English-learning infants developing sensitivity to vowel phonotactic cues to word segmentation

Hironori Katsuda & Megha Sundara

Department of Linguistics, UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543, USA

Correspondence
Hironori Katsuda, Present address: Department of Linguistics, University of Kansas, Blake Hall, 1541 Lilac Lane, Lawrence, KS 66045
Email: katsuda1123@gmail.com

Acknowledgements
Research was funded by NSF BCS-2214017 to MS; part of the results was presented at the 47th Boston University Conference on Language Development (BUCLD) and the International Congress of Infant Studies, Ottawa.

Declaration of Competing Interest
The authors have no competing interests to declare.

Ethics Approval
The research described was approved by UCLA's IRB#10-001562

Data availability
Experimental stimuli, data, and analysis scripts are available on the Open Science Framework: https://osf.io/45dp9/?view_only=4c5c35cd49b442b483429f53a0af30fd.

ORCID
Hironori Katsuda: https://orcid.org/0000-0002-0631-3951

English-learning infants developing sensitivity to vowel phonotactic cues to word segmentation

Research highlights

- Research indicates that when transitional probability conflicts with stress cues for word segmentation, English-learning 9-month-olds rely on stress, whereas younger infants rely on transitional probability
- In two artificial languages, we evaluated English-learning infants' sensitivity to transitional probability versus vowel phonotactic cues for word segmentation
- When these cues conflicted, 10-month-olds relied on vowel phonotactics, whereas 5-month-olds relied on transitional probability
- These findings align with statistical bootstrapping accounts, where infants first utilize domain-general distributional information for word segmentation, and then identify language-specific patterns from segmented words

Abstract

Previous research has shown that when domain-general transitional probability cues to word segmentation are in conflict with language-specific stress cues, English-learning 5- and 7-month-olds rely on transitional probability, whereas 9-month-olds rely on stress. In two artificial languages, we evaluated English-learning infants' sensitivity to transitional probability cues to word segmentation vis a vis language-specific vowel phonotactic cues - English words do not end in lax vowels. These cues were either consistent or conflicting. When these cues were in conflict, 10-month-olds relied on the vowel phonotactic cues, whereas 5-month-olds relied on transitional probability. These findings align with statistical bootstrapping accounts, where infants initially use domain-general distributional information for word segmentation, and subsequently discover language-specific patterns based on segmented words.

Keywords: artificial language, word segmentation, statistical learning, transitional probability, vowel phonotactics

1. Introduction

When learning their native language, infants face the initial challenge of accurately segmenting words from fluent speech. This task is not simple because the linguistic input consists of uninterrupted sequences of multiple words, lacking any pauses between them (Cole & Jakimik, 1980; van de Weijer, 1998). Moreover, cues to word boundaries are not universal and deterministic; their effectiveness varies from language to language. Consequently, there has been a long-standing question concerning when and how infants efficiently identify word boundaries in their native language.

Starting with the seminal work by Saffran, Aslin, and Newport (1996), we know that infants employ statistical learning as a language-general strategy for word segmentation (Aslin, Saffran, & Newport, 1998; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003; 2007; Pelucchi, Hay, & Saffran, 2009; Johnson & Tyler, 2010; Thiessen & Erickson, 2013). That is, infants track the dips in transitional probability of syllables, which refers to the likelihood of a syllable being followed by another, to detect word boundaries. Consistent with this account, English-learning infants have been shown to segment words by exploiting differences in syllable transitional probabilities even at 5-months (Thiessen & Erickson, 2013). An early reliance on syllable transitional probability cues for word segmentation has also been demonstrated in infants learning French (Mersad, Goyet & Nazzi, 2011; Mersad & Nazzi, 2011) and Spanish (Bosch, Figueras, Teixidó, Ramon-Casas, 2013), but not German (Marimon, Langus & Höhle, 2024).

Speech directed to infants also has language-specific cues that typically provide converging evidence for word boundaries. This includes information about prosody (Jusczyk, Cutler, & Redanz, 1993; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003; 2007; Johnson & Seidl, 2009), phonotactics (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994; Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001), coarticulation (Johnson & Jusczyk, 2001; Curtin, Mintz, & Byrd, 2001) and allophonic variation (Jusczyk, Hohne, & Bauman, 1999). Infants have been shown to use these cues as well to segment words.

How might infants integrate language-general cues, in particular transitional probability, with language-specific cues? The most extensive evidence on cue weighting for word segmentation comes from infants learning English. When transitional probability cues are pitted against either coarticulation (Johnson & Jusczyk, 2001) or stress cues (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), English-learning infants have been reported to rely on transitional probabilities earlier in development. For instance, 5- (Thiessen & Erickson, 2013) and 7-month-olds (Thiessen & Saffran, 2003) listened significantly longer to part words (i.e., words not matching transitional probability cues) over statistical words (i.e., words matching transitional probability cues), whether transitional probability cues were consistent with or in conflict with stress cues. These results indicate that English-learning infants initially rely on transitional probability cues. In contrast, older English-learning infants, specifically 8- (Johnson & Jusczyk 2001) and 9-month-olds' (Thiessen & Saffran 2003) behaved differently when stress and transitional probability cues were in conflict. Unlike younger infants, 8-month-olds listened longer to statistical words when stress and transitional probability cues were in conflict (Johnson & Jusczyk, 2001). Thiessen and Saffran (2003) as well report a switch in preference when older infants were presented with an artificial language where transitional probability cues to word boundaries were congruent with or in conflict with stress cues. In this case, infants listened longer to part words when the cues conflicted and statistical words when the cues were congruent (see Thiessen and Saffran's discussion of why their results are different from those of Johnson and Jusczyk, 2001).

Based on these results, Thiessen and colleagues propose a statistical bootstrapping account where infants initially use transitional probabilities to identify word forms and then utilize those word forms to uncover language-specific regularities, in this case, the distribution of stress cues in English (see Mersad & Nazzi, 2011, for a similar proposal for French-learning infants). In this paper we investigated how English-learning infants weight transitional probability cues to word boundaries with respect to a language-specific cue based on phonotactics, that is segment position and sequencing restrictions. We focused on one specific restriction in the positioning of English vowels - typical English words do not end in lax vowels like [ɛ] as in *pet* or [ɪ] as in *pit*. We know that adult speakers of English can use this phonotactic restriction when segmenting novel adjective-noun phrases (Skoruppa, Nevins, Gillard, & Rosen, 2015; see also, Newman, Sawusch,

& Wunnenberg, 2011). Therefore, English-learning infants must learn this vowel phonotactic constraint and use it to segment words from continuous speech at some point during development.

Pitting transitional probabilities against segment position and sequencing restrictions provides an informative and novel test of the statistical bootstrapping account for three reasons. First, prior investigations have primarily focused on stress, which is used to contrast meaning in languages like English, German and Spanish, but not French or Japanese. In contrast, segment position and sequencing restrictions can signal word boundaries in all languages. Second, stress cues are often acoustically complex, involving multiple covarying cues that differ across languages, and must also be learned. In contrast, acoustic cues involved in phonotactic restrictions tend to be simpler, making it easier to control the stimuli and interpret the results. Finally, lax vowels in English are never observed at the edges of words, and thus utterances, and as such exemplify a clear case of edge alignment, in contrast to the relatively flexible placement of stress. Pitting segment position restrictions that are edge aligned and, therefore, perceptually salient, against transitional probabilities thus provide a strong test of the statistical bootstrapping hypothesis.

Following Thiessen, Kronstein, and Hufnagle's (2013) statistical bootstrapping account, infants are expected to rely on transitional probability cues to segment words early in development, and only subsequently acquire language specific patterns.

2. Experiment 1

The purpose of this experiment was to determine which cue, vowel phonotactics (VP) or transitional probability (TP), English-learning 10-month-olds prioritize in segmenting words from speech stream. To examine this, we pitted VP and TP cues against each other, employing a modified version of Thiessen and Saffran's (2003) experiment implemented with the central fixation procedure. Infants were assigned to one of two conditions, differing only in the structure of the familiarization stream. In the cooperating-cue condition, both VP and TP cues indicated the same word boundaries: tense vowels were immediately followed by low TP boundaries (more likely to represent between-word boundaries), while lax vowels were immediately followed by high TP boundaries. In contrast, in the conflicting-cue condition, the two cues signaled different word boundaries: tense vowels were immediately followed by high TP boundaries, and lax vowels were immediately followed by low TP boundaries.

If infants prioritize VP, their word segmentation should be guided solely by the phonotactic constraint in English, which dictates that words do not end with lax vowels. Thus, syllables containing tense vowels would be perceived as word-final, regardless of the statistical properties of the speech stream. For example, syllable sequences like [...bɛdi|dɪbi|gɪbu…] ("|" indicates a low TP boundary) in the cooperating-cue condition would be segmented as [...bɛdi#dɪbi#gɪbu…]. In contrast, sequences in the conflicting-cue condition, such as [...dudɪ|digɪ|bigɛ…], would be segmented as [...du#dɪdi#gɪbi#gɛ…].

Alternatively, if infants prioritize TP, their word segmentation should be guided by the statistical patterns of the familiarization stream. In this scenario, syllables immediately followed by low TP boundaries would be considered word-final, irrespective of the vowel identity. Syllable sequences in the cooperating-cue condition like [...bɛdi|dɪbi|gɪbu…] would be still segmented as

[...bɛdi#dɪbi#gɪbu…], identical to the VP-priority scenario. However, in the conflicting-cue condition sequences such as [...dudɪ|digɪ|bigɛ…] would be segmented as [...dudɪ#digɪ#bigɛ…].

Given previous reports that infants learning many different languages begin to tune into phonotactic restrictions in their native language between 8 and 10 months of age (Jusczyk, Luce & Charles-Luce, 1994; see also Sundara, Zhou, Breiss, Katsuda, & Steffman, 2022 for a meta-analysis), we expected English-learning 10 months olds to exhibit sensitivity to VP cues.

2.1. Methods

2.1.1. Participants

The final sample in Experiment 1 included 56 (25 female) English-learning 10-month-olds (average age = 309 days; range = 286:339). All infants were reported to be healthy, full-term and had no ear infection on the day of testing. Each infant's language background was assessed using a detailed parental language questionnaire (Sundara & Scutellaro, 2011), and only infants with more than 90% exposure (mean = 99%; range 90:100) to American English were included. Seven additional infants were tested but their data were excluded from the final sample due to experimenter error (6) and a software problem (1).

2.1.2. Stimuli

Acoustics
Familiarization and test items were created by concatenating the following eight CV syllables: [di, du, bi, bu, dɪ, bɛ, gɪ, gɛ]. The first four syllables have a tense vowel while the last four have a lax vowel. To prevent coarticulation between syllables, each of the eight syllables was individually recorded. A phonetically trained female native speaker of English produced these syllables with a slightly falling pitch.

In English, lax vowels are typically shorter than tense vowels (Hillenbrand, Getty, Clark, & Wheeler, 1995). However, in our study, we equated the duration of the tense and lax vowels in the stimuli. We did this because differences in syllable duration in English correlate with stress, and we did not want infants to respond based on perceived differences in stress (Hay & Saffran, 2012). The duration of each syllable was adjusted to ~250 ms, by adding or deleting pitch periods at zero crossings at the vowel midpoints. This duration falls roughly between the durations of stressed and unstressed syllables (averaging around 310 ms and 185 ms respectively) in Thiessen and Saffran's (2003) stimuli, which are based on production data from American English speakers recorded by Crystal and House (1990). Additionally, we normalized the intensity of each syllable to 70 dB and matched the average fundamental frequency (without altering pitch contours) of each syllable to about 205 Hz (SD = 0.7 Hz).

Statistical structure of the speech stream
These eight syllables were combined to create two sets of four disyllabic nonce target words: [bɛdi, gɪbu, dɪbi, gɛdu] and [dudɪ, bigɛ, digɪ, bubɛ]. The words in the first group end with tense vowels, consistent with English phonotactic restrictions. These four words served as target words for the cooperating-cue condition. In contrast, the second group of words end in lax vowels, going against

the same phonotactic restriction. These four words served as target words for the conflicting-cue condition.

The four target words in each condition were then combined to form the familiarization stream for each condition. As TPs between words are lower than TPs within words, and thus likely to align with word boundaries, the statistical properties of the familiarization streams signaled the presence of word boundaries after tense vowels in the cooperating-cue condition, and after lax vowels in the conflicting-cue condition.

An issue involved in construction of these types of familiarization streams is that target words, serving as statistical words in the test phase, occur more frequently than part words. This introduces a potentially confounding factor, as infants might simply pay more attention to frequently-occurring sequences rather than segmenting words based on TPs. To control for this asymmetrical frequency effect, following Thiessen and Saffran (2003), we doubled the occurrence of the last two target words for each condition. Specifically, the last two target words appeared 90 times while the first two target words appeared 45 times in each familiarization stream. The target words were combined such that the TP within a target word was 1.0, whereas the TP across target words ranged between 0.0 and 0.6. Table 1 shows the beginnings of the two familiarization streams and their segmentation based on TP cues and VP cues. Both familiarization streams lasted approximately 2 min 30 s (full stream is available on the project OSF page).

	Cooperating-cue condition	Conflicting-cue condition
Target words	[bɛdi, gɪbu, dɪbi, gɛdu]	[dudɪ, bigɛ, digɪ, bubɛ]
Beginning of stream	[bɛdidɪbigɪbugɛdugɪbu…]	[dudɪdigɪbigɛbubɛbigɛ…]
TP-based segmentation	[bɛdi#dɪbi#gɪbu#gɛdu#gɪbu#…]	[dudɪ#digɪ#bigɛ#bubɛ#bigɛ#…]
VP-based segmentation	[bɛdi#dɪbi#gɪbu#gɛdu#gɪbu#…]	[du#dɪdi#gɪbi#gɛbu#bɛbi#gɛ…]

Table 1. Design of the familiarization stream.

Four disyllabic sequences, namely [bɛdi, gɪbu, bigɛ, dudɪ], were used as test items. Note that the first two sequences were the infrequent target words in the cooperating-cue condition, while the latter two were the infrequent target words in the conflicting-cue condition.

For infants in the cooperating-cue condition, the first two test items (i.e., [bɛdi, gɪbu]) were statistical words, while the last two test items were part words consisting of the two frequent target words (i.e., dɪ**bi#gɛ**du and gɛ**du#dɪ**bi). For infants in the conflicting-cue condition, the last two test items (i.e., [bigɛ, dudɪ]) were statistical words, whereas the first two test items were part words consisting of the two frequent target words (i.e., bu**bɛ#di**gɪ and di**gɪ#bu**bɛ). Thus, the TP between the two syllables in statistical words in both conditions was 1, while the TP in part words was 0.5.

In the cooperating-cue condition the statistical words were phonotactically legal while the part words were phonotactically illegal, aligning with the TP cues. Crucially, in the conflicting-cue condition, the relationship between the two cues was reversed. The statistical words were illegal whereas the part words were legal, deviating from the TP cues. This relationship between the two cues is summarized in Table 2.

Condition	Trial Type	
	Legal phonotactic sequences [bɛdi], [gɪbu]	Illegal phonotactic sequences [bigɛ], [dudɪ]
Cooperating-cue	Statistical words (TP = 1)	Part words (TP = 0.5)
Conflicting-cue	Part words (TP = 0.5)	Statistical words (TP = 1)

Table 2. Design of the test stimuli. Each of the four sequences appeared 45 times in each familiarization stream.

2.1.3. Procedure

Infants were tested while seated on a parent's lap, using the central fixation procedure. An experimenter outside the room monitored the infants' looking behavior on a screen connected to a camera inside the room and coded the infants' gaze online. To eliminate bias, the parent and the experimenter listened to masking music, ensuring they remained unaware of the stimulus being presented to the infant.

All trials started with a looming light accompanied by a baby giggle to attract the infant's attention. During the familiarization phase, a Tetris-like video was displayed on the screen alongside one of the two familiarization streams (either cooperating-cue or conflicting-cue). The fixed duration familiarization phase lasted approximately 2 min 30 s, and was not contingent on infant looking.

Immediately following the familiarization phase, all infants heard the same 12 test trials. Each test trial occurred three times during the test phase, resulting in a total of six statistical-word trials and six part-word trials.

In each trial, a test item was repeated 15 times, with a 500 ms inter-item-interval. The test trials were entirely contingent on the infant's looking behavior. The subsequent test trial began either when the infant looked away from the screen for more than two seconds or when the test trial ended (maximum duration ~15 s). If the infant looked away during the initial one second of the trial, the trial was repeated. Infant looking time was the dependent variable for the analysis.

2.1.4. Analysis

We analyzed looking time with a linear mixed effects model in R using *lmerTest* (data and R script also on the project OSF page). In the analysis, we included several fixed effects: Condition (between-subjects; cooperating-cue vs. conflicting-cue), Trial-type (within-subjects; statistical words vs. part words), and Block (within-subjects; 1-3), and all interactions. We included random intercepts for participants to account for baseline differences in looking time across infants; this was the highest random effect structure that converged (Barr, Levy, Scheepers, & Tily, 2013). Significant interactions were examined using *emmeans* (Lenth, Singmann, Love, Buerkner, & Herve, 2018). We also replicated the same pattern of results with log-transformed looking time as the dependent variable (see Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016 for evidence that looking time data are not normally distributed). We only report statistical analyses for raw looking times to allow comparisons to previously published results.

2.2. Results and Discussion

The looking time data are presented in Figure 1. The analysis revealed only a significant interaction between Condition and Trial-type ($F(1,606) = 8.68$, $p < 0.005$). That is, 10-month-olds behaved differently in the cooperating- vs the conflicting-cue condition. A post-hoc analysis with *emmeans* showed that the effect of Trial-type was significant in the cooperating-cue condition ($t(606) = 2.26$, $p = 0.02$) and near significant for the conflicting-cue condition ($t(606) = -1.90$, $p = 0.05$). Notably, the effect was positive in the cooperating-cue condition, indicating that in the cooperating-cue condition, 10-month-old infants showed a greater interest in part words over statistical words. However, in the conflicting-cue condition, there was a marginally significant tendency for them to pay more attention to statistical words than to part words. That is, across both conditions, 10-month-old infants displayed a tendency to listen longer to phonotactically illegal words than legal words.

These results show that at 10 months, infants prioritize vowel phonotactic cues over transitional probability cues to segment words. This aligns with the cross-linguistic developmental trajectory showing that infants begin to tune into the phonotactic restrictions in their native language between the ages of 8 and 10 months (Sundara et al., 2022). We confirmed this timeline using a previously unexplored phonotactic restriction, specifically one involving vowel phonotactics in English.

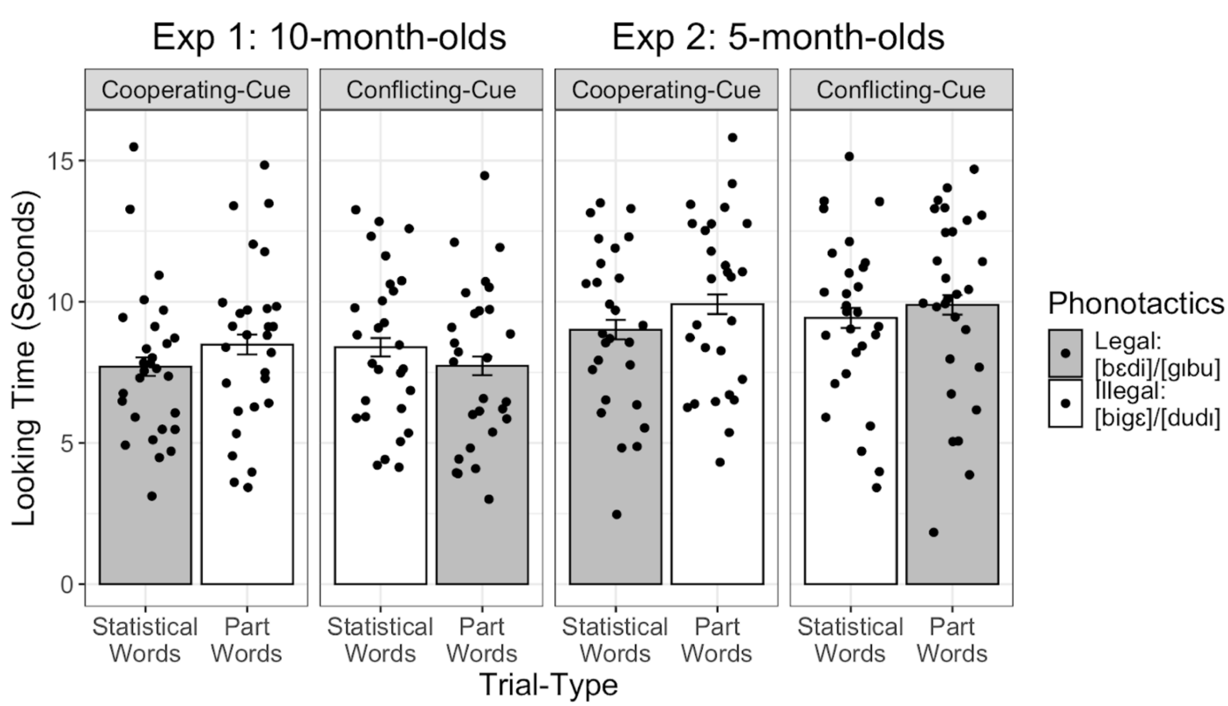

Figure 1: Looking times to statistical words and part words in the cooperating-cue and conflicting-cue conditions in Experiment 1 (left) and Experiment 2 (right).

3. Experiment 2

The results of Experiment 1 revealed that English-learning 10-month-olds utilize VP cues to segment words in fluent speech, regardless of whether the TP cues were consistent with or in conflict with English phonotactic cues to word boundaries. This finding raises a further question: how do infants initially learn that English words typically do not end in lax vowels? One possibility, in line with the statistical bootstrapping account, is that infants initially segment word forms from the speech stream using TP cues. As they segment a sufficient number of word forms, they detect that word forms in English do not end with lax vowels. If this is correct, younger infants should prioritize TP cues more than VP cues for word boundaries.

However, infants might rely on VP cues even before TP cues. Prior studies have indicated that speech material located at the edges of words, phrases, or utterances is more salient to infants, resulting in more precise encoding, compared to material in the middle (Seidl & Johnson, 2006; Gervain, Nespor, Mazuka, Horie, & Mehler, 2008; Endress, Nespor, & Mehler, 2009; Johnson, Seidl, & Tyler, 2014; Benavides-Varela & Mehler, 2015; Ferry, Fló, Brusini, Cattarossi, Macagno, Nespor, & Mehler, 2016). Consistent with the greater salience of speech material at utterance edges, effect sizes in experiments testing infants sensitivity to phonotactic patterns at edges are greater than those involving patterns that are word medial (Sundara, et al., 2022). Further, infants segment words earlier at edges of utterances than utterance medially (Johnson, et al., 2014). Based on these results Johnson, et al. (2014) argue that by attending to word boundaries at utterance edges, infants could begin to learn language-specific cues. Because the vowel phonotactic constraints in question are edge-aligned, it is possible that infants at 5 months old are already attuned to VP cues, having noticed the absence of lax vowels at the end of utterances.

Support for such a timeline, where even 6-month-olds are sensitive to language-specific cues to word segmentation, comes from research on German-learning infants. German-learning 6-month-olds exhibit a preference for the predominant, trochaic stress pattern (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009). They also prioritize stress cues over TP cues for word segmentation, to the extent that they fail to segment words using TP cues in the absence of stress cues (Marimon, et al, 2024).

In sum, whether statistical bootstrapping can account for the developmental differences in cue weighting between TP and VP remains an open question. Experiment 2 was designed to determine whether English-learning 5-month-olds prioritize TP cues over VP cues when segmenting words. We tested 5-month-olds because it is the earliest age at which infants have been shown to segment words in English (Thiessen & Erickson, 2013; Johnson & Tyler, 2010). In this experiment as well infants were exposed to the same speech stream as in Experiment 1.

3.1. Methods

3.1.1. Participant

The inclusion criteria were identical to that in Experiment 1. The final sample in Experiment 2 included 55 (26 female) English-learning 5-month-olds (average age = 153 days; range = 125:175). The average percent exposure to English was 99% (range 90:100). Six additional infants in this group were tested but excluded from the final sample due to fussiness (2), inability to see the infant's eyes due to the infant covering their face (1), and experimenter error (3).

3.1.2. Stimuli

Identical to Experiment 1.

3.1.3. Procedure

Identical to Experiment 1.

3.1.4. Analysis

Identical to Experiment 1, except that the final model also included a random slope of Trial-Type by subject.

3.2. Results and Discussion

There was only a significant effect of Trial-type ($F(1,53) = 6.00$, $p < 0.05$), showing that 5-month-olds looked longer to part words than to statistical words, regardless of the condition. That is, consistent with transitional probability cues, 5-month-olds listened longer to part words whether or not they were phonotactically illegal in English.

These results show that 5-month-olds rely more on transitional probability than vowel phonotactics for word segmentation. This is consistent with previous studies showing that English-learning infants, whether at 7 months (Thiessen & Saffran, 2003) or at 5 months (Thiessen & Erickson, 2013), prioritize transitional probability over conflicting language-specific (stress) cues.

4. General discussion

This study was designed to investigate when (if at all) English-learning infants rely on vowel phonotactic (VP) cues compared to transitional probability (TP) cues to word segmentation. In Experiment 1, we demonstrated that 10-month-olds looked longer to phonotactically illegal words compared to legal words, whether they were statistical words or part words. That is, they weighted VP cues more heavily than TP cues when segmenting words, confirming their sensitivity to the phonotactic constraint that English words do not end in lax vowels. In contrast, results from Experiment 2 revealed that 5-month-olds looked longer to part words compared to statistical words, whether they were phonotactically legal or illegal. Thus, 5-month-olds prioritized TP over VP cues.

Upon a closer examination of our results, it is evident that in the cooperating-cue condition, infants at both 5 months and 10 months segmented words in line with the converging cues of TP and VP, demonstrating increased looking times to part words. In contrast, the two age groups behaved differently in the conflicting-cue condition: 5-month-olds segmented words based on TP cues, showing a preference for part words, whereas 10-month-olds relied on VP cues, favoring statistical words. We interpret these results as a clear indication of a developmental shift in English-learning infants' reliance from language-general to language-specific cues.

It is possible that 5-month-olds did not prioritize VP cues over TP cues because they were insensitive to the vowel phonotactics. Alternately, it is also possible that 5-month-olds are aware of the English phonotactic restriction on the positioning of lax vowels but did not weight it over TP cues to segment words. Recall that we equated the duration of each syllable, despite the tendency for syllables with lax vowels in English to be shorter than ones with tense vowels. This adjustment was crucial to prevent infants from conflating differences in vowel quality with differences in stress, as syllable duration consistently cues stress in English (e.g., Lieberman, 1960). However, since duration is also a consistent cue to distinguish tense and lax vowels in English, the acoustic-phonetic cues for vowel tenseness might not have been distinct enough, at least for the 5-month-olds. Note that the 10-month-olds relied on the VP cues for word segmentation, despite the absence of duration cues to signal lax vowels. If indeed there is a developmental shift in English-learning infants' reliance on durational versus spectral cues to distinguish English tense and lax vowels, also remains to be determined. More generally, we need converging evidence to abstract away from specific acoustic instantiations of stimuli in artificial languages to draw inferences about acquisition in a natural setting.

To summarize, using an artificial language paradigm, we evaluated whether English-learning infants rely on domain-general distributional information or language-specific phonotactic cues to word segmentation. Results of our experiments align with the statistical bootstrapping account, where English-learning infants initially rely on domain-general distributional information to segment words before using language-specific cues. Whether this developmental timeline is also observed under more ecologically valid conditions, for instance when the length of the words is variable and not fixed, or with other language-specific cues in English, or other languages, is a question for future research. Finally, how these findings from infants can be reconciled with proposals that learners successfully extract statistical generalizations at multiple levels (syllables and segments) simultaneously (Rasin, Lan, & Katzir, 2019; Benitez & Saffran, 2021; Durvasula & Liter, 2020) remains an open question.

Acknowledgements
Research was funded by NSF BCS-2214017 to MS; part of the results was presented at the 47th Boston University Conference on Language Development (BUCLD) and the International Congress of Infant Studies, Ottawa.

References

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. 10.1016/j.jml.2012.11.001.

Benavides-Varela, S., & Mehler, J. (2015). Verbal positional memory in 7-month-olds. *Child development*, *86*(1), 209-223.

Benitez, V.L., & Saffran, J.R. (2021). Two for the price of one: Concurrent learning of words and phonotactic regularities from continuous speech. *PLoS ONE, 16*(6): e0253039.

Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Frontiers in psychology, 4*, 38830

Cole, R. & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NK: Erlbaum.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, *88*(1), 101-112.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology, 52*(4), 521-536. http://dx.doi.org/10.1037/dev0000083

Curtin, S., Mintz, T. H., & Byrd, D. (2001). Coarticulatory cues enhance infants' recognition of syllable sequences in speech. In *Proceedings of the 25th Annual Boston University Conference on Language Development* (Vol. 1, pp. 191-201). Cascadilla.

Durvasula, K., & Liter, A. (2020). There is a simplicity bias when generalizing from ambiguous data. *Phonology, 37*(2), 177-213. doi:10.1017/S0952675720000093

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in cognitive sciences*, *13*(8), 348-353.

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., & Mehler, J. (2016). On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain. *Developmental science*, *19*(3), 488-503.

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.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive psychology*, *57*(1), 56-74.

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, *17*(6), 610-641.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099-3111.

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.

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., 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., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental science*, *13*(2), 339-345.

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., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of memory and language*, *32*(3), 402-420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of memory and Language*, *33*(5), 630-645.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & psychophysics*, *61*, 1465-1476.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). "emmeans: Estimated marginal means, aka least-squares means," https://CRAN.R-project.org/package1⁄4emmeans

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451-454.

Marimon, M., Langus, A., & Höhle, B. (2024). Prosody outweighs statistics in 6-month-old German-learning infants' speech segmentation. *Infancy*, 1-21.

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

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive psychology*, *38*(4), 465-494.

Mersad, K., Goyet, L. & Nazzi, T., (2011) Cross-linguistic differences in early word form segmentation: a rhythmic-based account. *Journal of Portuguese Linguistics, 10*(1), 37-65. doi: https://doi.org/10.5334/jpl.100

Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory and Cognition*, 39, 1085-1093.

Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech.

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.

Rasin, E., Lan, N., & Katzir, R. (2019). Simultaneous learning of vowel harmony and segmentation. In Jarosz, G., Nelson, M., O'Connor, B., and Pater, J. (Eds.), *Proceedings of SCiL 2019*, vol. 2, pp. 353-357.

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

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

Skoruppa, K., Nevins, A., Gillard, A., & Rosen, S. (2015). The role of vowel phonotactics in native speech segmentation. *Journal of Phonetics*, *49*, 67-76.

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505-513.

Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., & Steffman, J. (2022). Infants' developing sensitivity to native language phonotactics: A meta-analysis. *Cognition*, *221*, 104993.

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

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychological bulletin*, *139*(4), 792-814.

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

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language learning and development*, *3*(1), 73-100.

Van de Weijer, J. (1998). *Language input for word discovery*. Unpublished Ph.D., University of Nijmegen, Nijmegen.