

## Original Articles

## Infants' sensitivity to vowel harmony and its role in segmenting speech

Toben H. Mintz<sup>a,b,\*</sup>, Rachel L. Walker<sup>b</sup>, Ashlee Welday<sup>a,1</sup>, Celeste Kidd<sup>b,2</sup><sup>a</sup> Department of Psychology, University of Southern California, United States<sup>b</sup> Department of Linguistics, University of Southern California, United States

## ARTICLE INFO

## Keywords:

Language acquisition

Infants

Vowel harmony

Word segmentation

Speech processing

Learning biases

## ABSTRACT

A critical part of infants' ability to acquire any language involves segmenting continuous speech input into discrete word forms. Certain properties of words could provide infants with reliable cues to word boundaries. Here we investigate the potential utility of vowel harmony (VH), a phonological property whereby vowels within a word systematically exhibit similarity ("harmony") for some aspect of the way they are pronounced. We present evidence that infants with no experience of VH in their native language nevertheless actively use these patterns to generate hypotheses about where words begin and end in the speech stream. In two sets of experiments, we exposed infants learning English, a language without VH, to a continuous speech stream in which the only systematic patterns available to be used as cues to word boundaries came from syllable sequences that showed VH or those that showed vowel disharmony (dissimilarity). After hearing less than one minute of the streams, infants showed evidence of sensitivity to VH cues. These results suggest that infants have an experience-independent sensitivity to VH, and are predisposed to segment speech according to harmony patterns. We also found that when the VH patterns were more subtle (Experiment 2), infants required more exposure to the speech stream before they segmented based on VH, consistent with previous work on infants' preferences relating to processing load. Our findings evidence a previously unknown mechanism by which infants could discover the words of their language, and they shed light on the perceptual mechanisms that might be responsible for the emergence of vowel harmony as an organizing principle for the sound structure of words in many languages.

## 1. Introduction

A fundamental problem that infants face from the earliest stages of language acquisition is identifying the sequences of sounds that are the words in their language. Pauses are poor predictors of lexical boundaries in continuous speech. Thus, infants must rely on other cues to identify beginnings and ends of words. Cues that have been identified include word stress (Curtin, Mintz, & Christiansen, 2005; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), transitional probabilities between syllables (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003), and phonotactic cues involving representations of sound sequences that are more or less likely to occur at word boundaries (Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999). Both within a language and across languages, none of these cues alone is guaranteed to identify word boundaries. Hence, infants must make use of a number of probabilistic cues to segment words from fluent speech.

Here we present research concerning perceptual biases that could

aid infants in segmenting continuous speech into words. Specifically, we present evidence that infants are sensitive to *vowel harmony*, a property of many languages whereby vowels within a word systematically exhibit similarity (or "harmony") for some aspect of the way they are pronounced. For example, in a language with vowel harmony for the property of lip rounding, within a word, all vowels are either produced with rounded lips (e.g., [u, o]) or with unrounded lips (e.g., [i, e]), but these two vowel types are never combined. Our research shows that 7-month-old infants learning English are sensitive to vowel harmony patterns, even though English does not exhibit vowel harmony. Our experiments demonstrate that infants segment words from continuous speech at locations where harmony patterns are disrupted. Since the infants in our study were never exposed to a natural language with vowel harmony, they never experienced the correlation of disruptions in harmony patterns with word boundaries, yet they segmented speech at junctures of disharmony. Our findings, thus, contribute to two important issues in language development. First, our finding of a previously unknown capability of human infants to rapidly

\* Corresponding author at: Department of Psychology, University Park Campus, SGM 501, University of Southern California, Los Angeles, CA 90089-1061, United States.  
E-mail address: [tmintz@usc.edu](mailto:tmintz@usc.edu) (T.H. Mintz).

<sup>1</sup> Present address: Social Thinking Stevens Creek, San Jose, CA, United States.

<sup>2</sup> Present address: Department of Brain and Cognitive Sciences, University of Rochester, United States.

detect vowel harmony patterns contributes to our understanding of the perceptual and representational capacities that could be a basis for human language learning. Second, our finding that infants use harmony patterns to extract words from continuous speech contributes to our understanding of the mechanisms by which infants can discover the fundamental building blocks of language: the words.

Before presenting our experiments and findings, we first provide a conceptual outline describing how a sensitivity to vowel harmony could be useful for word segmentation in infants, and how such a sensitivity might change with experience with a specific language or languages. We also provide a more in depth discussion of the typological and acoustic aspects of vowel harmony.

### 1.1. Vowel harmony and its potential role in infant word segmentation

Vowel harmony phenomena are widely attested in many of the world's major language families, including the Uralic languages (e.g., Finnish, Hungarian, etc.), the Altaic languages (e.g., Mongolian, Turkish), the Niger-Congo languages (e.g., Kikongo, Swahili, Yoruba), and the Nilo-Saharan languages (e.g., Maasai, Turkana). Yet, languages vary in the specific kinds of harmony constraints they impose. For instance, Turkish exhibits vowel harmony for the front/back dimension, in which 'front' and 'back' refer to the generalized relative anteriority of the highest point of the tongue during production of different vowels. In Turkish, the vowels in a word are either all front vowels ([i, y, e, ø]) or all back vowels ([u, ʊ, o, a]) (although loanwords may be disharmonic). Turkish also displays vowel harmony for the property of lip rounding, preventing certain intra-word combinations of round and unround vowels. In contrast, Swahili shows restrictions that involve harmony for the dimension of peak tongue height in vowels. In Swahili, forms of verbal suffixes containing a mid vowel [e] are selected following a syllable with a mid vowel ([e] or [o]); elsewhere forms of these suffixes containing the high vowel [i] are selected.

These generalized articulatory dimensions of vowels can be equated to acoustic properties of their sound waves. The sound wave for each vowel is characterized by resonant frequencies of its vocal tract configuration, known as formants. Shifts in the height dimension are reflected in the frequency of the first formant, in the backness dimension in the frequency of the second formant, and in the rounding dimension chiefly in the frequency of the second and third formants. Thus, adjacent syllables that exhibit some form of vowel harmony will have similar spectral properties due to their acoustic-phonetic realization. In principle, such similarities could constitute a perceptual basis for infants to treat adjacent syllables as a unit when segmenting speech.

Thus, vowel harmony patterns result in both phonological and acoustic similarities between syllables within a word,<sup>3</sup> which provides a powerful cue that a learner could exploit to segment speech into words: Learners could posit a boundary when they notice disharmony in the vowels of adjacent syllables, i.e. when the vowels differ in a particular aspect of their pronunciation. Indeed, research on adult speakers of languages with vowel harmony has shown that they actively use disharmony as a cue to word boundaries (Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998).

### 1.2. Vowel harmony and the role of language experience

Studies have shown that adult native speakers of languages that lack vowel harmony rules do not use disharmony as a cue to word boundaries (Vroomen et al., 1998), so adults' linguistic background clearly influences how they treat harmony patterns. Moreover, the fact that languages differ on the particular set of vowel harmony restrictions they implement, and indeed whether they implement vowel harmony

restrictions at all, means that some aspects of harmony must be learned through experience, and there is evidence that this experience has an effect early on in development. For example, children acquiring a harmony language show productive mastery of its harmony system by the time they start to combine words, generally before two years of age (Ketrez & Aksu Koç, 2003). Infants acquiring a harmony language show a sensitivity to the patterns as early as 6 months (Altan, Kaya, & Hohenberger, 2016; Hohenberger, Kaya, & Altan, 2017; van Kampen, Parmaksiz, van de Vijver, & Höhle, 2007).

Yet learning could occur in several computationally distinct ways. One possibility is that infants need to learn a sufficiently large number of word forms and then detect the consistent patterns that regularly occur. Another possibility is that infants are predisposed to attend to harmony patterns. On this alternative, the development of vowel harmony might be parallel to the development speech sound categories. From birth, infants can discriminate virtually all the speech sounds that serve as distinctive in any of the world's languages; experience narrows down and refines those representations to the ones relevant to the infant's linguistic environment (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl, 1979; Streeter, 1976; Werker & Lalonde, 1988; Werker & Tees, 1992). Infants may be inherently sensitive to harmony patterns and learn those patterns when they are present in their natural language input, but they may lose that sensitivity over development. If infants are predisposed to detect harmony patterns, they could also use vowel harmony as a source of information for segmenting words from continuous speech.

In our study, we ask whether English learning 7-month-olds who have had no experience with a vowel harmony language nevertheless use harmony patterns as a segmentation cue when the patterns are present. If they do, it demonstrates that infants learning a harmony language could detect word boundaries from vowel harmony patterns without building a prior lexicon (even a very small one) in which words conform to vowel harmony constraints. More broadly, such a finding would indicate that detecting and using harmony patterns in processing speech relies on mechanisms that do not require extensive exposure to a harmony language, but rather are available to all young infants.

To test this, we exposed infants to a continuous sequence of syllables that consisted of stretches with harmonizing vowels ("words") punctuated by points of disharmony (at "word" boundaries). We hypothesized that if infants treat vowel harmony as a linguistically significant property of the speech input, they should parse the harmonizing sequences as cohesive units upon hearing the continuous syllable stream, treating the junctures of disharmony as boundary points between units. That is, we hypothesize that infants would treat the harmonizing sequences as proto-word forms. We tested this by probing whether, after hearing the continuous speech stream, infants showed a systematic difference in their attention to syllable sequences that corresponded to the words compared to sequences of adjacent syllables that contained a word boundary, which we call part-words.

Since infants' linguistic experience has given them no evidence of an association between vowel harmony and word forms, if they parse the continuous speech stream into sequences that adhere to the harmony constraints, it would indicate that infants have an experience-independent sensitivity to vowel harmony patterns. It would further demonstrate that infants can use vowel harmony cues, when available, to detect word boundaries in fluent speech. We note, however, that this experiment—as many others that have investigated other segmentation cues (e.g., Aslin et al., 1998; Curtin et al., 2005; Mattys & Jusczyk, 2001; Saffran et al., 1996)—addresses the question of segmentation of sequences from fluent speech. When we refer to these segmented sequences as 'words,' we mean that they are word-sized units segmented from the continuous speech stream, not that infants have necessarily created a kind of lexical entry for them (Graf Estes, Evans, Alibali, & Saffran, 2007).

As the first step in this investigation, we describe a set of experiments that assess 7-month-old infants' response to the simplest form of

<sup>3</sup> Some cases are attested where harmony operates in domains that are larger than the word.

vowel harmony to detect: *vowel identity*. In this form of harmony, syllables within a word are constrained to be identical. Thus, word boundaries are signaled by a switch in vowel from one syllable to the next. A natural language that exhibits this kind of harmony pattern is the dialect of Servigliano, spoken in Italy (Camilli, 1929; Walker, 2011); additional languages are noted in van der Hulst and van de Weijer (1995). Once establishing that infants attend to this form of vowel harmony and use it to segment speech, we present experiments that target a subtler form of vowel harmony.

## 2. Experiment 1a

The goal of this experiment was to test infants' sensitivity to the most basic form of vowel harmony, which is vowel identity. The 'words' were three syllables long and contained identical vowels—for example, *tokobo*. We dubbed these *identity* patterns and call this the *identity language*.

### 2.1. Method

#### 2.1.1. Stimuli

The familiarization sequence for the experimental condition consisted of 45 repetitions of the fixed sequences of syllables *pidigitokobogetepedubuku*, consisting of the words *pidigi*, *tokobo*, *getepe*, and *dubuku*. Transitional probabilities between syllables within and across word boundaries were thus equal to 1, and the frequencies of each word and part-word were equal (45). Hence, there was no statistical information in the distribution and sequencing of syllables that provided cues to word boundaries. We also controlled consonant characteristics to ensure that they did not provide word boundary information. Specifically, half the words started with voiced consonants, and half with voiceless consonants. Similarly, the consonants in word-final syllables were equally likely to be voiced as to be voiceless, providing no reliable word boundary information. Consonants in word-initial (and word-final) syllables also varied with respect to place of articulation, and thus provided no reliable boundary information. Words were defined simply as syllable sequences containing identical vowels.

So that the beginning or end of the entire stream could not function as a word-segmentation cue, we appended several syllables that were not involved in the experiment to the beginning and end of the complete familiarization streams.

The test words for both conditions were *tokobo* and *getepe*, and the test part-words were *tepedu* and *gitoko*.

#### 2.1.2. Synthesis of stimuli

The familiarization streams were synthesized with a Macintosh G4 computer running OS 10.4 using the native text-to-speech synthesizer set to the Victoria voice. We programmed the synthesizer not to modulate pitch, to produce a flat intonational contour across the entire stream; pitch was still modulated somewhat for a given vowel, but there was no suprasegmental pitch modulation. The streams were synthesized at a rate of 216 words per minute. The four-word sequence was repeated 45 times, for a total length of approximately 84 s. The digital output from the synthesizer was captured to a file using Audio Hijack, sampling at 44.1 kHz.

Test stimuli were synthesized individually in the same voice and also with no suprasegmental pitch modulation.

#### 2.1.3. Subjects

Sixteen infants from monolingual English-speaking homes participated in the experiment. Data from four subjects were not included in the analysis due to experimenter error (1), the infant's failure to orient towards the lights during the testing phase (1), and infant fussiness/crying (2), resulting in data for 12 subjects (4 female; mean age 7 months and 9 days; range, 6 months and 17 days to 7 months and 25 days).

#### 2.1.4. Apparatus and procedure

Each infant was tested separately while seated on a caretaker's lap in the center of a sound-attenuated room. An experimenter observed the infant's looking behavior through a closed-circuit television monitor in an adjacent room. The experimenter entered the infant's head turn responses into a computer that controlled all aspects of the experiment.

At the start of the familiarization phase, a red light positioned on the wall directly in front of the infant at eye level flashed repeatedly. When the infant oriented towards the light, one of the familiarization streams was played continuously through two loudspeakers mounted on the walls to the left and right of the infant. When the familiarization stream started, the center light was extinguished and a yellow light mounted above one of the loudspeakers (chosen at random) started flashing. It continued to flash until the infant first looked towards it, then looked away for 2 consecutive seconds. The side light was then extinguished and the center light flashed again until the infant oriented to the neutral center position. This process was repeated for the duration of the familiarization phase, randomizing the side on which the light flashed. This kept the infants engaged and established the contingency between their looking behavior and the activation of the lights. As in Saffran et al. (1996), the familiarization stream played without interruption during the entire familiarization phase so that no silences were introduced that could signal a word boundary.

Because the auditory presentation in the familiarization phase was not contingent on infants' looking behavior, a brief *contingency training phase* immediately followed the familiarization phase, in which auditory information as well as the light behavior was contingent on infants' looking behavior. It was important to introduce infants to this contingency because it is leveraged in the testing phase. In the contingency training phase, no sound was presented until the infant looked towards the flashing yellow side light, at the beginning of a trial. When the infant oriented toward the flashing side light, a 440 Hz pure tone lasting 1 s was repeated until the infant looked away for 2 contiguous seconds. This phase consisted of two such trials. Its purpose was to prepare the infant for the test phase that immediately followed, in which auditory stimulus presentation was similarly contingent on orienting to the flashing light.

The test phase was similar to the previous phase except that one of the four (two words and two part-words) test items was played during each trial. Each test item occurred in two trials across two pseudorandom blocks, for a total of 8 trials. On each trial, the stimulus was repeated with a 500 ms pause between repetitions until either the infant looked away continuously for 2 s or after 16 s had elapsed. The randomized selection of the side light was constrained such that a given side was selected in no more than two consecutive trials. Infants' orientation time to each trial was recorded by the control computer. For inclusion in the final data analysis, we stipulated a listening time criterion of at least 2 consecutive seconds for at least two of each type of test trial (word and part-word).

## 2.2. Results and discussion

Listening times to each test trial type were averaged within infants, yielding a mean orientation time to words and a mean orientation time to part-words for each subject. We performed  $2 \times 2$  ANOVAs with block (first and last) and word type (word and part-word) as within-subject variables. Here, and in the remaining experiments, our interest in examining block effects was to guard against the effect of interest (word type) washing out because of infants' overall habituation to test items during the test phase; that is, we wanted to ensure that there were no interactions between block and word type.

There were main effects of block ( $F(1,11) = 7.53, p = .019$ ) and word type ( $F(1,11) = 13.38, p = .004$ ), and no interaction between the two variables ( $F(1,11) = 0.41$ ). Infants listened longer to the (dis-harmonizing) part-words ( $M = 9.87$  s,  $SE = 0.51$  s) compared to words ( $M = 8.15$  s,  $SE = 0.46$  s), indicating that infants distinguished the two

types of test items. Infants' looking times also tended to uniformly decrease towards the end of the experiment, as is typical in this procedure.

Infants' preference for part-words mirrors the preference pattern reported in Saffran et al. (1996), for infants segmenting speech using statistical information. These results are consistent with the hypothesis that infants segmented the familiarization stream at junctures of disharmony, extracting word forms that contained syllable sequences with identical vowels. This interpretation supports the conclusion that, at least for this simple form of vowel harmony, infants are sensitive to harmony and use it as a cue to the location of word boundaries in continuous speech.

However, there are two alternative explanations: (1) infants may have had an inherent preference for the test sequences with disharmony over completely harmonic test sequences, independent of prior exposure and segmentation, or (2) infants may have segmented sequences that contained the disharmonic transitions (i.e., part-words) and showed a familiarity preference for those sequences. If the first alternative is true, we cannot draw conclusions about infants' use of vowel harmony in segmenting continuous speech, but it would nevertheless indicate that infants are sensitive to vowel harmony—in this case, vowel identity—and disharmony when processing word forms. The second alternative seems unlikely, on linguistic or perceptual grounds, as it requires infants to systematically posit word boundaries between syllables with identical vowels, but it is logically possible.<sup>4</sup> We tested these alternatives in Experiments 1b and 1c.

In Experiment 1b we assessed whether infants show an inherent preference for part-word test sequences over the word sequences from Experiment 1a, in the absence of the familiarization stream. In Experiment 1c we assessed whether infants were segmenting words (resulting in a novelty preference in Experiment 1a), or part-words. We did this by reducing the exposure to the familiarization material by approximately half and showing that infants shift from a novelty to a familiarity preference with this reduced exposure. We argue that this is consistent with segmenting words, but not part-words.

### 3. Experiment 1b

In this experiment, we tested the hypothesis that infants have an inherent preference for the part-word test items compared to the word test items in Experiment 1a, in the absence of prior exposure to a familiarization stream that contained those sequences. If infants show such a preference, then one cannot interpret the results of Experiment 1a as evidence of harmony-based word segmentation. On the other hand, if infants show no preference for part-words over words, then the part-words preference in Experiment 1a must have been due to infants' exposure to the familiarization stream, and in particular to the sequences they segmented from it.

#### 3.1. Methods

##### 3.1.1. Materials

The materials were the test items from Experiment 1a.

##### 3.1.2. Subjects

Sixteen infants from monolingual English-speaking homes participated in the experiment. Data from four subjects were not analyzed because they did not meet the predetermined criterion of listening for at least 2 consecutive seconds to at least two of each type of test trial (word and part-word), yielding data from 12 infants (7 females; mean age 7 months and 4 days; range 6 months and 23 days to 7 months and

19 days).

##### 3.1.3. Procedure

The procedure was identical to that in Experiment 1a, but without the familiarization phase.

#### 3.2. Results and discussion

Listening times to each test trial type were averaged within infants, yielding a mean orientation time to words and a mean orientation time to part-words for each subject. A paired *t*-test revealed that listening times to words was not significantly different than listening time to part-words ( $M = 7.9$  s,  $SE = 0.73$  s and  $M = 8.4$  s,  $SE = 0.77$ , respectively;  $t(11) = 1.40$ ,  $p = .190$ ). Thus, there was no evidence that infants showed a preference for one type of test item over the other.

This result rules out the possibility that infants' part-word preference in Experiment 1a was due to an inherent bias towards items in which a greater variety of vowels occurred over items that contained only one vowel across the three syllables. Nevertheless, it is still possible, though, we think, unlikely, that infants in Experiment 1a segmented what we are calling part-words, rather than words, from the familiarization stream, and showed a familiarity preference for the part-words. We examine this possibility further in Experiment 1c.

### 4. Experiment 1c

In the test phase in Experiment 1a, infants preferred to listen to part-words compared to words. Following Saffran et al. (1996), we interpreted this as a novelty preference for part-words, indicating that the harmonic words were more familiar to the infants because they segmented them from the syllable stream during the familiarization phase, positing boundaries between syllables with contrasting vowels. However, it is logically possible that they segmented part-words, placing word boundaries between syllables with identical vowels. In that case, infants' preference towards part-words would have been a familiarity preference.

Since the results from Experiment 1a cannot resolve this question empirically, in this experiment, we manipulated infants' exposure to the familiarization stream by reducing it to approximately half the number of exposures of the four-word sequence and examined whether infants then showed a preference for words instead of part-word, in contrast to their preference in Experiment 1a. The logic and predictions regarding this manipulation are as follows. Since the question is whether infants in Experiment 1a were showing a familiarity or novelty preference, reducing familiarization is expected to have a different pattern of results depending on the type of preference in the full exposure condition of Experiment 1a. If we assume that a novelty preference is the result of habituation to a stimulus, then reducing exposure to that stimulus will, at some point, result in a familiarity preference (i.e., a preference for the other test stimulus type). This is because a familiarity preference can be expected when the infant has processed—here, segmented—the stimulus but has not habituated to it (Hunter & Ames, 1988; Kidd, Piantadosi, & Aslin, 2012). Thus, in this experiment we predicted that infants would show a preference towards word test items, in contrast to their behavior in Experiment 1a. On the other hand, reducing exposure to a stimulus that infants respond to with a familiarity preference should not make them switch to a different test item type (which would be a switch to a novelty preference). Rather, infants should maintain preference to the same test type, or even show no preference, indicating that they have not sufficiently processed the structure of the familiarization stream. Thus, if the alternative interpretation of Experiment 1a were correct, and infants were segmenting part-words, then we would still expect to see a preference towards part-words here, or perhaps no preference at all. Crucially, a preference for words in this experiment is not compatible with the interpretation that infants were segmenting part-words, because there is no clear reason for them to show a novelty

<sup>4</sup> While consonants in a word may be constrained to show disharmony in the world's languages, vowel patterns of this kind are much more limited and might actually be understood as quantitative effects involving metrical rhythm (Suzuki, 1998). Further, where dissimilation patterns among vowels have been reported, they primarily involve low vowels (Blevins, 2009), which were not used in our study.



preference here, with reduced exposure, if they showed a familiarity preference with greater exposure (Experiment 1a).

#### 4.1. Methods

##### 4.1.1. Materials

The repeated four-word sequence was the same as in Experiment 1a, but here the sequence was repeated only 20 times during the familiarization phase (compared to 45 in Experiments 1 & 2a), making the entire familiarization sequence (including the ‘lead-in’ and ‘lead-out’ syllables) approximately 38 s long.

##### 4.1.2. Subjects

Fourteen infants from monolingual English-speaking homes participated in the experiment. Data for two subjects were not analyzed because they failed to maintain a head-turn of at least two seconds on each trial, yielding data for 12 infants (3 female; mean age 7 months, 12 days; range 7 months 1 days to 8 months 1 day).

##### 4.1.3. Procedure

The procedure was identical to Experiment 1a.

#### 4.2. Results

Listening times to each test trial type were averaged within infants, yielding a mean orientation time to words and a mean orientation time to part-words for each subject. We performed a  $2 \times 2$  ANOVA with block (first and last) and word type (word and part-word) as within-subject variables. There was a main effect of word type ( $F(1,11) = 8.67$ ,  $p = .013$ ), a marginal effect of block ( $F(1,11) = 3.83$ ,  $p = .076$ ), and no interaction between the two variables ( $F(1,11) = 0.20$ ). Infants listened longer to the words ( $M = 11.52$  s,  $SE = 1.10$  s) compared to part-words ( $M = 9.92$  s,  $SE = 0.98$  s).

#### 4.3. Discussion

In this experiment in which we shortened the familiarization time to approximately 50% of the familiarization in Experiment 1a, subjects showed a preference for words, in contrast to their preference for part-words in Experiment 1a. The most coherent interpretation of these results is that infants across both experiments segmented out harmonic words during the familiarization phase, and showed a familiarity preference (for the words) when the familiarization period was brief (Experiment 1c) and a novelty preference (for the part-words) when the familiarization period was approximately twice as long (Experiment 1a). An alternative interpretation in which infants segmented *part-words* during familiarization would have to explain why infants with less exposure to the segmented sequences would show a novelty preference, but with more exposure would show a familiarity preference. We do not find any support for this interpretation.

Taken together, Experiments 1a–c demonstrated that 7-month-old infants who have had no significant exposure to a language with vowel harmony nevertheless are sensitive to an extreme form of vowel harmony—i.e., vowel identity—and segment continuous speech at junctures of disharmony. Experiment 1b demonstrates that English-reared infants at this age do not have inherent preference for nonce words that have identical vowels over those that do not (or vice versa), which shows that infants’ systematic preferences in the other two experiments cannot be accounted for by any such inherent bias.

However, one must be cautious in generalizing infants’ sensitivity and use to this kind of harmony (i.e., identity) to vowel harmony in general, since the processing of identical elements (in this case, the vowels of the syllables) might be privileged (e.g., Gervain, Berent, & Werker, 2012). Therefore, having established infants’ sensitivity to vowel identity, we now present experiments that probe whether English-reared infants are sensitive to a subtler form of vowel

harmony that is more widespread across the world’s languages. We used the same testing procedure as in Experiment 1, but with a modified artificial language.

#### 5. Experiment 2a

The results from Experiment 1 indicate that infants with no exposure to a natural language with vowel harmony, nonetheless use vowel identity—a kind of vowel harmony—as a basis for segmenting continuous speech. In Experiment 2 we investigated whether 7-month-olds would detect and use more subtle realizations of vowel harmony. Similar to Experiments 1a and 1c, in this experiment we exposed infants to a continuous sequence of syllables that composed words in an artificial lexicon. However, unlike in those experiments, the syllables within a word were not identical. Rather, they harmonized on the dimension of *backness*, while they varied on the dimension of *height*. This form of vowel harmony is more common than identity harmony. We then tested whether infants use disharmonic vowels in adjacent syllables as a basis for segmenting continuous speech into word-like units, thereby grouping syllables with harmonizing vowels into the same word form.

To illustrate in schematic terms, consider a speech stream consisting of a sequence of syllables [...  $S_A S_A S_B S_B S_A S_A S_B S_B S_A$  ...]. Each  $S$  symbolizes a variable syllable and subscripts indicate that the vowel in that syllable has either property ‘A’ or property ‘B’ for some dimension (e.g., A = ‘front’ and B = ‘back’). If infants had no inherent bias to segment on the basis of harmony or disharmony, we would expect them as a group to show no systematicity in segmenting harmonic sequences of the form [ $S_A S_A$ ] and [ $S_B S_B$ ] compared to the disharmonic sequences, [ $S_A S_B$ ] and [ $S_B S_A$ ]. If, however, infants showed a bias towards segmenting harmonic sequences, that would indicate they favor using vowel harmony as a basis for segmentation. That is, they favor similarity among vowels within segmented units rather than dissimilarity.

We first present results from a version of the experiment in which the familiarization exposure was similar to the short familiarization period in Experiment 1c, in which infants showed a familiarity preference. Because the harmony patterns in the present experiment are more subtle than the vowel identity patterns in Experiment 1, it was not clear whether infants would have sufficient exposure to detect them. If they did detect them, we thought that they would show a familiarity preference (for words) as they did in the case of vowel identity. But it also seemed plausible that infants would show no preference due to the relatively brief exposure to the lexicon and the relatively increased complexity of the harmony used in that lexicon.

##### 5.1. Method

###### 5.1.1. Materials

**5.1.1.1. Familiarization stream.** The familiarization stream consisted of four bisyllabic words, each composed of unique consonant + vowel (CV) syllables. The vowels varied on the phonetic dimensions of height (mid or high), referring to the relative height of the highest point of the tongue during a vowel, and backness (front or back), referring to the relative anteriority of the highest point of the tongue during a vowel. In English, backness in mid and high vowels correlates with lip rounding: Back vowels are produced with rounded lips (rounded), whereas front vowels are produced with spread lips (unrounded). We henceforth refer to the harmonizing dimension as *backness*, to subsume these correlated qualities, although we do not imply that backness is necessarily primary.

As discussed earlier, shifts in the height dimension are reflected in the frequency of the first formant, in the backness dimension in the frequency of the second formant, and in the rounding dimension chiefly in the frequency of the second and third formants. The vowels that were used and their conventional phonetic description were /i/ (high front, as in *heed*), /e/ (mid front, as in *hayed*), /u/ (high back, as in *who’d*),

and /o/ (mid back, as in *hoed*). Consonants were all stops that varied on the dimensions of place (alveolar or bilabial) and voicing (voiced or voiceless): /d/ (voiced alveolar), /b/ (voiced bilabial), /t/ (voiceless alveolar), and /p/ (voiceless bilabial). In this experiment, words harmonized in backness and rounding, but varied in height. Therefore, two words contained the vowels /e/ and /i/ and two other words contained the vowels /o/ and /u/. Syllables that flanked word boundaries were fully disharmonic for the dimensions in question; that is, they contained vowels that differed in height, backness, and rounding. This constraint necessitated that all words conformed to the same pattern of vowel-height alternations, either all high to mid (/i/ syllables before /e/ syllables, and /u/ syllables before /o/ syllables), or all mid to high. To control for artifact effects of alternation direction on segmentation, two different familiarization streams were created, one with mid-to-high words (MH) and one with high-to-mid words (HM). Each of the four consonants began a word, ensuring that neither voicing nor place of articulation (the two distinctive phonemic dimensions for our consonants) correlated with word boundaries. Across MH and HM streams, the order of consonants within words was held constant.<sup>5</sup>

A complete familiarization stream consisted of a fixed sequence of the four words that was repeated 22 times. Transitional probabilities between syllables within and across word boundaries were thus equal, as were the frequencies of each word and part-word. Hence, there was no statistical information in the distribution and sequencing of syllables that provided cues to word boundaries. For the MH stream, the repeated sequence was *detipubeditopu*, and for the HM stream the sequence was *ditepubobidetupo*. So that the beginning or end of the entire stream could not function as a word-segmentation cue, we appended several syllables that were not involved in the experiment to the beginning and end of the complete familiarization streams, as in Experiments 1a and 1c. Fig. 1a and b shows the words and word sequences for the HM and MH stimuli.

**5.1.1.2. Test stimuli.** Four test items were associated with each experimental familiarization stream: Two words, and two part-words. Part-words were syllable sequences that occurred in the familiarization stream but that crossed a word boundary. For the MH stream, the words were *deti* and *topu*, and the part-words were *pude* and *dito*; for the HM stream, the words were *dite* and *tupo* and the part-words were *podì* and *detu* (Fig. 1).

The stimuli were synthesized and assembled as in Experiment 1. With the 22 repetitions of the four words, plus the lead-in and lead-out syllables, the duration of the training stream was approximately 25 s.

### 5.1.2. Procedure

The procedure was identical to that in Experiments 1a & 1c.

### 5.1.3. Subjects

Twenty-seven infants from monolingual English-speaking homes participated in the experiment. Data from three subjects were not analyzed due to failure to maintain a head turn for at least 2 s on each trial (2) and experimenter error (1). The remaining 24 infants (8 female; mean age 7 months and 14 days; range 7 months and 3 days to 7 months and 28 days) were divided randomly into the between-subjects counterbalancing conditions, HM and MH, resulting in 12 subjects per subgroup. Subgroup assignment determined the familiarization and test materials heard (see Fig. 1a for high-to-mid and Fig. 1b for mid-to-high).

<sup>5</sup> We used two syllable words in Experiment 2 because we wanted the vowels within a word to be the most clearly distinct in height, while maintaining the back/front dimension. In English, this restricted us to two front vowels and two back vowels. Since we also wanted to avoid repetitions of vowels within a word, words necessarily could only be two syllables long.

## 5.2. Results

Overall, infants' listening time to words and part-words was similar ( $M = 8.8$  s,  $SE = 0.30$  s for words,  $M = 8.5$  s,  $SE = 0.31$  s for part-words). To assess the effects of experimental variables, we conducted a  $2 \times 2 \times 2$  ANOVA with counterbalancing condition (HM and MH) as a between subjects variable, and block (1 and 2) and word type (word and part-word) as within subjects variables. There was a main effect of block ( $F(1,22) = 10.27$ ,  $p < .005$ ), and a significant interaction of word type by counterbalancing condition ( $F(1,22) = 23.92$ ,  $p < .001$ ); there were no other main effects of interactions. Thus, overall, infants did not show any evidence of segmenting words based on vowel harmony, as we suspected might be the case with such short exposures. However, we did not predict the influence of the counterbalancing condition. To assess this, we ran separate  $2 \times 2$  ANOVA on the HM and MH counterbalancing groups. Both analyses yielded significant main effects of word type ( $F(1,11) = 11.63$ ,  $p < .01$  for MH;  $F(1,11) = 18.62$ ,  $p < .005$  for HM), but in the opposite direction (Fig. 2a). Infants in the MH group listened significantly longer to words over part-words (mean word minus part word difference  $M = 1.4$  s,  $SE = 0.4$  s) whereas infants in the HM group listened significantly longer to part-words over words (mean word minus part word difference  $M = -0.8$  s,  $SE = 0.2$  s). There was a main effect of block in the HM group, with listening times decreasing across blocks ( $F(1,11) = 7.45$ ,  $p = .02$ ), but no interaction with word type.

## 5.3. Discussion

We predicted that the exposure to the stream of words might have been too short for infants to detect and use the vowel harmony pattern to segment the continuous speech stream, and that prediction was borne out: Infants showed no overall preference for harmonic words or non-harmonic part-words. Moreover, infants responded differently to the two test item types depending on their counterbalancing condition, further indicating that vowel harmony was not a major influence on subjects' word segmentation. We did not predict any interaction of the counterbalancing variable and word type, however it is consistent with the interpretation that both counterbalancing groups showed a strong preference for segmenting words with mid-to-high vowel sequences. Mid-to-high sequences were the words in the MH group and part-words in the HM group, and infants listened longer to these types (Fig. 2b). Thus, our interpretation of these results is that infants were segmenting based on a mid-to-high template, and were not (yet) segmenting based on vowel harmony.

Given the *post hoc* nature of this account, we wanted to find some independent support for it. We reasoned that a mid-to-high segmentation bias might arise if English words that infants were likely to have heard tend to have mid-to-high versus high-to-mid sequences. If infants had internalized these tendencies, such tendencies may have influenced how subjects processed the artificial sequence. To explore this possibility, we analyzed the word forms in a large sample of child directed speech, taken from the English corpora of the CHILDES database (Macwhinney, 2000), using the word frequency counts derived by Ping Li (<http://chilides.talkbank.org/derived/parentfreq.cdc>). We then used the CMU pronouncing dictionary (<http://svn.code.sf.net/p/cmuspinx/code/trunk/cmudict/cmudict.0.7a>) to derive pronunciations based on Standard American English, and searched for the vowel sequences that were relevant for our materials. Specifically, we searched for words that contained the mid-to-high sequences /e/ → /i/, /o/ → /u/, /o/ → /i/, and /e/ → /u/, and the high-to-mid sequences /i/ → /e/, /u/ → /o/, /i/ → /o/, /u/ → /e/. We tallied both type and token frequencies of words with these sequences (Table 1). Overall mid-to-high sequences were more frequent than high-to-mid sequences by type counts (270 versus 175) and token counts (11,817 versus 1983).

The higher frequency of words with mid-to-high sequences compared to high-to-mid sequences gives a degree of support to our

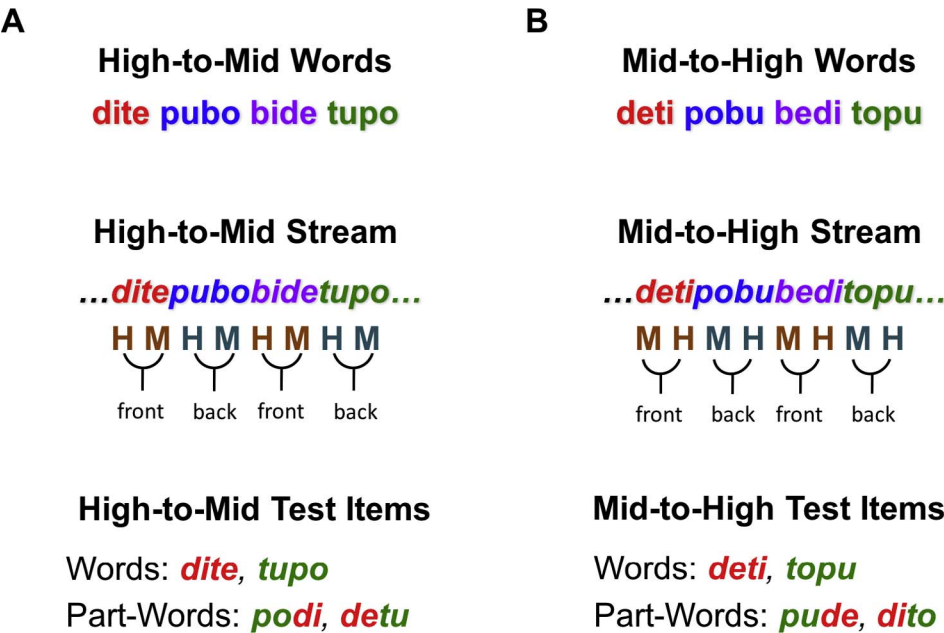


Fig. 1. Stimuli and designs for Experiments 2a and 2c. Panels A and B depict the HM and MH conditions, respectively. The top portion of each panel shows the words, the middle panel shows one iteration of the familiarization sequence, and the bottom portion shows the test items. Infants heard 22 repetitions of the strings in Experiment 2a, and 45 in Experiment 2c.

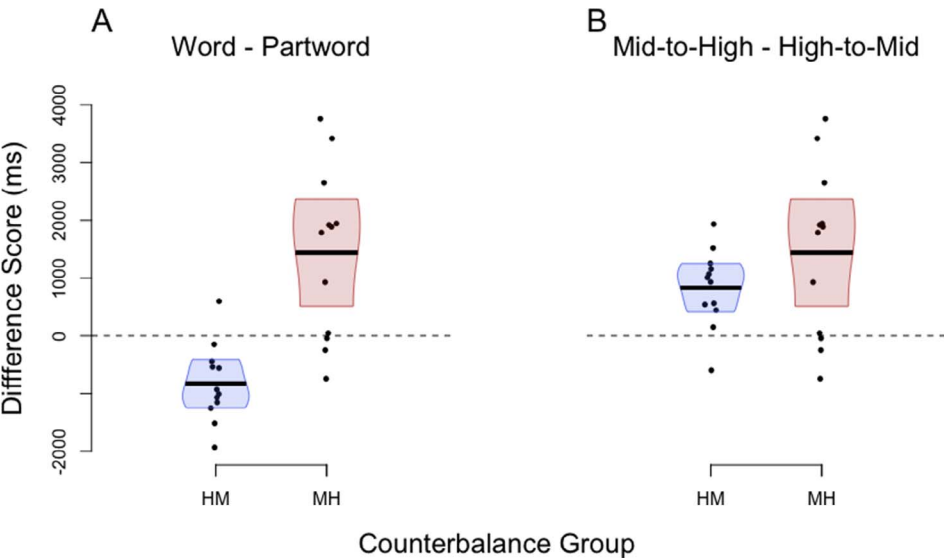


Fig. 2. Difference scores by counterbalance group for Experiment 2a. Each point represents the mean difference score for an individual subject. Solid lines represent the mean difference score for the group. Shaded boxes depict the 95% confidence interval around the group mean. Panel A difference scores are word subtracted by part-word; Panel B replots the same data as Mid-to-High subtracted by High-to-Mid.

**Table 1**  
Token and type counts of words in a large sample of child-directed English that contain the indicated vowels. Rows designate the vowels in question, and columns designate the type and token counts, organized by the direction of the height change.

Vowels	Mid → High		High → Mid	
	Token	Type	Token	Type
i,e	7994	124	869	58
o,u	6	3	42	12
o,i	3780	138	710	73
e,u	37	5	362	32
Sum	11,817	270	1983	175

speculation that infants in Experiment 2a were segmenting based on an internalized mid-to-high bias, when the familiarization phase was relatively short. It is also interesting to note that the magnitude of infants' preference for mid-to-high over high-to-mid test trials was numerically—although not reliably—greater when the sequences corresponded to the vowel harmony pattern (i.e., in the MH group). Thus,

while infants here showed no evidence of using vowel harmony to segment speech, it is possible that with more exposure to the vowel harmony pattern it could dominate infants' segmentation strategies. However, before presenting an experiment in which we used longer exposures, we will first present evidence that infants' preferences in the present experiment were not due to an inherent preference for the mid-to-high test items, irrespective of the familiarization stream. This assessment follows the logic of Experiment 1b.

## 6. Experiment 2b

We interpreted the results of Experiment 2a as showing that infants initially segment the continuous speech stream with a bias to segment word forms in which the vowels across the syllables conform to a mid-to-high template. We also speculated that with greater exposure to the vowel harmony patterns in the speech stream, infants would segment based on the front/back harmony pattern. The present experiment was designed to ensure that the results in Experiment 2a, and expected results with greater familiarization, do not involve inherent preferences for particular test items. To that end, this experiment tested infants on

the word and part-word test items from Experiment 2a, but without prior exposure to the familiarization sequence. We tested for a preference for either type, and, given the results of Experiment 2a, we also tested for a preference based on the sequence of vowel height (mid-to-high versus high-to-mid).

## 6.1. Methods

### 6.1.1. Materials

The materials were the test items from Experiment 2a.

### 6.1.2. Procedure

The procedure was identical to that in Experiment 2a, but without the familiarization phase.

### 6.1.3. Subjects

Data from 24 infants were analyzed (average age 7 months 0 days, range 6 months 20 days to 7 months 23 days).<sup>6</sup> Five additional infants were tested, but their data were not analyzed because they did not meet the predetermined criterion of listening time of at least 2 consecutive seconds for at least two of each type of test trial (word and part-word). Subjects were randomly assigned to one of two groups. The group assignment determined whether the test words and part-words were taken from the Experiment 2a test trials in which vowel height within words went from high to mid (HM) or from mid to high (MH).

## 6.2. Results and discussion

A  $2 \times 2$  ANOVA with the between-subjects variable of group (HM and MH) and within-subjects variable of trial type (harmonic and non-harmonic) showed no significant main effects or interactions (all  $F$ s < 1). Confirming this, when subjects' mean looking times were collapsed across group (HM and MH), a paired  $t$ -test revealed that listening times to words was not significantly different than listening time to part-words ( $M = 8.3$  s,  $SE = 0.4$  s and  $M = 8.4$  s,  $SE = 0.6$  s, respectively;  $t(23) = 0.190$ ,  $p = .85$ ). Similarly, when subjects' mean looking times were collapsed across harmonic word status (word and part-word), listening times to mid-to-high sequences was not significantly different from their listening times to high-to-mid sequences ( $M = 8.3$  s,  $SE = 0.3$  s, and  $M = 8.4$  s,  $SE = 0.4$  s, respectively;  $t(23) = 0.26$ ,  $p = .80$ ). Thus, infants did not appear to have any systematic inherent preference for the test items used in Experiment 2a.

Recall that our interpretation of the results in Experiment 2a was that infants were segmenting mid-to-high sequences because of the prevalence of these word forms (compared to high-to-mid sequences) in their experience with English. We speculated that the familiarization exposure may have been too brief for infants to have noticed and started to segment speech based on the vowel harmony patterns, and predicted that with more exposure, infants might start to segment based on backness harmony. (We also noted that the infants showed a non-significant stronger preference for mid-to-high sequences when they also corresponded to vowel harmony patterns.) This prediction was tested in the next experiment, in which the familiarization phase provided approximately twice the exposure to each of the words, as in Experiment 1a.

## 7. Experiment 2c

In this experiment, we provided 7-month-old infants with approximately double the exposure to the continuous speech stream as in Experiment 2a. We conjectured that with the increase in exposure,

infants would be able to detect the backness harmony patterns and use them to segment the continuous stream of speech. As in Experiment 2a, we were interested in whether 7-month-olds infants exposed only to English would be sensitive to the backness and rounding harmony that defines the words in the sequence, and segment out word forms based on vowel harmony. In Experiment 2a we found that infants' segmentation behavior was best described as a strategy to segment the stream into sequences of syllables with a mid vowel followed by a high vowel, perhaps mirroring tendencies in English. However, with twice the exposure, in the present experiment we hypothesized that infants might detect the harmony pattern and use it to segment speech, overriding the apparent mid-to-high bias revealed in Experiment 2a. However, whereas in Experiment 1a infants showed a novelty preference (i.e., for part-words), with the subtler harmony patterns here, we might expect a familiarity preference. This expectation rests on the assumption that, all else being equal, preferences that arise from familiarization should progress from no preference with minimal familiarization, to a familiarity preference, to a novelty preference with increased familiarization. Whereas infants exposed to simpler vowel identity patterns with brief exposures (Experiment 1c) showed a familiarity preference, infants showed no harmony-based preference with similar exposure durations with the more nuanced materials (Experiment 2a).

We also added a new condition in this experiment that tested an additional group of infants on their ability to generalize learned harmony patterns to new strings that were not in the familiarization stream, but that adhered to the harmony patterns. This type of generalization is suggestive of a more abstract, rule like representation of the harmony pattern, which goes beyond a representation based on the particular items in the familiarization stream. We call this the *generalization condition* and we call our traditional version where harmony-based segmentation is assessed the *segmentation condition*. The generalization condition followed the same design of familiarizing infants with a continuous stream of speech and testing on words and part-words. But for infants in the generalization condition, we altered the familiarization streams by changing the voicing of the consonants to produce a new set of syllables while maintaining the vowel harmony patterns (because the vowels were not changed from the original stream). While we altered the familiarization syllables in this condition, the test items (words and part-words) were identical to those in the segmentation condition. Thus, neither the 'words' nor the 'part-words' occurred in the generalization condition familiarization stream, but the 'words' conformed to the harmony pattern in the familiarizing sequence and the 'part-words' did not. If infants systematically discriminated these test item types, it would suggest that they generalized the backness harmony patterns to new syllable sequences.

### 7.1. Method

#### 7.1.1. Materials

**7.1.1.1. Segmentation condition.** The materials and design was the same as in Experiment 2a, except that the familiarization stream contained 45 repetitions of the 4-word sequence, for a total duration of approximately 50 s.

**7.1.1.2. Generalization condition.** The test items were identical to those in the segmentation condition (and Experiment 2a). However, we modified the familiarization streams (HM and MH) by changing the voicing value of each stop consonant (Fig. 3). This manipulation maintained all the general patterns between the corresponding segmentation and generalization streams (i.e., word-initial consonants, patterns of vowel alternations across syllables, etc.), but changed the actual syllables involved. If infants in the segmentation condition do not show evidence of harmony-based segmentation, then infants exposed to the generalization material would not be expected to differentiate the 'words', which conformed to the familiarized harmony patterns, from the 'part-words', which did not. On the other hand, if

<sup>6</sup> The age information for seven subjects was missing and the sex information for six subjects was missing (9 males and 9 females comprise the available information), but the recruitment criteria were the same for all infants.



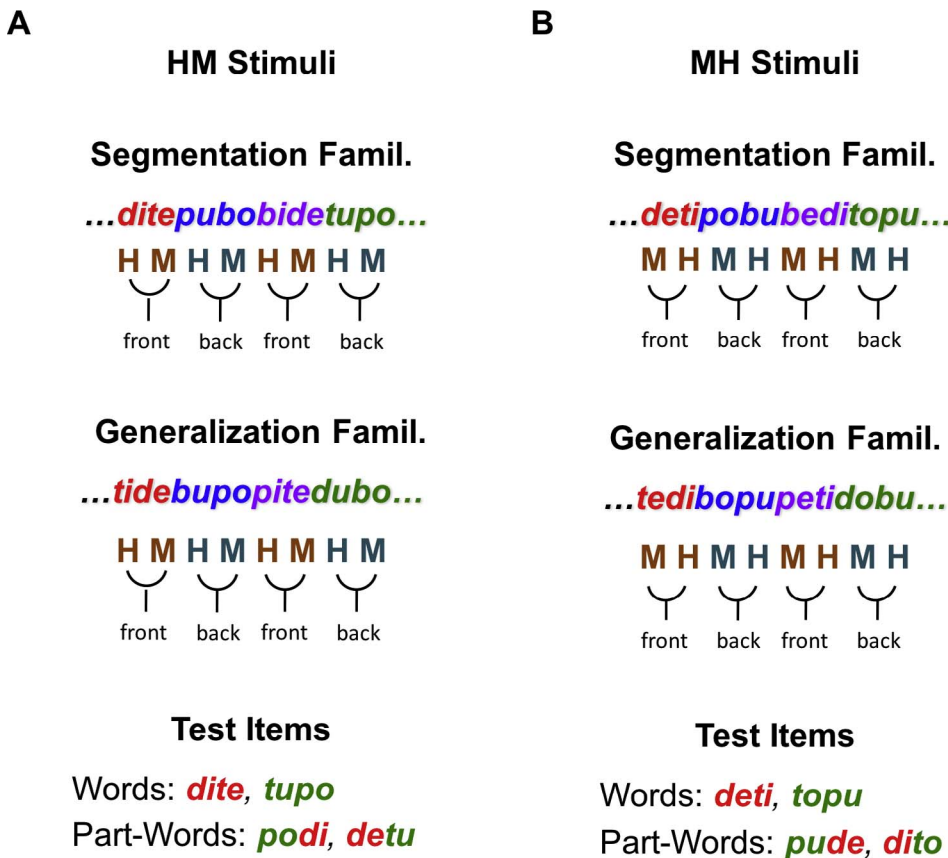


Fig. 3. Familiarization and test items for Experiment 2c. Panels A and B depict stimuli for the HM and MH groups, respectively. Within a group, the test items were the same for all subjects, but the familiarization material varied according to subgroup (segmentation or generalization).

infants do show evidence of harmony-based segmentation in the segmentation condition, then the results in the generalization condition would shed light on whether infants generalized a productive vowel harmony rule or constraint (Peña, Bonatti, Nespor, & Mehler, 2002), going beyond the particular items in the familiarization stream. Given the brevity of the familiarization materials and infants' lack of prior experience of a language with vowel harmony, we did not expect infants to generalize a harmony rule.

**7.1.1.3. Subjects.** Forty-nine 7-month-old infants from monolingual English-speaking homes participated in the experiment ( $M = 7$  months and 12 days; range 7 months and 2 days to 8 months and 5 days). Data from one additional subject were not analyzed due to failure to maintain a head turn for at least 2 s on each trial. Subjects were divided randomly and equally into two groups, HM and MH, then further into segmentation and generalization subgroups, resulting in 12 subjects per subgroup. Group assignment determined the familiarization and test materials heard. All infants hearing HM streams were tested on the same test items; all infants hearing MH streams were tested on the same test items (Fig. 3).

## 7.2. Results

We performed a  $2 \times 2 \times 2 \times 2$  ANOVA with block (first and last) and word type (word and part-word) as within-subject variables and harmony condition (MH and HM) and familiarization group (segmentation and generalization) as between subjects variables. There was a main effect of block ( $F(1, 44) = 8.67$ ,  $p = .005$ ), a marginal effect of word type ( $F(1, 44) = 3.632$ ,  $p = .063$ ) and a significant familiarization group by word-type interaction ( $F(1, 44) = 4.406$ ,  $p = .042$ ); there were no other significant main effects or interactions.

As expected, infants responded differently to words and part-words depending on whether they were in the segmentation or generalization

group. To better understand this interaction, we ran  $2 \times 2 \times 2$  ANOVAs separately on each group, with block (first and last) and word type (word and part-word) as within-subject variables and harmony condition (MH and HM) as a between subject variable. In the segmentation condition, there was a main effect of block ( $F(1, 22) = 7.071$ ,  $p = .014$ ) and a main effect of word type ( $F(1, 22) = 7.548$ ,  $p = .012$ ); there were no other significant main effects or interactions. Infants listened longer to words ( $M = 8.89$  s,  $SE = 0.76$  s) compared to part-words ( $M = 8.09$  s,  $SE = 0.74$  s;  $M_{diff} = 0.80$  s,  $SE_{diff} = 0.31$  s), and overall their listening time decreased from Block 1 ( $M = 8.97$  s,  $SE = 0.70$  s) to Block 2 ( $M = 8.01$  s,  $SE = 0.79$  s;  $M_{diff} = 0.95$ ,  $SE_{diff} = 0.36$  s).

In the generalization condition, none of the main effects or interactions reached significance (all  $F$ s  $< 2.23$ ), in particular, the main effect of word type was not significant ( $F(1, 22) = 0.02$ ,  $p = .89$ ).<sup>7</sup> Infants listened equally long to words and part-words ( $M = 8.67$  s and  $M = 8.59$  s, respectively;  $M_{diff} = 0.08$  s,  $SE_{diff} = 0.36$  s), and did not appear to generalize the harmony patterns to new syllables.

Fig. 4 shows the difference scores of infants' mean listening times to words minus mean listening times to part-words, as a function of familiarization group.

## 8. Discussion

Infants' preference for words over part-words in the segmentation group indicates that they segmented the words in the familiarization streams based on vowel harmony. The results support the hypothesis

<sup>7</sup> An anonymous reviewer suggested that the overall novelty of the test items in the segmentation condition could have overwhelmed any systematic preferences for words or part-words. However, if that were the case, one might expect to see such preferences emerge in the second block of test trials, after infants had multiple exposures to all test items in the first block. Analyzing just the second block, we still did not find any evidence of a word-type preference ( $t(23) = 0.50$ ,  $p = .62$ ).

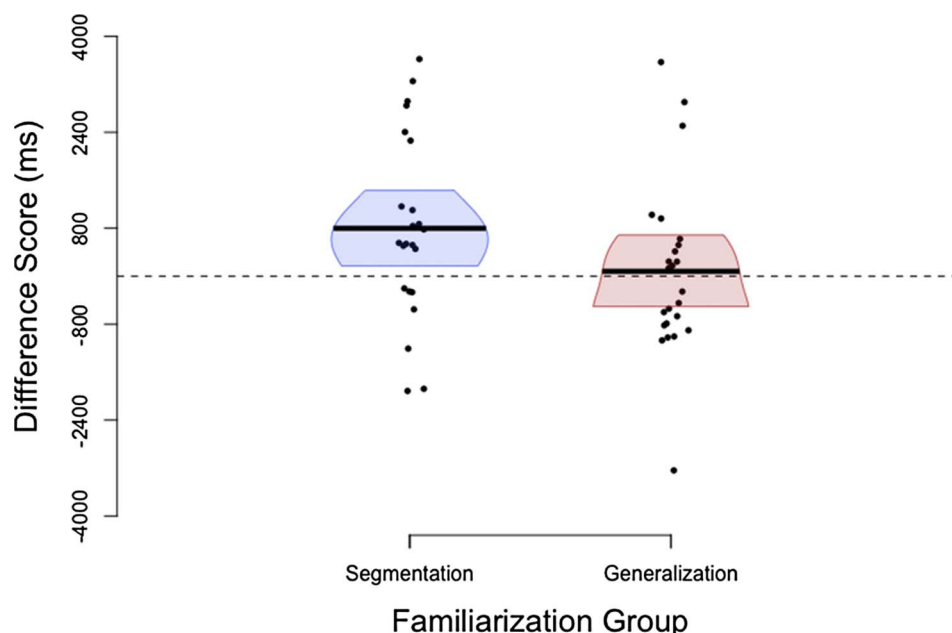


Fig. 4. Word minus part word listening scores for Experiment 2c. Data are organized by familiarization group. Points designate difference scores for individual subjects. Solid lines represent mean difference scores for each group. Shaded areas represent 95% CIs around the group means. The dashed line indicates chance (0 ms).

that 7-month-old infants can and do recruit vowel harmony cues in segmenting syllable sequences from fluent speech, and suggest that the absence of evidence for harmony-based segmentation in Experiment 2a was due to the very brief exposure to the sequences that contained the harmony patterns. There was also no evidence of an interaction between the particular pattern of height alternation (high-to-mid or mid-to-high), suggesting that the harmony pattern superseded the mid-to-high segmentation strategy that infants seemed to have in the first 25 s of processing the familiarization streams (Experiment 2a). Thus, infants apparently changed their segmentation strategy midstream, as it were, as the evidence for the stream-internal harmony pattern accumulated.

Interestingly, infants' behavior in the generalization condition indicates that the 50 s exposure to the harmony patterns in the familiarization stream did not result in the infants learning harmony rules. Specifically, they did not abstract a productive generalization of backness harmony from the familiarization stream and evaluate test items based on these generalizations. If they had, they would have shown a preference difference for words and part-words in the generalization condition. This is because the test 'words' in the generalization group showed vowel harmony even though the sequence of syllables was different from those the infants heard in the familiarization stream, resulting from the switch in voicing in the stop consonants.

The finding of harmony-based segmentation is especially noteworthy for our sample of English-exposed infants, as their language does not contain harmony patterns. In order to segment words as they did, infants first had to detect harmony on a particular phonetic dimension, and then they had to use that information to posit word boundaries; they accomplished this within the 50 s familiarization phase. Because infants in our study had no prior experience of vowel harmony and its correspondence to word boundaries, our results strongly suggest that infants are inherently predisposed to rapidly detect these relatively subtle patterns and employ them in ways that are advantageous for the critical process of discovering the words of their language.

## 9. General discussion

### 9.1. Summary

In a series of six experiments we demonstrated that 7-month-old infants, who had no regular exposure to a language with vowel

harmony, nonetheless used vowel harmony patterns as a basis for segmenting a continuous sequence of syllables into proto-word forms. In Experiment 1, we tested infants' sensitivity to vowel identity harmony, a version of harmony attested in the world's languages that is most complete and simplest to detect. In Experiments 1a and 1c, we showed that 7-month-olds segment out contiguous sequences of syllables that contain identical vowels, using a change in vowels as cue to a word boundary. Even with a very brief exposure of 38 s (20 repetitions of the sequence), infants showed evidence of harmony-based segmentation. Experiment 1b tested for the possibility that infants had an inherent preference for nonsense words with identical vowels or ones with different vowels, and showed they did not. Moreover, infants' preferences differed based on the duration of their exposure to the familiarization syllable sequence, from a familiarity preference (i.e., for words) with a very brief exposure (Experiment 1c) to a novelty preference (i.e., for part-words) with a longer exposure (Experiment 1a). This preference pattern is consistent with theories about how infants' listening preferences correspond to representation and processing (Hunter & Ames, 1988; Kidd et al., 2012) and further confirms that the units that infants had segmented were indeed the harmonic words. An alternative view where infants segmented out what we called the part-words would have to explain a preference pattern of novelty with a brief exposure and familiarity with a longer exposure, which does not appear to have any support in the literature, and would appear difficult to explain.

In Experiment 2, we tested a more subtle (and more widespread) form of harmony. Syllables within words were equated on the dimension of vowel backness (and lip rounding), but differed on the dimension of vowel height. Across word boundaries, syllables differed along both dimensions. Experiment 2a showed that with a relatively short exposure duration (22 repetitions of the familiarization sequence), 7-month-olds did not appear to use vowel harmony as a basis of segmentation, but rather appeared to segment such that the resulting units started with a syllable with a mid vowel followed by a syllable with a high vowel. We speculated that this bias was the result of infants' prior exposure to child-directed English, which tends to have more words with mid-to-high sequences versus high-to-mid sequences. Importantly, though, when infants' exposure to the familiarization sequence was doubled, they abandoned the English template and segmented based on the vowel harmony patterns, positing boundaries at junctures of disharmony. In Experiment 2b, we showed that infants' test preferences in

Experiments 2a and 2c were not due to inherent preferences for particular test items, and rather resulted from their processing of the familiarization stream. Finally, Experiment 2c also showed that, although infants detected and used the vowel harmony patterns to segment the continuous speech stream, they did not appear to generalize phonological rules: When the familiarization sequence was modified to implement the same harmony patterns but over different syllables, infants showed no systematic preference for test items based on those harmony patterns.

These findings make important contributions to our understanding of the mechanisms infants bring to bear on the word segmentation problem. More broadly, they contribute to our understanding of infants' capacities to detect linguistic relevant patterns—in this case, vowel harmony. At the same time, they raise a number of interesting questions about the details of this capacity.

### 9.2. Infants' representation of vowel harmony in an artificial language

The infants in these experiments experienced a relatively brief exposure—less than a minute—to a novel sequence of syllables in which there were systematic vowel harmony patterns, yet they had no prior experience with vowel harmony. What was the nature of the representations that allowed them to detect the patterns and use them as a basis for segmenting the speech stream?

There are two possible characterizations of these representations, either of which holds important theoretical implications. Phonological patterns, such as vowel harmony, operate over abstract phonological features (e.g., backness, rounding). If infants' representation of similarity between harmonizing vowels is similarly structured, then our results would provide evidence that representations of speech composed of abstract units influence language processing in 7-month-olds. On the other hand, vowel harmony causes acoustic similarities between vowels that harmonize: As we discussed in the Introduction, backness and rounding harmony largely involve similarities in the frequencies of the second and third formants. Indeed, a statistical analysis of the first three formants in our materials in Experiment 2 (Table A1) shows that, overall, frequencies of the second and third formants were significantly lower in the back vowels compared to the front vowels (F2 means for front vowels and back vowels were 2327 Hz and 1407 Hz, respectively,  $t(11) = 10.75$ ,  $p < .001$ ; F3 means for front vowels and back vowels were 2964 Hz and 2706 Hz, respectively,  $t(11) = 6.80$ ,  $p < .001$ ; the test items in Experiment 2 patterned similarly, Table A2). More importantly, in the actual familiarization sequences, F2 and F3 differences

between adjacent syllables were much greater at word boundaries compared to within words (Fig. 5). Thus, infants' segmentation behavior might have been due to lower-level acoustic-phonetic similarities rather than more abstract phonological representations. Harmony-driven segmentation might then engage more general auditory grouping mechanisms that respond to the spectral similarity of adjacent syllables that have harmonizing vowels (McAdams & Bertoncini, 1997). If so, our study provides important insights into the perceptual biases that could support the eventual acquisition of harmony patterns. The study also sheds light on the perceptual mechanisms that might be responsible for the emergence of vowel harmony as an organizing principle for the sound structure of words in many languages.

Although our findings cannot decisively rule out either possibility, the fact that we did not observe generalization of the familiarized harmony pattern to new syllables suggests that, at least with the relatively brief exposures in our study, infants' representations were more likely to be based on acoustic rather than phonological properties. Of course, it is also possible that both types of representations play some role. Further studies are needed to address this question more directly.

Another question is whether, under similar exposure conditions as in the present study, infants would treat other kinds of harmony patterns attested in the world's languages in a similar way. In Experiment 2, words harmonized on the dimension of backness and rounding, but languages vary as to the dimensions over which harmony restrictions apply (e.g., some involve vowel height or tongue root position). Would harmony-naïve infants respond to vowel harmony for other dimensions in the way they do to backness/rounding harmony? Comparisons of different harmony patterns would shed light on the question of mechanisms and representations just discussed, as the acoustic similarities across different dimensions of vowel harmony differ.

### 9.3. The development of vowel harmony representations

Another question concerns the role of experience in shaping sensitivity to vowel harmony during development. As previously mentioned, studies with adults indicate that native speakers of languages that lack vowel harmony rules do not use disharmony as a cue to word boundaries. Our findings, taken together with those, suggest that experience is necessary to maintain sensitivity to vowel harmony in speech processing tasks. The question naturally arises as to the developmental time course of this sensitivity loss, as well as the mechanisms of change. Moreover, if the language the infant is learning is one with vowel harmony, it would be revealing to understand how exposure to

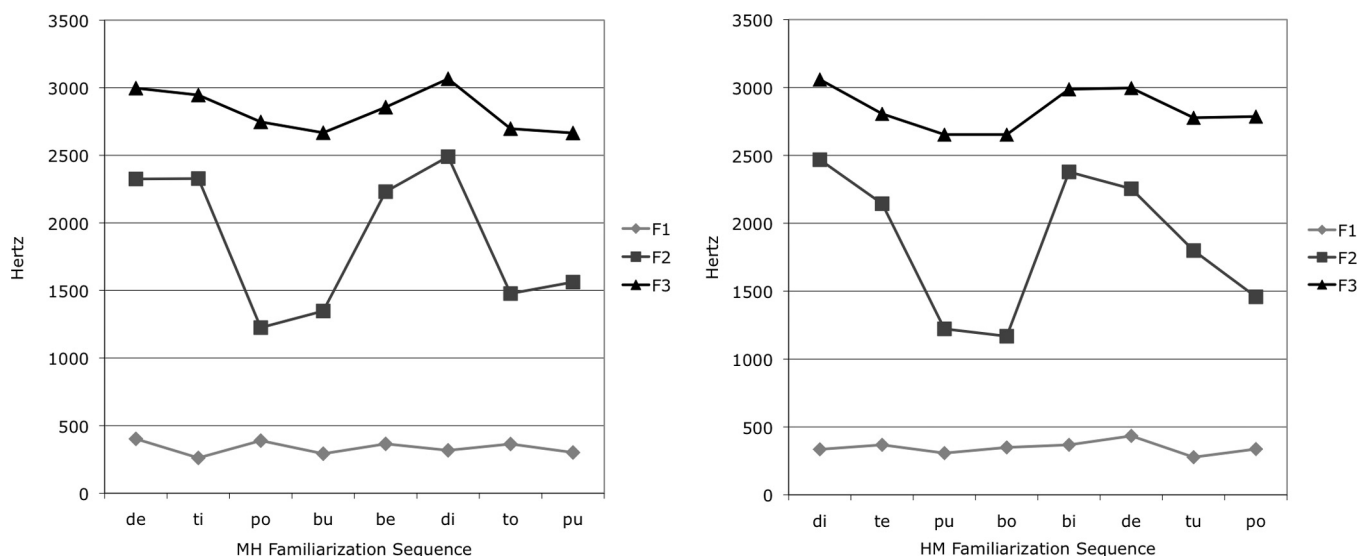


Fig. 5. Plot of first three formants in the MH and HM segmentation familiarization sequences (Experiments 2a & 2c). Formants were extracted using Praat (Boersma & Weenink, 2010).

harmony constraints that operate over one dimension (e.g., backness) affects sensitivity in other dimensions (e.g., vowel height). Are sensitivities to vowel harmony on other dimensions lost more rapidly than if the exposure language has no vowel harmony patterns? Or is global sensitivity to vowel harmony maintained longer, even for dimensions that are not relevant for the language being learned? Furthermore, it would be valuable to explore when a child integrates harmony as part of the phonological component of the grammar of his/her language and implements it in a productive fashion. The answers to these questions would reveal much about the nature of the mechanisms involved in the detection and representation of vowel harmony. More broadly, they would provide additional data on how experience modulates maturational change during development.

#### 9.4. The interplay of long term and short term learning

Although we did not set out to test what word segmentation strategies 7-month-old English-learners had acquired before they came into the lab, we did find preliminary evidence that infants are biased to segment continuous speech based on the within-word vowel patterns of child-directed English—i.e., syllable sequences with mid-to-high vowels. The general phenomena that infants acquire segmentation strategies from exposure to their language is well established (Curtin et al., 2005; Jusczyk, Houston, & Newsome, 1999; Mattys & Jusczyk, 2001; Mattys et al., 1999; Nazzi, Mersad, Sundara, Iakimova, & Polka, 2014; Polka & Sundara, 2011), however we are not aware of any prior reports of this particular sensitivity in English-reared 7-month-olds. An additional point that this finding revealed is that new patterns that infants encounter in relatively short samples of speech can dominate infants' segmentation strategies, at least in the short term. The factors that determine how multiple cues and sources of evidence interact is no doubt complex, having to do with the strength of the evidence, and the utility of the information. It is also likely that the way in which new

information influences behavior changes across development (Thiessen & Saffran, 2003). This study provides a small piece of evidence that can inform these larger questions, by showing that the word segmentation strategies of 7-month-olds are influenced by new patterns in the input. After a brief yet systematic exposure to vowel harmony patterns, infants began segmenting based on vowel harmony instead of the more established probabilistic patterns in their native language (i.e., mid-to-high vowel sequences). This finding also emphasized the importance of considering infants' language background, even when designing experiments that use artificial languages.

#### 10. Conclusion

We have shown that 7-month-old infants who have never been exposed to a language with vowel harmony nevertheless rapidly detect vowel harmony patterns when they occur. Moreover, when no other cues to word boundaries are present, infants segment proto-word forms from continuous speech based on these harmony patterns. Our results indicate that infants are predisposed to detect harmony patterns and to use vowel harmony to segment continuous speech into words. Our findings thus provide evidence for a previously unknown mechanism by which infants could discover the words of their language.

#### Acknowledgements

We thank all the parents and children who participated in this study. We also thank Dick Aslin, Justin Wood, and Laura Siegel for their comments on earlier versions of this paper.

This research was supported by a grant from the National Institutes of Health – United States (HD-040368), and from the Provost's Undergraduate Research Program and the Advancing Scholarship in Humanities and Social Sciences Initiative at the University of Southern California.

#### Appendix A

Tables A1 and A2 provide detailed acoustic measurements for stimuli in Experiment 1 and Experiment 2, respectively. Measurements were carried out using Praat (Boersma & Weenink, 2010).

**Table A1**

Measurements for each syllable in the MH and HM experimental familiarization streams in Experiment 1.<sup>a</sup> All frequency measurements were made on the vowel portion.

Syllable	Duration (ms)	Pitch Change (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/bi/	126	23	368	2378	2987
/be/	100	65	365	2231	2856
/di/	100	23	326	2479	3063
/de/	98	47	418	2290	2997
/ti/	130	27	261	2328	2946
/te/	138	35	368	2144	2806
Means and standard deviations for front vowels	115 (18.1)	37 (16.6)	351 (52.9)	2308 (116.4)	2943 (95.5)
/bo/	123	20	349	1168	2653
/bu/	124	26	292	1348	2667
/po/	138	32	363	1342	2767
/pu/	144	19	304	1392	2659
/to/	140	36	364	1477	2697
/tu/	128	22	277	1799	2777
Means and standard deviations for back vowels	133 (8.77)	26 (6.9)	325 (38.4)	1421 (211.0)	2703 (55.4)

<sup>a</sup> Note. Measurements for /di/, /de/, /po/, and /pu/ are averages from the two familiarization streams.



Table A2

Measurements for each syllable in the test items for Experiment 2. All frequency measurements were made on the vowel portion. The first four items are words and part-words in the MH condition; the last four items are words and part-words in the HM condition.

Test item	Syllable	Duration (s)	Pitch change (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
deti	/de/	0.1469	45	421	2352	3041
	/ti/	0.1812	24	466	2669	3172
topu	/to/	0.1575	41	374	1442	2668
	/pu/	0.1817	32	384	1271	2589
pude	/pu/	0.153	21	319	1553	2681
	/de/	0.1633	45	462	2423	2984
dito	/di/	0.1089	19	363	2528	3100
	/to/	0.1984	33	465	1420	2743
dite	/di/	0.1086	23	346	2469	3028
	/te/	0.1858	45	455	2403	3024
tupo	/tu/	0.1561	25	294	1705	2691
	/po/	0.1973	34	482	1228	2761
podi	/po/	0.1511	26	347	1438	2791
	/di/	0.148	29	352	2604	3134
detu	/de/	0.1011	33	403	2328	2983
	/tu/	0.1848	35	430	1767	2728

## References

- Altan, A., Kaya, U., & Hohenberger, A. (2016). Sensitivity of Turkish infants to vowel harmony in stem-suffix sequences: Preference shift from familiarity to novelty. In J. Scott, & D. Waughal (Eds.), *BUCLD 40 online proceedings supplements*. < In: [http://www.bu.edu/buclD/files/2016/09/BUCLD\\_Proceedings\\_2016\\_01\\_25\\_altan1.pdf](http://www.bu.edu/buclD/files/2016/09/BUCLD_Proceedings_2016_01_25_altan1.pdf) > .
- Aslin, R., Saffran, J. R., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.
- Blevins, J. (2009). Low vowel dissimilation outside of Oceanic: The case of Alamblak. *Oceanic Linguistics*, 48, 477–483.
- Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer [Computer program]*. Camilli, A. (1929). Il dialetto di Servigliano. *Archivum Romanicum*, 13, 220–271.
- Curtin, S., Mintz, T., & Christiansen, M. (2005). Stress changes the representational landscape: Evidence from word segmentation in infants. *Cognition*, 96, 233–262.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306.
- Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, 24, 564–574.
- 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, 254–260.
- Hohenberger, A., Kaya, U., & Altan, A. (2017). Discrimination of vowel-harmonic vs vowel-disharmonic words by monolingual Turkish infants in the first year of life. In M. LaMendola, & J. Scott (Eds.), *Proceedings of the 41st annual Boston University conference on language development* (pp. 309–322). Somerville, MA: Cascadia Press.
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69–95.
- Johnson, E., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.
- Ketrez, F., & Aksu Koç, A. (2003). Early verbal morphology in Turkish: Emergence of inflections. In D. Bittner, W. U. Dressler, & M. Kilani-Schoch (Eds.), *Development of verb inflection in first language acquisition: A cross-linguistic perspective* (pp. 27–52). Berlin: Mouton De Gruyter.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, 7, e36399.
- Kuhl, P. K. (1979). The perception of speech in early infancy. In N. J. Lass (Vol. Ed.), *Speech and language: Advances in basic research and practice: Vol. 1*, (pp. 1–47). New York: Academic Press Inc.
- Macwhinney, B. (2000). *The Childes project: Tools for analyzing talk: The database* (Vol. II, 3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mattys, S., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91–121.
- Mattys, S., Jusczyk, P. W., Luce, P., & Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494.
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *The Journal of the Acoustical Society of America*, 102, 2945–2953.
- Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: Task-dependent and dialect-specific aspects. *Journal of Child Language*, 41, 600–633.
- Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607.
- Polka, L., & Sundara, M. (2011). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, 17, 198–232.
- Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 259, 39–41. <http://dx.doi.org/10.1038/259039a0>.
- Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36, 422–444.
- Suzuki, K. (1998). *A typological investigation of dissimilation*. Doctoral dissertation. University of Arizona.
- 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, 706–716.
- van der Hulst, H., & van de Weijer, J. (1995). Vowel harmony. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 495–534). Cambridge, MA: Blackwell Publishers.
- van Kampen, A., Parmaksız, G., van de Vijver, R., & Höhle, B. (2007). Metrical and statistical cues for word segmentation: Vowel harmony and word stress as cues to word boundaries by 6- and 9-month-old Turkish learners. In A. Gavarró, & M. J. Freitas (Eds.), *Language acquisition and development: Proceedings of GALA 2007* (pp. 313–324). Newcastle: Cambridge Scholars Publishing.
- Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, 38, 133–149.
- Walker, R. (2011). *Vowel patterns in language*. Cambridge: Cambridge University Press.
- Werker, J., & Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24, 672–683.
- Werker, J., & Tees, R. (1992). The organization and reorganization of human speech-perception. *Annual Review of Neuroscience*, 15, 377–402.