the cue in a segmentation setting it would theoretically strengthen representations for words and make it more likely for a cue transfer effect to be observed during testing. However, creating a stream with repeating 3-syllable sequences that do not have transitional probabilities that are predictive of word boundaries presents a special challenge. In fact, earlier versions of the experiment included a stream that satisfied these constraints. Unfortunately, a side effect of these particular statistical structure constraints is that the stream contained a large proportion of syllables that did not belong to words, which may interfere with word learning. Additionally this stream was a total of over 40 minutes long, which would make it extremely difficult for subjects to maintain attentive listening. Because there was no evidence of segmentation and subjects reported extreme difficulty and displeasure with the stream, we removed the stream from the following versions of the experiment.

It is important to note that no previous experiment to our knowledge has shown learning of a novel segmentation cue from transitional probabilities. However, Thiessen and Saffran (2007) were able to train 9 month-old infants to successfully segment iambic words from a continuous stream, despite the fact that infants at this age show strong preference for segmenting words following a trochaic pattern even when it contradicts statistical cue. First, and most importantly the infants learned the new association between stress and word boundaries by listening to isolated exemplars. Learning from isolated words, as discussed earlier, may not be a realistic model for segmentation cue learning in natural speech. However, what is important to the discussion of the current experiment is that they were also able to teach 7-month olds, who have been shown to rely mostly on transitional probabilities, to use iambic stress patterns as a word segmentation cue. This shows that it is possible to teach new acoustic segmentation cues in the artificial language learning paradigm. However it is still not clear whether the observed effects are specific to infants. It is possible that infants are more adept at learning new segmentation cues, which could partly explain the infant's incredible language learning prowess relative to adults.

The acoustic cue was chosen because of the similarity to Mandarin tones (see Wang, Spence, Jongman, & Sereno, 1999), which we know are perceivable, rapidly processed and most importantly unfamiliar to most native English speakers. However because Mandarin tones are only used to distinguish lexical meaning, the viability of these pitch changes as segmentation cues is not know. Future investigations should strive to find an acoustic manipulation that is known to be a segmentation cue in a language other than English.

The results of the position preference tests may also provide a clue to why the novel cue wasn't learned. It was found that after familiarization the bias for preferring words with the cue on the initial syllable disappeared. If the listeners learned that the pitch change was not an informative segmentation cue and began to ignore it on some level, this would be problematic for learning the cue in following blocks. Follow-up studies may want to assess bias differently or emphasize that the language heard in the first session is different from the second session language.

Finally one possible complication has to do with the identification of the pitch changes as a distinct cue that is separate from the syllable on which it appears. The first speech stream was designed such that the transitional probabilities between syllables within a word was 1. This was done in hopes that statistical learning would be strongest; however, one issue this design raises is that the same 4 word final syllables will be repeated with the cue, and never without it. Therefore, the pitch changes may become linked to the syllables and perceived as an acoustic feature of the syllable rather than as a distinct acoustic change occurring on the syllable. Another side effect of the high transitional probabilities within words may be that the statistical cues are too strong and the listener may have no motivation to pay attention to the pitch changes during the learning stream. There is evidence that segmentation cues can compete and it is likely that if one cue provides a much more reliable method for segmentation other cues will not be relied on as heavily (Mattys, White, & Melhorn, 2005)

It is possible that statistical learning and the learning of structural regularities arise from two separate computational processes. A set of experiments by Peña, Bonatti, Nespor, and Mehler (2002) suggest that this may be the case. In their experiments they investigated the ability to learn simple AXC word patterns, that is if syllable A is presented, syllable C should follow after an intervening random syllable X. While participants were able to learn words by using the distant transitional probabilities between nonadjacent syllables, they did not generalize the pattern to new stimuli. However, in another experiment Peña et al. (2002) added a subtle 20 ms silence before each word in the stream. Participants were not consciously aware of these pauses but there was a dramatic shift in the results. This change in the speech stream allowed the listener to switch from computing transitional probabilities to analyzing structural regularities. During testing subjects were more likely to pick "rule words" (words that fit the AXC pattern but were not actually heard) over part-words (sequences heard during the stream but did not fit the pattern). Because the participants were only able to pickup on the structural generalization when they no longer had to calculate transitional probabilities to segment, this suggests that these are two distinct computation processes.

The competition stream and test results provided a novel finding in that it is one of the only artificial language learning designs to our knowledge that suggests that it is possible to use transitional probabilities to segment words from a stream of speech which contains what could be interpreted by listeners as words of variable lengths. Because the results of the cue transfer test suggest that the novel segmentation cue was not learned, we could argue that these cue words, which make up 40% of the syllables in the stream, were not segmented. Although only the high performer group was significantly higher than chance on selecting words, the overall subject average was marginally significant. This is an important finding given that transitional probabilities in natural speech may not always be predictive of word boundaries. However, the novel structure of the second stream did appear to have a deleterious effect on segmentation as seen in the drops in performance for both the overall group and the high performer group. This drop in performance makes sense given that the additional nonword material interrupts the continuous segmentation procedure that may arise from familiarity, as the listener is able to use the end of a word as a cue for the beginning of another word. There is also a potentially important difference between within-word transitional probabilities for the first and second stream. Although the stream was designed so that the transitional probability within words was 1 for both blocks it should be noted that this refers only to *forward* transitional probability and not *backward* transitional probability. *Backward* transitional probability can be defined as the probability of X given Y in an XY pair. So in the first stream both the backward and forward transitional probabilities between the second and third syllables of a word are 1 but in the second stream only the *forward* transitional probability is 1 while the *backward* transitional probability is .33. Perruchet and Desaulty (2008) have found that *backward* transitional probabilities are equally important as *forward* transitional probabilities. They created a speech stream in which *backward* transitional probabilities were the only cue for segmentation and found that performance on word and part-word discrimination (M=67%) was comparable to segmentation when *forward* transitional probabilities were the only cue (M=60%). This suggests that the change in *backward* transitional probabilities is at least partially responsible for the drop in performance, thus further reducing the effect of extra syllables on speech segmentation.

While the major objective of this study did not receive a conclusive finding, this study did bring to light a few important questions regarding artificial language learning paradigms and language learning in general that need to be addressed. For instance, a better understanding of the constraints on what type of acoustic properties of speech can be used for segmentation is in order. Additionally, a better understanding of developmental constraints on learning new segmentation cues needs to be better understood. Particularly a better understanding of the differences between adults and infants in segmentation cue learning may contribute to our knowledge of language acquisition and the difficulties of second language learning later in life. So far, Saffran and Thiessen (2007) is the only study to our knowledge that has shown learning of a new acoustic segmentation cue (or rather relearning the relationship between word boundaries and stress) and they only tested infants and used an acoustic cue that is known to be perceivable and usable to segment. As with many other aspects of learning, infants may be more adept at learning novel acoustic segmentation cues. It may be useful to take a step back and look at what new acoustic cues, if any, adults are able to learn from isolated words. Once that the learnability of an acoustic cue in isolation is established we can then better assess statistical bootstrapping.

Future studies should attempt to tease apart a few of the potential issues with the present study. For instance a follow-up study is currently underway to test the possibility that the difficulty with learning the novel acoustic cue was a result of listeners associating the novel cue with stress. The procedure of the follow-up study is nearly identical to the preceding study but syllables were replaced with synthesized tones that resembled notes being played on familiar musical instruments. The novel acoustic cue is again a modulated pitch change. Additionally to get a more definitive and possibly more sensitive assessment of any benefit of the novel acoustic cue during segmentation of the word learning stream, a follow-up study will be designed that compares word learning of participants listening to nearly identical streams with and without the novel acoustic cue.

FIGURES

Figure 1. Test results for all 32 subjects. Mean Proportion of words with final-syllable novel cue selected on the 2AFC test is plotted on the y-axis. Tests arranged in order in which they were taken. The position preference tests (cue and familiarized cue), word learning, cue transfer and competition tests are all shown. Dashed line indicates chance performance. The word-learning test was significantly different from chance and the competition test was near significance (p=.055).

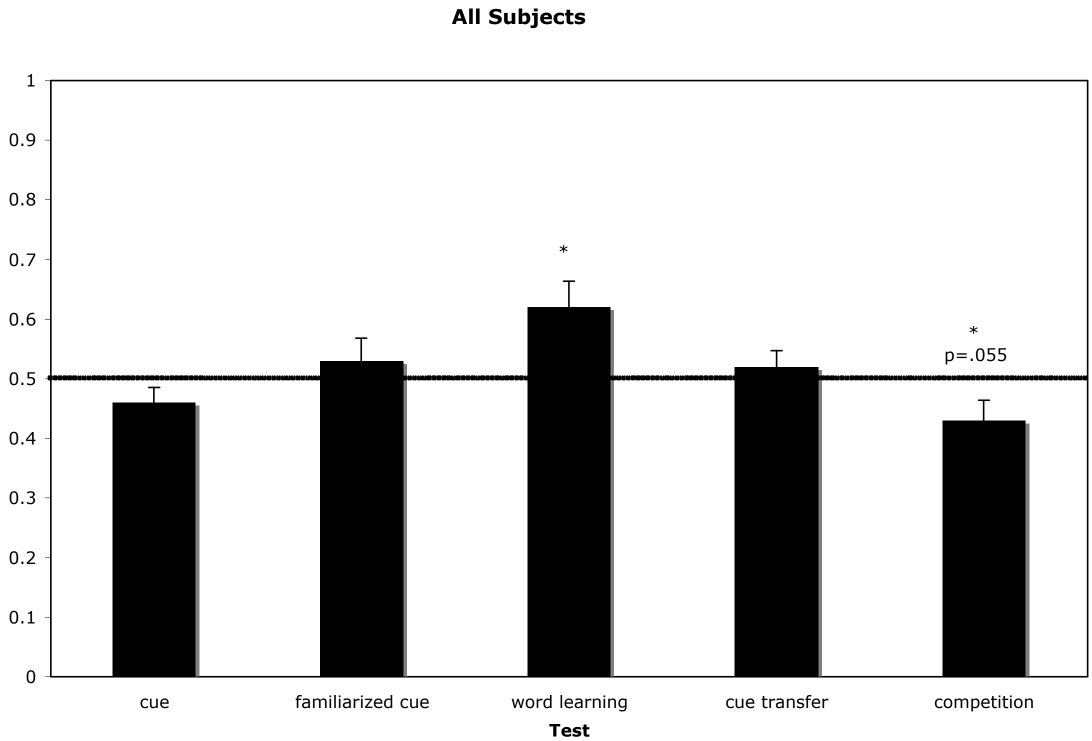


Figure 2. Test results for the 17 high-performers. Mean proportion of selected words with the novel cue on the last syllable is plotted on the y-axis. The dashed line represents chance performance. Both the word-learning and competition test are significantly different from chance. Performance on the cue transfer test does not differ from chance

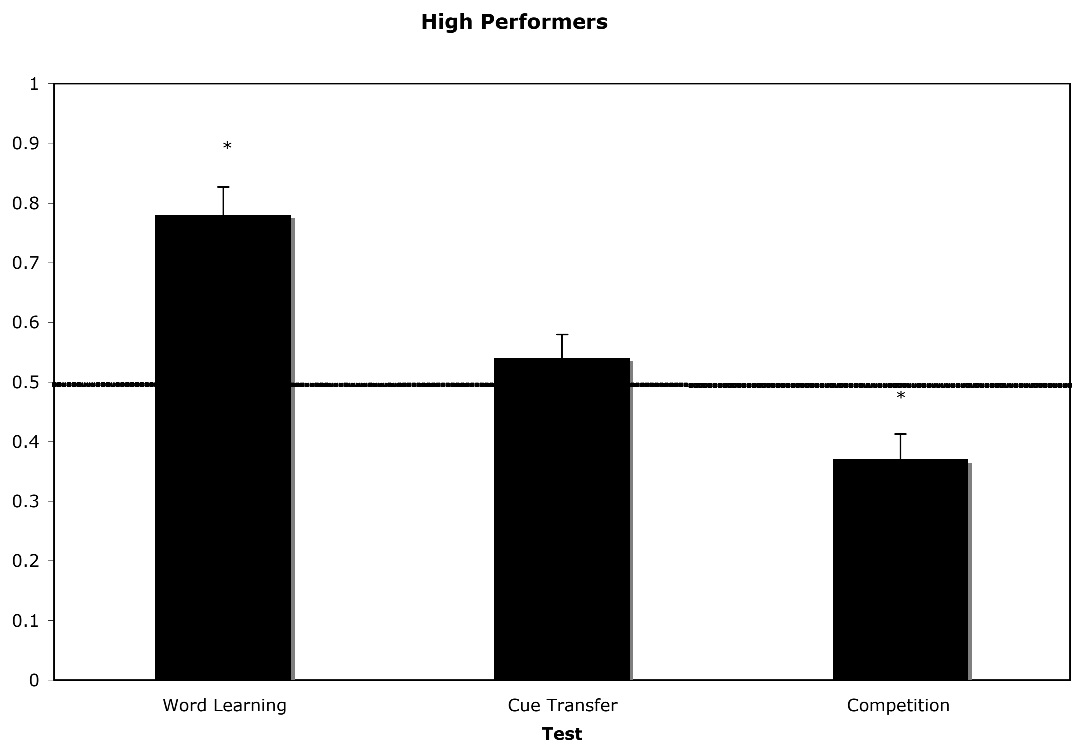


Figure 3. Individual results for the high performer group. Mean proportion of selected words with the novel cue on the last syllable is plotted on the y-axis. The dashed line represents chance performance. Each of the colored lines represents an individual's performance across the 5 tests.

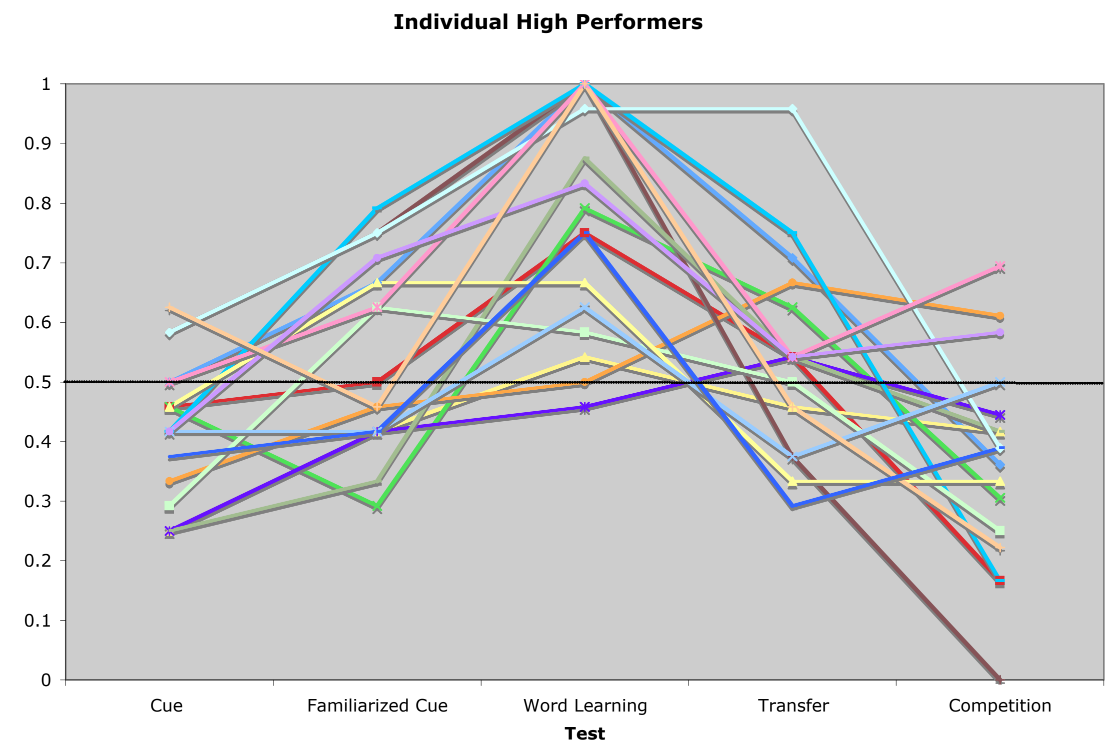
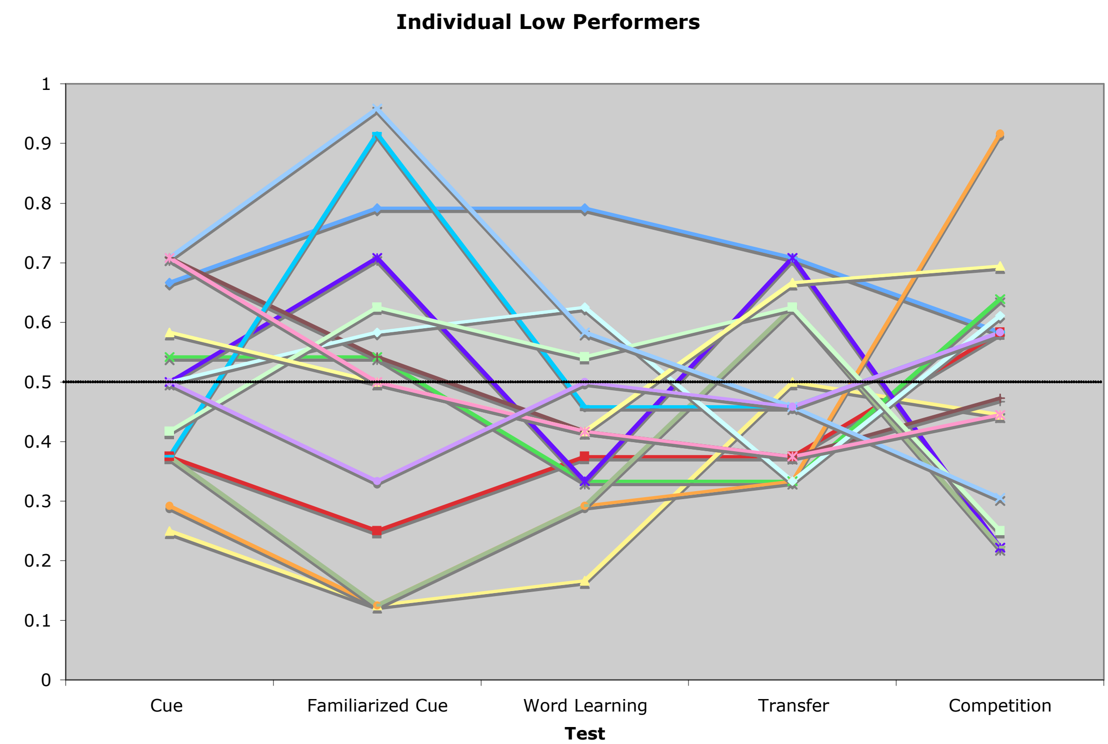


Figure 4. Individual results for the low performer group. Mean proportion of selected words with the novel cue on the last syllable is plotted on the y-axis. The dashed line represents chance performance. Each of the colored lines represents an individual's performance across the 5 tests.



REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20(6),* 952-964.

Aslin, R.N., Saffran, J.R., Newport, E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9(4),* 321-324.

Aslin, Woodward, LaMendola & Bever (1996). Models of word segmentation in fluent maternal speech to infants. In J.L. Morgan & K. Demuth (eds.), *Signal to Syntax* . Mahwah, NJ: LEA (pp. 117-134).

Bertoncin, J. & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior & Development, 4(3),* 247-260.

Biro, S., & Leslie, A.M. (2007). Infants' perception of goal-directed actions: development through cue-based bootstrapping. *Developmental Science, 10(3),* 379-398.

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

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33(2),* 111-153.

Costa, A. & Santesteban, M. (2004). Lexical access in bilngual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491-511.

Cutler, A., & Carter D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385-400.

Cutler, A., Mehler, J., Norris, D., Segui, J. (1983). A language-specific comprehension strategy. *Nature, 304(5922),* 159-160.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14(1),* 113-121.

Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33,* 824-844.

Dempster, F.N. (1981). Memory Span: sources of individual and developmental differences. *Psychological Bulletin, 89(1),* 63-100.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(5),* 1322-1338.

Echols, C.H., & Newport, E.L. (1992). The role of stress and position in determining first words. *Language Acquisition, 2(3),* 189-220.

Finn, A., & Hudson Kam, C.L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition, 108,* 477-499.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America. 27(4), 765-768.*

Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in human primate: statistical learning in cotton-top tamarins. *Cognition, 78,* B53-B64.

Johnson, E.K., Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language, 44,* 548-567.

Johnson, J.S., & Newport, E.L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

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

Jusczyk, P.W. & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics, 23,* 105-109

Jusczyk, P.W., Cutler, A.C., & Redanz, N.J. (1993). Infant's preference for the predominant stress patterns of English words. *Child Development, 64,* 675-687.

Katada, F. (1990) On the representation of moras: evidence from a language game. *Linguistic Inquiry, 21*, 641-646.

Kersten, A.W. & Earles, J.L. (2001). Less is really more for adults learning a miniature artificial language. *Journal of Memory and Language, 44,* 250-273.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. syllable rhythm. *Language and Speech, 51(4),* 343-359.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42.

Kubozono, H. (1989). The mora and syllable structure in Japanese: evidence from speech errors. *Language and Speech, 32,* 249-278.

Levy, B.J., McVeigh, N.D., Marful, A., & Anderson, M.C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18(1),* 29-34.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74(6),* 431- 461.

Marcus, G.F., Vijayan, S., Rao, S.B., Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science, 283(5398),* 77-80.

Mattys, S.L., White, L., & Melhorn, J.F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology, 134(4),* 477-500.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Mehler, J., Dommergues, J.Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298-305.

Nagai, Y., Hosoda, K., & Asada, M. (2003). Joint attention emerges through bootstrap learning. *Proceedings of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, 168-173

Newport, E.L. (1988). Constraints on learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Language Sciences, 10,* 147-172.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Newport, E.L. & Aslin, R.N. (2004). Learning at a distance: I. statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127-162.

Newport, E.L. & Supalla, T. (1990). A critical period effect in the acquisition of a primary language. Unpublished manuscript. University of Rochester, Rochester, NY.

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

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993) Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 258-278.

Peña, M., Bonatti, L., Nespor, M., Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298(5593),* 604-607.

Perruchet, P. & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition, 36(7), 1299-1305.*

Pinker, S. (1984). *Language Learnability and Language Development,* Cambridge, MA: Harvard University Press.

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.

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

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70(1),* 27-52.

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

Streeter, L.A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64(6), 1582-1592.*

Suomi, K., McQueen, J.M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36,* 422-444.

Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues 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: infant's acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3(1),* 75-102.

Toro, J.M., Sinett, S., & Soto-Franco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97, B25-B34.*

Toro-Soto, J.M., Rodríguez-Fornells, A., & Sebastián-Gallés, N. (2007). Stress placement and word segmentation by Spanish speakers.

Turk-Browne, N.B., Isola, P.J., Scholl, B.J., & Treat, T.A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(2),* 399-407.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, *38, 133-149.*

Wang, Y., Spence, M., Jongman, A., & Sereno. J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106(6),* 3649-3658.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119(1),* 597-607.

Yale, M.E., Messinger, D.S., Cobo-Lewis, A.B., & Delgado, C.F. (2003). The temporal coordination of early infant communication. *Developmental Psychology, 39(5),* 815-824.