



DR. ANGELINE SIN MEI TSUI (Orcid ID : 0000-0003-0333-7234)

Accepted Article  
Article type : Paper (under 8000 words)

## Dual Language Statistical Word Segmentation in Infancy: Simulating a Language-Mixing Bilingual Environment

Angeline Sin Mei Tsui<sup>1,2\*</sup>

Lucy C. Erickson<sup>3</sup>

Amritha Mallikarjunn<sup>4</sup>

Erik D. Thiessen<sup>5</sup>

Christopher T. Fennell<sup>1</sup>

1. School of Psychology, University of Ottawa
2. Department of Psychology, Stanford University
3. Society for Women's Health Research
4. University of Maryland, College Park
5. Department of Psychology, Carnegie Mellon University.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/DESC.13050](https://doi.org/10.1111/DESC.13050)

This article is protected by copyright. All rights reserved

**\*Corresponding author:** Angeline Sin Mei Tsui, atsui029@uottawa.ca

**Conflict of interest statement:** The authors do not have any conflicts of interest to report.

**Acknowledgments:** We thank Sarah Fitzpatrick for help in creating stimuli, and thank Maxine Berthiaume, Anika Talukder, Emma-Lee Peller and Caroline Zamor for coding the data. We also thank the parents and infants who participated in this study. This work was supported by a Natural Sciences and Engineering Council of Canada Discovery Grant (2014-04590) and the Official Languages and Bilingualism Institute Research Chair in Language Learning and Acquisition to Christopher T. Fennell.

**Data Availability Statement:** The data and the codes of the analysis reported in the current paper are available in the Open Science Framework at OSF.IO/U5VWK.

### Research highlights

- Infants at 9.5 and 12 months readily used phonetic cues to segment one of the two intermixed artificial languages.
- Monolingual and bilingual infants did not statistically differ in segmenting words from one of the two artificial languages.
- Infants' success of segmenting only one of the two languages reflects the challenges of learning words in a mixed-language bilingual environment.

## Abstract

Infants are sensitive to syllable co-occurrence probabilities when segmenting words from fluent speech. However, segmenting two languages overlapping at the syllabic level is challenging because the statistical cues across the languages are incongruent. Successful segmentation thus relies on infants' ability to separate language inputs and track the statistics of each language. Here, we report three experiments in investigating how infants statistically segment words from two overlapping languages in a simulated language-mixing bilingual environment. In the first two experiments, we investigated whether 9.5-month-olds can use French and English phonetic markers to segment words from two overlapping artificial languages produced by one individual. After showing that infants could segment the languages when the languages were presented in isolation (Experiment 1), we presented infants with two interleaved languages differing in phonetic cues (Experiment 2). Both monolingual and bilingual infants successfully segmented words from one of the two languages—the language heard last during familiarization. In Experiment 3, a conceptual replication, we replicated the findings of Experiment 2 with a different population and with different cues. As before, when 12-month-old monolingual infants heard two interleaved languages differing in English and Finnish phonetic cues, they learned only the last language heard during familiarization. Together, our findings suggest that segmenting words in a language-mixing environment is challenging, but infants possess a nascent ability to recruit phonetic cues to segment words from one of two overlapping languages in a bilingual-like environment.

*Keywords:* infant statistical learning, word segmentation, bilingualism, language mixing

## Introduction

When learning a new language, learners rarely receive explicit information about word boundaries within fluent speech. Words in speech are not consistently marked by pauses and rarely produced in isolation (Aslin, Woodward, LeMendola & Bever, 1996; Brent & Siskind, 2001). Given the lack of explicit cues, how do infants, who are novice language learners, extract words from speech? One possibility is that infants can discover words based on the transitional probabilities across syllables in the speech stream. Syllable-based transitional probability refers to the probability of syllable X given the occurrence of syllable Y. Syllable transitions within words are characterized by higher transitional probabilities, because syllables within words tend to more reliably co-occur together than syllables crossing word boundaries. Consider the phrase “pretty baby”: the probability that “pre” precedes “tty” is high because “pre” in English speech is typically followed by relatively few syllables (e.g., prepare, pretend). However, the probability that the final syllable of “pretty” is followed by the first syllable of “baby” is much lower, as “tty” commonly precedes many other syllables (e.g., “pretty dog”, “pretty cat”). Indeed, the transitional probability that “pre” will be followed by “tty” is approximately 80% in English speech to young infants whereas the “tty-ba” transition is much less probable—approximately 0.03% (Saffran, 2003). These statistical structures can provide clues to young infants that “pretty” is a word but “tty#ba” is not. This sensitivity to co-occurrence information—statistical learning—can aid in speech segmentation.

Statistical learning enables learners to detect statistical regularities by tracking stable patterns in the environment. Many studies (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) report that 7- to 8-month-olds can extract novel word forms based on syllable transitional probabilities after brief exposure to a fluent speech stream (see Thiessen & Erickson, 2013, for evidence that younger infants also possess this ability). Further, research shows that older infants readily map statistically-segmented new words to referents in word learning tasks (Graf Estes, Evans, Alibali & Saffran, 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011), suggesting that statistical learning supports early lexical acquisition.

However, most laboratory infant statistical segmentation studies have used a single input stream. Although valid for monolingual learning, it is not clear how this applies to bilingual infants, who must segment two language streams. Statistical segmentation of two input streams is especially

challenging when languages share syllables—which describes nearly all language pairings. For example, suppose a French-English bilingual learner hears the following syllable sequence: *mo#to*. In French, *moto* is a word (i.e., motorcycle), and is characterized by a high-probability transition. In contrast, *moto* is not a word in English, and the low transitional probability between these syllables might signal a word boundary (e.g., This *limo* totally rocks.). To successfully segment words from two different languages using statistical information, learners must separately represent and track the transitional probabilities of each language system.

To date, research on how learners statistically segment words from two language inputs (i.e., a dual-input segmentation task) is relatively scarce. Pioneering work by Weiss and colleagues suggests that monolingual adult learners can statistically segment words from two languages, even when they share some syllables. However, they found that adult learners could only successfully segment each language when strong contextual cues marked the difference between the languages. Specifically, learners only succeeded when each language was uniquely associated with either a male or female voice (Weiss, Gerfen & Mitchel, 2009) or two different faces (Mitchel & Weiss, 2010). While adult learners demonstrate some success in segmenting two languages with overlapping syllables without contextual cues (Bulgarelli & Weiss, 2016; Gebhart, Aslin & Newport, 2009), they are less successful in such context-free situations. For example, Gebhart et al. (2009) reported that adults only segmented the first-presented language and not the second when they lacked clear contextual cues. These studies suggest that monolingual adult learners segment words from two artificial languages best when provided with strong, salient contextual cues that differentiate the languages. But what of infants, who are less cognitively mature? Can they statistically segment two overlapping languages when provided with contextual cues to distinguish the languages? This question is important to understand how infants overcome the challenges of differentiating two languages and discovering word forms in bilingual environments.

Three recent studies have shed light on whether infants can statistically segment words from two language inputs. Benitez, Bulgarelli, Byers-Heinlein, Saffran and Weiss (2020) reported that 8-month-old monolingual infants failed to segment words in a dual-input segmentation task that lacked contextual cues to differentiate the two languages. In follow-up studies, they investigated whether adding distinguishing contextual cues (e.g., differences in pitch, speakers, and person-based accent)

could help infants succeed. Even with these cues, infants failed to learn. In contrast, Antovich and Graf Estes (2018) found that older 14-month-old bilingual infants successfully segmented two languages using statistical cues when the languages were clearly marked by voices of different genders; yet, monolingual infants failed in the same task. These results suggest that older bilingual infants may be better equipped to process and segment two language inputs than age-matched monolinguals. However, the languages in Antovich & Graf Estes (2018) did not share syllables. The statistical cues across the two languages were compatible, even if learners combined languages. This raises the possibility that bilingual infants may simply outperform monolingual infants when segmenting words from a longer, single language input rather than two language inputs.

Antovich & Graf Estes (2020) replicated their study by testing monolingual and bilingual 16-month-olds with two languages that overlap in syllabic inventory (i.e., the statistical cues across languages are incongruent). They again found that only bilinguals succeeded in segmenting two languages. Importantly, they reported that bilinguals' experience with more speakers, but not their cognitive ability, was correlated with better segmentation performance. Together, these studies suggest that older bilingual infants have an advantage in statistically segmenting two languages and this advantage stems from their daily dual-language experience. Further, it appears that dual-input statistical segmentation is challenging, as only older infants succeed in the task. This contradicts studies reporting bilinguals can segment words in their native languages during their first year of life (e.g., Bosch, Figueras, Teixido, Ramon-Casas, 2013; Polka, Orena, Sundara & Worrall, 2017). Thus, it is still unknown (i) whether younger infants can statistically segment words from two overlapping languages; (ii) what contextual cues can support infants' segmentation.

The first aim of the present study is to investigate whether young infants in their first year can segment words from two languages that share some syllables, resulting in incompatible statistical cues across languages. Further, we extend the current literature by exploring whether natural phonetic cues within a single individual's speech can facilitate statistical segmentation of two overlapping languages. Previous studies have explored how speaker-specific contextual cues (e.g. two individual speakers differing in gender (Weiss et al., 2009) or accent (Benitez et al., in press), or languages associated with two distinct faces (Mitchel & Weiss, 2010) facilitate word segmentation in two languages. However, in naturalistic bilingual environments, different speakers do not always speak

distinct languages. After all, a bilingual parent must switch between languages as a function of context (e.g., speaking English to a partner, but French to a neighbor). Furthermore, bilinguals often code-switch—alternate between languages within conversations—when speaking with other bilinguals (e.g., Poplack, 1980; see Heredia & Altarriba, 2001, for a review), including with their own infants (Byers-Heinlein, 2013). It is thus necessary and ecologically valid to examine whether contextual cues within a single person's speech effectively support statistical segmentation from two language inputs.

To address this concern, we investigated whether phonetic cues within a single individual's speech can support segmentation of two languages. Phonetic realizations of syllables typically vary across languages, such that languages often sound different at the syllable level. Recall our previous example about the syllable combination “*mo-to*”. Although it is the same phoneme sequence in both French and English, the phonetic properties clearly differ between languages. “*Mo-to*” would be produced with tenser vowels and shorter duration in French than in English. Hence, we argue that phonetic variation is a relevant contextual cue that can signal a language change within a single individual's speech. This cue may facilitate infants' segmentation of words from two languages.

In the current paper, we presented infants with two interleaved artificial languages differing in phonetic cues. The statistical structures of these two languages were incongruent, as they shared syllables. Prior studies (e.g., Weiss et al., 2009; Benitez et al., in press) reported that infant and adult learners failed to segment either of two interleaved languages with incongruent statistical cues when receiving no cues to differentiate the two languages. Thus, while an obvious measure of success is segmenting words from both languages, we argue that segmentation of one of the two languages would also indicate success in the dual-input statistical segmentation task for two reasons. First, successful segmentation of one of the two languages would also require infants to differentiate between two language inputs and track transitional probabilities correctly for that specific language. Second, it is an increase in performance from the established baseline of failing to segment either language in the studies above.

In Experiment 1, we first presented 9.5-month old monolingual Canadian-English infants with each language presented in isolation, and tested whether they could segment words from the languages when presented alone. In Experiment 2, we presented both 9.5-month-old monolingual

Canadian-English and bilingual (Canadian-French and -English) infants with two interleaved languages differing in French and English phonetic cues. We included bilingual infants to test whether early bilingual exposure may facilitate statistical segmentation of two languages with overlapping syllables. Bilinguals reportedly outperform monolinguals in several statistical learning tasks, including segmenting words from two languages (infants: Antovich & Graf Estes, 2018; adults: Bartolotti, Marian, Schroeder, & Shook, 2011), segmenting words from a tonal language (Wang & Saffran, 2014), and learning words across different scenarios (Poepsel & Weiss, 2016). Thus, bilingual infants may outperform monolingual infants in our segmentation task. In Experiment 3, we conducted a conceptual replication of Experiment 2 to investigate the stability of our results, using a different population, set of stimuli, and age group. Specifically, we examined whether 12-month-old monolingual American-English infants can segment words from two interleaved languages differing in Finnish and English phonetic cues.

## Experiment 1

Before investigating whether infants can segment words from two interleaved artificial languages, we investigated if infants can segment words from an artificial language presented in isolation in our experimental set-up. We familiarized 9.5-month-old monolingual Canadian-English infants with a single artificial language from our set in which the only cue to segmentation was transitional probabilities between syllables. If infants can use transitional probability to segment words when a language is presented in isolation, we predicted that they would attend differently to words (trisyllable sequences with high transitional probabilities between all syllables) and part-words (trisyllable sequences containing a low-probability transition) at test.

### Method

**Participants.** We tested 26 typically developing 9.5-month-old infants (Mean age: 9.41,  $SD = 0.65$ , 13 females) from monolingual Canadian-English homes (i.e., parents reported that infants were only exposed to English). Four additional infants were tested but excluded from analysis because of crying or fussiness (3) or not attending to the stimuli during the experiment (1).

**Stimuli.** We created two artificial languages, AL1 and AL2, based on the scripts used in the incongruent condition of Experiment 2 in Weiss et al (2009), with the only change being that the vowel *a* replaced the vowel *æ* in our script. This ensured that all phonemes in the final script are

present in both Canadian English and Canadian French, and consequently that the speaker would not add, substitute or delete phonemes when producing the syllables in the scripts. Each language comprised four trisyllabic CV.CV.CV words presented an equal number of times with no word repeated twice in a row. The transitional probabilities signaling the words and word boundaries in each language were 100% and 33% respectively. For each language, test items consisted of two words<sup>1</sup> (i.e., trisyllabic sequences where all adjacent syllables within the word itself had a transitional probability equal to 1.0) and two part-words (i.e., syllable sequences that spanned word boundaries such that adjacent syllables had low transitional probabilities).

For the stimuli, we recorded a balanced Canadian French-English bilingual female speaker. When recording the stimuli, the speaker produced all possible trisyllabic sequences in each language (AL1 and AL2). Only the middle syllable of each trisyllabic sequence was extracted when concatenating the syllables for the creation of the artificial languages. This ensured that all syllables in the languages were fully co-articulated with their surrounding syllables, including low-probability transitions. Thus, coarticulation was not a possible cue to word segmentation from the language speech streams, as it could not be used to differentiate the words from part-words. We recorded the speaker in two different sessions, producing all the stimuli with Canadian-English phonetic cues on the first day and all the stimuli with Canadian-French phonetic cues on the second. See Table 1 for the stimuli, Table 2 for acoustic measurements and Table 3 for adults' rating of the stimuli.

**Apparatus.** Testing took place in a 2.38m X 1.82m sound-attenuated room dimly-lit by a 60W lamp that was placed 80cm to the left of the infant and parent. Visual stimuli were projected onto a SmartBoard screen from a NEC Duocom LT280 projector. The auditory stimuli were played at 65 dB (+/- 5 dB) by two speakers below the screen. A digital camera, hidden under a table draped with black cloth below the screen, captured infants' performance. The camera lens poked out of a hole in the black cloth 10 cm below the screen.

**Procedure.** We used a single-screen procedure, similar to previous adaptions of the Headturn Preference Procedure for word segmentation studies (Altvater-Mackensen & Mani, 2013, Thiessen &

---

<sup>1</sup> Although each language was recorded to consist of four words, we only tested infants in the current paper on two of the words and two of the part-words. We reduced the number of words because a number of infants in a pilot study fussed out when tested on the comprehensive set of test items containing all four words and all four part-words. Importantly, the pilot study also revealed that infants may have preferences for some words and part-words in the two languages. The two words and two part-words that yielded the longest and shortest looking times in the pilot study were not used in Experiment 1 in order to minimize the influence of infants' pre-existing preferences for acoustic features of the test items.

Erickson, 2013). Infants were seated on their parent's lap approximately 150cm away from the screen. To eliminate parental interference, parents were instructed not to talk to the infants and listened to female vocal music via headphones to mask the stimuli. The experimenter observed the infants from a different room via a closed-circuit video system, using Habit 2 software (Oakes, Sperka & Cantrell, 2015) to present stimuli and record looking time. The experimenter was blind to the nature of the stimuli being presented.

There were two phases in the experiment: familiarization and test. An animated cartoon baby drew infants' attention to the display before familiarization and before each test trial. When the infants oriented to the screen, the familiarization phase began. During familiarization, a colorful Winnie the Pooh video accompanying the artificial language helped infants stay engaged in the task.

Infants heard a single language during familiarization. They were randomly assigned to one of the four conditions: (i) AL1 with Canadian-English phonetic cues, (ii) AL1 in with Canadian-French phonetic cues, (iii) AL2 with Canadian-English phonetic cues, or (iv) AL2 with Canadian-French phonetic cues. For speech with French phonetic cues, the duration of AL1 was approximately 133.76 sec and that of AL2 was 130.38 sec. For speech with English phonetic cues, the duration of AL1 was 141.62 sec and that of AL2 was 139.15 sec. Infants immediately proceeded to the test phase after familiarization. In this phase, infants received 12 test trials: three blocks of four items. Within each block, infants heard 2 words and 2 part-words from the familiarized language. The order of test items was counterbalanced across infants. Phonetic cues were consistent across familiarization and test phases (e.g. if infants heard AL1 with French phonetic cues, then they would be tested on AL1 with French phonetic cues). When listening to the audio stimuli during test trials, infants watched an animated green circle (changing in size over time) onscreen. Thus, infants' attention to test items was measured by looking time to the screen. Test trials ended when infants looked away for more than 1 sec or reached the maximum length of the trial (20 sec).

**Coding.** Videos of all test trials were coded offline using frame-by-frame analysis (1 frame = 30ms). Two research assistants, blind to condition, coded all data. We calculated a difference score (i.e., the difference in average looking time between words and part-words) and used this score to test the reliability of the coding of the two assistants. A Pearson's correlation test revealed a high correlation between the two assistants' coding ( $r = 0.96, p < 0.001$ ).

## Results and Discussion

One infant whose difference score deviated by more than  $2.50\text{ SDs}$  from the group mean was excluded from the following analyses. If infants segmented the artificial language, their average looking time to words and part-words should significantly differ. First, we ran a preliminary analysis to test whether infants' gender (male or female), language phonetic cues (French or English cues) and language at test (AL1 or AL2) influenced infants' looking preference to words and part-words. We entered these between-subject factors into an ANOVA and found that none of these factors significantly explained the variance of infants' looking preference to words and part-words ( $p > 0.21$ ). Therefore, we collapsed across these variables in the following main analysis. Using a within-subjects t-test, we found that infants looked significantly longer to words ( $M = 10.14\text{s}$ ,  $SD = 3.05\text{s}$ ) compared to part-word items ( $M = 7.86\text{s}$ ,  $SD = 2.76\text{s}$ ) at test [ $t(25) = -3.40$ ,  $p = 0.002$ ,  $d = -0.67$ ] (see Figure 2). Hence, our findings indicate that infants can segment words from the language when it was presented alone. However, we noted that infants showed a familiarity preference for words in this segmentation task. The direction of infants' preference for words and part-words has implications for their underlying learning processes in this experiment (Hunter & Ames, 1998). Normally, infants typically show a preference for novel stimuli (i.e., novelty preference) when the familiarization stimuli are fully processed. In contrast, infants show a preference for familiar stimuli (i.e., familiarity preference) when they have learned aspects of the familiarization stimuli without fully processing the information, which may be the case in our experiment. However, while prior statistical-learning word-segmentation experiments (e.g., Aslin et al., 1998; Saffran et al., 1996) predominantly have demonstrated that infants pay more attention to part-words than words, the direction of preference reported in the statistical segmentation literature is not uniform (e.g., Graf Estes, Gluck & Bastos, 2015; Graf Estes & Lew-William, 2015; Thiessen, Hill, & Saffran, 2005). Nevertheless, the significant difference between word and part-words in Experiment 1 clearly demonstrates that monolingual English infants can successfully segment words based on the syllabic transitional probability in the artificial languages presented in isolation. We next tested whether infants can segment words from two language inputs presented within the same learning session.

## Experiment 2

This experiment investigated whether monolingual English and French-English bilingual 9.5-month-old infants can learn two interleaved artificial languages when the syllable inventory overlaps across the two languages. Importantly, these overlapping cues would lead to a failure to properly segment words if learners cannot separate the two languages, as they would be collapsing the statistics across the languages (see Apparatus and Procedure, below, for a precise description about how collapsing across languages alters transitional probabilities).

To help infants differentiate the languages produced by a single talker, the two interleaved artificial languages differed in terms of phonetic cues based on two natural languages (i.e., English and French). This allows us to explore how infants segment words in a language-mixing context. For example, if AL1 was presented with French phonetic cues, AL2 would be presented with English phonetic cues, or vice-versa. Further, we selected English and French phonetic cues to test whether daily experience of hearing the phonetic cues of their native languages could bolster bilingual infants' processing of language inputs marked by those specific differences. Phonetic cue changes within a single individual's speech are potentially subtler than a change from one speaker's accent to another speaker's accent, as in previous research. But within-speaker phonetic cues are arguably more ecologically valid for studying bilingual infants because bilingual parents often mix two languages together when talking to infants (Byers-Heinlein, 2013).

Our hypotheses are as follows. First, if infants are able to separate the two languages, they could then learn the statistical structure of each language independently. If learners successfully tracked each language independently, they would demonstrate different looking times for items that conform to high syllable transitional probabilities in one of the languages at test than for those items that conform to low probabilities. Second, if bilingual infants outperform monolingual infants in this task, bilinguals are expected to show a bigger effect size of learning than monolinguals, which would be indicated by a significant interaction between infants' language background (bilingual or monolingual) and infants' looking preference to words and part-words.

### Method

**Participants.** Sixty-nine infants (Mean age = 9.54,  $SD = 0.67$ , 37 females) were tested: thirty-four monolinguals (Mean age = 9.60,  $SD = 0.64$ , 19 females) and thirty-five bilinguals (Mean age =

9.47,  $SD = 0.69$ , 18 females). Monolingual English infants were only exposed to English as reported by parents. Bilingual infants were exposed to English and French since birth and had a maximum of 80% exposure to one language and a minimum of 20% exposure to the other language (Mean English exposure = 53.40%,  $SD = 17.01\%$ ; Mean French exposure = 46.38%,  $SD = 17.11\%$ ), as determined by the Language Exposure Questionnaire (Bosch & Sebástian-Gallés, 1997). Twelve additional infants were tested but not included in the final analysis because of crying/fussiness (4 monolingual, 7 bilingual), or exposure to English fell outside our limits (1 bilingual).

**Stimuli.** The artificial language stimuli were identical to those used in Experiment 1.

**Apparatus and procedure.** The apparatus and general procedure were almost identical to Experiment 1, with the only difference being the addition of another language during familiarization. The total exposure to an individual language was identical to the length of its exposure in Experiment 1. There were four language exposure trials during familiarization; AL1 and AL2 were presented in an interleaved order with half the duration of each language presented per trial.

Following Antovich and Graf Estes (2018), we added a 5 sec pause between languages. Previous research (Antovich & Graf Estes, 2018; Benitez et al., in press) has shown that infants were unsuccessful in segmenting words from two languages even with pauses between languages. Thus, we argue that pauses between languages alone would not support infants' word segmentation, and additional contextual cues (i.e., phonetic cues in our study) would remain the key driver of successful word segmentation of the languages. In addition to the 5 sec pause between the two languages, the major transition cue between AL1 and AL2 was a shift between English and French phonetic cues (i.e., AL1 and AL2 always had different phonetic cues). AL1 with English phonetic cues was paired with AL2 with French phonetic cues, whereas AL1 with French phonetic cues was paired with AL2 with English phonetic cues.

As a reminder, the transitional probability signaling the word boundaries in each language were 1.0 within word and 0.33 across word. However, this presumes that participants can separate the two languages and calculate the statistical set of each. If participants combined the two languages, the transitional probabilities would be inconsistent. The transitional probability within words would vary from 0.5 to 1.0 whereas the transitional probability across word boundaries would vary from 0.17 to 0.33 (see Figure 1). Apart from statistical cues, there were no pauses or other acoustic cues (e.g.

stress) between syllables within each language that might signal any word boundaries. The order of phonetic cues (hearing a language with English phonetic cues first or second) and presentation of the language (AL1 presented first or second during familiarization) were counterbalanced across infants.

At test, infants were only tested on one of the artificial languages, allowing for a direct comparison to the test phase of Experiment 1. Half of the infants were tested on words and part-words from AL1 and the other half on items from AL2. Phonetic cues for a language were consistent across familiarization and test (e.g. if infants heard AL1 with French phonetic cues and AL2 with English phonetic cues, then they would be tested on either AL1 with French phonetic cues or AL2 with English phonetic cues).

**Coding.** As in Experiment 1, infants' looking times at test were coded by two research assistants. The inter-coder reliability (measured by Pearson correlation of the difference scores between coders) was high ( $r = 0.97$ ,  $p < 0.0001$ ).

## Results and Discussion

If infants can separate the two artificial languages and successfully segment each language, they should look differentially to word compared to part-word test trials. In the following analysis, we added one more factor, presentation order of the tested language, based on the results of Gebhart et al. (2009). In that study, adults only acquired the first-presented language when learning two overlapping languages in a difficult context without overt cues distinguishing the languages. Although we provided infants with phonetic cues to help them segment the artificial languages, it is still possible that infants found the task difficult due to their relative cognitive immaturity and the subtlety of our cues. This may lead to a similar primacy bias as seen in adults.

Prior to analysis, we found and removed the data from one infant whose difference score was more than 2.50 *SDs* from the group mean. In a preliminary analysis, we first conducted a mixed ANOVA to check whether infants' gender (male or female), language phonetic cues (French or English cues) and language at test (AL1 or AL2) influenced infants' looking preference to words and part-words. We did not find any of these factors influenced infants' looking preference ( $p > 0.3$ ). Thus, we did not control for these factors in the subsequent main analysis.

In the main analysis, we conducted a 2 (language background: monolingual or bilingual) X 2 (test trial: words or part-words) X 2 (presentation order: tested language presented first or second

during familiarization) mixed ANOVA. All main effects and two-way interaction terms were entered into the model. We found a significant main effect of test trial type [ $F(1, 65) = 6.20, p = 0.02, \eta^2 = 0.02$ ], suggesting that infants looked significantly longer to part-words ( $M = 10.01\text{s}, SD = 3.44\text{s}$ ) than words ( $M = 9.11\text{s}, SD = 3.34\text{s}$ ) (See Figure 2). There was no significant interaction between the test trial type and infants' language background [ $F(1, 65) = 1.62, p = 0.21, \eta^2 = 0.005$ ], suggesting that bilinguals and monolinguals both showed a novelty preference and did not significantly differ in the size of looking preference for part-words over words at test (see Figure 3). There was a significant interaction between test trial type and presentation order [ $F(1, 65) = 5.89, p = 0.02, \eta^2 = 0.02$ ]. Further analyses suggested that infants looked significantly longer to part-words ( $M = 10.50\text{s}, SD = 3.42\text{s}$ ) than words ( $M = 8.69\text{s}, SD = 3.3\text{s}$ ) when tested with the second-heard language from familiarization [ $t(33) = -3.48, p = 0.0009, d = 0.60$ ]. However, there was no difference between looking to part-word ( $M = 9.52, SD = 3.43\text{s}$ ) and word test trials ( $M = 9.52\text{s}, SD = 3.37\text{s}$ ) when infants were tested with the first language during familiarization [ $t(33) = -0.041, p = 0.967, d = 0.007$ ] (see Figure 4).

We performed an informal post-hoc analysis exploring whether bilinguals and monolinguals performed similarly when taking into account the tested language order (see Figure 3). There is reasonable justification for this analysis, given the recent studies demonstrating that bilinguals perform differently from monolinguals (e.g., Antovich & Graf Estes, 2018; Polka, et al., 2017). However, we note that this analysis is not statistically justified by the null main and interaction effects in the ANOVA analysis, thus the following inferential test results should be interpreted with caution. When tested with the last language from familiarization, both individual groups looked longer to part-words (bilinguals:  $M = 11.41\text{s}, SD = 4.13\text{s}$ ; monolinguals:  $M = 9.69\text{s}, SD = 2.5\text{s}$ ) than words (bilinguals:  $M = 9.83\text{s}, SD = 3.82\text{s}$ ) [ $t(16) = 2.06, p = 0.043, d = 0.5$ ; monolinguals:  $M = 7.68\text{s}, SD = 2.44\text{s}$ ] [ $t(16) = 3.65, p = 0.0005, d = 0.88$ ]. When tested with the first language from familiarization, both individual groups did not look longer to part-words (bilinguals:  $M = 8.65\text{s}, SD = 3.79\text{s}$ ; monolingual:  $M = 10.49\text{s}, SD = 2.78\text{s}$ ) than words (bilinguals:  $M = 9.32\text{s}, SD = 4.12\text{s}$ ) [ $t(16) = -0.72, p = 0.477, d = -0.17$ ; monolinguals:  $M = 9.75\text{s}, SD = 2.36\text{s}$ ] [ $t(16) = 0.75, p = 0.453, d = 0.18$ ].

These results indicate that as a group, infants successfully segmented words from the interleaved dual-input speech stream, suggesting that phonetic cues effectively supported segmentations from interleaved artificial languages. Bilinguals performed similarly to monolinguals,

suggesting daily experience with French and English phonetic cues did not enhance learning in the task. The ability to use phonetic cues to separate new languages does not appear to depend on prior experience with specific cues. There were two surprising results in this experiment. First, infants only successfully segmented words from the language that they heard last during familiarization. This raises the concern that infants have not differentiated the two language inputs, as infants could still succeed in segmenting words if they only focused on segmenting the last heard language without separating the two languages. Second, infants in this experiment demonstrated a novelty preference, looking longer to part-words than to words. This contrasts with the familiarity preference exhibited by monolingual infants tested with a single language in Experiment 1, although we note that novelty preferences are more common in statistical learning segmentation tasks (as discussed earlier). Infants typically show novelty preferences when the information from familiarization has been fully encoded (Hunter & Ames, 1988). Accordingly, Experiment 2's findings seem to suggest that monolingual infants processed the statistical cues to segmentation more thoroughly with two language inputs than with one language input. This is puzzling, because the task demands in Experiment 2 (i.e., segmenting words from two overlapping languages) are arguably higher than those of Experiment 1 (i.e., segmenting words from one language). As the results in Experiment 2 are perhaps counterintuitive and have a potential explanation that does not require differentiating the languages, we conducted a conceptual replication study and an additional control study—Experiment 3—to confirm that these findings were not due to chance. Further, given the importance of replicating findings across populations, stimuli, and laboratories, we designed our conceptual replication study to increase the generalizability of our findings in several key ways.

### Experiment 3

Experiment 3 aimed to replicate Experiment 2's findings. We tested whether infants can use phonetic cues to separate and segment words from two artificial languages with a new population, and a different and arguably stricter experimental setup. This experiment also allowed us to explore another possible explanation for Experiment 2's findings. Originally, we hypothesized that bilinguals' daily experiences hearing French and English phonetic cues may enhance their processing of language inputs differentiated by those cues. Yet, we found that both monolinguals and bilinguals succeeded in the task, and inferred that the ability to separate languages using phonetic cues does not

depend on previous experience. However, another possibility is that our English monolinguals may have been exposed to some French by virtue of living in a bilingual English-French area. Although they were not actively learning French at home, it is possible that some degree of exposure to French in the ambient environment enhanced their use of that language's phonetic cues. To negate infants' pre-exposure to our target cues, we tested English monolingual infants living in an English monolingual culture (the USA) and chose not to use French phonetic cues as prior research suggests that syllabic structures in Spanish and French can be processed in a similar way (e.g., Pallier et al., 1993) and Spanish is the most common second languages in the USA (U.S. Census Bureau, 2015). We therefore used Finnish phonetic cues, which are very unlikely to be familiar to American-English monolingual infants.

Experiment 3 also allowed us to explore the discrepancy between our monolingual findings in Experiment 2 and those reported in the past. Both 9.5-month-old monolinguals and bilinguals succeeded in our dual-input segmentation task, whereas Antovich and Graf Estes (2020) found that only 16-month-old bilinguals succeeded in a similar task. One possibility is that language-based phonetic cues are strong natural cues for monolingual infants to segment words from one of two languages. The failure of 16-month-old monolinguals in Antovich et al.'s study may be due to the use of non-language-based cues (voices of different genders). To address this question, we extended our study to older infants. Given that monolinguals as young as 12 months show some ability to differentiate between two different linguistics rule structures (Kovács & Mehler, 2009), we decided to test 12-month-old infants in Experiment 3 to examine whether monolinguals' success in a dual-input segmentation task were stable beyond 9.5 months, rather than constituting an odd blip of success at a younger age before native language skills have become more developed. Further, bilingual infants succeeded in the dual-input segmentation task in both our study and in Antovich et al.'s studies, demonstrating stable results. In addition, because of the difficulty recruiting a homogeneous bilingual population in the USA where bilinguals come from different places, we did not include bilinguals in this experiment.

The current experiment comprises two studies—Experiments 3a and 3b. Experiment 3a aimed to replicate our findings in Experiment 2 by presenting infants with two languages that alternated with different phonetic cues. Experiment 3b was a control study that tested whether infants can segment

words successfully when phonetic cues did not differentiate the languages. Considering our results in Experiment 2 where infants only succeeded in segmenting the last-heard language, this control is especially relevant to test whether infants would simply focus on and remember the last heard language regardless of any language-based cues. If phonetic cues facilitate infants' separation of two languages for successful word segmentation, infants should succeed in Experiment 3a but not Experiment 3b.

Finally, we highlight that this extension was conducted with a different set of stimuli, a different infant population, and at a different site. If the findings of Experiment 2 were subject to variations in one of these factors, it is unlikely that similar results would be found in Experiment 3a. In contrast, if similar results are found in Experiment 3a, it would suggest these findings reflect something stable about infant performance. In addition, replicating our finding under these circumstances would increase the generalizability of our findings, a key goal in infancy research (Frank et al., 2017; ManyBabies Consortium, 2020).

## Method

**Participants.** We tested 120 typically developing 12-month-old monolingual American-English infants (Mean age: 11.94,  $SD = 0.28$ , 56 females). An additional 21 infants were tested but excluded from the final analysis because of fussing/crying (14), experimenter error (4), or average test trial looking times below two seconds (3).

**Stimuli.** Following Experiments 1 and 2, we created the two artificial languages based on the scripts used Weiss et al (2009). The only differences were slight phonological changes to ensure that all phonemes in the languages were present in both English and Finnish, and consequently that the speaker would not add, substitute, or delete phonemes when producing the syllables.

Each language contained four trisyllabic CV.CV.CV words. The transitional probabilities across syllables in each language were 100% within a word and 33% at word boundary. As in our previous experiments, we tested infants only on two words and two part-words in one of the two languages.

The stimuli were recorded by a balanced Finnish-English bilingual female speaker. During recording, the speaker produced all possible trisyllabic sequences in each language (Artificial Language 1 (AL1) and Artificial Language 2 (AL2)). As before, we extracted the middle syllable of

each trisyllabic sequence only when concatenating the syllables for the creation of the artificial languages, which ensures full co-articulation between surrounding syllables in the languages. The speaker recorded the stimuli in two sessions, producing the stimuli with Finnish phonetic cues and the stimuli with English phonetic cues on different days. See Table 4 and 5 for the acoustic measurements and adult ratings of our stimuli.

**Apparatus.** Testing took place in a 3.2m X 2.1m sound-attenuated room dimly-lit by a 60W light fixture above the infant and parent. Visual stimuli were displayed on a 42" monitor. The auditory stimuli were played at 65 dB (+/- 5 dB) by two speakers on either side of the monitor. A camera mounted above the monitor recorded infants' performance.

**Procedure.** The procedure used in this experiment was similar to that used in Experiments 1 and 2. The only differences were due to the stimuli presented in Experiments 3a and 3b. Half the infants heard one language with Finnish phonetic cues and the other language with English phonetic cues (Experiment 3a), whereas half the infants heard both AL1 and AL2 with the same phonetic cues (e.g., both AL1 and AL2 were produced with Finnish phonetic cues; Experiment 3b). Language order and language/phonetic cues pairing were counterbalanced across infants.

**Coding.** We coded videos online using frame-by-frame analysis. The coding of the two coders were highly correlated, based on a sample of 25% of the sessions that were coded by both coders ( $r = 0.93, p < 0.001$ ).

## Results and Discussion

We first identified and removed the data from one infant in Experiment 3b whose difference score was more than 2.50 *SDs* from the mean difference score of the group. Next, we conducted a 2 (test trial: word or part-word) by 2 (condition: Experiment 3a vs. Experiment 3b) ANOVA and found a significant interaction between test trials and conditions [ $F(1, 117) = 4.47, p = 0.04, \eta^2 = 0.005$ ], suggesting that infants performed differently between Experiments 3a and 3b. Thus, in the following, we ran two separate analyses according to the conditions. For both sets of analyses, we first conducted a preliminary analysis by running a mixed ANOVA to check whether infants' gender (male or female) or language phonetic cues (English or Finnish), and language at test (AL1 or AL2) influenced infants' looking preference to words and part-words. None of these factors influenced infants' looking

preferences (all  $p > 0.15$  in both Experiment 3a and 3b), we therefore did not control for these factors in the subsequent main analyses.

In the main analysis for Experiment 3a, we performed a 2 (test trial: word or part-word)  $\times$  2 (presentation order: tested language presented first or second during familiarization) ANOVA. We found a main effect of test trial type [ $F(1, 58) = 5.49, p = 0.02, \eta^2 = 0.01$ ] indicating infants looked significantly longer at part-words ( $M = 12.88\text{s}, SD = 3.4\text{s}$ ) than to words ( $M = 11.99\text{s}, SD = 4.19\text{s}$ ), consistent with the pattern seen in Experiments 2. Importantly, we also found a significant interaction between test trial and presentation order [ $F(1, 58) = 12.16, p < 0.001, \eta^2 = 0.03$ ]. Follow up analyses indicated that infants looked significantly longer to part-words ( $M = 14.05\text{s}, SD = 2.96\text{s}$ ) than words ( $M = 11.83\text{s}, SD = 4.28\text{s}$ ) when tested on the language they heard second,  $t(29) = 4.12, p < 0.001, d = 0.75$ . However, when tested on the first-heard language, infants showed no significant difference in looking to words ( $M = 12.16\text{s}, SD = 4.18\text{s}$ ) and part-words ( $M = 11.72\text{s}, SD = 3.45\text{s}$ ) [ $t(29) = -0.81, p = 0.422, d = -0.15$ ].

In the main analysis for Experiment 3b, we performed a 2 (test trial: word or part-word)  $\times$  2 (presentation order: tested language presented first or second during familiarization) ANOVA. We found no main effect of test trials [ $F(1, 57) = 0.26, p = 0.61, \eta^2 = 0.0003$ ] or presentation order [ $F(1, 57) < 0.001, p = 0.98, \eta^2 = < 0.0001$ ], and no interactions [ $F(1, 57) = 0.17, p = 0.68, \eta^2 = 0.0002$ ]. Unlike infants in Experiments 2 and 3a, infants did not prefer part-words ( $M = 12.16\text{s}, SD = 3.53\text{s}$ ) over words ( $M = 12.28\text{s}, SD = 3.46\text{s}$ ). There were no significant differences between infants tested with test items from the first-heard language (word  $M: 12.34\text{s}$ ; part-word  $M: 12.12\text{s}$ ) [ $t(29) = -0.66, p = 0.512, d = -0.12$ ], and infants tested with items from the second-heard language (word  $M: 12.22\text{s}$ ; part-word  $M: 12.2\text{s}$ ) [ $t(29) = -0.068, p = 0.945, d = -0.012$ ].

Taken together, the results of Experiments 3a and 3b showed that infants succeeded in differentiating two language inputs differing in Finnish and English phonetic cues. Further, our overall null findings in Experiment 3b, including infants' failure to segment the second-heard language during familiarization, suggest that infants needed the phonetic cues to separate the two language inputs for successful word segmentation.

## General Discussion

The current paper investigated whether infants can segment word forms from two interleaved artificial languages using transitional probability cues. Importantly, these two languages overlapped partially at the syllable level. Thus, successful learning required infants to represent the two languages separately in order to statistically segment words. In our studies, we used phonetic cues from two different natural languages as the marker for infants to separate the two overlapping languages. Both types of phonetic cues were produced by the same person, allowing us to explore whether infants can segment words from two languages within a single individual's speech. Our study design closely approximates how infants segment words in naturalistic bilingual environments. Bilingual parents commonly alternate between languages when speaking to their infants (Byers-Heinlein, 2013; De Houwer, Bornstein & Putnick, 2014), and each language is typically produced with all or many of the associated language-appropriate phonetic cues (Bosch & Ramon-Casas, 2011, except cases with very strong accents). Thus, our findings have strong ecological validity with respect to infants' word segmentation in real-world bilingual environments.

Results from Experiment 1 demonstrated that monolingual infants can statistically segment words from an individual language speech stream, suggesting that infants can segment words from the languages in Experiment 2 when the languages were presented alone. In Experiment 2, we showed that both monolingual and bilingual infants can successfully segment words from one of two interleaved languages. Unlike some previous studies (Antovich & Graf Estes, 2018; Bartolotti et al., 2011), we did not find any bilingual advantage. The common success across monolingual and bilingual groups may relate to young infants' heightened sensitivity to phonetic cues. Infants under one year of age are still highly sensitive to both native and non-native phonetic sounds (Kuhl et al., 2006; Werker & Tees, 1984). Additionally, our follow up analyses indicated that infants' learning in Experiment 2 depended on the presentation order of the language stimuli. Specifically, infants successfully discriminated words and part-words only when tested on words from the language heard last during familiarization, raising the concern that infants may simply segment words from the last-heard language without differentiating the languages. In a conceptual replication (Experiment 3) using different language cues, older monolingual infants similarly successfully segmented words from the language they heard last during familiarization. However, older infants could not segment words from

two languages that did not differ in phonetic cues. Together, our findings show that infants, even those with no previous bilingual experience and who are older, readily recruit phonetic cues to separate two languages and successfully segmented one of the two languages. This has strong implications for early bilingual language acquisition.

By showing that phonetic cues can facilitate infant statistical segmentation in a bilingual environment, our findings align with work showing bilingual infants are sensitive to the language-specific cues of their two languages when segmenting words from natural speech. Past research showed that bilingual infants can segment monosyllabic and bi-syllabic words from their native languages (Bosch, Figueras, Teixido, Ramon-Casas, 2013; Orena & Polka, 2019; Polka et al., 2017; Singh & Foong, 2012). Overall, bilingual infants' segmentation success hinges on their language experience. For example, bilingual infants segment words better from their dominant language (Bosch et al., 2013; Orena & Polka, 2019). Singh and Foong (2012) revealed that it is not until 11-months of age that Mandarin-English bilingual infants can identify segmented words in correct language contexts (e.g., tones in words matters in Mandarin but not English contexts). Importantly, Polka and colleagues (2017) have shown that 8-month-old French-English infants only successfully segmented words from English with longer exposure to English sentences or longer exposure to both English and French within a language-mixing context (Orena & Polka, 2019). These findings suggest that infants segment words better with more bilingual language experience over time (i.e., increasing exposure to the languages). This point also provides a potential explanation of why our lab-based studies found that infants can segment one of the two languages whereas another similar study failed (Benitez et al., in press). A key methodological difference between the two studies is that we presented infants with each language twice, but Benitez's study only presented infants each language once. With longer exposure, infants in our study were able to better process information and succeed in segmenting words from one of the two languages.

Our findings suggested that infants generally find word segmentation in a language mixing context challenging (i.e., infants only showed evidence of learning the last-presented language stream during familiarization). Limits in infants' cognitive capacity may only allow them to retain the latest language input heard during familiarization. This argument is consistent with prior research reporting that infant word learning in a mixed language context is cognitively demanding (Byers-Heinlein et al.,

2017; Potter et al., 2018). Despite the fact that word segmentation in a language-mixing environment is challenging, young infants can still use within-speaker cues to separate two interleaved artificial languages and segment words from one language successfully in a short lab-based study (i.e., < 10 minutes). This may reflect a similar success in bilingual homes where parents individually produce two languages to their children. For example, children whose parents often mixed two languages at home successfully learn both languages (De Houwer, 2007; Place & Hoff, 2011). Bilingual children can also track languages in multilingual environment and speak the correct language (e.g., English) in accordance with the interlocutor's language background (e.g., monolingual English speaker) (Genesee, Nicoladis, & Paradis, 1995; Lanza, 1992). These behaviors clearly demonstrate that young bilingual children are highly sensitive to the context and speakers' language background when using their languages. Thus, while language mixing contexts may present challenges to bilingual infants in terms of information processing, such contexts may not necessarily hinder infants from segmenting words in a natural bilingual environment.

### Conclusion

In conclusion, the current paper demonstrates that phonetic cues can effectively support infants in a dual-input statistical segmentation task. Both monolingual and bilingual infants recruited phonetic cues to separately represent two language inputs and segment words. However, our finding that infants only successfully segmented words from one of the two languages (i.e., the last presented stream during familiarization) also highlights that learning words in a mixed-language bilingual environment can be challenging.

(Word count: 7901)

## References

- Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science*, 16(6), 980-990. doi: 10.1111/desc.12071
- Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, 21(2), e12548. doi: 10.1111/desc.12548
- Antovich, D. M., & Graf Estes, K. (2020). One language or two? Navigating cross-language conflict in statistical word segmentation. *Developmental Science*, e12960. doi: 10.1111/desc.12960
- 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. doi:10.1111/1467-9280.00063a
- Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117-134). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: The one-parent-one-language approach*. Clevedon, UK: Multilingual Matters.
- Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, 2, 324. doi: 10.3389/fpsyg.2011.00324
- Benitez, V.L., Bulgarelli, F., Byers-Heinlein, K., Saffran, J.R., & Weiss, D. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. *Developmental Science*, 23(2), e12896. doi: 10.1111/desc.12896
- Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, 39(4), 514-526. doi: 10.1016/j.wocn.2011.02.001

- Bosch, L., & Sebastián-Gallés, N. (1997). Native language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65, 33-69. doi: 10.1016/S0010-0277(97)00040-1
- 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, 106. doi: 10.3389/fpsyg.2013.00106.
- Brent, M. R., & Siskind, J. Mark. (2001). The Role of Exposure to Isolated Words in Early Vocabulary Development. *Cognition*, 81(2), B33-B44. doi: 10.1016/S0010-0277(01)00122-6
- Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1621-1631. doi: 10.1037/xlm0000263
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16(1), 32-48. doi: 10.1017/S1366728912000120
- Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences*, 114(34), 9032-9037. doi: 10.1073/pnas.1703220114
- U.S. Census Bureau (2015, November 03). *Census Bureau Reports at Least 350 Languages Spoken in U.S. Homes*. Retrieved from <https://www.census.gov/newsroom/press-releases/2015/cb15-185.html>
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(03), 411–424. doi:10.1017/S0142716407070221
- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A Bilingual-Monolingual Comparison of Young Children's Vocabulary Size: Evidence from Comprehension and Production. *Applied Psycholinguistics*, 35(6), 1189–1211. doi: 10.1017/S0142716412000744
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306. doi: 10.1126/science.171.3968.303

- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, 22, 421–435. doi: <https://doi.org/10.1111/infa.12182>
- Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, 33(6), 1087-1116. doi: 10.1111/j.1551-6709.2009.01041.x
- 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. doi: 10.1111/j.1467-9280.2007.01885.x
- Graf Estes, K., Gluck, S. C. W., & Bastos, C. (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
- Graf Estes, K., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, 51(11), 1517. doi: 10.1037/a0039725
- Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, 22(3), 611-631. doi: 10.1017/S0305000900009971
- 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. doi: 10.1016/j.cogpsych.2011.06.002
- Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, 10(5), 164–168. doi: 10.1111/1467-8721.00140
- Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in Infancy Research* (pp. 69–95). Stamford, CT: Ablex.

- Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, 325(5940), 611-612. doi: 10.1126/science.1173947
- 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-F21. doi: 10.1111/j.1467-7687.2006.00468.x
- Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, 19(3), 633-658. doi: 10.1017/S0305000900011600
- ManyBabies Consortium (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1), 24-52. doi: 10.1177/2515245919900809
- Mitchel, A. & Weiss, D.J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, 25(4), 456-482. doi: 10.1080/01690960903209888
- Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual & Multicultural Development*, 13(1-2), 19-39. doi: 10.1080/01434632.1992.9994481
- Oakes, L.M., Sperka, D. J., & Cantrell, L. (2015). Habit 2. *Unpublished software. Center for Mind and Brain, University of California, Davis.*
- Orena, A. J., & Polka, L. (2019). Monolingual and bilingual infants' word segmentation abilities in an inter-mixed dual-language task. *Infancy*, 24(5), 718-737. doi: <https://doi.org/10.1111/infa.12296>
- Pallier, C., Sebastiangular, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, 32(3), 373-389. doi: <https://doi.org/10.1006/jmla.1993.1020>
- Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, 152, 9-19. doi: 10.1016/j.cognition.2016.03.001
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics*, 18(7/8), 581-618. doi: 10.1515/ling.1980.18.7.8581

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence two-year-olds' bilingual proficiency. *Child Development*, 82(6), 1834–1849. doi:10.1111/j.1467-8624.2011.01660.x

Polka, L., Orena, A. J., Sundara, M., & Worrall, J. (2017). Segmenting words from fluent speech during infancy—challenges and opportunities in a bilingual context. *Developmental science*, 20(1), e12419. doi: <https://doi.org/10.1111/desc.12419>

Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., & Lew-Williams, C. (2018). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, e12794. doi: 10.1111/desc.12794

Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. Paris: Champion.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110–114. doi: 10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928. doi: 10.1126/science.274.5294.1926

Schmale, R., Cristia, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, 15(6), 650-662. doi: 10.1111/j.1532-7078.2010.00032.x

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, 12(4), 583-601. doi: 10.1111/j.1467-7687.2009.00809.x

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, 124(2), 128-142. doi: 10.1016/j.cognition.2012.05.008

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. doi: 10.3389/fpsyg.2012.00590

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53-71. doi: 10.1207/s15327078in0701\_5

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5, 953. doi: 10.3389/fpsyg.2014.00953

Weiss, D.J., Gerfen, C. & Mitchel, A.D. (2009) Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development* 5, 30-49. doi: 10.1080/15475440802340101

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: the role of cue strength and general cognitive processes. *Language and Cognitive Processes*, 25(3), 402–422. doi: 10.1080/01690960903212254

Weiss, D. J., Poepsel, T., & Gerfen, C. (2015). Tracking multiple inputs: The challenge of bilingual statistical learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*, pp. 167-190. Amsterdam, the Netherlands: John Benjamins Publishing Company.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49-63. doi: 10.1016/S0163-6383(84)80022-3

Table 1. *Stimuli in the Experiment 1, 2 and 3*

Stimuli for Language 1	
Experiment 1 and 2	Experiment 3
Words	
bə-ti-gu	bə-tu-gu
sɪ-tʃə-vi	sɪ-tʃə-vi
vʊ-bo-sa	vʊ-bɔɪ-sa
to-gu-tʃa	ta-gɔɪ-tʃa
Part words	
sa-to-gʊ	sa-ta-gɔɪ
gu-vʊ-bo	gu-vʊ-bɔɪ
tʃa-si-tʃə	tʃa-sɪ-tʃə
vi-bə-tɪ	vi-bə-tu
Stimuli for Language 2	
Words	
gu-pa-tə	gu-pa-tə
dʒi-ga-pʊ	ji-ga-pɔɪ
sa-dʒu-bo	sa-ju-bae
ta-bi-sɪ	ta-bi-sae
Part words	

bo-dʒi-ga	bae-ji-ga
si-ga-pʊ	sae-ga-pɔɪ
pʊ-ta-bi	pɔɪ-ta-bi
tə-sa-dʒu	tə-sa-ju

Table 2. *Acoustic Measurement of the languages with English and French phonetic cues in Experiment 1 and 2.*

	Language produced with English phonetic cues	Language produced with French phonetic cues	Paired t test analysis between languages produced with English and French phonetic cues		
	Mean	SD	Mean	SD	
<b>Mean duration (ms)</b>	247.65	89.56	207.63	55.30	$t = 4.14, p < 0.001$
<b>Mean fundamental frequency (Hz)</b>	249.96	32.47	250.93	39.17	$t = -0.14, p = 0.889$
<b>Min fundamental frequency (Hz)</b>	212.60	39.33	221.29	39.75	$t = -1.26, p = 0.212$
<b>Max fundamental frequency (Hz)</b>	289.98	47.11	282.91	51.33	$t = 0.73, p = 0.471$
<b>Mean intensity (dB)</b>	80.92	1.98	79.93	1.96	$t = 3.46, p = 0.001$

*Note.* The mean and standard deviations of the acoustic measurements were computed across syllables.

Table 3. *English-French bilinguals' rating of the languages with English and French phonetic cues in Experiment 1 and 2.*

	French phonetic		English phonetic	Overall
	Naturalness	cues rating	cues rating	sound quality
<b>Languages with English</b>				
phonetic cues	3.15 (1.14)	4.38 (0.77)	1.75 (0.83)	3.38 (0.77)
<b>Languages with French</b>				
phonetic cues	3.17 (0.93)	1.90 (0.88)	4.04 (0.91)	3.59 (0.56)

*Note.* Twelve local English-French bilinguals rated the concatenated artificial languages in Experiment 1 and 2. The stimuli were rated on a scale of 1 to 5 in four aspects: (i) naturalness (1 = unnatural-sounding to 5 = natural-sounding); (ii) French phonetic cues rating (1 = very French-sounding to 5 = very non-French-sounding); (iii) English phonetic cues rating (1 = very English-sounding to 5 = very non-English-sounding); and (iv) overall sound quality of the language (1 = bad to 5 = excellent). Table 3 describes the mean and standard deviations of ratings for each language with English and French phonetic cues across all raters. We found that local English-French bilinguals rated the stimuli with French phonetic cues as clear and French-sounding, and they also rated stimuli with English phonetic cues as clear and English-sounding.

Table 4. *Acoustic Measurement of the languages with English and Finnish phonetic cues in Experiment 3.*

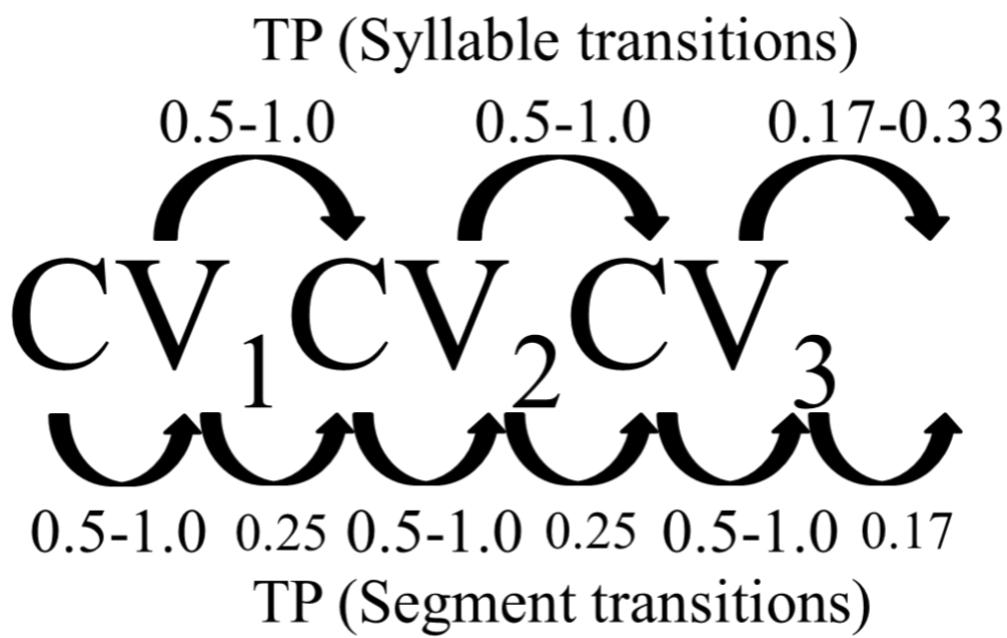
	Language produced with English phonetic cues	Language produced with Finnish phonetic cues	Paired t test analysis between languages produced with English and French phonetic cues		
	Mean	SD	Mean	SD	
<b>Mean duration (ms)</b>	269.78	33.71	269.83	30.77	$t = -0.03, p = 0.974$
<b>Mean fundamental frequency (Hz)</b>	204.21	9.88	211.77	14.53	$t = -9.77, p < 0.001$
<b>Min fundamental frequency (Hz)</b>	187.99	10.11	191.92	13.39	$t = -5.92, p < 0.001$
<b>Max fundamental frequency (Hz)</b>	225.38	16.86	233.96	15.26	$t = -9.49, p < 0.001$
<b>Mean intensity (dB)</b>	69.39	1.49	60.14	1.70	$t = 2.90, p = 0.004$

Table 5. *Monolingual English speakers' rating of the languages with English and Finnish phonetic cues in Experiment 3.*

	Naturalness	Foreign phonetic cues rating	English phonetic cues rating	Overall sound quality
<b>Languages with English</b>				
<b>phonetic cues</b>	3.17 (0.94)	1.67 (0.49)	4.25 (0.97)	3.42 (0.67)
<b>Languages with Finnish</b>				
<b>phonetic cues</b>	3.08 (1.08)	4.17 (0.72)	1.92 (0.67)	3.50 (0.80)

*Note.* Twelve local monolingual English American speakers rated the concatenated artificial languages in Experiment 3. The stimuli were rated on a scale of 1 to 5 in four aspects: (i) naturalness (1 = unnatural-sounding to 5 = natural-sounding); (ii) Foreign phonetic cues rating – participants were asked to rate how foreign the speech sounds like (1 = not foreign-sounding to 5 = very foreign-sounding); (iii) English phonetic cues rating – participants were asked to rate how much the speech sounds like English (1 = very non-English-sounding to 5 = very English-sounding); (iv) overall sound quality of the language (1 = bad to 5 = excellent). Table 5 describes the mean and standard deviations of ratings for each language with English and Finnish phonetic cues across all raters. We found that local monolingual English speakers rated the stimuli with Finnish phonetic cues as clear and sound like foreign to them, and they also rated stimuli with English phonetic cues as clear and English-sounding.

Accepted Article



*Figure 1.* Transitional probabilities (transitional probability) of the underlying structures (i.e., inconsistent statistical cues) of the two languages when combined.

## BILINGUAL INFANT STATISTICAL LEARNING

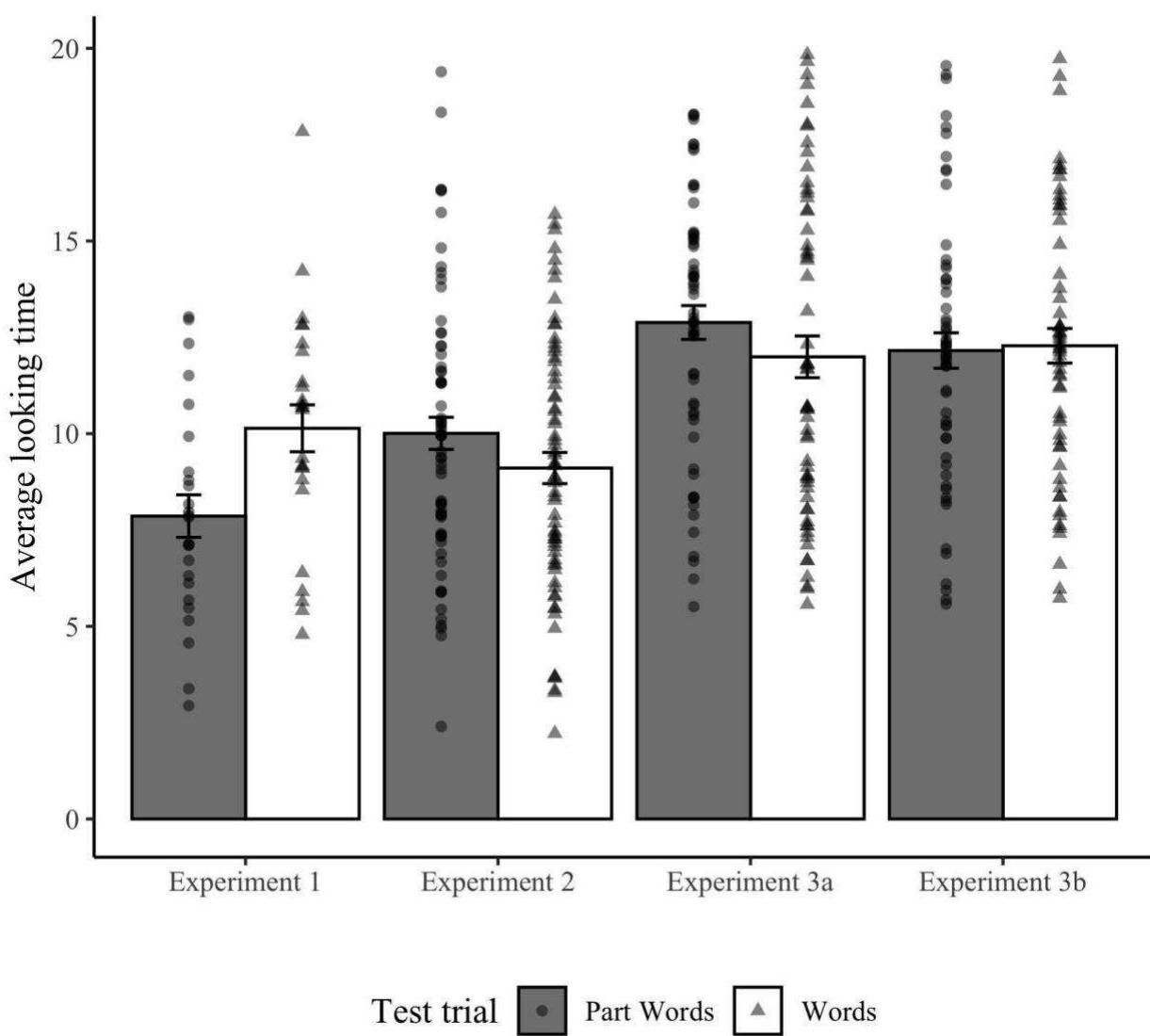


Figure 2. Infants' mean looking time to words and part-words across three experiments. Error bars represent the standard error of the mean looking time. Circles plot values for individual participants' looking time to part words and triangle plot values for individual participant's looking time to words.

## BILINGUAL INFANT STATISTICAL LEARNING

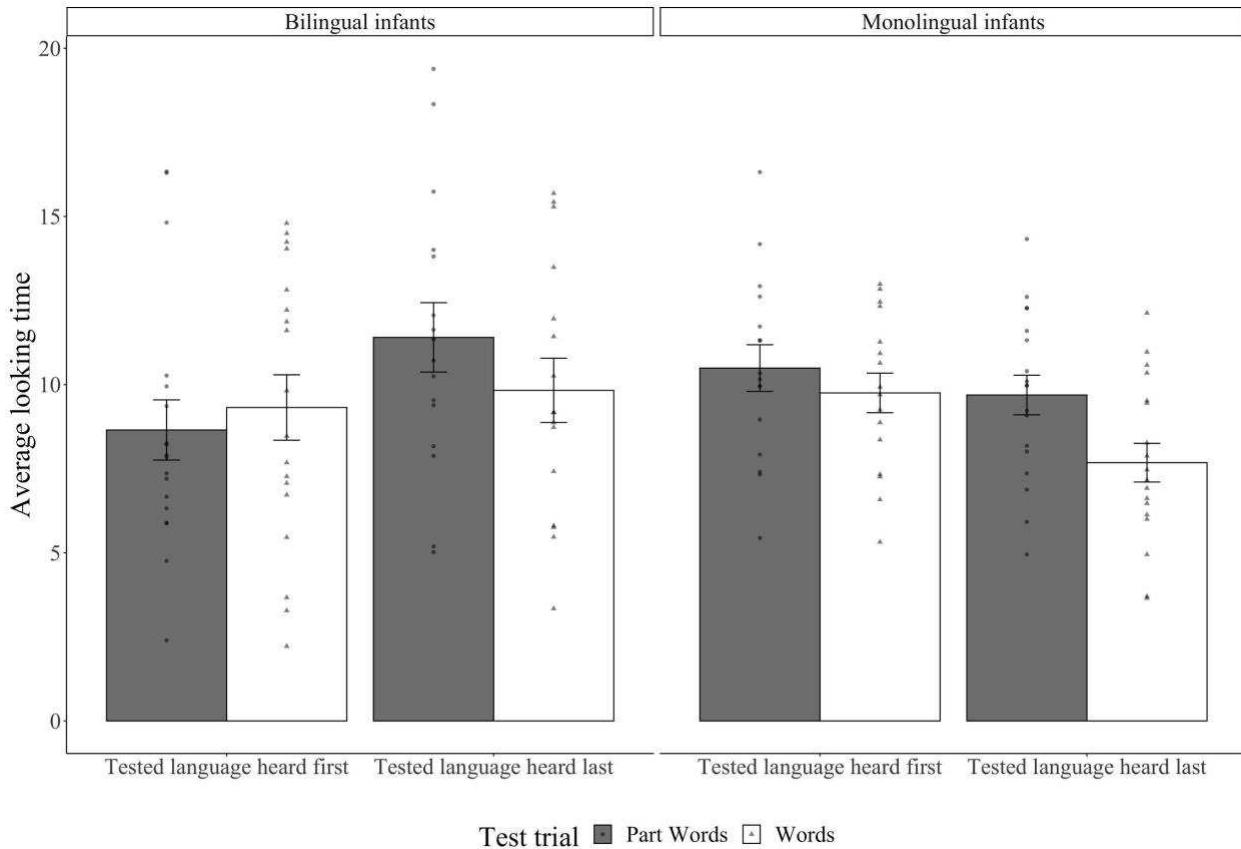
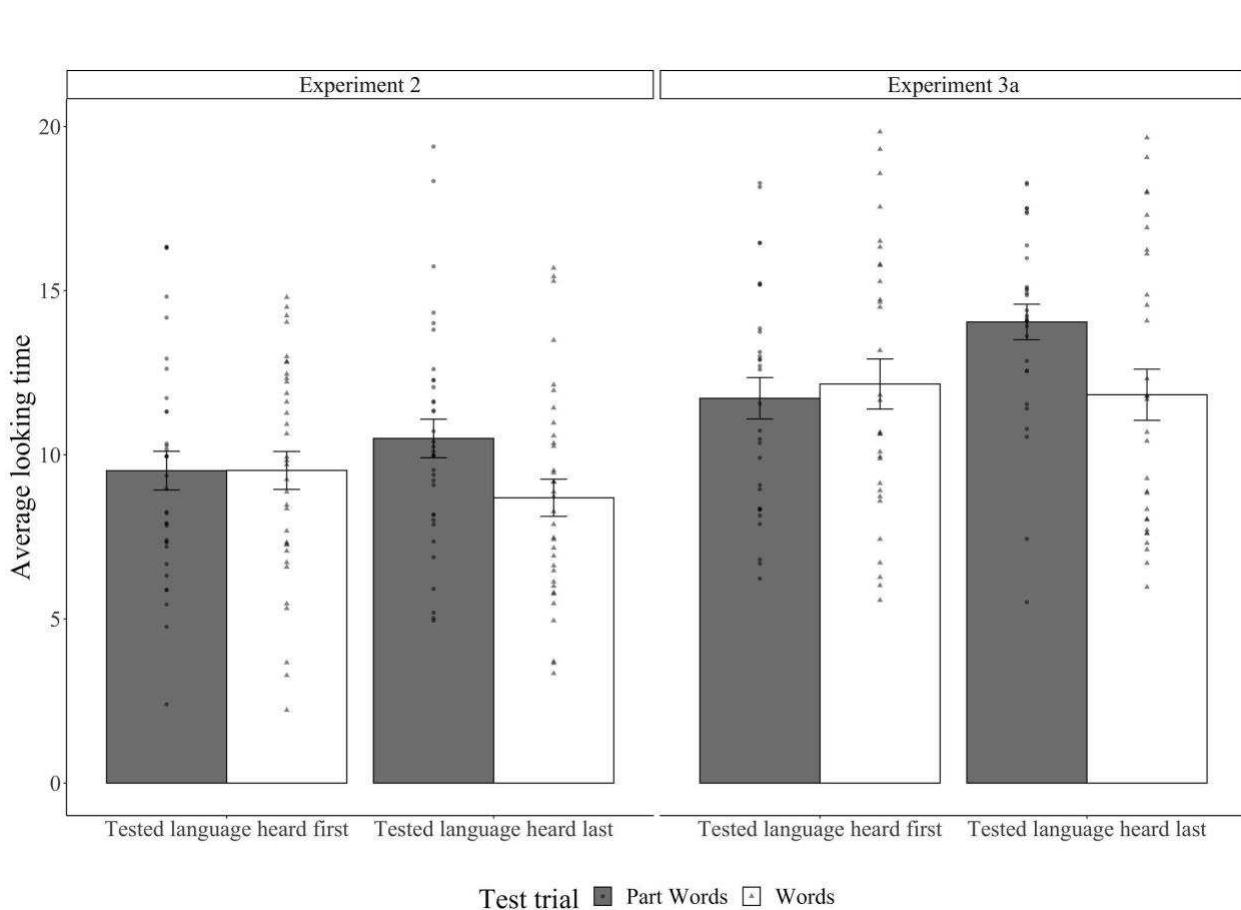


Figure 3. Bilingual and monolingual infants' mean looking time to words and part-words in Experiment 2. We plotted infants' performance as a function of which language is being tested as we found significant tested language order effects. Error bars indicate the standard error of mean looking time.

## BILINGUAL INFANT STATISTICAL LEARNING



*Figure 4.* Infants' mean looking time to words and part-words as a function of which language is being tested across Experiment 2 and 3a. Error bars indicate the standard error of mean looking time.