

Specificity of Infant Statistical Word Segmentation

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Sara Parvanezadeh Esfahani

December 2022

Abstract

By their first birthday, infants develop sensitivity to language-general (e.g., transitional probability (TP) between syllables) and language-specific (e.g., lexical stress pattern) cues to identify word boundaries. Across 5 experiments, I investigated two questions: 1) how do English learning 8-month-old infants segment high TP (TP=1; HTP) words in non-native languages that have a different rhythm and prosody? 2) how do English learning 8-month-old infants represent nascent word representations across indexical (i.e., talker's voice), segmental (i.e., onset consonant) and suprasegmental (i.e., stress pattern) information? To that end, in each experiment, using HTPP, I familiarized infants with a familiarization corpus with two embedded HTP words and at test, infants were evaluated on their ability to discriminate between the target words versus foils (either low TP or Novel words). I found that while the redundant language-specific cues can facilitate infants' ability to identify word boundaries, a different language-specific cue can create a competition in infants' attention to statistically segmented words versus words with familiar language-specific cues. For the second question, I found that English learning infants can generalize nascent word representation across indexical and suprasegmental information but not across segmental information that is essential to word meaning. Together, the findings suggest that while English-learning infants may need redundant information from stress cues to segment nonnative speech, they do not necessarily encode stress information in nascent word representations as they tend to only store information that is more readily relevant to word meaning (i.e., segmental information).

Table of Contents

Chapter 1: Introduction.....	1
Word Segmentation Accounts	2
Word Representations	8
Chapter 2: Infant Statistical Word Segmentation in Languages with Different Rhythm and Prosody	14
Abstract	15
Introduction.....	16
Experiment 1	20
Method	21
Participants	21
Stimuli	21
Procedure.....	22
Results and Discussion.....	23
Experiment 2	24
Method	25
Participants	25
Stimuli	26
Procedure.....	27
Results and Discussion.....	27
Discussion	28

Chapter 3: Specificity of Infant Speech Segmentation.....	33
Abstract.....	34
Introduction.....	35
Experiment 1	38
Method	41
Participants.....	41
Stimuli	42
Procedure.....	43
Results and Discussion.....	44
Experiment 2	46
Methods.....	48
Participants	48
Stimuli	48
Procedure.....	49
Results and Discussion.....	49
Experiment 3	50
Methods.....	53
Participants	53
Stimuli	53
Procedure.....	54
Results and Discussion.....	54
Discussion	55

	v
Chapter 4: General Discussion	64
Study Limitations and Future Research.....	69
Conclusion	70
References	72
Appendix	82
Vita	98

List of Tables

Table 1. Words used in familiarization and testing phases of Chapter 2 Experiments 1 and 2. ...	88
Table 2. Acoustic characteristics of syllables of tokens in the familiarization languages of Experiment 1 of Chapter 2.	89
Table 3. Acoustic characteristics of syllables of test tokens in Experiment 1 of Chapter 2.....	90
Table 4. Acoustic characteristics of syllables of tokens in the familiarization language of Experiment 2 of Chapter 2.	91
Table 5. Acoustic characteristics of syllables of test tokens in Experiment 2 of Chapter 2.....	92
Table 6. Words used in familiarization and testing phases of Experiments 1, 2, and 3 (Chapter 3)	93
Table 7. Mean (SD) Pitch (Fundamental Frequency, F0) and Mean Word Length (SD) for each word in Languages 1A/B and Test words used in Experiment 1.	94

List of Figures

Figure 1. Mean looking time (sec) to HTP and Novel words in Experiment 1 (Italian) and Experiment 2 (Persian; Language A and Persian; Language B), Chapter 2.....	95
Figure 2. Mean looking time (sec) to HTP vs LTP (Exp 1) or vs. Novel (Exp 2 & 3) words (Chapter 3). Error bars represent SEM.	96
Figure 3. Forest plot providing a visualization of effect size comparisons between the current studies and Pelucchi et al.'s (2009a) experiments from which our familiarization languages were drawn (grouped according to familiarization languages used). Each experiment is represented by a line in the plot. TE = treatment effect (effect size); seTE = estimated treatment effect; CI = confidence interval; diamond = overall effect estimate; width of diamond = CI for overall effect estimate; point/box size = inverse variance weight based on effect size SE; width of line = CI for effect estimate for each study—lines are plotted in black if the CI is larger than that point/box size and in white if the CI falls within the point/box. The Forest Plot was the output of a mini meta-analysis performed in R (R Core Team, 2013), using a MetaLab script provided by Bergmann and colleagues (Bergmann, Tsuji, Piccinini, Lewis, Braginsky, Cristia, & Frank, 2018) and used previously by Karaman & Hay (2018).	97

Chapter 1: Introduction

To learn language, infants need to build a lexicon, and in order to build a lexicon, they have to overcome multiple challenges. One of the earliest challenges infants encounter is referred to as the word segmentation problem, which poses the question of how infants locate word boundaries in fluent speech. This is a particularly challenging problem for young infants who do not yet know any words in their native language(s) as even in infant-directed speech, words do not occur in isolation (Brent & Siskind, 2001). Further complicating the problem is that word boundaries are not marked by acoustic cues like pauses after each word (Cole & Jakimic, 1980). To locate word boundaries and extract words from continuous speech, infants use different sources of information present in the language. When infants' experience with language is still limited, they tend to use language-general cues (e.g., transitional probabilities between syllables; Saffran et al., 1996; Pelucchi et al., 2009a; 2009b; Thiessen & Saffran, 2003), and as they gain more experience and learn about language-specific cues (e.g., language rhythm, lexical stress, etc.) present in their native language, they start to incorporate those cues into their word segmentation routine. Another challenge infants face in developing a lexicon is the ability to store the word forms they segment from continuous speech in order to be able to use them later as object labels.

In this two-prong dissertation, across 5 experiments, I will dive into how English learning 8-month-old infants use statistical information to both extract and represent words found in non-native natural language input. In Chapter 2, I conduct two experiments to investigate infants' statistical segmentation in languages with different rhythm and stress patterns. In Chapter 3,

across 3 experiments, I look at the status of infants' representations of recently segmented words across indexical, segmental, and suprasegmental information. In general discussion of the dissertation, I will aggregate the findings and discuss explanations and future directions.

Word Segmentation Accounts

Unlike written language, spoken speech is not marked by clear pauses between words and identifying word boundaries pose a challenge for young infants. To overcome this challenge, there are certain cues in the language input that infants learn to use to parse continuous speech. By their first birthday, infants demonstrate sensitivity to language-general (e.g., Saffran et al., 1996; Aslin et al., 1998) as well as language-specific cues (e.g., Johnson & Seidl, 2009; Jusczyk et al., 1993; Jusczyk et al., 1999; Hay & Saffran, 2012) in word segmentation. In this section I list four different word segmentation accounts based on both language-general and language-specific cues present in natural languages.

Regardless of the language, natural speech has a great number of regularities, and there is ample evidence showing that infants are remarkable at finding regularities that can facilitate word segmentation. For example, young infants who do not yet know much about their native language can leverage language-general cues in the input, such as the transitional probability (TP) between syllables, to start finding words in fluent speech (Saffran et al., 1996). Transitional probability (TP) between syllables, computed as the frequency of XY (a syllable sequence) given the frequency of the syllable X (Swingley, 2005), is generally higher within words and lower across word boundaries. Impressively, using an artificial language that contained no cues to word boundaries other than differences in transitional probability (TP) within versus between words, Saffran and colleagues (e.g., Saffran et al., 1996; Aslin et al., 1998) found that infants can track this type of co-occurrence statistic in order to locate words in fluent speech. These findings have

been replicated numerous times across many different labs (for a review, see Cannistraci et al., 2020; Krogh et al., 2013; Romberg & Saffran, 2010)

One critique of artificial language studies is that the languages themselves are artificially simplistic and do not represent the complexity of natural language input. In order to address this concern, Pelucchi and colleagues (Pelucchi et al., 2009a) set out to test statistical learning in complex natural language input. They familiarized 8-month-old English-learning infants with naturally produced Italian sentences that had four embedded trochaic target words. Two of the words were high TP (HTP) because their syllables never occurred anywhere else in the corpus (TP = 1.0). Two of the words were low TP (LTP) because their first syllable occurred in many other words throughout the corpus (TP = .33). Following familiarization, infants were tested using the head-turn preference procedure on their ability to discriminate HTP words from LTP words. Infants preferred listening to the relatively familiar HTP words, even though both HTP and LTP words were heard an equal number of times in the speech stream, suggesting that 8-month-olds can also use co-occurrence statistics to extract words from natural language input (Pelucchi et al., 2009a).

As infants gain more experience with their native language, their word segmentation ability is more and more impacted by their native language's specific characteristics. For example, English learning 8-month-old infants can segment bisyllabic words that are stressed on the initial syllable (i.e., trochaic words) such as *hamlet* and *kingdom* but fail to segment bisyllabic words that are stressed on the final syllable (i.e., iambic words) such as *guitar* and *beret* (Jusczyk et al., 1999). A corpus analysis by Cutler and Carter (1987) demonstrates that 90% of the English content words are trochaic (i.e., stressed on the initial syllables). Therefore, English learning infants develop a bias to treat stressed syllables as word onset to segment

speech. This segmentation strategy has been referred to as Metrical Segmentation Strategy (Cutler & Norris, 1988).

There has been a fair amount of debate in the literature about whether infants rely more heavily on statistical cues versus these types of prosodic or metrical cues to word boundaries. In one study, Johnson and Jusczyk (2001) pitted stress cues against statistical cues (i.e., transitional probability between consecutive syllables and stress pattern of the words indicated conflicting cues to word boundaries). They showed that 8-month-old infants rely more heavily on stress cues to find the word boundaries, suggesting that stress cues carry more weight than statistical cues during word segmentation. In another study using artificial language materials, Thiessen and Saffran (2003) familiarized English-learning infants with either a trochaic or an iambic speech stream. In the trochaic language, the initial syllables of the statistically defined bisyllabic words were stressed. Thus, for these English-learning infants, the stress pattern and statistical cues provided consistent cues to word boundaries. However, in the iambic language, the second (or final) syllables of statistically defined bisyllabic words were stressed, leading to stress patterns and statistical cues that provided conflicting information about word boundaries. Both 7- and 9-month-old English learning monolingual infants successfully segmented words from trochaic language (i.e., distinguished words from part-words at test). However, in the iambic language condition, while 7-month-old infants performed similarly to previous groups and continued to rely on statistical cues to extract words, 9-month-olds paid greater attention to stress cues and mis-segmented words (i.e., they considered the stressed syllable as the onset of the words pulling out part-words instead of words). Thiessen and Saffran (2007) demonstrated that a brief exposure to iambic words results in the successful segmentation of iambs (i.e., iambic words) even in 7-

month-old infants. These results indicate the significance of linguistic experience in learning language-specific cues such as stress in identifying word boundaries (Hay & Saffran, 2012).

Another language-specific account that has been used to account for infants' word segmentation ability is language rhythm (e.g., Cutler et al., 1992). Languages are categorized into three rhythmic classes: stress-timed languages (e.g., English, Dutch), syllable-timed languages (e.g., Italian, French, Persian), and mora-timed languages (e.g., Japanese). Each language is classified into these categories based on the efficiency in the processing of the languages. For example, in English the majority of content words are stressed at the onset syllable, and a segmentation strategy based on stressed onset syllables works very well. However, for a language like French that does not use stress at the word level, there should be another unit of segmentation (i.e., syllables). Why are syllables not an efficient unit of processing in English? English adult native speaker ratings have shown that some syllable boundaries are ambiguous in English (Cutler et al., 1986). For instance, when a single consonant happens between a stressed vowel and an unstressed vowel, (e.g., /l/ in *balance*), such consonants are ambisyllabic (Cutler et al., 1986) which means that it can belong to both syllables. This syllable-based segmentation account makes English processing inefficient relative to stress-based segmentation. On the contrary, in French, which is syllable-timed, native speaker ratings demonstrate that syllable boundaries are unambiguous for the most part, making syllable-based segmentation an efficient processing strategy.

Cutler and colleagues suggested that language rhythm is more like a continuum where stress-timed languages with a predominant stress pattern (e.g., English) fall on one end and syllable-timed languages with no word-level stress (e.g., French, Persian) fall on the other end.

Languages such as Italian and Spanish that are syllable-timed but also have word-level stress fall somewhere in between (Cutler et al., 1992).

In infant word segmentation literature, there is evidence showing that infants can segment nonnative speech that is in the same rhythm class of their native language. For example, English and Dutch are both stress-timed languages with a predominant trochaic stress pattern. Houston and colleagues (Houston et al., 2000) tested both English learning and Dutch learning 9-month-old infants with Dutch natural speech. Infants in both groups could segment Dutch. However, English learning and French learning 8-month-olds could not segment the other language (Polka & Sundara, 2012), suggesting that infants' rhythm-based segmentation ability does not transfer to a nonnative language that belongs to other rhythm categories than that of infant's native language.

Despite these findings, previous research suggests that English learning 8-month-old infants can segment Italian trochaic words (Pelucchi et al., 2009; 2009b) and Spanish trochaic and iambic words (Sundara & Mateu, 2018). Italian and Spanish are both syllable-timed and if infants rely on a rhythm-based segmentation strategy, these English learning infants should have failed segmenting Italian and Spanish. It is worth noting, however, that although Italian and Spanish are syllable-timed, both languages have word-level stress. Italian bisyllabic content words are predominantly trochaic and Spanish bisyllabic words are split between trochaic and iambic word stress, with many multisyllabic words having penultimate stress. Thus, going back to the language rhythm continuum, Italian and Spanish fall somewhere in between and probably closer to English than a language like French that is syllable-timed but does not use word-level stress. It is also noteworthy that in both Italian and Spanish corpora used in these studies, transitional probabilities between syllables were also manipulated to have a perfect TP (TP=1;

Italian; Pelucchi et al., 2009a; 2009b), (TP=1; Spanish trochaic words; Sundara & Mateu., 2018) or a relatively high TP (TP=.75; Spanish iambic words; Sundara & Mateu, 2018), and incorporating language-general and language-specific cues may have helped infants to segment words.

Given these findings, it would be interesting to see how English learning infants would perform word segmentation in a language on the end of the continuum, a language that is essentially too unfamiliar and does not share any language-specific cues with English. It would be especially interesting to see if infants' statistical segmentation ability can override any language-specific cues that are either in conflict or absent in the nonnative speech. Previous work shows that English learning 8-month-old infants cannot segment French (e.g., Polka & Sundara 2012; Sundara & Mateu, 2018). However, in these corpora transitional probability between the target words was probabilistic (i.e., TP=.79, not TP=1), and extracting statistically-defined words if the TP is less than 1 in the continuous speech is much harder for infants. So, it remains unclear if infant could segment neutrally stressed French target words if the TP between syllables was perfect (i.e., TP=1).

In Chapter 2 of this dissertation, I address this knowledge gap and investigate English learning 8-month-old infants' ability to segment statistically-defined words (TP=1) in a language that is syllable-timed and does not use stress at word level. To attain this aim, I use Persian, a language that is very similar to French and thus we can compare the findings with those of previous work on French. Persian bisyllabic words, like French, are neutrally stressed unless they appear at the end of a phrase, in which case they take final syllable stress (Rahmani et al., 2015). Persian bisyllabic words, however, when produced in isolation, are iambic. Thus, in Chapter 2, after replicating previous findings by Pelucchi and colleagues (Pelucchi et al., 2009a;

Experiment 1) in our lab setting using the same language material and procedure, I familiarize infants with a naturally produced Persian language where two neutrally stressed target words are embedded with high TP (TP=1; HTP) and test infants with their ability to discriminate the HTP words from Novel words. Since all test words are produced in isolation, they should all have a subtle iambic stress pattern. Thus, if infants do not need the redundant information from stress cues to track TP between syllables, they should be able to segment Persian.

Word Representations

Word representations refer to the information that one knows about a word, including which sounds make up a word (phonology), what a word means (semantics) and how words are used in sentences (morpho-syntax). From the phonological perspective, words can be represented at three different levels: indexical (i.e., properties of the speaker's voice), segmental (i.e., phonotactic rules of language), and suprasegmental (i.e., prosody of language). At the indexical level, it is adaptive for infants to learn to ignore variation in words because whoever says a given word does not modify its meaning. Segmental changes are lexically contrastive across the world's languages (e.g., cat→bat) and thus, it is adaptive to learn to attend to variation at the segmental level. While suprasegmental information plays a vital role in word meaning in some languages (e.g., tone languages where varying the pitch contour can alter the word's meaning), not all languages use suprasegmental information this way. Indeed, English-learning infants sometimes need to pay attention to suprasegmental cues (e.g., to differentiate between the noun record /'rekərd/ and the verb record /rə'kôrd/) and sometimes need to ignore them (e.g., when someone makes a pronunciation error). Although the fact that people rarely make stress errors during speech production (Cutler & Isard, 1980) suggests that suprasegmental information is represented on some level, it is unclear the extent to which it is adaptive for English-learning

infants to represent suprasegmental information at the word level as stress does not typically differentiate word meanings.

A significant body of research suggests that infants home in on the relevant features of their native language relatively early on (see Werker & Gervain, 2013 for a review; Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Werker & Tees, 1984) and provide a basis for their word representations of familiar words, but much less is known about what infants encode about words when they first encounter them in fluent speech. Despite our knowledge of infants' statistical learning ability and the development of the word representation of familiar words, we know very little about how novel words, newly segmented from continuous speech, are represented. There is evidence that later in development, these recently segmented novel words with strong statistics (i.e., $TP=1.0$), which have no meaning or reference associations, can make better object label candidates than those with weaker statistics (e.g., $TP=.33$) (Graf Estes et al., 2007; Hay et al., 2011). However, it is not clear how these nascent word forms are represented at different phonological levels. In Chapter 3, I aim to investigate how newly segmented words are represented at 3 levels: indexical (i.e., talker's voice), segmental (i.e., consonant change), and suprasegmental (i.e., stress pattern).

Phonologically, words can be represented at the indexical, segmental, and suprasegmental levels. Although 7.5-month-olds appear to have a difficult time recognizing target words when some features of the voice, such as gender (Houston & Jusczyk, 2000), affect (happy vs neutral; Singh et al., 2004, and pitch (Singh et al., 2008) are changed, as infants gain experience with their language, they should come to ignore features that do not alter the meaning of a word (e.g., 'cat' produced by a male and 'cat' produced by a female voice mean the same

thing). Indeed, 11-month-old infants can recognize words recently segmented from an artificial language when the speaker changed from familiarization to test (Graf Estes, 2012).

At the segmental level, however, which sounds make up a word is essential to word meaning – a change in either a consonant or a vowel can change the meaning of the entire word. From birth, infants show a remarkable ability to discriminate many, if not most, of the sounds from across the world's languages (see Werker & Tees., 1984 for a review; Aslin et al., 1981; Eimas et al., 1971; Werker & Gervain, 2013). However, between 6 to 10 months as they gain more experience, they attenuate this broad sensitivity to their native language sounds (for a review, see Maurer & Werker, 2014), and demonstrate improved ability to discriminate some difficult native contrasts (Kuhl et al., 2005). Given this early perceptual narrowing, we might expect that even young infants would be sensitive to changes at the segmental level. Because 12 months is often thought of as the onset of word learning (although see Burgelson & Swingley, 2012), we do not know very much about the segmental specificity of young infants' early lexical representation. Interestingly, work by Werker and colleagues suggest that at 14 months, infants ignore segmental changes in minimal pair novel word learning (Stager & Werker, 1997). However, if the words are familiar (Fennell & Werker, 2003), 14-month-olds appear to be sensitive to these segmental features of words. Future work is needed to determine the degree of segmental specificity of infants' early lexical representations.

In contrast to indexical features, which infants should learn to ignore, and segmental features, which infants should learn to attend to, suprasegmental features fall in a gray zone for infants learning intonation languages, such as English. Suprasegmental features are prosodic cues such as lexical tone, pitch contour or stress pattern. In tonal languages (e.g., Mandarin Chinese) lexical tones or pitch contours are used to contrast meaning. However, in intonation

languages (e.g., English), stress pattern plays an important role in the speech signal. Stressed syllables typically have a longer duration, and higher pitch and amplitude (Hayes, 1995).

Although changing the stress pattern of words in English can sometimes change their meaning (e.g., record /'rekərd/ as a noun and record /rə'kôrd/ as a verb), by and large, suprasegmental features are not relevant to word meaning in English.

Nonetheless, from birth, infants begin to respond to native language prosodic information (Mehler et al., 1988), and by two months, they are sensitive to stress pattern changes (Jusczyk & Thompson, 1978). While 9-month-olds prefer listening to words that are consistent with the dominant stress pattern in their native language, neither 6-month-olds (Jusczyk et al., 1993), nor 7-month-olds (Echols et al., 1997) have shown such a preference, suggesting that perceptual narrowing of prosodic features is the result of more experience with language happening sometime between 6 to 9 months of age.

Research on whether stress information is encoded in lexical representations or whether it is stored separately during infancy does not yield conclusive findings. One camp of researchers (e.g., Halle, 1998) does not place stress as part of the lexical representation. They argue that lexical stress is stored separately and assigned according to a set of rules. In support of this account, Vihman and colleagues (Vihman et al., 2004) explored the role of prosodic and segmental features of lexical representation in infants. They tested both 9- and 11-month-olds on their ability to differentiate familiar words (e.g., baby) from more rare words (e.g., bridle). Across conditions, words were either produced correctly or they were mispronounced at the segmental level (e.g., bunny → vunny) or at the suprasegmental level (e.g., BAby → baBY). Nine-month-olds failed to differentiate familiar from novel words even when the words were correctly pronounced, suggesting rather weak word representation at this age. In contrast, 11-

month-olds listened longer to familiar words than rare words both when the words were correctly pronounced and when there was a change in stress pattern. However, they failed to differentiate familiar words from rare words when there was a segmental change. These findings suggest that by 11 months, infants have a relatively robust mental representation of familiar words at the segmental level that may allow infants to disregard stress change (Vihman et al., 2004).

In contrast, Cutler and colleagues (Cutler, 1979; Cutler & Isard, 1980) argue that stress information is integrated into word representations for familiar words. They support their account by providing evidence from word stress errors. Cutler and Isard (1980) discuss that stress errors occur because same stem words are stored together, and when people access their lexicon, confusions cause pronunciation errors. However, they could not provide similar evidence for unfamiliar words (Cutler & Isard, 1980). Thus, there remains ambiguity about whether and when in the process of learning words, stress information becomes integrated into the representation.

In line with findings of Cutler, Curtin and colleagues (Curtin et al., 2005) found that English-learning 7- and 9-month-olds are sensitive to stress cues during speech segmentation. Specifically, after listening to a naturally produced artificial language that contained both stressed and unstressed versions of the same syllables, English-learning infants preferentially segmented words that had initial stress (i.e., SWW) over words that had medial (i.e., WSW) or final stress (i.e., WWS). After demonstrating that stress information can be used to posit word boundaries, Curtin and colleagues (Curtin et al., 2005) explored whether stress information is represented in recently encountered words. They familiarized a group of 7-month-olds with the same artificial familiarization language but tested them with English sentences that contained target words from previous experiments that were stressed either on the initial syllable or the

middle syllable (e.g., ‘*I like your DObita option*’) and control words that did not occur in the familiarization. Infants showed sensitivity and listened significantly longer to the sentences containing initially stressed words extracted from continuous speech, suggesting that infants preferred a certain parsing strategy (trochaic bias). Furthermore, infants maintained stress information in lexical representation after word segmentation. These findings suggest that stressed syllables are represented differently than their unstressed counterparts. Further, these findings suggest that this difference facilitates infants’ ability to use transitional probability information to segment words from a fluent speech stream by helping infants identify word boundaries. Together, this work suggests that infants may represent stress information even in newly encountered sound sequences (Curtin et al., 2005).

Although we now have a better understanding of infants’ ability to perceive these various types of acoustic information (see Chapter 3 for more in-depth review) and a significant body of research suggests that infants home in on the relevant features of their native language relatively early on (see Werker & Gervain, 2013 for a review; Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Werker & Tees, 1984) forming the basis for their lexical representations of familiar words, much less is known about what infants encode about words when they first encounter them in fluent speech. Thus, the goal of Chapter 3, presented across three experiments, is to address this gap in knowledge by investigating indexical, segmental, and suprasegmental specificity of newly encountered, statistically-defined words.

Chapter 2: Infant Statistical Word Segmentation in Languages with Different Rhythm and Prosody

Abstract

Well before their first birthday, infants demonstrate sensitivity to language-general (e.g., transitional probability (TP) between syllables) and language-specific cues (e.g., stress pattern of their native language) during word segmentation. In natural languages, language-general and language-specific cues often provide redundant information regarding word boundaries. To understand whether English-learning 8-month-old infants need the redundant cues to segment a nonnative speech, we test whether infants are able to use TPs to extract words from complex natural language input if redundant language-specific cues to word boundaries are not present. In two Experiments we tested infants' ability to statistically segment either a natural language that provided English-learning infants with redundant cues to word boundaries (i.e., Italian; Experiment 1) or one that did not (i.e., Persian; Experiment 2). In Experiment 1, infants statistically segmented Italian trochaic words, replicating previous findings. In Experiment 2, infants showed a preference for words that were produced with a stress pattern more similar to their native language, suggesting that infants benefit from redundant information between TP information and stress cues in speech segmentation and fail to segment an unfamiliar natural speech with no redundant cues to word boundaries.

Introduction

Finding words in continuous speech is one of the many challenges young infants face when learning their native language. Infants appear to overcome this problem by relying on both language-general and language-specific cues to identify word boundaries. One of the most studied language-general cues is between syllable transitional probability (TP), which is computed as the frequency of XY (a syllable sequence) given the frequency of the syllable X (Swingley, 2005). Importantly, between-syllable TPs tend to be higher within words and lower across word boundaries. There is ample evidence in support of infants' use of transitional probability to segment both artificial (e.g., Saffran et al., 1996; Aslin, et al., 1998) and natural (e.g., Pelucchi et al., 2009a; 2009b) languages from an early age. A statistical segmentation account predicts that infants can successfully segment unfamiliar languages if they are provided with sufficient information about the co-occurrence statistics between syllables (for a review, see Cannistraci et al., 2020; Krogh et al., 2013; Romberg & Saffran, 2010).

As infants gain more experience with language, their segmentation ability is modulated by the characteristics of their native language (e.g., Thiessen & Saffran, 2003). Work by Jusczyk and colleagues (Jusczyk et al., 1999) demonstrated that while English-learning 8-month-old infants can successfully segment English bisyllabic words that have initial syllable stress (i.e., trochaic or strong/weak stress pattern), they cannot do so when the final syllable of the word is stressed (i.e., iambic or weak/strong stress pattern). This finding is interesting as 90% of English content words are stressed on their first syllables (Cutler & Carter, 1987). These findings suggest that infants develop a bias toward their native language prosodic features.

In addition to developing sensitivities to the prosodic features of their native language, infants also develop a sensitivity to the rhythmic patterns of their native language. Generally,

languages are classified under one of these three main rhythm categories, stress-timed (e.g., English, Dutch, German), syllable-timed (e.g., French, Italian, Persian, Spanish) and mora-timed (e.g., Japanese). Previous research has shown that infants are better able to segment languages that share the same rhythmic pattern as their native language. For example, both English and Dutch are stress-timed and have a predominant trochaic stress pattern (Houston et al., 2000), and both monolingual English learning and Dutch learning 9-month-old infants can segment Dutch trochaic words from fluent speech. Conversely, when languages do not share the same rhythmic class, segmentation is sometimes hindered. For example, although Canadian French (syllable-timed) learning and Canadian English (stress-timed) learning infants successfully segmented bisyllabic words in their respective native language, they failed to segment words in the other language (Polka & Sundara, 2012). Together, these findings suggest that overlapping rhythmic properties may help infants to segment non-native languages.

Despite these findings, it appears that when infants are provided with sufficiently robust cues, they can overcome the rhythmic differences to segment non-native languages with different rhythmic patterns. For example, although English (stress-timed) and Italian (syllable-timed) do not share the same rhythmic pattern, work by Pelucchi and colleagues (Pelucchi et al., 2009a; 2009b) has demonstrated that monolingual English learning 8-month-old infants can successfully segment words from an Italian corpus if those words contain other cues to word boundaries. Specifically, Pelucchi and colleagues (Pelucchi et al., 2009a) designed the language in which they inserted two high TP (HTP; TP=1) bisyllabic target words. Importantly, in addition to strong internal co-occurrence statistics, the target words were always produced with first syllable stress, thus providing English-learning infants with redundant cues to word boundaries. Across a number of experiments, following familiarization with the languages, infants listened longer to

familiar HTP words compared to foils (Pelucchi et al., 2009a). Similarly, work by Sundara & Mateu (2018) shows that English learning 8-month-old infants can segment both trochaic and iambic target words in Spanish after being familiarized with Spanish corpora that either had a high TP (TP=1) for trochaic words or a relatively high TP (TP=0.75) for iambic words. Together these findings suggest that infants can use TP information to segment words from a non-native language even if the language does not belong to the same rhythmic class as their native language.

Languages fall on a rhythm continuum, where stress-timed languages that have a predominant stress pattern (e.g., English, Dutch) fall on one end, and syllable-timed languages with neutral word-level stress (e.g., French, Persian) fall on the other end. Languages that have some shared rhythmic and prosodic characteristics with these two ends fall somewhere in between (Cutler et al., 1992). For example, Italian and Spanish are both syllable-timed but they also have word-level stress. In fact, Italian trochaic stress structure is very similar to English – the stressed syllable is longer than the unstressed syllable, causing a vowel reduction in the unstressed syllable that facilitates stress perception (White et al., 2009). Thus, the presence of statistical information as well as prosodic cues may have facilitated English learning 8-month-old infants' success in segmenting target words in both Italian (Pelucchi et al., 2009) and Spanish (Sundara & Mateu, 2018). This is interesting because English learning 8-month-olds cannot segment target words from French continuous speech after either 30 seconds (Polka & Sundara, 2012) or 60 second accumulated familiarization time (Sundara & Mateu, 2018). French words take a neutral stress structure in fluent speech unless they appear at the end of a phrase. In this case, they take an iambic stress pattern, similar to when they are produced in isolation. Thus, infants have to extract neutrally-stressed target words and generalize them to iambic words at

test. In these studies, the TP between syllables of the French target words was only 0.79 with a standard deviation of 0.21, making TP between the syllables of some of the target words as low as 0.58. Given that there is no evidence of infants' ability to track TP information in the laboratory when the TP is less than 1, and ample evidence showing that English learning 8-month-olds can segment artificial languages with only statistical cues (i.e., transitional probability between syllables) and no stress information (e.g., Saffran et al., 1996), it is unclear whether English learning infants would still fail to segment French target words if the TP between syllables of the target words was consistently 1, similar to Pelucchi's study.

It is important to note that language general cues, such as TP information, and language-specific cues often provide redundant information regarding word boundaries. And indeed, there is ample evidence that infants may come to rely on the redundant information provided by language-general cues such as TP between syllables to discover language-specific cues (e.g., Sahni et al., 2010; Thiessen & Saffran, 2003). There is also evidence that sometimes the redundant cues do not even need to be consistent with infants' native language and as long as they are present, infants can utilize them to segment speech stream. For instance, English learning infants can segment Spanish iambic words (Sundara & Mateu, 2018), suggesting that infants' ability to segment a nonnative language with the presence of the word-level stress even if the stress pattern is not consistent with the dominant stress pattern of their native language (i.e., English trochaic stress pattern).

In the current study, we aim to examine whether English learning 8-month-old infants can segment target words in a language that does not share any language-specific cues with English (i.e., no redundant information is present) and the only cue to word boundaries is transitional probability between syllables. To that end, in Experiment 1, we set out to first

replicate previous findings by Pelucchi and colleagues (2009) that English learning 8-month-old infants can use TP information to segment high TP words from a non-native Italian corpus.

Recall, Italian is rhythmically different from English but has similar lexical stress patterns. For Experiment 2, which followed the same familiarization and testing procedures as Experiment 1, we designed and developed a novel Persian (Farsi) familiarization corpus. Like French and Italian, Persian is syllable-timed. Like French, but unlike Italian, Persian does not implement stress at word level; words can only carry stress on the final syllable if they occur at the end of a phrase, similar to when they are produced in isolation (Rahmani et al., 2015). Thus, Persian is a good language to test the question of whether English-learning infants are still able to use TP information to segment a complex natural language that lacks overlapping language-specific features.

Experiment 1

In Experiment 1, using the material from the Pelucchi and colleagues (Pelucchi et al., 2009a; Experiment 1), we examine English learning 8-month-olds' ability to segment target words that are characterized by high transitional probability ($TP=1$; HTP) by discriminating HTP words from Novel words. Although this experiment was a replication of Pelucchi's first experiment, we re-recorded the speech material and used a different visual stimulus (a spinning pinwheel instead of flashing lights) to ascertain that our experimental setting did not confound the results and to set the stage for our Experiment 2.

Method

Participants

Twenty-four 8-to-9-month-old ($M_{\text{age}} = 8.64$ months; 11 female) monolingual English learning infants participated in this study. All infants were full-term with no record of hearing and/or vision problems and were recruited from monolingual English-speaking families with no consistent exposure to another language especially Italian or Spanish. Participants were recruited from the Child Development Research Group database based in the Department of Psychology at the University of Tennessee, Knoxville.

Infants were familiarized with one of the two counterbalanced languages. Eleven other infants participated in the study but were excluded from analysis due to experimental error ($n = 3$), fussiness ($n = 5$), crying ($n = 3$). All parents gave informed consent before the experiment. All infants received a t-shirt as a gift for their participation.

Stimuli

Speech materials were taken from Pelucchi et al. (2009a; Experiment 1). A new female native speaker of Italian who was naïve to the purpose of the study recorded the stimuli in a lively manner. The familiarization corpus included 12 syntactically correct and semantically meaningful Italian sentences that were repeated 3 times to create a speech stream of about 2 minutes and 20 seconds. All sentences were normalized using Praat software (Boersma & Weenink, 2001) to an average of $\sim 65\text{dB}_{\text{SPL}}$. See Appendix for a transcript of all familiarization languages.

Four bisyllabic Italian words (*fuga*, *melo*, *pane*, and *tema*) with trochaic (strong/weak) stress patterns were also recorded to use in the testing phase (see the target words used in each

language in Table 1). The syllables of the target words were all phonotactically legal in English. In each of the two counterbalanced languages, two of these words were embedded in a way that the internal TP of the words were 1.0 (HTP words) while the other two words were novel words. For example, in Language A, *fuga* and *melo* were HTP words, and *pane* and *tema* were novel words. In Language B, *pane* and *tema* were HTP words, and *fuga* and *melo* were novel words. The syllables of the novel words did not appear anywhere in the corpus. All the test words were recorded with their original trochaic stress pattern maintaining the stress pattern of the HTP words in the familiarization phase (see Table 2 for the acoustic characteristics of syllables of tokens in the familiarization languages and see Table 3 for the acoustic characteristics of syllables of test tokens). All the words were edited in Praat to have an equal length of 500 ms and intensity of approximately 65 dB_{SLP}.

Procedure

There were two phases: a familiarization phase and a testing phase. Infants were exposed to one of the two counterbalanced languages in the familiarization phase and then were tested using the Head Turn Preference Procedure (HTPP). The study was conducted in a sound-attenuated booth that had a center screen, two side screens and two side audio speakers. During the study, the caregiver held the infant on their lap while listening to masking music to lower any possible bias. The experimenter observed the infant's looking behavior in another room over a closed-circuit camera. In the Familiarization phase, a video of a spinning pinwheel appeared on the screens depending on the infant's looking behavior while the language was played continuously.

Right after familiarization, infants heard 12 test trials. All infants were presented with the same items regardless of the counterbalanced languages. Each of the test words (i.e., two familiar

HTP words and two Novel words) were repeated three times, randomized by block. Test trials began with the appearance of the spinning pinwheel on the center screen. When the infant oriented to the center screen, the experimenter began the trial which transferred the pinwheel from the center to one of the side screens. As soon as the infant made a head turn of at least 30° to the side screen showing the pinwheel, one of the test trials played continuously until the infant looked away for at least 2 seconds or 15 seconds had passed. The next trial began with the appearance of the pinwheel on the center screen. This process repeated for all 12 trials. Therefore, in this procedure, the infant was in control of how long they heard the target words. Trials which lasted for less than one second were repeated at the end. The dependent variable was the infant's total looking time toward each type of test trial. After the experiment, the caregiver also filled out a demographic information survey and a Communicative Development Inventory (CDI).

Results and Discussion

Initial analyses revealed no significant differences in performance based on familiarization language, $t(22) = -.149, p = .84, d = .06$. Thus, all subsequent analyses were collapsed across this variable. To examine whether 8-month-old English-learning infants were able to discriminate statistically segmented words from Novel counterparts, we conducted a paired-sample t test on the looking time to the HTP words versus Novel words, $t(23) = 3.192, p = .004, d = .648$. The results indicate that infants (17 out of 24 infants) listened significantly longer on HTP (Mean = 10.08 s, SD = 2.11) than on Novel word (Mean = 9.52 s, SD = 2.02) test trials (see Figure 1). This suggests that the infants were successful at pulling the HTP words out of an unfamiliar natural fluent speech stream and had a familiarity preference for listening to these recently segmented words at test. Thus, we replicated Pelucchi and colleagues' (Pelucchi et al., 2009) Experiment 1

with a new female voice and visual stimulus (using a spinning pinwheel instead of flashing lights).

Infants' successful segmentation of an unfamiliar language with a different rhythmic pattern from that of their native language suggests that variability across rhythm may not block TP-based speech segmentation. However, while Italian and English are rhythmically different, they have very similar prosodic characteristics (White et al., 2009). For example, like English, Italian stressed vowels are longer compared to vowels in unstressed syllables. This vowel reduction is an essential cue in stress realization. Therefore, infants in Pelucchi's study and in our Experiment 1 may have benefited from the redundant information provided by the overlap between the TP and stress pattern cues to word boundaries.

It would be interesting to examine how English learning 8-month-old infants perform in TP-based segmentation if the language is stripped of any familiar language-specific cues. To address this knowledge gap, in Experiment 2, we explore this question by familiarizing infants with a naturally produced Persian corpus where the only cue to word boundaries is the TP between syllables. Indeed, in order to segment HTP words at test, infants have to overcome a challenging task- extracting neutrally stressed HTP words and generalize them to iambic HTP words at test.

Experiment 2

In Experiment 2, we familiarize infants with a Persian natural speech stream with two embedded high TP (TP=1; HTP) words. Persian is syllable-timed and does not use stress at word level unless they appear at the end of a phrase. When words appear at the end of a phrase or produced in isolation, they are iambic (Rahmani et al., 2015). So Persian words in Persian are typically either neutrally stressed or are iambic. At test, we gauge infants' ability to discriminate

between the HTP words versus Novel words. To be successful at this challenging task, infants should first segment neutrally stressed HTP words in a language where the only available cue to word boundaries is TP between syllables. We know that in artificial languages infants are successful at extracting statistically-defined words when TP is the only cue to word boundaries. However, in addition to the complexity of natural languages and having to segment language with a different rhythm from that of infants' native language, here infants have to generalize the segmented neutrally stressed HTP words to iambic HTP words at test.

If English learning 8-month-old infants successfully segment the Persian language and differentiate between HTP words from Novel words, we can conclude that infants' statistical segmentation ability is not language bound and overrides rhythm-based and stress-based segmentation strategies. However, if infants fail to segment Persian corpus, it might be an indication that infants' natural speech segmentation ability requires redundant information such as stress information in addition to TP cues to successfully segment words from a rhythmically different natural language.

Method

Participants

Twenty-two 8-to-9-month-old ($M_{\text{age}} = 8.41$ months; 10 female) monolingual English learning infants participated in this study. All infants were full-term with no record of hearing and/or vision problems and were recruited from monolingual English-speaking families with no consistent exposure to another language, especially Persian or French. Participants were recruited from the Child Development Research Group database based in the Department of Psychology at the University of Tennessee, Knoxville.

Nine other infants participated in the study but were excluded from analysis due to fussiness ($n = 5$) and crying ($n = 4$). All parents gave informed consent before the experiment. All infants received a t-shirt as a gift for their participation.

Stimuli

We designed and developed speech materials following the material design used in our Experiment 1 and Pelucchi et al. (2009a, Experiment 1). A female native speaker of Persian who was naïve to the purpose of the study recorded the stimuli with a lively manner. The familiarization corpus included 12 syntactically correct and semantically meaningful Persian sentences that were repeated 3 times to create a speech stream of about 2 minutes and 20 seconds. All sentences were normalized using Praat software (Boersma & Weenink, 2001) to an average of $\sim 65\text{dB}_{\text{SPL}}$.

Four bisyllabic Persian words (*bazi*, *limu*, *sudi*, and *give*) were also recorded to use in the testing phase (see Table 1 for the target words used in each language). The syllables of the target words were all phonotactically legal in English. In each of the two counterbalanced languages, two of these words were embedded in a way that the internal TP of the words were 1.0 (HTP words) while the other two words were Novel words. For example, in Language A, *bazi* and *limu* were HTP words, and *sudi* and *give* were Novel words. In Language B, *sudi* and *give* were HTP words, and *bazi* and *limu* were Novel words. Persian words are neutrally stressed in fluent speech unless they appear at the end of a phrase (Rahmani et al., 2015). When words appear at the end of a phrase, they take an iambic stress pattern, similar to when they are produced in isolation. None of the HTP words within the corpus were produce at the end of the phrase. Thus, the HTP words in the corpus are pronounced with a neural stress pattern while the test tokens were recorded with an iambic stress pattern. All the words were edited in Praat to have an equal

length of 825 ms and intensity of approximately 65 dB_{SLP}. See Table 4 for the acoustic characteristics of syllables of tokens in the familiarization languages and Table 5 for the acoustic characteristics of syllables of test tokens.

Procedure

The procedure was identical to that of Experiment 1.

Results and Discussion

Initial analyses demonstrated a significant effect of counterbalanced language on the looking time difference for HTP words versus Novel words, $t(20) = -3.53, p = .002, d = 1.51$. A 2 (Language: A vs B) $\times$ 2 (Word Type: HTP vs Novel) repeated measures Analysis of Variance (ANOVA) with infants' looking time across trials as the dependent variable revealed a significant interaction between Language and Word Type, $F(1, 21) = 12.49, p = 0.02$ (see Figure 1). This significant interaction suggests that the difference between our familiarization languages was driven by word preferences. We also conducted post hoc pairwise comparisons to identify the source of difference. The results showed that in Language A, the difference between HTP and Novel words was not significant, $t(9) = -2.06, p = 0.07$ although there was a trend (8 out of 10 infants) toward preferring the Novel words (i.e., *give, sudi*). In Language B, however, we found a significant difference between HTP and Novel words, $t(11) = 2.99, p = .01, d = .91$ (see Figure 1), demonstrating that infants (10 out of 12 infants) listened significantly longer to HTP words (i.e., *give, sudi*) compared to Novel words (i.e., *bazi, limu*).

Thus, overall, infants showed a statistically significant preference for HTP words in Language B (i.e., *give, sudi*) but did not show a statistically significant preference for either HTP or Novel words in Language A. To examine the source of the word preference, we conducted an

acoustic analysis of the tokens in the familiarization languages and of the test words (see Table 4 for the acoustic characteristics of syllables of tokens in the familiarization languages and Table 5 for the acoustic characteristics of syllables of test tokens). An examination of the acoustic characteristics of test tokens shows that while *bazi* and *limu* had longer second syllables with higher intensity and pitch, meeting the characteristics of iambic words, the other target words were either less iambic (i.e., *give*) or had relatively equal values for both syllables, making it sound neutrally-stressed (i.e., *sudi*). Thus, infants' shorter looking to *bazi* and *limu* might have been due to their less familiar iambic stress pattern. Similarly, we looked at the acoustic properties of target words tokens in the familiarization languages and extracted the duration, pitch, and intensity of first and second syllables. Each language contained 6 repetitions of the HTP words, so we listed the values for all tokens and reported the averages in Table 4. Note that in natural speech words are produced at a much faster rate, and hence have shorter overall duration, than isolated words. The acoustic analyses show that for while there was a great deal of variability in the production of *bazi* and *limu*, on average, the first syllable was louder, but the second syllable was longer and higher pitched. For *give* and *sudi*, on average, first syllable was longer, louder, and higher pitched, making them sound more like trochaic words when produced within the familiarization corpus. This was especially more severe for *sudi* tokens, resulting in infants' preference to listen to this target word over the other target words.

Discussion

The current study was an attempt to test English learning 8-month-old infants' ability to statistically segment words in a nonnative language that has a different rhythm but similar stress pattern (i.e., Italian; Experiment 1) and a nonnative language that has both different rhythm and different word-level stress structure (i.e., Persian; Experiment 2). In Experiment 1, we tested TP-

based segmentation in a language (i.e., Italian) that is rhythmically different but prosodically very similar to English, providing redundant prosodic and TP information during familiarization. We replicated previous findings by Pelucchi and colleagues (Pelucchi et al., 2009, Experiment 1) and other works using similar Italian corpora (e.g., Karaman & Hay, 2018; Lew-Williams et al., 2011) that English learning 8-month-olds can use statistical information (i.e., transitional probabilities between syllables) to segment high TP words from a nonnative language. Experiment 2 was an investigation to test English-learning 8-month-olds' ability to segment high TP target words (HTP) in Persian, a language that is both rhythmically and prosodically different from English, and thus the TP information was not supported by any other redundant cues to word boundaries. However, in this Experiment, infants showed a difference in performance in the two counterbalanced languages, with an overall word preference for *give* and *sudi* making it difficult to assess the impact of the familiarization language on segmentation.

One possible explanation for these finding is that infants' preference for *give* and *sudi* test tokens may have overridden infants' sensitivity to the statistical information available in the corpus. Indeed, a follow-up acoustic analysis on the duration, intensity, and pitch of the syllables of the test tokens demonstrated that although *bazi* and *limu* had values consistent with an iambic stress pattern (longer duration and higher intensity and pitch of the second syllable; see Table 5), the other target words were either less iambic (i.e., *give*) compared to *bazi* and *limu*, or had an equal duration in the first and second syllables, making it sound more neutrally stressed (i.e., *sudi*). Thus, infants may have preferred *give* and *sudi* simply because they were less unfamiliar (i.e., they were less iambic).

However, if this was the entire explanation, we would have expected infants in both counterbalance conditions to have demonstrated the same preference for *give* and *sudi*. This is

not what we found; Infants who listened to Language A during familiarization (where *bazi* and *limu* were HTP words) did not show any significant difference between listening to HTP versus Novel words. However, infants who listened to Language B (where *give* and *sudi* were HTP words) listened significantly longer to HTP words compared to Novel words. Thus, experience with the familiarization language does appear to have impacted infants' word preferences at test.

Our acoustic analysis of the target word tokens in the familiarization languages revealed that while *bazi* and *limu*, on average, had relatively longer and higher pitched second and longer first syllables, *give* and *sudi*, on average, had mostly longer, louder, and higher pitched first syllables, more consistent with a trochaic stress pattern. Thus, it is possible that infants familiarized with Language B were better able to segment *give* and *sudi* from the speech stream because these words were produced with a more familiar stress pattern. Conversely, infants may have failed to segment HTP words from Language A where *bazi* and *limu* were produced with a neutral-iambic stress pattern and thus the lack of redundant stress cues may have impacted infants' ability to track TP information in the corpus.

It is also worth noting that infants familiarized to the two different languages had to accomplish different degrees of generalization between familiarization and test. For example, infants familiarized with Language B had to extract trochaic-like HTP words in the familiarization and then generalized them to a more neutrally-stressed HTP word at test. Thus, compared to Experiment 1, where infants had to segment HTP words from a totally trochaic Italian corpus and discriminated trochaic HTP words from Novel words, this task was harder because infants had to segment the trochaic-like words from corpus and generalize them to neutrally stressed words. Nevertheless, infants still appeared to benefit from the more trochaic-like stress cues available in Language B and perform similarly to infants in Experiment 1. Infants

familiarized with Language A were exposed to neutrally-stressed HTP words in the corpus and then were asked to generalize to an iambic realization of those words at test which they were unable to. Thus, it appears that although the presence of redundant stress cues is essential to nonnative speech segmentation (as shown in Experiment 1 and Language B of Experiment 2), the degree of difference in the stress pattern of the familiarization tokens and test tokens may not be as important and infants tend to generalize across stress cues.

One potential explanation for the pattern of findings in Experiment 2 can be the degree of variability in the production of familiarization tokens in both languages. In this Experiment, although all the syllables used in the Persian target words (i.e., *ba*, *zi*, *li*, *mu*, *gi*, *ve*, *su*, & *di*) were both legal and had similar phonotactic probabilities in English, each word token was produced with a certain degree of variability in the acoustic characteristics. Although this is usually a typical phenomenon in using natural language stimuli, in our stimuli in Language A, we observed a much higher degree of variability in the production of *bazi* and *limu* compared to *give* and *sudi* (see Table 4) in Language B in a way that the duration, intensity and pitch of the syllables were consistent with different stress pattern structures. Thus, while variability should help infants form more robust representations (Rost & McMurray, 2008), in Language A too much variability in the acoustic characteristics of *bazi* and *limu* might have prevented infants from forming single representations of the HTP words. We know that when the stress shifts between the syllables, vowel reduction in the unstressed syllable creates a much different instance of the word and thus, there is a possibility that too much variability in the production of these words may have prevented infants from recognizing them at test. However, in Language B, both *give* and *sudi* were produced with more trochaic-like stress pattern in the familiarization and with more neutral stress at test. Therefore, infants' experience with English may have facilitated

infants' word segmentation by utilizing redundant cues between TP and stress pattern to extract trochaic sounding target words from continuous speech.

Previous research has shown that infants utilize word-level stress to segment nonnative speech even if they are inconsistent with the stress pattern of their native language (Sundara & Mateu, 2018). They showed that English learning 8-month-old infants can segment Spanish iambic words but fail to segment French target words, suggesting that presence of stress cues in the corpus helps infants segment the nonnative speech. In our Experiment 2, we used Persian, and Persian is very similar to French in terms of word-level stress. In both languages, depending on where the target words appear in the phrase, they may be pronounced stress-free or iambic. Therefore, if we did not have acoustic inconsistencies in our stimuli and *give* and *sudi* were not produced with a trochaic stress pattern, infants may have totally failed to segment Persian, similar to previous findings with French that English learning infants were unable to segment French (e.g., Polka & Sundara, 2012; Sundara & Mateu, 2018). Future studies should repeat this study with language material that is controlled for acoustic inconsistencies.

Interestingly, there is ample evidence that infants segment artificial languages that the only cue to word boundaries is TP between syllables (Saffran et al., 1996). However, with the complexity of natural speech infants might still need redundant information from language-specific cues to segment a non-native natural language (e.g., Sahni et al., 2010). Future studies should repeat this study to test TP information versus word-level stress but control for natural language variabilities that may act as confounding variables. Together, our findings suggest that experience with language-specific stress structure that is redundant with statistical information facilitates word segmentation in complex fluent natural speech. However, if language-specific cues are absent, infants may be unable to segment the speech.

Chapter 3: Specificity of Infant Speech Segmentation

Abstract

Infants' ability to segment continuous speech is well-documented, yet representational specificity of these nascent lexical forms is less-well understood. We tested English-learning 8-month-old's indexical, segmental, and suprasegmental representations of newly segmented words using a Headturn Preference Procedure. Following familiarization with a female-produced Italian corpus, infants (N=76) successfully discriminated previously embedded Target words produced by a male (Experiment 1; $d=.62$) and those produced with a mispronunciation in stress (Experiment 3; $d=.3$) from foils but failed to recognize Target words with a mispronounced first consonant (Experiment 2). Findings demonstrate infants can generalize representations of recently segmented words across acoustic forms less relevant to word meaning in English, but not across phonemic changes fundamental to meaning.

Introduction

Learning language presents infants with numerous cognitive and perceptual challenges. Two notable challenges involve segmenting candidate words from continuous speech and representing those nascent word forms with the appropriate level of phonological detail. Research over the past three decades suggests that by 9 months of age infants rely on a number of language-specific strategies for segmenting the speech stream such as phonotactic regularities (e.g., Friederici & Wessels, 1993; Mattys & Jusczyk, 2001) and native language prosodic patterns (e.g., Dutch: Houston et al., 2000; English: Jusczyk et al., 1993; Jusczyk et al., 1999; Thiessen & Saffran, 2003; French: Goyet et al., 2013; German: Höhle & Weissenborn, 2003; Spanish/Catalan: Bosch et al., 2003), however, younger infants may rely more heavily on language-general cues (e.g., Aslin et al., 1998; Saffran et al., 1996; Thiessen & Saffran, 2003). Indeed, one of the most prominent hypotheses is that infants may begin to discover boundaries between words by tracking the probability that two syllables will co-occur, commonly referred to as the transitional probability (TP) between syllables (Saffran et al., 1996; Swingley, 2005; for a review, see Cannistraci et al., 2019). Indeed, there is considerable evidence to suggest that infants, as young as 8 months of age, are able to find word boundaries by tracking TPs in both artificial (e.g., Saffran et al., 1996) and natural languages (e.g., Goyet et al., 2013; Nishibayashi et al., 2015; Pelucchi et al., 2009a, 2009b). Further, the output of these statistical learning processes appears to feed into subsequent word learning by 17 months (e.g., Graf Estes et al., 2007; Hay et al., 2011; Shoaib et al., 2018).

We also know that infants as young as 6 months represent familiar words with an adaptive level of phonological detail, for example, ignoring less relevant speaker information (e.g., Bergleson & Swingley, 2018; Johnson et al., 2014), while attending to vowel and

consonant changes relevant to word meaning (e.g., Johnson et al., 2014; Nishibayashi & Nazzi, 2016). These familiar words (e.g., milk, nose, apple, mouth) become known to infants through a wide variety of experience with hearing them produced by multiple talkers (Bergelson & Swingley, 2018), often in the context of their referents (Yu et al., 2021), in both fluent speech and in isolation (Brent & Siskind, 2001). This variability appears to produce the optimum conditions for learning what to attend to in the speech signal (Graf Estes & Lew-Williams, 2015). However, we know much less about how infants, who are in the midst of developing sensitivities to native language cues to word boundaries, represent the phonological features of words when they first segment them from the speech stream. In particular, we do not know whether the representations of these newly segmented sequences share properties with more familiar lexical items.

If the output of these segmentation processes are word-like units, we might expect the phonological representations of those newly segmented word forms to share common features with actual words. Specifically, we would expect infants to attend to the features that are most relevant to word meaning and generalize across the features that are less relevant to meaning, thus reducing the language learning problem space. Alternatively, the representation of these newly segmented sequences, while recognizable, may not share representational similarities with actual words, limiting the potential importance of segmentation processes in the forming of lexical representations. Here we address this question from a new angle, testing 8-month-old English-learning infants' representations of words newly segmented from a novel (i.e., Italian) language. What makes our study unique is that we look at indexical (Experiment 1), segmental (Experiment 2) and suprasegmental (Experiment 3) differences across three studies using stimuli

that are novel to the infants and that are also well-controlled in terms of their statistical structure (see below for further description).

From a phonological perspective, words can be represented at three different levels: indexical (i.e., properties of the speaker's voice), segmental (i.e., phonotactic rules of language), and suprasegmental (i.e., prosody of language). Examples of indexical information are sex of the speaker (male vs. female) and affect (happy vs. sad voice). Although talker characteristics can probabilistically cue infants to a given language (as when each parent speaks a different language) and hence lead to a narrower subset of word candidates, by and large, variations in talker characteristics do not alter word meaning. Thus, in most circumstances, it may be advantageous for infants to ignore this type of variation during word-form recognition.

Segmental information includes how sounds combine to make words. For example, the acoustic forms of cat and bat differ only in their onset sounds, yet they have different meanings and refer to different entities. Segmental information is lexically contrastive across the world's languages, and thus, it is adaptive for infants to learn to attend to variation across this category.

Suprasegmental information includes prosodic cues, such as lexical tone, pitch contour, and stress pattern. While suprasegmental information plays a vital role in word meaning in some languages (e.g., tone languages where varying pitch contour can alter word meaning), not all languages use suprasegmental information this way. Indeed, for English-learning infants, suprasegmental cues are sometimes important to attend to (e.g., to differentiate between the noun record /'rekərd/ and the verb record /rə'kôrd/), but often they are not. We now have a better understanding of infants' ability to perceive various types of phonological information (see below for an overview). Further, a significant body of research suggests that infants home in on the relevant features of their native language relatively early on (see Werker & Gervain, 2013 for

a review; Bosch et al., 2003; Kuhl et al., 1992; Werker & Tees, 1984), forming the basis for their lexical representations of familiar words. Nevertheless, much less is known about what information infants encode when they first encounter words in fluent speech. To bridge this gap, in the experiments that follow we test the indexical (Experiment 1), segmental (Experiment 2), and suprasegmental (Experiment 3) specificity of infant speech segmentation.

Experiment 1

In Experiment 1, we examine the specificity of infants' indexical representations of newly segmented words. Infants demonstrate a sensitivity to indexical information from birth (e.g., De Casper & Fifer, 1980; Mehler et al., 1978), which appears to persist across the lifespan (for a review, see Strori et al., 2018). Nevertheless, as indexical features of speech are not typically used to differentiate word meaning, learning to generalize across talkers is adaptive at the lexical level. Consistent with this idea, work by Houston and Jusczyk (2000) suggests that infants' representations of indexical information may become specialized across development. In their task, infants were given experience with two familiar words (e.g., cup & bike) produced by either a female or a male talker and were then tested on 4 passages produced by a speaker of the opposite sex (e.g., infants familiarized with a female speaker were tested with passages produced by a male speaker and vice-versa): 2 passages contained the familiarized words, and 2 passages did not. At 7.5 months, infants failed to recognize familiarized words in continuous speech when the sex of the speaker changed. However, by 10.5 months of age, infants were able to generalize across sex of speaker, suggesting a shift in how infants represent talker information between 7 and 11 months.

Other work suggests that infants may be able to generalize across indexical information even earlier in development. For example, Bergelson and Swingley (2018) found that 6- to 7-month-olds can understand some names for food and body parts (i.e., they look at a named picture) even when the label is produced by an unfamiliar talker. Similarly, van Heugten & Johnson (2012) found that 6- to 7.5-month-old infants were able to recognize words segmented from natural language when the sex of the talker was changed between familiarization and test (see also Singh, 2018), in line with classic findings on cross-sex speech perception in infancy (e.g., Jusczyk et al., 1992; Kuhl, 1979). Although the reason for the discrepant developmental trajectory between Houston & Jusczyk's (2000) early work and that of van Heugten & Johnson (2012) is not fully understood, methodological differences in word-passage presentation order may play a role. Work by Nazzi and colleagues (Nazzi et al., 2014) shows that the passage-word order (i.e., familiarization with a language and test with isolated words) like the one used by van Heugten & Johnson (2012) is an easier task compared to the type of word-passage order (i.e., familiarization with isolated words and test with passages) used by Houston & Jusczyk (2000), and thus this might be a driving factor in the discrepancies between these findings. Further, familiarization with isolated words, as in Houston and Jusczyk (2000), may have provided these younger infants with insufficient acoustic and phonetic variability, as compared to the natural speech passage familiarization used by van Heugten & Johnson (2012), and thus may have limited infants' ability to form generalizable representations at such an early age.

Variability also appears to play a vital role in talker generalization in statistical speech segmentation tasks using artificial language materials. Specifically, Graf Estes and colleagues found that although 11-month-olds were able to recognize statistically-defined words when a single talker's voice changed from familiarization to test (Graf Estes, 2012), younger 8-month-

old infants required significant variation in talkers' voices during familiarization to recognize words produced by a new talker at test (Graf Estes & Lew-Williams, 2015). These studies suggest that although variability is essential for generalizing across talker information early in development, as infants gain experience with language, they may come to broadly ignore changes in talker information at the lexical level. Thus, while their word representations may still contain talker-specific information, changes in talker may not impact the recognition of recently segmented words.

Together, these studies suggest that infants begin to generalize across indexical information early in development, at least when the words are either highly familiar or potentially familiar (e.g., Bergelson & Swingley, 2018; Houston & Jusczyk, 2000; van Heugten & Johnson, 2012) and/or when words are embedded in native-language sentences (e.g., Houston & Jusczyk, 2000; van Heugten & Johnson, 2012). However, when unfamiliar word forms are embedded in an unfamiliar artificial language, as in Graf Estes and Lew-Williams (2015), infants appear to require a significant degree of between-talker variability during familiarization to recognize the same words produced by a new speaker.

In our first experiment we focused on 8-to-9-month-old infants' ability to generalize across talker information following familiarization with a naturally-produced statistical learning language. Unlike artificial languages, natural speech is characterized by substantial within-talker variability and contains more diverse phonological, prosodic, and acoustic information. Thus, using a naturally-produced language provides a more ecologically valid test of infants' representations of newly segmented words. Further, as we are interested in infants' nascent representations, we needed to ensure that the Target words were unfamiliar to infants. To that end, in Experiment 1, English-learning infants were exposed to an unfamiliar natural language

that had overlapping prosodic and phonetic features (see below for a description of the Italian familiarization corpus) used in previous research (i.e., Karaman & Hay, 2018; Pelucchi et al., 2009a, Language 3A/B) and were tested on their ability to discriminate Target words from the corpus that had high TP (HTP; TP=1) from those that had lower TP (LTP; TP=.33). Key to this experiment is that the familiarization language was produced by a female speaker, whereas the Target words used at test were produced by a male speaker. Statistical learning studies that familiarize infants with these natural language materials consistently find that infants have a familiarity preference for the HTP words over LTP or Novel words (e.g., Pelucchi et al., 2009a, 2009b). A familiarity preference suggests that infants not only recognize the HTP words, but that they remain interested in them (see Pelucchi et al., 2009a and Graf Estes, 2012 for an extensive discussion of novel versus familiarity preferences in the speech segmentation literature). Thus, we predict that if infants are able to generalize across talker information, they should listen longer to the male-produced HTP Target words at test, consistent with previous findings where there was no change in talker's voice (Pelucchi et al., 2009a). The specificity of segmental and suprasegmental representations is explored in Experiment 2 and 3, respectively, using the same task and very similar familiarization languages, also drawn from Pelucchi et al. (2009a).

Method

Participants

Twenty-six 8- to 9-month-old ($M_{\text{age}} = 8.66$ months; 13 female) monolingual English-learning infants participated in Experiment 1. In line with previous work (Pelucchi et al., 2009a, 2009b; Karaman & Hay, 2018) and a power analysis assuming a medium effect size and power of .8, our intended sample was 24 participants. The two additional participants were run in error

but are included here in order to represent the entirety of the data collected. All infants were born full-term with no hearing and/or vision problems, according to parental reports.

Infants were recruited from a participant database maintained in the Department of Psychology at the University of Tennessee. Eleven other infants participated in Experiment 1 but were excluded from analysis due to excessive fussiness or crying ($n = 9$) and parental interference ($n=2$). Infants received a lab t-shirt or a book as a gift for their participation.

Stimuli

The familiarization corpus was taken from Pelucchi et al. (2009a, Experiment 3; see Appendix for familiarization corpora used across all three experiments). A female native speaker of Italian who was naïve to the purpose of the study re-recorded the stimuli in a lively manner. The familiarization corpus was composed of 12 syntactically correct and semantically meaningful Italian sentences that were repeated 3 times, in the same order, to create a speech stream of approximately 2 minutes and 20 seconds. All sentences were normalized using Praat software (Boersma & Weenink, 2001) to an average of $\sim 65\text{dB}_{\text{SPL}}$.

Four Italian trochaic (strong/weak) Target words (*casa*, *bici*, *fuga*, and *melo*) were embedded in the familiarization corpus. The syllables of the Target words were all phonetically and phonotactically legal in English. Two of the words had high TP (HTP; $\text{TP}=1.0$) because their syllables never occurred anywhere else in the corpus. Two of the words had low TP (LTP; $\text{TP}=0.33$) because their first syllable occurred 12 additional times in other words throughout the corpus. Each Target word appeared six times in the 12-sentence corpus. In Language 1A, *fuga* and *melo* were HTP words and *casa* and *bici* were LTP words. In a counterbalanced Language 1B, HTP and LTP words were switched (see Table 6). A male native speaker of Italian recorded the Target words in isolation for the test phase. The female-produced HTP Target words in the

corpus had a higher mean fundamental frequency (F_0) than the male-produced HTP Target words produced in isolation (see Table 7 for F_0 values). To ascertain listening preferences at test were not derived from unrelated acoustic characteristics, isolated words were digitally edited in Praat to have similar length (500ms) and intensity of 65dB_{SPL}.

Procedure

Infants were exposed to one of the two counterbalanced languages in the familiarization phase and then were tested using the Head-turn Preference Procedure (HTPP). The study was conducted in a sound-attenuated booth that had a center monitor, positioned 1 meter directly in front of the infant at eye level. Two side monitors along with two side audio speakers were positioned 180 degrees apart from one another and 1 meter to the right and left of the infant at ear level. During the study, the caregiver held the infant on their lap. To reduce the likelihood of inadvertently biasing their infant's looking behavior, caregivers listened to masking music over headphones during the study. The experimenter observed the infant's looking behavior in another room over a closed-circuit camera. During familiarization, a video of a spinning pinwheel appeared on the monitors depending on the infant's looking behavior (see below) while the language played continuously.

Immediately after familiarization, infants heard 12 test trials across 3 testing blocks. Within each block, infants were tested on all four Target words (i.e., two HTP words and two LTP words), randomized by block. All infants were presented with the same test words produced by the male speaker regardless of the counterbalanced language heard during familiarization. Test trials began with the appearance of the spinning pinwheel on the center monitor. When the infant faced the center monitor, the experimenter initiated the test trial, which transferred the pinwheel from the center to one of the side monitors. For each trial, the side of presentation was

randomized by the experimental computer, with the constraint that no more than two consecutive trials on the same side were permitted. As soon as the infant made a head-turn to the side monitor showing the pinwheel, one of the test words played constantly until the infant looked away for at least 2 seconds or after 15 seconds had passed. The next trial began with the appearance of the pinwheel on the center monitor. This sequence repeated for all 12 trials. Therefore, in this procedure, the infant essentially controlled how long they heard the Target words. Trials lasting less than one second were repeated at the end and thus the looking times under one second were excluded. To prevent potential bias, the experimenter was blind to the test words being presented. The dependent variable was the infant's total looking time on each test trial. After the experiment, the caregiver filled out a demographic information survey.

Results and Discussion

Preliminary analyses revealed no significant effect of counterbalanced language on the difference in looking time between HTP and LTP test trials, $t(24)=-.7$, $p=.48$, $d=.27$ (all t tests are 2-tailed; effect sizes are Cohen's d). Thus, all subsequent analyses were collapsed across Language. We conducted a paired sample t test to examine whether 8-month-old infants were able to discriminate HTP from LTP Target words when the sex of the talker's voice was changed between familiarization and test. Infants listened significantly longer to HTP words ($M=9.76$, $SD=1.61$) relative to LTP words ($M=8.97$, $SD=1.82$), $t(25)=2.297$, $p=.03$, $d=.62$ (see Figure 3). These data not only replicate previous findings that infants are able to track TP information in natural language input, but importantly extend them to show that 8-month-olds are able to generalize their representations of HTP words recently segmented from natural language to a novel talker. Indeed, changing the speaker from female to male, as in Experiment 1, did not appear to meaningfully alter the effect size between previous studies that used a female voice for

both familiarization and test ($d=.70$ in Pelucchi et al., $d=.76$ in Karaman & Hay) and the current study ($d=.62$). This suggests that infants' indexical representations of naturally produced statistically-defined words may be quite abstract (see Figure 4 for a Forest Plot which provides a visualization of effect size comparisons between the current findings and those of Pelucchi et al., 2009, using the same familiarization languages). This work adds to the growing body of evidence that recently segmented words with high TP may have a special status for the young language learner; not only are infants able to generalize their representation of words with high TP across speaker information, as demonstrated here, but previous research has also found that these words make better candidate object labels (e.g., Hay et al., 2011) and are remembered after a delay (e.g., Karaman & Hay, 2018).

In the experiments that follow, we probe the segmental (Experiment 2) and suprasegmental (Experiment 3) specificity of infants' representations of newly segmented words. In comparison to indexical changes tested in Experiment 1 which are instantiated at the whole word-level, generalizing across segmental and suprasegmental features may be more challenging as infants need to attend to often more subtle changes that unfold over a shorter timeframe (i.e., at the phoneme and syllable levels, respectively). Thus, in Experiments 2 and 3, we decided to test infants on an easier contrast – namely, Target words (TP=1.0) embedded in the corpus versus Novel words (i.e., syllable sequences that were unattested in the language and thus had a TP=0). This contrast provides infants with a less stringent test of word recognition as successful discrimination only requires that infants recognize that the Target words that they have segmented from the corpus are different from words that they have never heard before. Thus, if infants have any ability to recognize the mispronounced Target words it should be evidenced in our task. The familiarization languages used in Experiments 2 and 3 were very similar to those

used in Experiment 1. However, in order to ensure that our materials were counterbalanced (i.e., Novel words for one group of infants served as Target words for the other, and vice versa), we used Languages 1A/B and 2A/B from Pelucchi et al (2009a). The primary differences between the familiarization languages used in Experiment 1 (and Languages 3A/B from Pelucchi et al, 2009) and those selected for use in Experiments 2 and 3 (Languages 2A/B and 1A/B, respectively, from Pelucchi et al., 2009a) is that there was no manipulation of LTP words or their component syllables. Thus, by using Languages 1A/B and 2A/B from Pelucchi et al, (2009a), Experiments 2 and 3 do not provide a strict test of statistical learning per se. However, we can have a high degree of confidence that the Target words will have been successfully segmented from the corpus by our participants. This in turn will allow us to assess the segmental and suprasegmental specificity of infants' representations of these newly segmented words.

Experiment 2

In Experiment 2 we focus on infants' ability to generalize across segmental information (i.e., the initial constant of the target words). Segmental and indexical information differ from each other in many ways, most importantly in their relevance to word meaning – unlike a change in talker, a change in either a consonant or a vowel can alter the meaning of the entire word. In tests of speech perception skills, infants show early sensitivities to segmental changes, which become tailored to their native language by their first birthday (Aslin et al., 1981; Eimas et al., 1971; Werker & Gervain, 2013; for a review see Maurer & Werker, 2014). However, how veridically infants represent segmental information in newly encountered words remains an open question.

Work by Johnson and colleagues (Johnson et al., 2014) and by Nishibayashi & Nazzi (2016) suggests early sensitivity to segmental information in newly segmented words. Specifically, in a speech segmentation task, Johnson and colleagues (2014) found that English-learning infants as young as 6 months failed to recognize words from an English familiarization corpus if the initial segment of the word was missing. Similarly, Nishibayashi & Nazzi (2016) found that European French learning infants develop a consonant bias between 6 and 8 months of age where they come to be more sensitive to consonant than vowel mispronunciations in recently segmented French words (see also Nespor et al., 2003 for a discussion of consonants to lexical processing). This work contrasts with Bergelson and Swingley (2018) who found that at 6- to 7-months infants recognize phonologically deviant familiar words produced by their mothers (e.g., nazz for nose); however, by 11 to 14 month of age infants appear to be more sensitive to these types of mispronunciations. These findings suggest that infants' representations become more adult-like and more specialized as they learn which forms of variation are most relevant to meaning (see also Vihman et al., 2004).

With these findings in mind, in Experiment 2 we examine the segmental specificity of words newly segmented from an unfamiliar natural language. To that end, following familiarization with a naturally-produced Italian corpus, English-learning infants were tested on their ability to discriminate mispronounced Target words (i.e., words with an initial consonant change, e.g., melo → belo) from Novel words. If infants form weak representations of the segmental information in the Target words, they should continue to recognize the mispronounced Target words, as evidenced by a familiarity preference. The prediction of a familiarity preference is consistent with previous literature using these natural language stimuli (e.g., Pelucchi et al. 2009a, b; Karaman & Hay, 2018; Lew-Williams et al., 2011), and with findings from

Experiment 1. However, if infants attend to segmental information and have a robust representation of the sound segments of the Target words, they should not treat mispronunciations as functionally equivalent to the Target words that they segmented from the corpus. In turn, they may fail to differentiate the mispronounced Target words from Novel words. A lack of familiarity preference here would suggest that the segmental changes made to the Target words interfered with their recognition.

Methods

Participants

Twenty-four 8- to 9-month-old ($M_{\text{age}} = 8.56$ months; 11 females) monolingual English-learning infants participated in Experiment 2. Recruitment procedures and exclusion criteria were identical to Experiment 1. Data from 6 additional infants were excluded due to excessive fussiness or crying.

Stimuli

The familiarization languages were very similar to those of Experiment 1 and were modeled after Pelucchi et al. (2009a, Experiment 2). As in Pelucchi et al. (2009a, Experiment 2), two high TP trochaic Target words (TP=1.0) and the syllables of the Novel words (TP=0) were embedded in the corpus with equal occurrence. We replaced a few syllables from the original languages to ensure that the initial syllables in the test words did not appear in the corpus (see below for details). As in Experiment 1, a female native speaker of Italian recorded the stimuli in a lively manner. In Language 2A, *fuga* and *melo* were Target words, and *pane* and *tema* were Novel words; in Language 2B, Target and Novel words were switched.

The same female speaker produced the test words: 2 mispronounced Target words and 2 Novel words. The initial consonant of both the Target and the Novel words was changed by one feature, either place of articulation or manner of articulation, to have the smallest change possible. Thus, for example, if infants heard *fuga* and *melo* as Target words and *pa*, *ne*, *te*, and *ma* as syllables of the Novel words, they were tested with *shuga*, *belo*, *tane*, and *sema* (see Table 6). To ensure that the changed consonants were equally unfamiliar to the infants, the syllables *shu*, *be*, *ta*, and *se* never appeared in the corpus.

Procedure

The procedure was identical to that of Experiment 1.

Results and Discussion

Initial analyses demonstrated no significant effect of counterbalanced language on the looking time difference for mispronounced Target words versus Novel words, $t(22) = .50$, $p = .62$, $d = .20$. Therefore, all subsequent analyses were collapsed across this variable. Interestingly, even though the segmental change was relatively small (i.e., one feature), there were no statistically significant looking time differences between mispronounced Target words ($M = 9.44$, $SD = 1.75$) and Novel words ($M = 9.42$, $SD = 2.22$), $t(23) = .079$, $p = .93$, $d = .01$ (see Figure 3). Consistent with the work of Johnson and colleagues (2014) and Nishibayashi & Nazzi (2016), these findings suggest that by 8 months of age, infants may have already formed fairly specific representations of segmental information in newly segmented words.

It is worth reiterating that Experiment 2 was essentially a replication of Pelucchi et al.'s (2009a) 2nd experiment with the caveat that we changed the first consonant of the Target and Novel test words between familiarization and test. As can be seen in the Forest Plot (see Figure

4), this type of segmental change meaningfully impacted the effect sizes between these two studies ($d=.49$ in Pelucchi et al., $d=.01$ in our Experiment 2), suggesting that the segmental change completely disrupted infants' ability to differentiate the mispronounced Target words from Novel words.

Experiment 3

In Experiment 3 we focused on infants' ability to generalize across suprasegmental (i.e., stress pattern) information following familiarization with a natural language. Suprasegmental features are prosodic cues such as lexical tone, pitch contour, or stress pattern. In contrast to indexical features, which infants should learn to ignore because they are typically irrelevant to word meaning, and segmental features, which infants should learn to attend to because they are highly relevant to word meaning, suprasegmental features fall in a gray zone for infants learning intonation languages, such as English. For example, changing the stress pattern of words in English can sometimes change their meaning (e.g., record /'rekərd/ as a noun and record /rə'kôrd/ as a verb). However, this is not always the case, and by and large, suprasegmental features are not relevant to word meaning in English.

Nonetheless, from birth, infants begin to respond to native language prosodic information (Mehler et al., 1988), and by two months, they are sensitive to stress pattern changes (Jusczyk & Thompson, 1978). While 9-month-olds prefer listening to words that are consistent with the dominant stress pattern in their native language, 6- to 7-month-olds (Jusczyk, et al., 1993; Echols et al., 1997) do not typically display a preference, suggesting that perceptual narrowing of prosodic features happens sometime between 6 and 9 months. Extensive work by Curtin (e.g., Curtin, 2009; 2010) has shown that by 12 months infants can successfully map trisyllabic novel

sequences that differ in only stress pattern (e.g., BEdoka vs. beDOka) to novel objects (Curtin, 2010) and by 14 months they are able to detect stress mismatches in newly learned trisyllabic sequences, suggesting that infants have access to stress pattern information from early on.

Infants learning languages with different rhythmic structures have demonstrated distinct patterns of sensitivity to stress cues. For example, Hohle and colleagues (Hohle, et al., 2009) found that while infants learning German (a stress timed language) have a trochaic bias by 6 months of age, infants learning French (a syllable timed language) did not show a preference for either trochaic or iambic stress patterns despite their ability to discriminate between them. Likewise, Skoruppa and colleagues (Skoruppa et al., 2009., 2009) have found that infants learning Spanish, a language with contrastive stress patterns, and French, which does not use stress contrastively, differ in their ability to distinguish between different stress patterns in pseudo-words at 9 months. These findings emphasize the role language-specific experience plays in shaping infants' sensitivities to stress cues.

In the word segmentation literature, there has been a fair amount of discussion about whether infants rely more heavily on statistical (e.g., TP between syllable sequences) versus suprasegmental (e.g., stress pattern) cues to word boundaries. Research by Thiessen and Saffran (2003) suggests that early in development (i.e., before 8 months) English-learning infants may weight statistical cues to word boundaries more heavily than stress cues (see also Hay & Saffran, 2012). However, by 8 to 9 months, infants rely more heavily on stress cues to find words in speech (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). Interestingly, younger infants can be encouraged to adopt a stress-based segmentation strategy if they are given prior experience with stress cues (Theissen & Saffran 2007). Further, although 7.5-month-old English-learning infants struggle to segment iambic sequences from natural language passages, when a familiar

word such as mommy/mama is inserted into the speech stream right before the iambic sequences, infants successfully segmented those iambic words (Sandoval & Gomez, 2016). These findings suggest that linguistic experience may gradually enable infants to rely on language-specific cues such as stress in addition to statistical cues in identifying word boundaries.

Although our understanding of the role suprasegmental cues play during speech segmentation is growing, we still know very little about how infants represent syllable stress in words newly segmented from natural continuous speech. In Experiment 3, we examine whether 8-month-olds are able to generalize across stress information. To that end, following familiarization with an Italian corpus that contained trochaic Target words (TP=1), infants were tested on their ability to discriminate mispronounced Target words (i.e., Target words with an iambic stress pattern; MElo → meLO) from Novel iambic words (e.g., paNe, teMA; TP=0). Our hypothesis is that if infants treat stressed and unstressed syllables as functionally equivalent, they should listen to the mispronounced Target words longer than Novel words, replicating previous results by Pelucchi and colleagues (2009a) who maintained the trochaic stress pattern of the Target words between familiarization and test. However, if infants do not treat stressed and unstressed syllables equally, they may not recognize the mispronounced Target words and thus fail to differentiate them from the Novel words. It is worth noting that theoretically, it may have been ideal to be able to test infants on correctly pronounced (i.e., trochaic stress) versus mispronounced (i.e., iambic stress) versions of the Target words (e.g., FUga vs fuGA). However, we know that, by this age, infants have already developed a trochaic bias (Jusczyk et al., 1993). Thus, had we compared iambic versus trochaic versions of the Target words, we would expect infants to show a preference for the trochaic versions irrespective of their familiarization experience or the resulting representations of the newly segmented words.

Methods

Participants

Twenty-six 8- to 9-month-old ($M_{\text{age}}=8.7$ months; 13 females) monolingual English-learning infants participated in Experiment 3. Recruitment procedures and exclusion criteria were identical to the previous experiments. Data from 6 additional infants were excluded due to excessive fussiness or crying.

Stimuli

The familiarization corpus was taken from Pelucchi et al. (2009a, Experiment 1). Again, a female native speaker of Italian recorded the stimuli in a lively manner. The familiarization language was very similar to those of previous experiments. Two high TP trochaic Target words (TP=1.0) were embedded in the corpus. In Language 3A, *fuga* and *melo*, and in Language 3B, *pane* and *tema* were HTP words (see Table 6).

In the test phase, we again tested infants with two mispronounced Target words and two Novel words; Only the stress pattern of the words was altered between familiarization and test.

Specifically, whereas all Target words in the familiarization corpus were trochees, infants were tested on their ability to discriminate iambic (weak/strong) versions of the Target words from Novel iambic words (see supplementary material for the acoustic features of the tokens in the corpus and the acoustic features of the iambic test words).

Procedure

The procedure was identical to those of previous experiments.

Results and Discussion

Initial analyses revealed no significant differences in performance based on familiarization language, $t(24)=1.34$, $p = .91$, $d = .52$. Thus, all subsequent analyses were collapsed across this variable. To examine whether 8-month-old English-learning infants are able to discriminate newly segmented words from Novel words when the stress pattern of the Target words was changed between familiarization and test, we conducted a paired sample t test on the looking time to the mispronounced Target words versus Novel iambic words. Infants listened significantly longer to the mispronounced (i.e., iambic) Target words ($M=9.16$, $SD=2.00$) relative to the iambic Novel words ($M=8.41$, $SD=1.98$), $t(25)=2.055$, $p=.05$, $d=.30$ (see Figure 2). These results suggest that our 8-month-olds may have treated recently segmented trochaic words and their iambic versions as functionally equivalent.

It is worth noting the effect size for Experiment 3 ($d=.30$) was considerably lower than the effect size from Pelucchi et al.'s (2009a) first experiment ($d=1.26$), which used the same familiarization language (see Figure 3). The only notable difference between these two experiments is that Pelucchi et al. (2009a) maintained the same trochaic stress pattern between familiarization and test (i.e., HTP and Novel test words were trochaic), whereas we changed the stress pattern (i.e., our HTP and Novel test words were iambic). These findings suggest that although infants may continue to recognize the Target words when they are produced with a different (i.e., iambic) stress pattern, they appear to be less well recognized than the original (i.e., trochaic) versions of the Target words segmented from the speech stream. Thus, changing the stress pattern might disrupt the strength of the HTP word representations, at least to some degree.

Discussion

The present study was an investigation of the specificity of infants' acoustic representations of newly segmented words. In a series of experiments, we examined English-learning 8-month-olds' ability to generalize representations of words recently segmented from continuous speech across indexical (i.e., talker voice; Experiment 1), segmental (i.e., consonant change; Experiment 2), and suprasegmental (i.e., stress pattern; Experiment 3) information. Importantly, we familiarized infants with an unfamiliar natural language to ensure that we were testing nascent, not previously formed, representations. In Experiment 1, we tested TP-based segmentation and replicated previous findings that 8-month-olds can use statistical cues to segment words from a female-produced unfamiliar natural language (Pelucchi et al., 2009a, 2009b, Karaman & Hay, 2018; Lew-Williams et al., 2011). Further, we found that infants successfully recognized the HTP words even when they were produced by a male voice. In Experiments 2 and 3, instead of testing statistical learning per se, we focused on the segmental and suprasegmental specificity of infants' representations of newly segmented Target words by testing their ability to discriminate mispronounced versions of those target words from Novel words. Although infants in Experiment 3 successfully recognized the Target words when they were produced with a violation of their original stress pattern (i.e., trochaic → iambic), infants in Experiment 2 failed to recognize the Target words when they were produced with a different onset consonant.

It is interesting that our 8-month-olds were able to successfully generalize across speaker information, whereas the 7.5-month-olds tested by Houston and Jusczyk (2000) and the 8-month-olds tested by Graf Estes and Lew-Williams (2015) were not. One possible account for the differences in findings between our study and Houston and Jusczyk (2000) may have to do with

differences in the methodologies used. While we familiarized infants with natural language corpora and subsequently tested them with isolated words, Houston and Jusczyk (2000) did the reverse – they familiarized infants with isolated words and tested them with passages produced by a speaker of the opposite gender. Thus, infants may have performed better in our task because the passage-word order is an easier task compared to word-passage order (Nazzi et al., 2014). A second possibility is that the distinction between our female and male talkers was less salient than between the talkers used by Houston and Jusczyk (2000), leading our infants not to generalize but instead to fail to differentiate our talkers.

However, an examination of the acoustic differences in the female and male voices used across the two studies suggests that this is unlikely to account for the observed differences. The average fundamental frequency (F_0) of the Target words produced by our female and male talkers was 247 Hz and 141 Hz, respectively. In comparison, the average F_0 of the Target words used by Houston and Jusczyk (2000) was 306 Hz and 227 Hz for female and male talkers, respectively. A final possibility is that infants may have been less good at differentiating our male and female voices because they were speaking an unfamiliar language. Indeed, Johnson and colleagues (2011) found that 7-month-olds show weaker talker recognition in foreign languages. However, they only tested infants on female voices, and all of the voices they used within each language had very similar acoustic properties (e.g., F_0 differences of less than 50 Hz). Although we acknowledge that F_0 is not the only feature that allows for the categorization of female versus male voices, it does represent a distinctive cue (e.g., Pernet & Belin, 2012). Thus, as the 106 Hz difference in F_0 between our two talkers was greater than the differences used by either Johnson and colleagues (2011) or Houston and Jusczyk (2000), it is unlikely that results from Experiment

1 reflect a failure to differentiate our female and male voices (stimuli are available here: <https://anonymous.4open.science/r/InfantSpeechSegmentation-94BD/>).

Although Graf Estes and Lew-Williams (2015) found that 8-month-old infants were able to recognize Target words produced by a male speaker following familiarization with an artificial language produced by female speakers, infants required considerable between-talker variability during familiarization to achieve this level of representational flexibility. Indeed, when infants were familiarized with just two female talkers, they failed to recognize Target words produced by a third female talker. Why then were our infants able to recognize Target words produced by a male speaker following familiarization with a single female speaker? We suggest that this might have to do with the difference in the amount of variability available in artificial versus natural language materials. In contrast to artificial languages, natural fluent speech contains varied phonological, prosodic, co-articulatory, and acoustic information that may better facilitate forming word-like representations (see van Heugten & Johnson, 2012). Thus, while we may have been able to provide infants with sufficient variability to facilitate generalization across indexical information with a single speaker producing natural language materials, cross-talker variation may be necessary to facilitate generalization of artificial language materials, at least at this age (see Graf Estes, 2012 for evidence of cross-sex generalization in 11-month-olds). Future work should replicate this finding and test the effects of within and between talker variability on newly formed representations that allow infants to generalize across talkers.

The idea that variability may help with generalization is not new. Examples of within-talker variation promoting generalization come from work by Maye and colleagues (Maye et al., 2008), who found that within-talker distributional information can promote phonemic category

learning in 8-month-old infants. Further, there is ample evidence that high between-talker variability promotes early word learning by converging the infants' attention on the features that are shared within different productions of the same word (Rost & McMurray, 2009). Here, we suggest that the variability present in our naturally produced corpus may have allowed our infants to generalize across indexical information. Another by-product of the variability present in the natural corpora employed in the current study is that it may have encouraged infants to focus their attention on segmental similarities of the phonetic features within different productions of the Target words (Rost & McMurray, 2009), facilitating the formation of proto-phonemic categories. This, in turn, may have led infants in Experiment 2 to fail to recognize the Target words when the onset consonant was changed by a single feature of place or manner of articulation.

Although Experiment 2 was not designed to test statistical learning per se, from a statistical learning perspective, infants' failure to recognize the mispronounced Target words is somewhat surprising. Even when the first consonants of the Target words were changed, the TP information between the other sounds within the words were maintained and this, in theory, could have driven recognition. For example, when the Target word *melo* from the corpus was changed to *belo* at test, the TP between the remaining segments (i.e., *e* and *l* and *l* and *o*) remained high and intact. Never-the-less, infants failed to differentiate these mispronounced Target words from never before heard foils, suggesting that the segmental information may have had a stronger impact on infants' recognition of the Target words than the between-segment TP information.

Our findings are consistent with previous research which demonstrates that 6- to 8-month-old infants are sensitive to consonant and vowel changes in recently segmented words

(Johnson et al., 2014; Nishibayashi & Nazzi, 2016), but extend this work to show that infants are able to form at least somewhat detailed phonemic representations of words just segmented from an unfamiliar natural language. This finding is particularly interesting given that while 8-month-olds generally demonstrate good discrimination of single feature segmental changes (e.g., a change from *bih* to *dih*), 14-month-olds typically fail to attend sufficiently to the phonetic details in novel isolated words necessary to successfully map minimal pair labels to novel objects (e.g., Fennell & Werker, 2003; Stager & Werker, 1997). Thus, auditory-only segmentation or discrimination tasks may support a higher degree of segmental specificity than tasks that require infants to map auditory and visual information (see PRIMIR; Werker & Curtin, 2005).

In Experiment 3 we found that infants recognized recently segmented Target words following a change in stress pattern between familiarization and test, suggesting that they are able to generalize across suprasegmental information. These findings are consistent with work by Singh et al. (2008) who found that by 9 months of age, infants can recognize familiar words across changes in suprasegmental cues of amplitude and pitch. However, when they tested younger infants of 7.5 months of age, they failed to recognize words across changes in pitch, but not amplitude (see also Bortfeld & Morgan, 2010), suggesting sensitivity to some suprasegmental cues early in development. Interestingly, our effects size comparisons (for more discussion see below) suggest that stress information may indeed be represented to some degree in these newly segmented words.

It is also important to acknowledge that similarities in the prosodic features between our infants' native language (i.e., English) and the language of the experimental materials (i.e., Italian) may have contributed to the observed patterns of results in Experiment 3. For example, based on data from Spanish-speaking adults tested with Italian stimuli, Alfano and colleagues

(Alfano et al., 2007) suggest that acoustic information of both the participant's native language and the test language play an important role in the recognition of words with mispronounced stress patterns. Although English is a stress-timed language, and Italian is a syllable-timed language, they have very similar prosodic characteristics. Specifically, the difference in the duration of stressed versus unstressed syllables – a cue that helps listeners interpret stress patterns – is similar across both English and Italian (Sundara & Mateu, 2018). Thus, our infants may have interpreted the lexical stress in the Italian stimuli differently had they come from a language background that shared fewer features with Italian. For example, French learning infants may be even more likely to generalize across stress changes following exposure to Italian because they are less sensitive to stress information in general (Hohle et al., 2009) and thus would be less likely to integrate stress cues into their representations of newly segmented words. Conversely, English learning infants may be less likely to generalize across stress changes following exposure to a language that was prosodically more similar, such as German. In this case, the similarities in the way stress is realized between the two languages may lead English-learning infants to better integrate stress cues into their representations. In addition to testing how the degree of similarity between the native language and the experimental stimuli might impact infant's representations of newly segmented words, future work could also explore how infants represent novel words embedded in a native language corpus.

Although our data demonstrate that infants continue to recognize Target words even when they are spoken by a different talker or are produced with a different stress pattern, a remaining question is whether indexical and suprasegmental changes have more subtle effects on recognition. The current set of studies was not designed to test this question explicitly; however, a Forest Plot that provides a visualization of effect sizes across studies using similar

familiarization languages is revealing. Interestingly, changing the speaker from female to male, as we did in Experiment 1, did not appear to meaningfully alter the effect size ($d=.62$) relative to previous studies (i.e., Pelucchi et al., 2009, $d=.70$; Karaman & Hay, 2018, $d=.76$) that have used the same materials but not changed the speaker between familiarization and test. These findings suggest that following segmentation infants may have fairly abstract indexical representations of naturally produced statistically-defined words.

As seen in the Forest Plot, a similar effect size comparison between our 3rd experiment, where we changed the stress pattern of the Target words between familiarization and test ($d=.30$), and Pelucchi et al.'s (2009a) 1st experiment, where the same familiarization language was used but no changes in word level stress were made ($d=1.26$), reveals that suprasegmental changes do appear to impact recognition. This is true even though infants were tested on a much easier contrast in Experiment 3 (i.e., Target words with high TP vs Novel words) than in Experiment 1 (i.e., high TP vs low TP words), which should have facilitated discrimination. The iambic realization of the Target words may have provided infants with less good exemplars of the newly segmented words, thereby causing a mismatch between the instances heard at test and the nascent suprasegmental representations formed during familiarization. These findings are consistent with work by Campbell et al. (2019) who demonstrated that while 17-month-olds can ultimately recognize familiar words with mispronounced stress patterns (e.g., BAby produced as baBY), they are slower to fixate the target object when words are mis-stressed.

Consistent with our findings in Experiment 2 that infants failed to differentiate mispronounced Target words from Novel words, a visual examination of our effect sizes (see Figure 2) reveals that the segmental mispronunciations tested in Experiment 2 meaningfully impacted the effect size between our study ($d=.01$) and that of Pelucchi et al.'s (2009a) 2nd

experiment in which no segmental modifications were made ($d=.49$). Together, these effect size comparisons suggest that at 8 months of age infants may form abstract indexical representations and specific segmental representations of newly segmented words, with suprasegmental representations falling in between.

Our findings not only replicate previous work demonstrating robust statistical learning abilities in young language learners, but for the first time gauge the specificity of nascent lexical representations. Whereas infants appear to generalize across indexical (talker) and to a lesser extent across suprasegmental (stress pattern) information, segmental information appears to be more veridically coded in young infants' representations of recently segmented words. This may be because English-learning infants have learned to attend to the characteristics in the input that are most relevant to word meaning (i.e., segmental information) and disregard acoustic features that are less relevant to word meaning (i.e., indexical, and suprasegmental information). Obviously, we are not suggesting that infants already have mature consonant-based phoneme categories at this age (we know that perceptual narrowing is just starting to be evident; Maurer & Werker, 2014). Nevertheless, their language-general perceptual skills combined with the variability afforded by the natural language stimuli used here may have helped infants notice a segmental change between the target words in familiarization and mispronounced versions of those words, much in the same way that they are able to differentiate familiar words (e.g., *doll* vs. *ball*). On this account, infants' representations of words newly segmented from natural language appear to contain information that will ultimately be relevant to word meaning.

Alternatively, segmental information may be weighted more heavily as a by-product of the segmentation task. Although Experiments 2 and 3 did not test statistical learning per se, Experiment 1 provides a clear demonstration that infants can track TP information to segment

words from natural language input (for further evidence of TP-based segmentation using these stimuli see Pelucchi et al., 2009a, 2009b; Karaman & Hay, 2018; Hay et al., 2011; Lew-Williams et al., 2011). The languages used in Experiments 2 and 3 were very similar to those used in Experiment 1. The primary difference between the experiments came during the test phases in which we varied indexical, segmental, or suprasegmental information in the Target words along with the type of foils used (i.e., low TP words or Novel words). Thus, there is no reason to think that infants would be engaging in a different strategy when segmenting words from the familiarization languages used in Experiments 2 and 3. Thus, to the extent that infants are tracking TP information over segments and syllables, the computational nature of the task may bias infants to form more specific segmental representations of recently segmented words. Future work is needed to disambiguate these alternatives. Nevertheless, our findings suggest that nascent lexical representations of words segmented from natural language share common characteristics with more established lexical entries to include information that is most relevant for word meaning in their language and support the notion that speech segmentation processes play an integral role in early language development through the formation of proto-lexical representations.

Chapter 4: General Discussion

To learn language, infants must not only be able to segment words from continuous speech but also to represent the output of word segmentation at the appropriate level of acoustic detail. Extensive work on word segmentation over the first year of life has identified multiple language-general and language specific speech segmentation strategies (for a review see Sundara & Mateu, 2018). While we know that English learning 8-month-old infants can track transitional probabilities between syllables to segment words in a non-native language that has different rhythmic but similar prosodic patterns (i.e., stress pattern; Pelucchi et al., 2009a; 2009b), we do not know if these infants can do so in a non-native language that is both rhythmically and prosodically different from their native language. I investigated this gap in knowledge in Chapter 2 of this dissertation. Similarly, there is an extant literature on familiar word representations. But we know much less about how infants represent nascent word forms, newly segmented from fluent speech. Specifically, I was interested in nascent word representations across indexical, segmental, and suprasegmental information. I examined these representations in depth in Chapter 3.

In Chapter 2, I explored English learning 8-month-old infants' ability to segment statistically-defined words in non-native languages with language-specific cues that either did or did not provide redundant cues to statistically-defined word boundaries. The results of Experiment 1 replicated previous work (Pelucchi et al., 2009a; 2009b, Karaman & Hay, 2018) showing that English learning 8-month-old infants can track TP information to segment an Italian corpus in which TP information and trochaic stress pattern provided redundant information about target words in the corpus. In Experiment 2, I observed an effect of counterbalanced language; infants listened longer to HTP words in Language B (*give* and *sudi*),

replicating results of Experiment 1, but in Language A, infants failed to discriminate between the HTP (*bazi* and *limu*) versus Novel (*give* and *sudi*) words. To uncover why infants listened longer to *give* and *sudi* when they were HTP words but did not show a significant preference when they were Novel words, I conducted an acoustic analysis of the test words (Table 5) and target word tokens in the familiarization language (Table 4). The results show that two of the test words (*give* and *sudi*) were produced with a more neutral stress pattern. In addition, an acoustic analysis of tokens of the same words in the familiarization language showed that these words, on average, had acoustic measures more similar to trochaic words (i.e., longer initial syllables with higher intensity and pitch).

Together, the results suggest that we cannot conclude from infants' preference to listen to *give* and *sudi* in Language B that they only used TP information to segment HTP words because *give* and *sudi* were acoustically more similar to trochaic words and provided English learning infants with redundant cues to find word boundaries. Approximately 90% of the content words in English are trochaic (Cutler & Carter, 1987) and infants' experience with English may have facilitated infants' segmentation in Language B. This explanation may also account for infants' performance in Language A, where they failed to segment Persian target words when words did not provide any redundant information and the only cue to find word boundaries was TP between syllables.

Another possibility for this pattern of results could be that there was too much variability in the production of *bazi* and *limu* in Language A. On average, these words were produced with louder initial syllables, but longer and higher pitched second syllables. Further, each individual token was produced with acoustic characteristics similar to either trochaic or iambic or neutral stress pattern. This high degree of variability that creates inconsistencies in the stress structure

can be problematic because for example, in a trochaic bisyllabic word, vowel reduction happens in the second syllable while in an iambic bisyllabic word the vowel in the second syllable is lengthened. Thus, infants may have failed to form a single representation of these tokens in the familiarization and subsequently were unable to recognize them at test.

In Chapter 3, I investigated the specificity of nascent word representations in English learning 8-month-old infants across indexical (i.e., talker voice; Experiment 1), segmental (i.e., consonant change; Experiment 2), and suprasegmental (i.e., stress pattern; Experiment 3) cues. In Experiment 1, I familiarized infants with a female-produced natural speech stream of Italian that had four embedded target words, two HTP and two LTP words. At test, I assessed infants' ability to differentiate between male-produced HTP and LTP words. The findings not only replicated previous findings that English learning 8-month-olds can segment a nonnative Italian language but also demonstrated that infants can recognize HTP words even when they were produced by a male speaker after being familiarized with a female talker's voice. These findings suggest that 8-month-olds can generalize the output of statistical word segmentation across talker's voice. In Experiments 2, I investigated whether infants' nascent word representations are generalizable across segmental information. To that end, I familiarized infants with a naturally produced Italian language that had two embedded HTP words, and at test, infants were tested on their ability to distinguish between mispronounced HTP (e.g., *melo* → *belo*) versus Novel words. The findings showed that infants did not recognize the mispronounced versions of the HTP words, suggesting that even after a short familiarization, infants have relatively robust representation of the segmental information. To gauge infants' representation of newly segmented words across suprasegmental information, in Experiment 3 infants were again familiarized with an Italian language with two embedded trochaic HTP words and were then

tested with stress mispronounced HTP words (i.e., iambic HTP words) versus Novel iambic words. Interestingly, infants could recognize the iambic HTP words although with a smaller effect size compared to when they were tested with the trochaic HTP words versus Novel words at test (Pelucchi et al., 2009a; Experiment 1). These findings suggest that these infants' nascent word representations may not yet include stress pattern cues and that although the change of stress pattern have affected their performance, it was not a strong enough effect to block word recognition.

Why were 8-month-olds infants successful at generalizing across talker voice in Chapter 3's first experiment? The answer may lie in the inherent variability present in natural speech. During the 2 min and 20s familiarization, infants heard the target words 18 times each time with some degree of within-talker variability. From previous research, we know that variability in the language input plays an important role in forming representations (Maye et al., 2008) by converging the infant's attention on the information that is shared within different instances of the same word (Rost & McMurray, 2009). Thus, the variability in the production of the target words in the naturally produced corpus may have allowed infants to generalize across the talker's voice in Experiment 1. This relatively high degree of variability (variability in the production of the tokens but maintaining the dominant pattern of acoustic characteristics) may have also led infants to attend more to the shared features (i.e., segmental information) within different production of the target words in Experiment 2, resulting in a more robust representation of the segmental information and thus failing to discriminate between the mispronounced HTP versus Novel words.

The findings of Experiment 3 that infants could generalize across stress pattern information should be interpreted with caution. Although the results show that infants could

generalize across stress information and recognize mispronounced HTP words (i.e., iambic HTP words at test), the effect size is smaller compared to the effect size reported by Pelucchi et al., (2009a) Experiment 1 where the stress pattern of the HTP words maintained form familiarization to test, suggesting that 8-month-old infants may represent stress information to some degree but are not robust enough to override segmental cues in the mispronounced version. It is essential to reiterate that English (infants' native language) and Italian (the test language) have a very similar stress structure. Specifically, most content words in both languages have a strong/weak stress pattern with a longer duration of the stressed syllable compared to the unstressed syllable. As both infants' native language and the test language have an essential role in the recognition of words with mispronounced stress (Alfano et al., 2007), the pattern of results could be different if a different language was used. For example, French is a language that does not use stress at word level, so if I had tested French learning infants with the same Italian stimuli, I might have observed a different pattern of results. French learning infants are not as sensitive to stress patterns as English learning infants are and thus were more likely to generalize across changes in the stress pattern. Overall, cross-linguistic differences should be taken into account when interpreting the findings and future research is needed to shed light on these language-specific differences.

Together, the findings replicated previous studies indicating robust statistical learning abilities in infants, and for the first time reveal how infants represent acoustic information in newly segmented statistically-defined words. While infants appear to generalize across indexical (i.e., talker) and suprasegmental (i.e., stress pattern) cues, segmental information appears to be more robust in young infants' representations of statistically segmented words. One reason might be that English-learning infants may have learned to attend to the properties in the input that are

most relevant to word meaning (i.e., segmental information) and disregard acoustic features that are less relevant to word meaning (i.e., indexical, and suprasegmental information).

Alternatively, infants may rely more heavily on segmental information because of the nature of the statistical segmentation tasks, in which infants appear to compute statistics over segments and syllables. Thus, the nature of the task may allow infants to form more specific segmental representations of statistically-defined words. Future work should address these alternatives. Nevertheless, the findings suggest that statistically-derived nascent word representations share common properties with familiar words to include information that is more essential for word meaning in their language, thereby bolstering the idea that statistical segmentation may play a crucial role in early language development.

One interesting finding from aggregating the findings of Chapter 2 and 3 is that English learning infants may need redundant stress cues to be able to segment nonnative speech. However, after successfully segmenting the speech stream, they do not appear to need stress pattern information to generalize the extracted words (as shown in Experiment 3 of Chapter 3). It appears that with the presence of redundant cues from TP and word-level stress, infants are able to form relatively robust representations of the newly segmented words and thus generalize these representations across stress information.

Study Limitations and Future Research

In the present dissertation, I explored infants' statistical word segmentation in the presence and absence of the redundant cues in nonnative speech and the specificity of nascent word representations. Although I used naturally-produced speech stimuli to increase the ecological validity of the test material, to control for any potential confounds, I had to test infants in lab-based settings where all natural environment noise was canceled. Therefore, to what extent

these findings are comparable to real-life settings is unclear. Future research should more systematically address this caveat in experimental design.

A limitation from Chapter 2 is that because of the observed acoustic inconsistencies in the Persian material (observing evidence for stress pattern while words were supposed to be produced with neutral stress), it is hard to draw any strong conclusions. In language development research, there is always a trade-off between using natural language material to increase ecological validity of the stimuli and the variabilities natural speech introduces into the research that may cause confounds. Future research is needed to control for such variabilities in natural language material and repeat this study to validate the suggested effects.

Another limitation of this work is the role of cross-linguistic differences in infants' behavior. Here I only tested English learning infants and used only Italian (except for Experiment 2 of Chapter 2) as the test language. The similarity between stress structure in both languages could be driving the observed effects and it is possible that if I had used another group of infants learning a different language like French, I would have observed different results. Cross-linguistic differences play a major role in infants' behavior and future research should replicate these effects using other test languages and infants learning other languages.

Conclusion

In the current dissertation, I investigated the English learning 8-month-old infants' word segmentation in languages that have language-specific cues different from those of English and the specificity of the output of word segmentation across the talker's voice, change of onset consonant, and stress pattern. Across 5 experiments and using natural speech produced to assess infants' statistical segmentation, I found that infants' statistical segmentation ability at 8 months can benefit from language-specific cues when they provide redundant information. Further, I

found that the 8-month-olds can generalize the output of infants' statistical word segmentation across talker's voice and stress pattern but not across segmental information that are key to word meaning.

This study provides evidence on first what sources of information infants utilize in non-native speech segmentation, and second what type of information they encode after segmentation. Future research should replicate these findings across different languages and developmental stages of young language learners.

References

- Alfano, I., Llisterri, J., & Savy, R. (2007). The perception of Italian and Spanish lexical stress: A first cross-linguistic study. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 6-10).
- Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, 52(4), 1135.
- 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.
- Bates, D. (2005). Fitting linear mixed models in R. *R news*, 5(1), 27-30.
- Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, 89(5), 1567-1576.
- Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253-3258.
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996-2009.
- Boersma, P., & Weenink, D. (2001). Praat speech processing software. Institute of Phonetics Sciences of the University of Amsterdam.
- Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive psychology*, 60(4), 241-266.

- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46(2-3), 217-243.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33-B44.
- Cannistraci, R. A., Dal Ben, R., Karaman, F., Parvanezadeh Esfahani, S., & Hay J. F. (2019). Statistical learning approaches to studying language development. In J. von Koss Torkildsen & J. Horst, *The Handbook of Language Development*. Abingdon, UK: Routledge.
- Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and Production of Fluent Speech*, 133–163. Hillsdale, NJ: Erlbaum.
- Curtin, S. (2009). Twelve-month-olds learn novel word–object pairings differing only in stress pattern. *Journal of Child Language*, 36(5), 1157-1165.
- Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, 105(4), 376-385.
- Cutler, A. (1979). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 167–180). New York: Academic Press.
- Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, 96(3), 233-262.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3-4), 133-142.

- Cutler, A., & Isard, S. D. (1980). The production of prosody. In *Language production* (pp. 245-269). Academic Press.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24(3), 381-410.
- 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(4), 385-400.
- 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.
- De Casper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 17(5), 481-491.
- Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36(2), 202-225.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.
- Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46(2-3), 245-264.
- Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, 54(3), 287-295.
- Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology*, 3, 447.

- Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18(3), 254-260.
- Graf Estes, K., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, 51(11), 1517.
- Goyet, L., Nishibayashi, L. L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *Plos one*, 8(11), e79646.
- Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, 63(2), 93-106.
- Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, 17(6), 610-641.
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32(3), 262-274.
- Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1570.
- Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, 7(3), 504-509.
- 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(4), 548-567.

- Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS one*, 9(1), e83546.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002-1011.
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675-687.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3-4), 159-207.
- Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Effects of talker variability on speech perception by 2-month-old infants. *Cognition*, 43(3), 253-291.
- Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, 23(2), 105-109.
- Karaman, F., & Hay, J. F. (2018). The longevity of statistical learning: When infant memory decays, isolated words come to the rescue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 221.
- Krogh, L., Vlach, H., & Johnson, S. P. (2013). Statistical learning across development: flexible yet constrained. *Frontiers in Psychology*, 3, 598.
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6), 1668-1679.

- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, 1(3-4), 237-264.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, 2(0), 734.
- Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14(6), 1323-1329.
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56(2), 154-178.
- Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2), 91-121.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1), 122-134.
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491-497.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143-178.
- Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54(3), 283-299.

- Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language*, 41(3), 600-633.
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, 2(2), 203-230.
- Nishibayashi, L. L., Goyet, L., & Nazzi, T. (2015). Early speech segmentation in French-learning infants: Monosyllabic words versus embedded syllables. *Language and Speech*, 58(3), 334-350.
- Nishibayashi, L. L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, 155, 188-203.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674-685.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244-247.
- Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, 3, 23.
- Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, 17(2), 198-232.
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Rahmani, H., Rietveld, T., & Gussenhoven, C. (2015). Stress “deafness” reveals absence of lexical marking of stress or tone in the adult grammar. *PloS one*, 10(12), e0143968.

- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. Wiley Interdisciplinary Reviews: Cognitive Science, 1(6), 906-914
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339-349.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Sahni, S. D., Seidenberg, M. S., & Saffran, J. R. (2010). Connecting cues: Overlapping regularities support cue discovery in infancy. *Child Development*, 81(3), 727-736.
- Sandoval, M., & Gómez, R. L. (2016). Overriding the metrical bias with lexical information: English-learning 7.5-month-olds use mommy to segment iambic words. *Language Learning and Development*, 12(4), 398-412.
- Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? Evidence from English-learning infants hearing Italian. *Cognitive Science*, 42(8), 3083-3099.
- Singh, L. (2018). Bilingual infants demonstrate advantages in learning words in a third language. *Child Development*, 89(4), 397-413.
- Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51(2), 173-189.
- Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4(2), 157-178.

- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., ... & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12(6), 914-919.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381.
- Storri, D., Zaar, J., Cooke, M., & Mattys, S. L. (2018). Sound specificity effects in spoken word recognition: The effect of integrality between words and sounds. *Attention, Perception, & Psychophysics*, 80(1), 222-241.
- Sundara, M., & Mateu, V. E. (2018). Lexical stress constrains English-learning infants' segmentation in a non-native language. *Cognition*, 181, 105-116.
- Swingle, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5), 432-443.
- 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.
- Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3(1), 73-100.
- van Heugten, M., & Johnson, E. K. (2012). Infants exposed to fluent natural speech succeed a cross-gender word recognition. *Journal of Speech, Language, and Hearing Research*. 55(2), 554-560.
- Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336-353.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197-234.

Werker, J. F., & Gervain, J. (2013). Speech perception in infancy: A foundation for language acquisition. *The Oxford Handbook of Developmental Psychology*, 1, 909-925.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49-63.

White, L., Payne, E., & Mattys, S. L. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. *Phonetics and phonology: Interactions and interrelations*, 137-158.

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, 118(52).

Appendix

Familiarization Languages

Chapter 2; Experiment 1

Language A

HTP words: fuga, melo Novel words: pane, tema

Torno a casa con le bici cariche di frutta in bilico sulla sella.

La zia Carola si è esibita in una fuga colla bici verde.

Se porti il melo sulla bici forse cali un po' di chili.

La bici ha subito un danno dentro la casa del capo di Lara.

La cavia Bida è in fuga da casa per aver giocato con le bilie blu.

La biscia in lenta fuga dal giardino capita in casa mia.

Il tuo melo arcano fuga l'afa che debilita la folla.

Arriviamo in bici fino al bivio del grande melo con un caro amico.

Il picchio si abitua a fare la sua casa in ogni melo cavo e alto.

Gusto i bigoli dentro casa o coricata all'ombra del melo verde.

Di rado una bici in rapida fuga rincorre la moto bigia e rossa.

Per ascoltare la fuga quasi cadi sul melo e inciampi sulla biro sull'erba.

Language B

HTP words: pane, tema Novel words: fuga, melo

Torno a casa con le bici cariche di frutta in bilico sulla sella.

La zia Carola si è esibita in una tema colla bici verde.

Se porti il pane sulla bici forse cali un po' di chili.

La bici ha subito un danno dentro la casa del capo di Lara.

La cavia Bida è in tema da casa per aver giocato con le bilie blu.

La biscia in lenta tema dal giardino capita in casa mia.

Il tuo pane arcano tema l'afa che debilita la folla.

Arriviamo in bici fino al bivio del grande pane con un caro amico.

Il picchio si abitua a fare la sua casa in ogni pane cavo e alto.

Gusto i bigoli dentro casa o coricata all'ombra del pane verde.

Di rado una bici in rapida tema rincorre la moto bigia e rossa.

Per ascoltare la tema quasi cadi sul pane e inciampi sulla biro sull'erba.

Chapter 2; Experiment 2, Language A

HTP words: bazi, limu Novel words: give, sudi

پرهام واسه بازی استخر رفت، شنا کرد و پتو برد

آرش یه دونه لیمو خرید چون سرما خورده بود

تارا هنگام پرنده بازی آشکارا مظلومه

پری در کلاس شنا داخل آب پرید و پاش شکست تا پتو پیچ بره دکتر

مامانش لیمو بهش داد تا بره پتو بیاره

ما از تماشای بازی پشیمانیم و روی پتو میپریم تا معاشرت کنیم

پیام شگفت زده از پدافند بیرون اومد و از شراکت شکایت کرد و به بازی رفت

آب لیمو پیش از اینکه شنا بری مفیده و شاپرک در بازی، شکار میکنه

پتو راه ششم مصرف بهینه است آن هم زمانی که شنا شنوایی رو کاهش میده
 مریم لیمو روی جو پرک ریخت تا وقتی شنا تموم شد، بخوره
 موقع بازی زیر شکوفه های لیمو شکفته در حیاط، نیکو دسر مورد پسند همه که لیمو، پنیر و شکر دارد رو میاره
 جای بعد از شنا زیر پتو لذت داره

Chapter 2; Experiment 2, Language B

HTP words: give, sudi Novel words: bazi, limu

پرهام گیوه به پا استخر رفت، شنا کرد و پتو برد
 آرشیو دونه گیوه خرید چون سرما خورده بود
 سودی هنگام پریدن آشکارا مظلومه
 پری در کلاس شنا داخل آب پرید و پاش شکست تا پتو پیچ بره دکتر
 مامانش گیوه بهش داد تا بره پتو بیاره
 ما از تماشای سودی پشیمانیم و روی پتو میپریم تا معاشرت کنیم
 پیام شگفت زده از پدافند بیرون اومد و از شراکت شکایت کرد و به همراه سودی رفت
 پیش از اینکه شنا بری سودی رو ببین و شاپرک در گیوه شکار میکنه
 پتو راه ششم مصرف بهینه است آن هم زمانی که شنا شنوایی رو کاهش میده
 سودی جو پرک ریخت تا وقتی شنا تموم شد، بخوره
 موقع گیوه پوشیدن زیر شکوفه های سیب شکفته در حیاط، سودی دسر مورد پسند همه که پنیر و شکر دارد رو میاره
 پوشیدن گیوه بعد از شنا زیر پتو لذت داره

Chapter 3

Language 1A

HTP words: fuga, melo LTP words: casa, bici

Torno a casa con le bici cariche di frutta in bilico sulla sella.

La zia Carola si è esibita in una fuga colla bici verde.

Se porti il melo sulla bici forse cali un po' di chili.

La bici ha subito un danno dentro la casa del capo di Lara.

La cavia Bida è in fuga da casa per aver giocato con le bilie blu.

La biscia in lenta fuga dal giardino capita in casa mia.

Il tuo melo arcano fuga l'afa che debilita la folla.

Arriviamo in bici fino al bivio del grande melo con un caro amico.

Il picchio si abitua a fare la sua casa in ogni melo cavo e alto.

Gusto i bigoli dentro casa o coricata all'ombra del melo verde.

Di rado una bici in rapida fuga rincorre la moto bigia e rossa.

Per ascoltare la fuga quasi cadi sul melo e inciampi sulla biro sull'erba.

Language 1B

HTP words: casa, bici LTP words: fuga, melo

Non è da me scendere dal melo in una futile fuga dalle api.

Torno a casa dalla futa con la bici piena di mele mature.

Il melo e diverse bici furono portate presso la mescita di vino.

Zio Luigi Medo è in fuga colla bici verde.

Vi fu l'età dei tentativi di fuga in bici verso il rifugio del melo antico.

Il fu Romero Rossi temeva di andare in gita colla bici nuova.

Dario fu l'ingenuo che portò una bici a casa il mese scorso.

Una fuga da casa è il sogno della topina Mela verso la libertà.

Il ratto Meco tentò la fuga da casa quando vi fu la tempesta.

Il micio Refuso medita in casa o dimena la coda sotto al melo ombroso.

Sui rami del melo che sembrano fusi c'è la casa del fuco solitario.

La fuga della stella cometa si è fermata sul melo che fu della zia.

Language 2A

HTP words: fuga, melo Novel words: pane, tema

Il giovane figlio di Marisa ha abbattuto male il melo per fare pali da lavoro per le porte.

Quella pazza di Teresa si mostra in fuga audace dal negozio.

Di solito cerco l'ombra del melo rosso presso la casa di Paco Rossi.

La biscia rossa in fuga viene da te per trovare riparo fra le macerie.

Il cane di Matilde gode il riso sotto al melo ombroso.

La zia passa le sue ferie alla spiaggia in fuga dall'afa.

Le sirene pedalano in bici fino al bivio del grande melo con un caro amico.

Ho visto la bici in rapida fuga sulla strada nevosa.

Lilla la fattucchiera si isola spesso dietro al melo antico.

Per udire la fuga quasi ho macchiato la stuoia di te bianco.

Qualora porti la terra per il melo con te sulla bici cali un po' di chili.

L'altro Lunedì` Marisa era in fuga da casa tra la paglia gialla.

Language 2B

HTP words: pane, tema Novel words: fuga, melo
 Ogni mese compro il pane e i rigatoni dal signore della Furia.
 La zia Me`da ha scritto un futile tema sul gatto artistico di Locarno.
 Gabriella ha messo il pane sulla bici per calare un po' di chili.
 Ieri ho preso in officino il tema tinto che fu della nonno Carlo.
 Tua sorella Carla ha preso il pane dalla ciotola ha usato il mestolo nuovo.
 Il gattino Refuso e' il simpatico eroe del tema che ho svolto.
 Purtroppo per comprare il pane sono andati in fumo tutti i soldi.
 Luigi vuole proprio ammirare l'orario fissato per un olare tema di storia.
 Dentro il sacco del pane c'e` anche un regalino che ti ha preso per il compleanno.
 Mescolo il tema musicale ad altra musica che era del fu Lorenzo Bianchi.
 I tuoi bimbi corrono presto a comprare il pane fresco.
 Questo mese ho scritto un tema sulla mostra gastronomica dell'anno scorso.

Language 3A

HTP words: fuga, melo Novel words: pane,tema
 Torno a casa con le bici cariche di frutta in bilico sulla sella.
 La zia Carola si è esibita in una fuga colla bici verde.
 Se porti il melo sulla bici forse cali un po' di chili.
 La bici ha subito un danno dentro la casa del capo di Lara.
 La cavia Bida è in fuga da casa per aver giocato con le bilie blu.
 La biscia in lenta fuga dal giardino capita in casa mia.
 Il tuo melo arcano fuga l'afa che debilita la folla.
 Arriviamo in bici fino al bivio del grande melo con un caro amico.
 Il picchio si abitua a fare la sua casa in ogni melo cavo e alto.
 Gusto i bigoli dentro casa o coricata all' ombra del melo verde.
 Di rado una bici in rapida fuga rincorre la moto bigia e rossa.
 Per ascoltare la fuga quasi cadi sul melo e inciampi sulla biro sull'erba.

Language 3B

HTP words: pane, tema Novel words: fuga, melo
 Torno a casa con le bici cariche di frutta in bilico sulla sella.
 La zia Carola si è esibita in una tema colla bici verde.
 Se porti il pane sulla bici forse cali un po' di chili.
 La bici ha subito un danno dentro la casa del capo di Lara.
 La cavia Bida è in tema da casa per aver giocato con le bilie blu.
 La biscia in lenta tema dal giardino capita in casa mia.
 Il tuo pane arcano tema l'afa che debilita la folla.
 Arriviamo in bici fino al bivio del grande pane con un caro amico.
 Il picchio si abitua a fare la sua casa in ogni pane cavo e alto.
 Gusto i bigoli dentro casa o coricata all'ombra del pane verde.
 Di rado una bici in rapida tema rincorre la moto bigia e rossa.

Per ascoltare la tema quasi cadi sul pane e inciampi sulla biro sull'erba.

Table 1. Words used in familiarization and testing phases of Chapter 2 Experiments 1 and 2.

Exp.	Language	HTP words (test token)	Novel words (test token)
1 Italian	A	fuga, melo	pane, tema
	B	pane, tema	fuga, melo
2 Persian	A	bazi, limu	give, sudi
	B	give, sudi	bazi, limu

Table 2. Acoustic characteristics of syllables of tokens in the familiarization languages of Experiment 1 of Chapter 2.

Target Words in Familiarization Language (Exp 1)	1 st syllable (Stressed)			2 nd syllable		
	Duration (s)	Intensity (dB)	Pitch (Hz)	Duration (s)	Intensity (dB)	Pitch (Hz)
fuga	0.291	64.12	354.67	0.124	71.12	226.82
	0.173	60.36	334.14	0.118	67.99	370.12
	0.305	68.97	271.56	0.100	73.13	257.78
	0.208	62.58	354.55	0.099	73.21	354.68
	0.341	69.52	429.89	0.209	75.74	271.71
	0.285	67.43	633.66	0.139	73.54	261.12
Average	0.267	65.49	396.41	0.131	72.45	290.37
melo	0.167	68.55	214.84	0.096	68.39	212.17
	0.136	71.10	230.24	0.110	71.63	226.45
	0.202	68.14	205.08	0.134	67.30	237.93
	0.168	68.53	220.94	0.100	67.69	195.47
	0.126	65.05	191.78	0.112	63.86	193.69
	0.258	64.46	199.22	0.152	65.47	233.49
Average	0.176	67.63	210.35	0.117	67.39	216.53
pane	0.124	74.80	206.26	0.100	64.15	195.74
	0.117	74.62	325.32	0.124	65.42	190.76
	0.210	72.86	158.80	0.149	61.60	208.70
	0.120	73.85	220.83	0.120	63.82	270.01
	0.136	74.07	462.12	0.136	64.14	236.30
	0.213	74.93	278.45	0.144	63.27	247.26
Average	0.153	74.18	275.29	0.128	63.73	224.79
tema	0.232	72.55	313.43	0.176	66.19	290.56
	0.157	73.16	219.55	0.113	70.45	196.81
	0.203	72.53	157.30	0.151	71.07	189.30
	0.091	70.59	301.27	0.132	71.64	322.32
	0.244	70.35	192.58	0.235	70.03	363.15
	0.222	74.44	272.95	0.195	69.90	227.13
Average	0.191	72.27	242.84	0.167	69.88	264.87

Table 3. Acoustic characteristics of syllables of test tokens in Experiment 1 of Chapter 2.

Chapter 2- Exp 1	1 st syllable (stressed)			2 nd syllable		
	Duration (s)	Intensity (dB)	Pitch (Hz)	Duration (s)	Intensity (dB)	Pitch (Hz)
FUga	0.316	77.46	334.98	0.186	78.09	178.94
MElo	0.272	78.77	368.66	0.230	75.73	196.22
PAne	0.271	79.41	318.75	0.225	67.63	201.15
TEma	0.253	78.93	304.09	0.246	74.74	205.40

Table 4. Acoustic characteristics of syllables of tokens in the familiarization language of Experiment 2 of Chapter 2.

Target Words in Familiarization Language (Exp 2)	1 st syllable (Stressed)			2 nd syllable		
	Duration (s)	Intensity (dB)	Pitch (Hz)	Duration (s)	Intensity (dB)	Pitch (Hz)
bazi	0.137	66.03	225.87	0.126	65.82	208.44
	0.153	68.97	282.60	0.144	63.10	269.50
	0.125	67.00	219.80	0.134	62.69	202.99
	0.185	63.66	266.75	0.196	61.56	249.85
	0.217	75.29	372.30	0.298	64.60	532.34
	0.211	75.40	371.92	0.292	64.71	538.50
Average	0.171	69.39	289.87	0.198	63.74	333.60
limu	0.127	68.30	218.46	0.137	64.35	183.37
	0.167	61.93	189.00	0.145	67.04	222.71
	0.186	63.02	383.85	0.240	63.00	366.95
	0.225	65.41	205.22	0.176	64.96	235.33
	0.147	64.61	197.39	0.162	61.69	187.10
	0.282	59.13	181.92	0.385	61.44	210.27
Average	0.189	63.73	229.30	0.207	63.74	234.28
give	0.147	66.27	215.26	0.137	66.45	203.96
	0.142	66.01	199.23	0.121	63.69	173.16
	0.232	65.75	266.14	0.248	65.77	261.85
	0.173	69.75	420.59	0.122	70.22	397.33
	0.187	69.91	350.14	0.236	67.61	338.58
	0.183	65.34	211.80	0.136	65.20	192.28
Average	0.177	67.17	277.19	0.166	66.49	261.19
sudi	0.332	71.87	806.96	0.291	67.77	474.67
	0.227	67.32	361.65	0.188	65.95	340.15
	0.234	64.64	646.62	0.181	61.99	259.33
	0.224	66.97	699.52	0.105	66.24	248.09
	0.355	73.58	978.56	0.265	66.20	516.72
	0.323	69.78	742.00	0.270	62.88	264.94
Average	0.282	69.02	705.88	0.216	65.17	350.65

Table 5. Acoustic characteristics of syllables of test tokens in Experiment 2 of Chapter 2.

Chapter 2- Exp 2	1 st syllable (stressed)			2 nd syllable		
	Duration (s)	Intensity (dB)	Pitch (Hz)	Duration (s)	Intensity (dB)	Pitch (Hz)
baZI	0.277	78.14	193.27	0.547	81.04	232.61
liMU	0.271	77.89	211.62	0.553	81.13	252.01
giVE	0.317	79.27	193.72	0.508	80.84	225.02
suDI	0.412	76.94	196.17	0.412	82.21	228.31

Table 6. Words used in familiarization and testing phases of Experiments 1, 2, and 3 (Chapter 3)

Exp.	Language	HTP words (test token)	LTP words (test token)	Novel words (test token)
1 Indexical	1A	fuga, melo (male voice)	bici, casa (male voice)	-
	1B	bici, casa (male voice)	fuga, melo (male voice)	-
2 Segmental	2A	fuga, melo (shuga, belo)	-	pane, tema (tane, sema)
	2B	pane, tema (tane, sema)	-	fuga, melo (shuga, belo)
3 Suprasegmental	3A	fuga, melo (fuGA, meLO)	-	pane, tema (paNE, seMA)
	3B	pane, tema (paNE, seMA)	-	fuga, melo (fuGA, meLO)

Table 7. Mean (SD) Pitch (Fundamental Frequency, F₀) and Mean Word Length (SD) for each word in Languages 1A/B and Test words used in Experiment 1.

Familiarization Language /Test	Speaker	Words	Pitch (F ₀)	Mean Word Length (SD)
1A	Female	bici	243.98 (31.07)	380 (59.7)
1A	Female	casa	246.98 (38.20)	405 (86.4)
1A	Female	fuga	290.15 (44.09)	407 (85.3)
1A	Female	melo	203.66 (21.79)	305 (65.3)
1B	Female	bici	247.70 (25.94)	308 (56.4)
1B	Female	casa	260.22 (34.57)	403 (73.0)
1B	Female	fuga	255.03 (19.70)	343 (86.2)
1B	Female	melo	231.98 (20.12)	329 (60.7)
Test	Male	bici	158.30 (80.55)	500
Test	Male	casa	134.92 (58.40)	500
Test	Male	fuga	145.34 (63.40)	500
Test	Male	melo	126.85 (39.44)	500

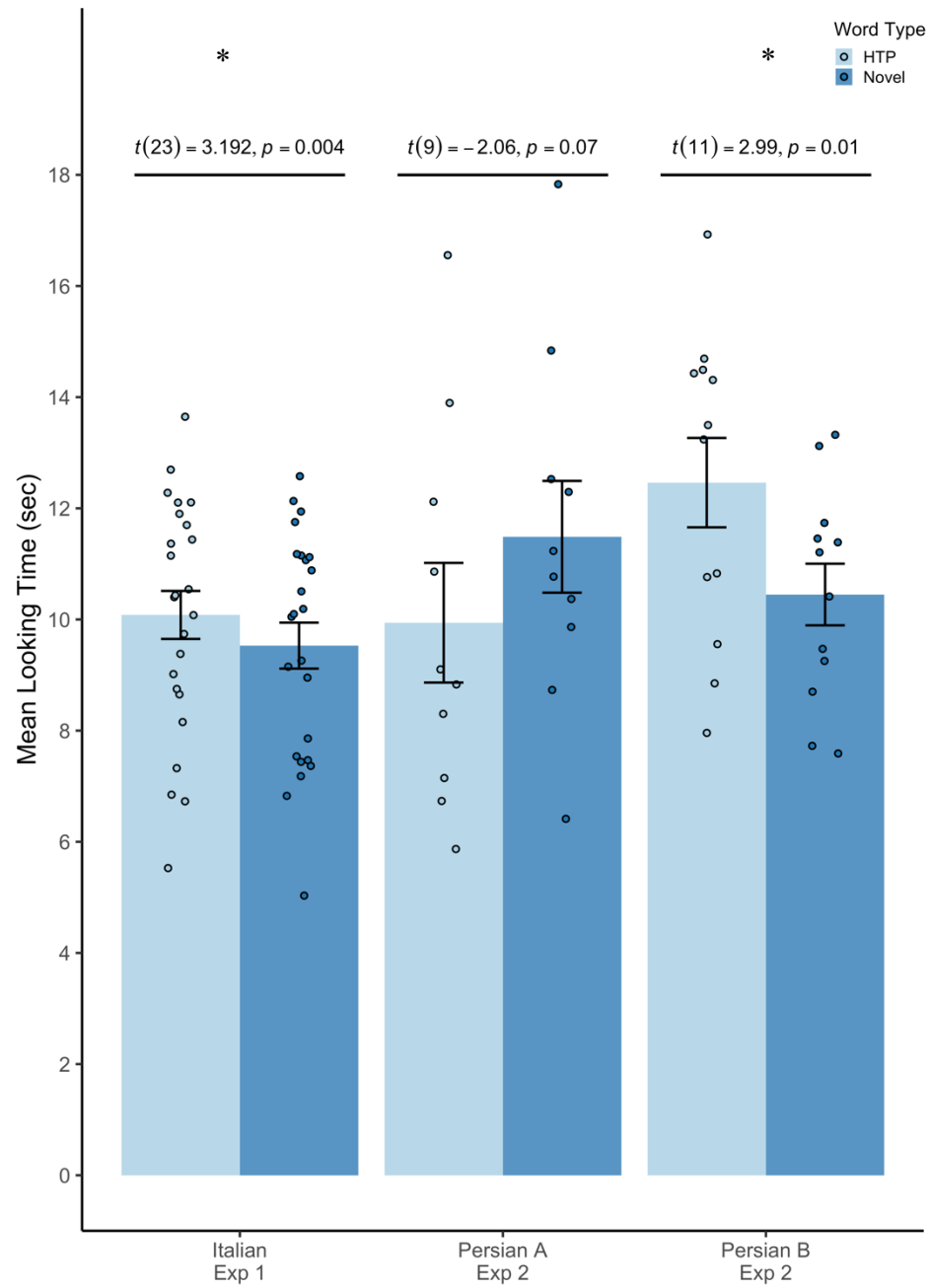


Figure 1. Mean looking time (sec) to HTP and Novel words in Experiment 1 (Italian) and Experiment 2 (Persian; Language A and Persian; Language B), Chapter 2.

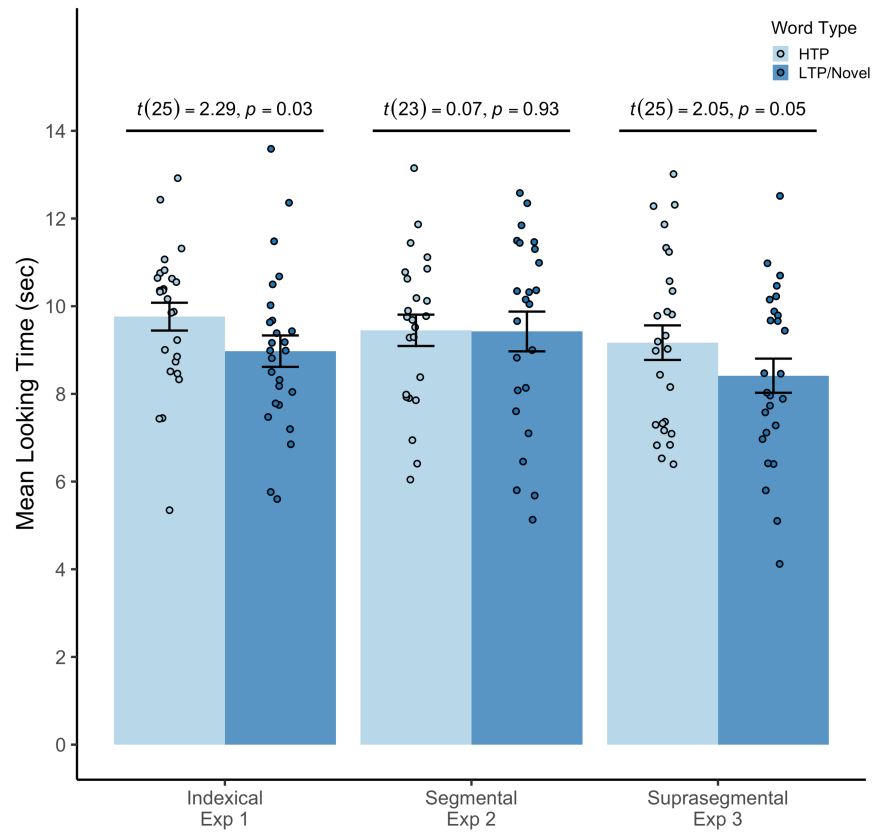


Figure 2. Mean looking time (sec) to HTP vs LTP (Exp 1) or vs. Novel (Exp 2 & 3) words

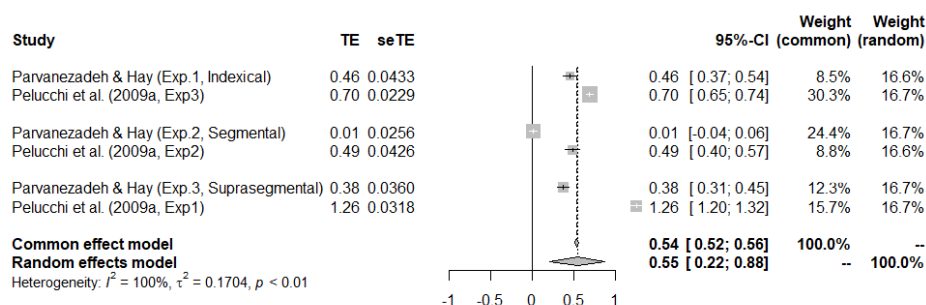


Figure 3. Forest plot providing a visualization of effect size comparisons between the current studies and Pelucchi et al.'s (2009a) experiments from which our familiarization languages were drawn (grouped according to familiarization languages used). Each experiment is represented by a line in the plot. TE = treatment effect (effect size); seTE = estimated treatment effect; CI = confidence interval; diamond = overall effect estimate; width of diamond = CI for overall effect estimate; point/box size = inverse variance weight based on effect size SE; width of line = CI for effect estimate for each study—lines are plotted in black if the CI is larger than that point/box size and in white if the CI falls within the point/box. The Forest Plot was the output of a mini meta-analysis performed in R (R Core Team, 2013), using a MetaLab script provided by Bergmann and colleagues (Bergmann, Tsuji, Piccinini, Lewis, Braginsky, Cristia, & Frank, 2018) and used previously by Karaman & Hay (2018).

Vita

Sara was born and grew up in Iran. After graduating from college with a bachelor's degree in English, she did a master's at Tehran University in Teaching English as a Foreign Language (TEFL). In 2015, she joined the Cognitive and Developmental Sciences group in the Experimental Psychology PhD program at the University of Tennessee, Knoxville. Her research interests include early speech segmentation and word representation. For her post graduate career, Sara is interested in the intersection of research and tech, and she is continuing research in tech industry as a user experience researcher. She is grateful for the support she received from her academic advisor, her spouse and family throughout this journey.