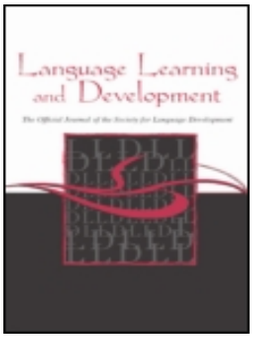

Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: https://www.tandfonline.com/loi/hlld20

Flexibility in Statistical Word Segmentation: Finding Words in Foreign Speech

Katharine Graf Estes, Stephanie Chen-Wu Gluck & Carolina Bastos

To cite this article: Katharine Graf Estes, Stephanie Chen-Wu Gluck & Carolina Bastos (2015) Flexibility in Statistical Word Segmentation: Finding Words in Foreign Speech, Language Learning and Development, 11:3, 252-269, DOI: 10.1080/15475441.2014.926730

To link to this article: https://doi.org/10.1080/15475441.2014.926730

Published online: 18 Jul 2014.

Language Learning and Development, 11: 252–269, 2015

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2014.926730

Flexibility in Statistical Word Segmentation: Finding Words in Foreign Speech

Katharine Graf Estes
Department of Psychology, University of California, Davis

Stephanie Chen-Wu Gluck
Department of Psychology, San Francisco State University

Carolina Bastos
Department of Psychology, University of California, Davis

The present experiments investigated the flexibility of statistical word segmentation. There is ample evidence that infants can use statistical cues (e.g., syllable transitional probabilities) to segment fluent speech. However, it is unclear how effectively infants track these patterns in unfamiliar phonological systems. We examined whether English-learning 14-month-olds could segment continuous speech in an artificial language consisting of Mandarin Chinese syllables. Syllable transitional probabilities provided the only reliable word boundary cues. In Experiment 1, we found that males segmented the language; they reliably discriminated words from the language versus across-word sequences. Females did not display reliable discrimination. In Experiment 2, when the test items were designed to have less overlap in their statistical structure, males and females displayed successful learning. Variation in direction of preference was revealing about infants' processing in the tasks. Taken together, the findings indicate that infants' statistical word segmentation abilities are sufficiently robust for processing speech streams containing foreign sounds.

From early in life, infants are adept at learning by tracking the statistical patterns that are present in their input. Powerful statistical learning mechanisms are hypothesized to help infants discover the underlying structure of many aspects of their environments (see recent reviews in Krogh, Vlach, & Johnson, 2013; Romberg & Saffran, 2010). One specific process for which infants may use statistical learning is to segment individual words from the fluent stream of speech. There is evidence that infants can use patterns of syllable probabilities to detect which syllables co-occur reliably and therefore can likely form words in the language (Graf Estes, Evans, Alibali, & Saffran, 2007; Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003). This is one type of statistical cue that is available for detecting words (see also, e.g., Curtin, Mintz, & Christiansen, 2005; Mattys & Jusczyk, 2001). Findings from many studies indicate that infants' ability to

Correspondence should be addressed to Katharine Graf Estes, Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, CA 95618. E-mail: kgrafestes@ucdavis.edu

use syllable probabilities to perform statistical word segmentation is quite strong, but there are remaining questions regarding how robust and flexible statistical segmentation is and how useful it might be for language acquisition (Graf Estes, 2012; Johnson, 2012; Johnson & Tyler, 2010; Mersad & Nazzi, 2012; Pelucchi, Hay, & Saffran, 2009; Yang, 2004). The present experiments examined the limits of statistical learning by testing infants' ability to apply statistical learning to segment words from linguistic input containing foreign language sound sequences.

Most prior studies of statistical word segmentation have presented infants with artificial languages (Johnson & Jusczyk, 2001; Johnson & Tyler, 2010; Saffran, Aslin et al., 1996; Thiessen & Erickson, 2013; Thiessen & Saffran, 2003). Although the languages are much less complex than natural foreign languages, they share important characteristics. Like foreign languages, the artificial languages consist of a vocabulary of unfamiliar words produced in a continuous speech stream where it is difficult to determine where each word begins and ends. In many languages designed to test segmentation, the only reliable word boundary cue is syllable sequence transitional probability information, which is calculated as the frequency of a syllable sequence XY given the frequency of the syllable X. The transitional probability patterns of artificial languages mimic a probability pattern that occurs in natural speech, though natural speech typically provides less consistent regularities (Harris, 1955; Swingley, 2005). In fluent speech, the transitional probability from one syllable to the next is high within words because these syllables consistently occur together in sequence. Across word boundaries, syllable-to-syllable probabilities are lower because these syllables occur together incidentally; a given word can be followed by a large range of other words. Infants can use probability patterns to detect the reliably-occurring word units from these unfamiliar languages, even with only a few minutes of listening experience (e.g., Saffran, Aslin et al., 1996).

Although artificial languages use novel words, they are typically based on the phonology and accent of the infants' native language. The syllable inventories also frequently include real words (e.g., tea, dough, boo, go, two, bee, row) (Graf Estes et al., 2007; Johnson & Jusczyk, 2001; Saffran, Aslin et al., 1996; Thiessen & Saffran, 2003). In phonological processing tasks with older children, similarity to native language sound patterns facilitates processing. For example, children are better at repeating novel words that are word-like compared with those with unusual phoneme combinations (e.g., Edwards, Beckman, & Munson, 2004; Gathercole, 1995). Similarly, during statistical word segmentation, infants may benefit from the familiarity of the phonological system when attempting to detect new words in continuous speech.

There is evidence that infants' statistical segmentation abilities are sufficiently flexible to support tracking sound sequence regularities in a foreign language. Pelucchi and colleagues (2009) presented 8-month-old English learners with Italian sentences. The sentences used real words in grammatically correct constructions, but they were carefully designed so that some words had high internal transitional probabilities and other words had low internal probabilities. The authors found that infants differentiated the high versus low probability words, indicating that they had detected the statistical structure of the foreign speech stream (see also Hay, Pelucchi, Graf Estes, & Saffran, 2011). However, to maintain the natural characteristics of Italian, the speech stream included its typical rhythmic patterns. Rhythmic information is a powerful segmentation cue (Curtin et al., 2005; Johnson & Jusczyk, 2001; Johnson & Seidl, 2009) that could contribute to English learners' successful segmentation of Italian. Houston, Jusczyk, Kuijpers, Coolen, and Cutler (2000) proposed that English-learning infants' success in segmenting Dutch words could be due to rhythmic similarities across languages. Although Dutch and English (both

stress-timed languages) are more rhythmically similar than Italian (syllable-timed) and English (Nespor, Shukla, & Mehler, 2011), there is overlap in English and Italian rhythmic characteristics. Specifically, Italian uses the strong-weak syllable stress pattern that occurs in many English words (Cutler, 2008). In Pelucchi et al.'s (2009) experiments, the target words followed this trochaic stress pattern. Infants' differentiation of high and low transitional probability sequences established that they detected the probability patterns in the speech stream. However, the possibility remains that infants could have applied language-specific segmentation cues, based on their experience with English words, in order to do so.

It is not clear whether infants can segment foreign language speech streams in which supplemental cues are eliminated. To do so, infants must take advantage of language-general cues, rather than prior knowledge of language-specific word boundary cues. This is an important issue because it addresses the range of the robustness and flexibility of statistical learning. Is statistical word segmentation based on syllable probabilities a sufficiently powerful process that infants can apply it to novel speech streams that do not follow directly from their prior linguistic experience? Or is statistical word segmentation limited to speech streams that contain only native language phonology?

In the present experiments, we addressed whether infants can use a language-general word segmentation cue, syllable transitional probabilities, to segment a speech stream with unfamiliar phonology when other native language-specific cues, such as rhythmic information, were unavailable. Segmenting a speech stream with unfamiliar phonology requires a learning mechanism that is sufficiently flexible to adapt to complex stimuli that differ from learners' prior experience. This work contributes to defining the power and limits of statistical word segmentation using transitional probabilities, and of statistical learning more broadly. Statistical learning has received substantial attention as an underlying mechanism of many aspects of language acquisition (see recent reviews in Krogh et al., 2013; Romberg & Saffran, 2010; Thiessen, Kronstein, & Hufnagle, in press), but there is limited understanding of how far statistical learning can go in explaining acquisition. In particular, there have been questions regarding whether statistical word segmentation using syllable probabilities is an effective learning process (Johnson, 2012; Johnson & Tyler, 2010; Mersad & Nazzi, 2012; Yang, 2004).

We tested English-learning infants using a novel manipulation of foreign language materials to investigate the flexibility of statistical learning. We designed an artificial language that consisted of a fluent stream of Mandarin Chinese syllables. Successful segmentation of this language requires flexibility because infants must track probabilities in sounds that are unfamiliar; they must learn from input that is dissimilar to their prior experience. There is some previous research indicating that segmenting naturally produced passages of Mandarin is challenging for English-learning infants. Tsay and Jusczyk (2003) familiarized 7.5-month-olds with repetitions of Mandarin words produced in isolation. During testing, the infants gave no indication that they detected the familiarized words when they were presented in fluent passages; they failed to differentiate passages with familiarized words from passages with unfamiliar words (see also Newman, Tsay, & Jusczyk, 2003). In contrast, Jusczyk and Aslin (1995) demonstrated that 7.5-month-olds succeeded in detecting native language words in a similar task. Segmenting Mandarin may be difficult for English-learning infants for many reasons, including its unfamiliar phonology, use of pitch contour, and the lack of English-like rhythmic patterns.

In the present experiments, we focused on how the presence of unfamiliar speech sounds affects infants' ability to detect a language-general segmentation cue. The goal in the present

research was to investigate learners' use of statistical regularities to segment speech, rather than English-learning infants' segmentation of Mandarin per se. Therefore, we used an artificial language, rather than natural speech, to isolate transitional probability cues to word boundaries. Mandarin differs from English in many ways. For example, unlike English, Mandarin is a tonal language, meaning that differences in pitch contour can determine word meanings. It also has different rhythmic and phonotactic patterns from English (Duanmu, 2000). A pertinent cross-linguistic difference for the design of the present experiments is that Mandarin and English phoneme inventories differ; Mandarin contains phonemes that do not occur in English (Dryer & Haspelmath, 2011; Duanmu, 2000; Li & Thompson, 1989). The Mandarin-based artificial language incorporated phonemes unfamiliar to English learners. In order to maintain transitional probability as the sole word segmentation cue, we did not incorporate language-specific pitch, rhythmic, or phonotactic patterns in the artificial language. Pelucchi et al. (2009) found that infants could detect transitional probabilities in a foreign language, at least when natural rhythmic patterns were also present. Other segmentation tasks with foreign languages have allowed language-specific and language-general segmentation cues to occur naturally rather than to isolate transitional probabilities (e.g., Houston et al., 2000; Tsay & Jusczyk, 2003). In the present experiments, the limited segmentation cues and the unfamiliar phonology of the speech stream established a substantial challenge for infants' ability to track syllable sequence regularities. We investigated whether infants can use statistical learning to segment words in an acoustically complex, perceptually distinct foreign language using only transitional probability cues to word boundaries.

We tested 14-month-old infants because at this age, speech perception has started to tune to native language phoneme categories. At 14 months of age, infants have already experienced language-specific perceptual narrowing (e.g., Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Werker & Tees, 1984). In order to perform a stringent test of whether infant statistical word segmentation is challenged by the presence of non-native phonology, it is crucial to test infants whose phonological systems have already begun to process native language sounds more effectively than non-native sounds (Kuhl et al., 2006). For younger infants, the presence of nonnative phonology during statistical segmentation could be irrelevant if they process native and non-native phonemes similarly.

During the segmentation task, infants listened to a fluent speech stream consisting of four statistically defined disyllabic words produced in Mandarin. During testing, infants heard repetitions of the words, produced in isolation, as well as repetitions of sequences that crossed word boundaries in the language, termed part-words (Experiment 1) or novel sequences of syllables from the language, termed nonwords (Experiment 2). If statistical word segmentation is sufficiently flexible that infants can apply it to foreign linguistic materials, they should recognize the segmented words during testing. In measures of listening time, infants should differentiate the high transitional probability words from sequences with low transitional probability.

EXPERIMENT 1

In Experiment 1, infants listened to an artificial language consisting of Mandarin syllables. An important characteristic of the language design was the frequency-balancing of the test items. That is, the word and part-word sequences that served as test items occurred with equal frequency

during segmentation (Aslin, Saffran, & Newport, 1998). As discussed in more detail below, the balancing was achieved by including highly frequent words that often occurred together in sequence during the speech stream (but did not act as test items), thereby forming across-word sequences that occurred regularly. These part-words were matched in frequency to the test words. Critically, the test items differed in transitional probability. During testing, if infants learned the details of the language structure, they should differentiate between the words (transitional probability = 1.0) and the sequences that crossed words (part-word transitional probabilities = .5), even though they heard them the same number of times. There were two particularly challenging components to Experiment 1. First, the lack of familiarity with Mandarin may hinder infants' ability to track syllable patterns. Second, to display learning, infants must discriminate between test items with equal frequency but different internal probabilities. The structure of the test items presented a stringent test of learning.

METHOD

Participants

Thirty 14-month-olds (16 females) participated. The mean age was 14.3 months (range 13.6–14.9 months). The infants were born full term and had no history of chronic ear infections or hearing or vision problems. All infants came from English-speaking homes. Parents of three infants reported exposure to a second language (fewer than four hours per week). No infants had exposure to Mandarin. Seventeen additional infants were excluded due to fussiness (n = 10), parental interferences (2), equipment/experimenter error (3), and moving out of view (2). Two additional infants were identified as outliers (listening preferences over 2 SD from the mean) and were excluded from the final analyses.

Stimuli

A female native Mandarin speaker produced the segmentation stimuli. She was instructed to maintain a consistent speaking rate and to use a monotone speaking style with consistent lexical tones based on Mandarin Tone 1, the high level tone. Tone 1 is characterized by a flat pitch contour (Duanmu, 2000). To produce natural coarticulation across syllable transitions, the speaker recorded a list of syllable triplets that included all of the syllable to syllable patterns that occurred in the language (similar to the methods used by Curtin et al., 2005; Graf Estes et al., 2007). She produced the syllable triplets in citation form, with a pause before each item. To prevent the tendency to produce words with word-initial stress or word-final lengthening from affecting the stimuli, we only used the middle syllables from each triplet. Each middle syllable was excised and then spliced in sequence with the appropriate coarticulation context to construct the fluent stream (i.e., the sequence *mabi* was spliced from the recording of *pao**ma**bi* and *ma**bi**shu*). This splicing method prevented the speaker from inadvertently adding supplemental boundary cues such as pauses or stress. The only reliable word markers were the transitional probabilities within words (1.0) versus across words (0 - .5). In the assembled fluent stimuli, the mean syllable duration was 275 msec for initial syllables and 264 msec for final syllables, the mean pitch was 205 Hz for

TABLE 1
Test Items Used in Experiments 1 and 2

	Experiment 1		*Experiment 2*	
	Words	*Part-words*	*Words*	*Nonwords*
Language 1	shunai, woka	bixie, paoma	shunai, woka	xiebi, mapao
Language 2	bixie, paoma	shunai, woka	bixie, paoma	naishu, kawo

TABLE 2
Phonetic Transcriptions (International Phonetic Alphabet) for Syllables in Segmentation Stimuli

Syllable (in pinyin)	*IPA*	*Notes on differences from English*
shu	ʂu	Consonant ʂ does not occur in English
nai	nai	
wo	wo	
ka	k^ha	Aspirated k, phoneme boundary differs from English /k/
xie	ɕiɛ	Consonant ɕ and dipthong iɛ not used in English
pao	p^hau	Aspirated p, phoneme boundary differs from English /p/
ma	ma	
bi	pi	Unaspirated p, phoneme boundary differs from English /p/

both initial and final syllables, and the mean amplitude was 77 dB for initial and final syllables, based on analyses performed in Praat (Boersma & Weenink, 2010).

Two counterbalanced versions of the artificial language were created to control for arbitrary listening preferences. The test items that were words in Language 1 were part-words in Language 2, and vice versa. Table 1 shows the word and part-word test items for each language. The words in Language 1 (written in Mandarin pinyin) were *mabi, xiepao, shunai*, and *woka*; the words in Language 2 were *bixie, paoma, naiwo*, and *kashu.* The syllables all represent real Mandarin words (when produced with the appropriate pitch contours). Table 2 shows the phonetic transcriptions and information about differences from English phonology, where applicable (Li & Thompson, 1989).

Both language versions were designed to equate the frequency of occurrence of the sequences that formed word and part-word test items. In order to produce the frequency balancing of the test items, each language contained two high frequency words (Language 1: *mabi* and *xiepao*; Language 2: *naiwo* and *kashu*) and two low frequency words (Language 1: *shunai* and *woka*; Language 2: *bixie* and *paoma)*. The low frequency words occurred 90 times in the speech stream, whereas the high frequency words occurred 180 times. Each high frequency word was followed by the other high frequency for half of its repetitions. This design formed two part-word sequences that occurred 90 times because they occurred across the conjunction of the high frequency words. For example, in Language 1, the high frequency word *mabi* preceded the other high frequency word *xiepao* 90 times, forming the sequence *bixie* across the word boundaries. *Bixie* was repeated the same number of times as the low frequency words in the language. The remaining repetitions of the high frequency words preceded the low frequency

words *shunai* and *woka (*e.g., in the sequence ***mabi****#xiepao#shunai#****mabi****#woka#****mabi****#shunai# woka#****mabi****#xiepao*). This design produced an internal transitional probability for the test part-words of .5, whereas the probability of the words was 1.0. Thus, the test words and part-words matched in frequency but differed in probability.

Previous research indicates that consistency of word length may promote tracking of syllable probabilities in word segmentation and inconsistent word length may hinder segmentation (Johnson & Tyler, 2010). Therefore, in this experiment (and many other statistical segmentation tasks) infants may take advantage of consistent word length to segment words. However, the test items were all the same length; successful discrimination required infants to detect precise transitional probability information.

The duration of each language was 4 min 48 sec and the average F0 was 205 Hz. Preliminary analyses revealed no differences in performance based on language, so subsequent analyses collapsed across languages.

The original speaker recorded the test items in citation form using a monotone speaking style based on Tone 1. The test items' average F0 was 206 Hz. There were four tokens of each test item, separated by 800ms of silence, repeating in a loop. To measure infants' listening time to the test items, all items were paired with a visual animation of red dots appearing in a clockwise rotation on a white background. The duration of infants' fixation to the visual stimuli allowed us to compare attention to the words and part-words.

All stimuli were played at approximately 65 dB.

Procedure

The infant and a parent first listened to the segmentation stimuli over loudspeakers for 4 min 48 seconds while playing quietly in a sound-attenuated booth. During testing, the infant sat on the parent's lap approximately 3.5 feet from a television screen with integrated speakers. The parent was briefly reminded about the testing instructions before the trials began. Because of the break before testing, the infant heard the language for 30 additional seconds, accompanied by a silent cartoon. To prevent bias, the parent listened to music on headphones and the experimenter was blind to the identity of the stimuli being presented. The experimenter monitored the infant's attention from a separate room via a camera mounted below the television. The program Habit X presented the test items and recorded listening times (Cohen, Atkinson, & Chaput, 2004).

We used an auditory preference procedure to measure infants' attention to the test items. Each infant began testing with one pre-test trial, consisting of a repeated novel word (*neem*) paired with the same visual animation as the test trials. The purpose was to give the infant experience with the audio-visual stimuli presentation.

Each test trial began with a short animated clip to capture the infant's attention. When the infant looked at the screen, repetitions of a word or part-word played. The stimuli played for as long as the infant attended, stopping when the infant looked away for 1s or after a maximum listening time of 20s. The four test items (two words, two part-words) were presented in four randomized blocks, for a total of 16 trials.

Habit X tallied the infant's attention to the test items. The dependent measure was based on how long the infant listened to the words versus part-words. This measure is similar to the central fixation procedure used by Shi and Werker (2001; Shi, Werker, & Cutler, 2006) and the

visual fixation-based auditory preference procedure used by Cooper and Aslin (1990, 1994). The listening time measure is also comparable to the headturn preference procedure frequently used in statistical word segmentation experiments (Aslin et al., 1998; Johnson & Jusczyk, 2001; Saffran, Aslin et al., 1996).

RESULTS AND DISCUSSION

A 2 (Trial type: word vs. part-word; within subjects) x 2 (Gender: male vs. female; between subjects) mixed ANOVA revealed no main effect of trial type and no main effect of gender (both Fs <1) on listening time. There was a significant interaction of trial x gender, $F(1, 28) = 4.25$, $p = .049$, $\eta_p^2 = .13$. To further examine the interaction, we performed paired samples t-tests broken down by gender. They revealed that males listened significantly longer to the words ($M = 9.36$, $SD = 2.95$) than to the part-words ($M = 8.62$, $SD = 2.69$), $t(13) = 2.73$, $p = .017$, $d = .26$. Females did not show a significant listening time difference to words ($M = 8.87$, $SD = 3.85$) versus part-words ($M = 9.28$, $SD = 3.72$), $t(15) = -.88$, $p = .39$, $d = .11$.

Figure 1 shows the listening time difference values for words minus part-words (positive values indicate longer listening to words) for individual infants. The scatterplot illustrates that the females had a wider spread in their listening times than males; their performance was quite heterogeneous. Females also showed a non-significant tendency toward longer listening to part-words, with 11 of 16 females showing this preference pattern. In contrast, the males showed reliably longer listening to words (10 of 14 males). It is somewhat surprising that females failed to show

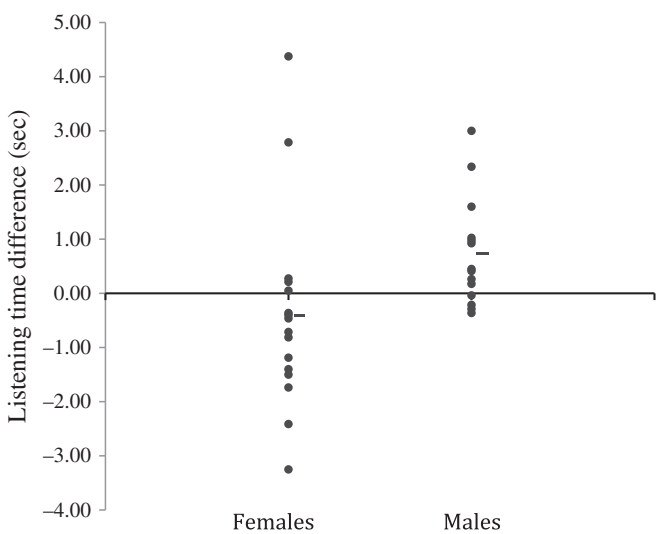

FIGURE 1 Infants' listening time differences to word—part-word test trials in Experiment 1. Dots represent individual infants, dashes represent group mean scores.

evidence of learning when males showed successful learning, given that females have outperformed males in other early language measures (Dale & Fenson, 1996; Hartshorne & Ullman, 2006; Lany & Gómez, 2008).

However, patterns in the directions of preference suggest that the statistical comparisons do not fully capture infants' learning in this task. The lack of a statistically reliable preference in the females does not necessarily indicate failure to learn (Houston-Price & Nakai, 2004; Hunter & Ames, 1988). According to Hunter and Ames's (1988) model of infant attention, as learning proceeds, infants' preferences shift from familiarity (early in learning) to novelty (late in learning). Not all infants proceed at the same pace. Therefore, the lack of a significant listening preference in the females may suggest that they differed in degree of learning and in the magnitude and direction of their preferences, yielding a null result. The presence of novelty preferences may suggest that some females learned the segmentation stimuli even more thoroughly than the males who showed familiarity preferences. This interpretation should be taken with significant caution. It is based on an examination of the distribution of data for a result that was not statistically reliable. The same distribution also could have occurred if the females displayed random variation in their listening time preferences due to a lack of learning.

Another notable aspect of the preference patterns is the word preference that the males displayed. Many statistical word segmentation experiments have reported greater attention to part-words, a pattern that is interpreted as a novelty preference following the detection of the statistically-defined words (Aslin et al., 1998; Johnson & Jusczyk, 2001; Saffran, Aslin et al., 1996; Thiessen, Hill, & Saffran, 2005, Experiment 2). Some experiments have also reported greater attention to words, a pattern that is viewed as a familiarity preference for the items that the infants recognized following segmentation (Saffran, 2001; Thiessen et al., 2005, Experiment 1; Thiessen & Saffran, 2003, Experiment 1). Hunter and Ames (1988) explained that novelty preferences occur when infants find a task relatively easy to process, such as when information has been thoroughly encoded during familiarization or when test items are readily distinguishable because they have many non-overlapping features (see also Houston-Price & Nakai, 2004). Conversely, familiarity preferences occur when a task is difficult or learning is incomplete, or when test items share features, making them hard to distinguish. Thus, the performance of the males indicates that they not only learned the structure of the artificial language but also that the task was challenging.

The design of Experiment 1 was demanding in two ways. First, the presence of Mandarin syllables required infants to process foreign language phonology to segment the speech stream. Second, the test items were inherently difficult to differentiate. The words and part-words occurred with equal frequency during segmentation. As discussed below, the frequency of exposure to both items may have led infants to consider both as possible lexical units. Furthermore, the items differed in internal probability, but the difference in probability between the words (1.0) and part-words (0.5) may have been less salient than other probability contrasts. The probabilities at other word boundaries ranged from 0 to .25, potentially forming more prominent word markers than the .5 probability boundaries. Thus, it is possible that the infants successfully tracked the statistical structure of the Mandarin syllable stream, but the test items were sufficiently difficult to differentiate that some participants failed to display their learning. To address this possibility, Experiment 2 tested whether infants can demonstrate learning when the test items provide a better opportunity to do so.

EXPERIMENT 2

In Experiment 2, infants listened to the Mandarin syllable stream from Experiment 1, but the test trials contrasted words with novel sequences of syllables from the artificial language. Models of word segmentation indicate that discriminating words versus part-words should be more difficult that discriminating words versus the novel sequences, often termed nonwords (e.g., Pelucchi et al., 2009; Saffran, Aslin et al., 1996; Saffran, Newport, & Aslin, 1996). Clustering (or chunking) segmentation models propose that infants extract reliably occurring sequences in continuous speech by tracking probabilistic information, yielding statistically coherent word-like units (see various instantiations by Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Giroux & Rey, 2009; Perruchet & Vinter, 1998; Swingley, 2005). In one influential model, PARSER, Perruchet and Vinter (1998) proposed that when associations between syllables are repeated, the sequence forms a representational unit that strengthens with each exposure. Because syllable sequences that form words are more likely to recur than syllable sequences that cross word boundaries, words emerge as strong units and across-word sequences remain weakly connected. Over time, representations of across-word sequences are lost due to memory decay and due to the interference that occurs because across-word sequences consist of components of other words.

Concepts from PARSER help explain why the part-words in Experiment 1 were difficult to discriminate from the words. The part-words occurred across word boundaries, but they occurred frequently in the speech stream, providing ample opportunity to build association strength between syllables. The input may have prevented the representations of the part-words from weakening as they typically would in a language without a frequency-balanced design. Therefore, at this point in learning, the part-words did not fully drop out of consideration as possible lexical units, making them difficult to distinguish from the words.

In contrast to the part-words, test items consisting of novel syllable sequences do not have representations strengthened by prior exposure. Therefore, words should be easy to distinguish from nonwords. Consistent with this idea, Saffran and colleagues (Saffran, Newport, et al., 1996) found that adults were more accurate in identifying words when they were paired with nonword test items than when they were paired with part-word test items. Accordingly, in Experiment 2, infants listened to the same exact Mandarin speech stream, but the test items were words from the language versus nonwords consisting of syllables from the language in a novel order. We predicted that this change in the test items would produce reliable evidence of learning with less heterogeneity in performance than in Experiment 1.

METHOD

Participants

Twenty-eight 14-month-olds (13 female) participated. The mean age was 14.0 months (range 13.3–14.8 months). All infants came from English-speaking homes. Five infants had exposure to a second language (less than 6h per week; no infants heard Mandarin). An additional 10 infants were excluded from the study due to fussiness (n = 9), and parental interference (1).

Stimuli

The artificial language was the same as the language used in Experiment 1. As shown in Table 1, the word test items were identical to those used in Experiment 1. The new test items, the nonwords, were created by reversing the test part-words (i.e., the part-word *bixie* became the nonword *xiebi*) to create a syllable sequence that never occurred during the language (transitional probability = 0). This manipulation allowed us to maintain a close match between the syllables used as words, part-words, and nonwords across experiments, since it was not possible to use the same segmentation stimuli across experiments and perform the complete counterbalancing of the test items. It is important to note that in Experiment 2, the word and non-word test items differed in frequency as well as in transitional probability. The nonwords never occurred during the segmentation phase, whereas the words occurred frequently. Both probability and frequency could contribute to the distinctiveness of the word versus nonword test items.

The original speaker recorded the test items using the same method described in Experiment 1. As in Experiment 1, preliminary analyses revealed no differences in performance across language versions (Language 1 vs. Language 2), so subsequent analyses collapsed across this factor. Given the lack of complete test item counterbalancing in Experiment 2, the equivalent performance across language versions helps to rule out the possibility that the results were due to arbitrary listening preferences for particular test items.

Procedure

The procedure was the same as in Experiment 1. The only difference was in the use of word and nonword test items. The mean pitch of the test items was 206 Hz.

RESULTS AND DISCUSSION

A 2 (Trial type: word vs. nonword; within subjects) × 2 (Gender: male vs. female; between subjects) mixed ANOVA revealed a significant main effect of trial type, $F(1,26) = 6.47$, $p = .017$, $\eta_p^2 = .20$, on listening time. There was no main effect of gender and no interaction (F's < 1). The main effect of trial type indicated that infants listened longer to the nonwords ($M = 8.35$, $SD = 3.36$) than the words ($M = 7.25$, $SD = 3.53$).

The scatterplot in Figure 2 illustrates the distribution of scores based on the listening time difference to words minus nonwords, separated by gender to allow for a comparison with Experiment 1. Both males and females tended to show novelty preferences for nonwords (21 of 28 infants). These results indicate that infants successfully learned the structure of the language. As discussed in more detail below, the direction of preference suggests that infants found the word versus nonword comparison easier than the word versus part-word comparison in Experiment 1.

GENERAL DISCUSSION

The results of the present experiments indicate that infants can use statistical word segmentation to detect words in continuous speech containing foreign language sounds. The infants

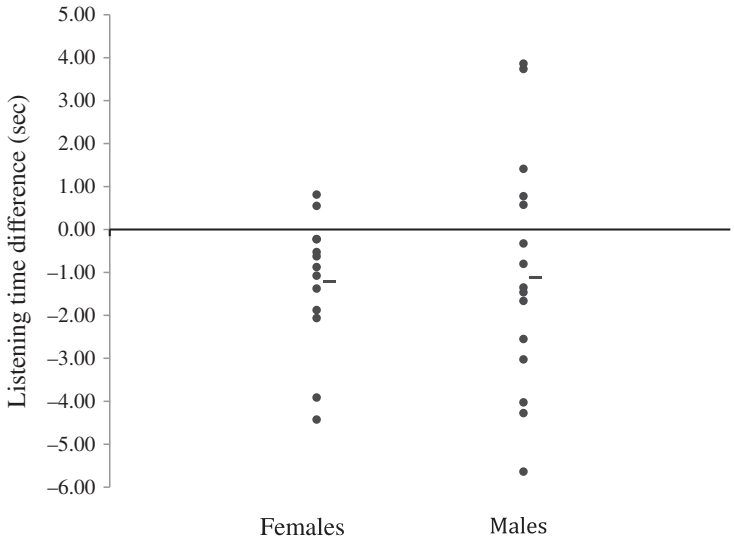

FIGURE 2 Infants' listening time differences to word—nonword test trials in Experiment 2. Dots represent individual infants, dashes represent group mean scores.

applied a language-general segmentation strategy in the presence of non-native phonology and in the absence of cues beyond transitional probabilities. In Experiment 1, 14-month-old males, but not females, reliably differentiated high transitional probability words versus low probability part-words that had equal frequency in the fluent Mandarin syllable stream. The males listened significantly longer to the familiar words, whereas females showed a wide spread in their listening patterns, with a non-significant tendency to listen longer to the part-words. In Experiment 2, infants differentiated words from novel syllable sequences. Across genders, they listened longer to the novel nonwords. The reliable discrimination of test items across experiments shows that infants learned the structure of the language, and the variation in direction of preference provides information about infants' processing during the tasks.

We have shown that infants can track statistical regularities in a highly controlled language consisting of foreign syllables. These findings support the hypothesis that at least some dimensions of statistical learning are robust. Specifically, tracking syllable transitional probabilities does not require the presence of native language phonology and the absence of foreign phonology. Our results expand on the findings from Pelucchi and colleagues (2009) that infants can track syllable probabilities to segment a foreign language. The present results indicate that infants can also learn when the speech stream is more distant from their native language in origin and in phonology (Dryer & Haspelmath, 2011), and when no natural language-specific rhythmic information is available to scaffold segmentation. This does not mean that foreign phonology has no effect on learning. The familiarity preference that the males in Experiment 1 displayed suggests that the task was challenging. Furthermore, the familiarity preference is inconsistent with the novelty preference that has been shown in even younger infants with similarly structured frequency-balanced languages that used native language phonology (Aslin et al., 1998; Graf

Estes, 2012; Johnson & Tyler, 2010). Although we cannot state definitively that learning from the Mandarin stimuli was more difficult than learning from English-based stimuli (because we did not test an English-only equivalent artificial language), we propose that the unfamiliarity of Mandarin phonology contributed to infants' performance in our tasks, as discussed below.

The present findings also build on Saffran, Johnson, Aslin, and Newport's (1999) demonstration that infants can track statistical patterns in nonlinguistic materials. They found that at 8 months of age, infants detected consistent tone triplets in continuous sequences of pure tones. This research established that statistical learning is sufficiently flexible to function with nonlinguistic material and is a domain general mechanism. The present experiments show another dimension of this flexibility. Here, infants learned from linguistic materials that contained unfamiliar sounds that were substantially more complex than pure tones—phonemes and syllables are signals that change over time and are defined by multiple acoustic dimensions. They present a naturalistic challenge to the limits of statistical learning.

Across experiments, variation in infants' preference patterns revealed characteristics of their information processing. According to Hunter and Ames's (1988) model of infant attention, several factors affect whether infants display familiarity or novelty preferences. For example, familiarity preferences are likely to emerge with younger infants, after brief exposure to learning material, or when the task requires learning complex information. In contrast, novelty preferences emerge when infants have thoroughly processed the new information. Thoroughness of learning affects how test items are encoded (i.e., when infants have encoded effectively during learning, they can readily detect what is consistent versus inconsistent with prior experience during testing). Test items characteristics can also interact with learning conditions to influence performance. Items that have many dimensions in common are harder to differentiate than dissimilar items. This produces a familiarity preference for the test items that are most consistent with the familiarization stimuli. Novelty preferences occur when, all else being equal, test items are clearly distinct. In Experiment 1, the test items were similar in key ways. The words and part-words both had higher transitional probability than the surrounding across-word sequences and they were both frequently occurring sequences in the speech stream. Consistent with Hunter and Ames's model, infants (specifically, males) demonstrated learning with a familiarity preference. The distribution of the females' scores suggests that some females tended to attend to novelty despite the difficulty of differentiating the test items (this is consistent with females' more advanced linguistic processing seen in other research; Dale & Fenson, 1996; Hartshorne & Ullman, 2006; Lany & Gómez, 2008). However, the females' response pattern was not strong or consistent enough to produce a reliable group level preference pattern; this may also reflect the difficulty of processing test items with similar structure.

In Experiment 2, the test items were more structurally distinct than in Experiment 1. Specifically, the words and nonwords differed substantially in frequency and in transitional probability. Accordingly, infants across genders showed a reliable novelty preference. There is no reason to suspect that *learning* during segmentation was different across experiments because the segmentation stimuli were identical, but performance during *testing* was quite different. When the test items were selected to promote their distinctiveness (i.e., differences in transitional probability and frequency), infants could robustly display what they learned. Importantly, across both experiments, infants had to detect structure in a speech stream containing foreign phonology to successfully differentiate the test items.

The differences in performance across test item types support clustering models of word segmentation, such as PARSER (Perruchet & Vinter, 1998). According to these models, potential lexical units emerge as syllable sequences gain association strength. Typically, as exposure to the speech stream proceeds, memory decay and interference weaken the representations of across-word sequences. However, in the present experiments, some across-word sequences (i.e., the test part-words) occurred frequently, producing relatively high association strengths. This could make it difficult for learners to eliminate the part-words as potential lexical units and to distinguish them from words, particularly early in learning. Eventually, the effects of interference should override the association frequency to make words and part-words readily discriminable in this task, but evidence with adults suggests that interference can require substantial time to fully override frequency effects (Giroux & Rey, 2009). In contrast, clustering models suggest that learners should have little difficulty distinguishing statistically-coherent words from nonwords that never occurred in the language. Syllable sequences in nonwords have no prior activation strength and are readily rejected as possible lexical units, providing a strong contrast with the words.

The model-based predictions about the difficulty of the word versus part-word and word versus nonword test items are consistent with the patterns of familiarity preference and uneven performance in Experiment 1 and the robust novelty preference in Experiment 2. It is interesting that infants in prior experiments showed novelty preferences for part-words even using frequency-balanced languages (Aslin et al., 1998; Graf Estes, 2012; Johnson & Tyler, 2010). We propose that the added challenge of the unfamiliar Mandarin phonology revealed vulnerability in processing words versus frequent across-word sequences that was not apparent in less demanding tasks with familiar phonology.

Given the test item designs, there are differences in the conclusions that can be drawn from Experiments 1 and 2. In experiments with frequency-balanced languages (Aslin et al., 1998), successful discrimination of words and part-words requires that infants tracked the *probability* of the syllable co-occurrences during segmentation, rather than only the *frequency* with which they occurred. We have evidence that the males in Experiment 1 performed this type of learning. The conclusions from Experiment 2 are somewhat more limited because, similar to many other statistical segmentation tasks (e.g., Johnson & Jusczyk, 2001; Pelucchi et al., 2009; Saffran, Aslin et al., 1996; Thiessen et al., 2005), the test items were not matched in frequency. At minimum, the discrimination of words and nonwords required infants to differentiate between frequent high probability sequences and sequences of familiar syllables in novel orders. This differentiation did not require precise transitional probability calculations, but could be achieved based on frequency information. It is important to keep in mind that detecting frequency patterns is a form of statistical learning, and both frequency and probability patterns contribute to language acquisition. In some cases, probability is a more reliable cue to structure. For example, the sheer frequency of a phrase like "the cup" could lead infants to identify "thecup" as a single word, but probabilistic information (e.g., about the contexts in which "the" occurs) marks each word as an independent unit. But frequency is also important. Word frequency affects adult speech processing (Allen, McNeal, & Kvak, 1992; Dahan, Magnuson, & Tanenhaus, 2001; Dell, 1990; Morrison & Ellis, 1995) and child lexical acquisition (Goodman, Dale, & Li, 2008; Hoff & Naigles, 2002). Furthermore, frequency is inherent to probability calculations. Thus, both experiments provide evidence of successful statistical learning from a speech stream with unfamiliar phonology.

In previous research, Tsay and Jusczyk (2003) reported that English-learning infants failed to segment words from Mandarin fluent speech (see also Newman et al., 2003). There are several

design differences that could have produced the difference in results across studies. First, the present research focused on a specific language-general word segmentation cue, transitional probability. Other dimensions of Mandarin were carefully controlled, preventing characteristics like lexical stress and tone from affecting (and potentially interfering with) infants' detection of the word patterns. Thus, the availability of a specific cue, and the absence of potentially confusing information, may have facilitated segmentation in our experiments. In addition, the more concentrated exposure during segmentation in the current task may have allowed infants to detect the language-general cue in the presence of unfamiliar speech sounds. Finally, the participants in Experiments 1 and 2 were nearly 7 months older than Tsay and Jusczyk's (2003) participants. If segmenting foreign speech is computationally demanding, older infants may have an advantage in some tasks because they have greater cognitive speed and capacity (but see below for more discussion of possible age-related effects). Future research will be necessary to investigate the effects of age, concentrated exposure, and natural Mandarin speech characteristics (e.g., rhythmic and phonotactic patterns), on English learners' segmentation performance.

In other infant speech processing tasks, infants exhibit a decline in processing non-native speech sounds as perception narrows to focus on native language phonology (e.g., Kuhl et al., 2005; Kuhl et al., 2006; Werker & Tees, 1984). Here, we found openness to segmenting a speech stream that contained foreign phonology. However, the present findings are not incongruous with the notion of perceptual narrowing. Rather, mechanisms used to explain losses in non-native phoneme perception can contribute to explaining infants' success in the current tasks. Best's Perceptual Assimilation Model (Best, 1991; Best & McRoberts, 2003; Best, McRoberts, & Goodell, 2001; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995) contends that listeners assimilate non-native phonemes to the most similar native phoneme categories (e.g., the Hindi retroflex stop /ɖ/ is likely to assimilate to the English /d/). The present experiments did not require precise non-native phoneme discrimination—infants could successfully track syllable patterns by assimilating unfamiliar phonemes to native categories. For example, adult American English-speakers reported hearing the Mandarin consonant /ʂ/ in our stimuli (in the syllable *shu*) as similar to the English /s/ and the consonant /ɕ/ (in the syllable *xie*) as the English /ʃ/. They heard /pʰ/ (in the syllable *pao)* as English /p/ and the Mandarin unaspirated /p/ (in the syllable *bi*) as the English /b/. While overlap between nonnative and native phonemes is problematic in phoneme discrimination tasks, it may not prevent infants from tracking syllable patterns. Infants may maintain representations of the nonnative sounds that allow them to detect the recurring tokens. In addition, each consonant was paired with a different vowel, which further reduced the demands on precise phoneme discrimination. Thus, we predict that older children and adults, whose phonological systems are even more committed to native language structure than the systems of 14-month-olds, could use the same processes to effectively segment the Mandarin-based stimuli.

Given that experience sharpens infants' processing of native sounds (Kuhl et al., 2006; Werker & Curtin, 2005), infants' representations of the foreign phonemes may be less precise than for highly familiar sounds. This could produce an additional processing load during segmentation. Thus, although infants clearly learned from the patterns of non-native sounds in the present experiments, we propose that segmenting speech with unfamiliar phonology was more challenging than segmenting with familiar sounds. This could be due to the presence of the foreign phonemes as well as the unfamiliar accent producing the sounds that overlap with native sounds. We expect that the segmentation task would be substantially more difficult if it required attending to phoneme

contrasts that infants no longer readily perceive. Younger infants who have not yet fully experienced native language perceptual tuning may be at a substantial advantage over older infants in such a task.

The results of the current experiments indicate that infants show flexibility in a statistical word segmentation task; they successfully tracked syllable patterns in a continuous speech stream that contained foreign speech sounds. Future research will be necessary to establish the full boundaries of flexibility in statistical learning.

ACKNOWLEDGMENTS

We would like to thank the members of the Language Learning Lab at the University of California, Davis for their assistance with this research. We would also like to thank Tianlin Wang for her comments on an earlier draft of this manuscript. We also thank the parents who generously contributed their time.

FUNDING

This research was supported by a grant to Katharine Graf Estes from the National Science Foundation (BCS0847379).

REFERENCES

Allen, P. A., McNeal, M., & Kvak, D. (1992). Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language*, *31*(6), 826.

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.

Best, C. T. (1991). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *Haskins Laboratories Status Report on Speech Research*, 107–108, 130.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior & Development*, *18*(3), 339–350.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25). Amsterdam, The Netherlands: University of Amsterdam.

Cohen, L. B., Atkinson, D. J., & Chaput, H. J. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies. (Version 1.0). Austin, TX: University of Texas.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, *65*(6), 1663–1677.

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. (2008). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Malden, MA: Blackwell Publishing.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*, *28*(1), 125–127.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349.

Dryer, M. S., & Haspelmath, M. (2011). Word atlas of language structures. Retrieved from http://wals.info/

Duanmu, S. (2000). *The phonology of standard Chinese*. Oxford, England: Oxford University Press.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*(2), 260–272.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531.

Graf Estes, K. (2012). Infants generalize representations of statistically segmented words. [Original Research]. *Frontiers in Psychology*, *3*.

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.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*, 190–222.

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science*, *9*(1), 21–32.

Hay, J., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*, 93–106.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433.

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.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infants preference procedures. *Infant and Child Development*, *13*, 341–348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Johnson, E. K. (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language aquisition*. Berlin, Germany: De Gruyter Mouton. Retrieved from http://www.degruyter.com/view/product/43619

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. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345.

Krogh, L., Vlach, H., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology, 3*.

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., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–21.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Los Angeles, CA: University of California Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121.

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303–315.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116–133.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable timed languages. In C. J. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1147–1159). Hoboken, NJ: Wiley-Blackwell.

Newman, R., Tsay, J., & Jusczyk, P. W. (2003). The development of speech segmentation abilities. In D. M. Houston, A. Seidl, G. Hollich, E. K. Johnson, & A. M. Jusczyk (Eds.), *Jusczyk lab final report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.

Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language*, *39*(2), 246–263.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 906–914.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169.

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

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

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

Shi, R., & Werker, J. F. (2001). Six-month old infants' preference for lexical words. *Psychological Science*, *12*(1), 70–75.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132.

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

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (in press). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Tsay, J., & Jusczyk, P. W. (2003). Detection of Words in Fluent Chinese by English-acquiring and Chinese-acquiring Infants. In D. M. Houston, A. Seidl, G. Hollich, E. K. Johnson, & A. M. Jusczyk (Eds.), *Jusczyk lab final report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk

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

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*(10), 451–456.