

Research Article

Cite this article: Omane, P.O., Boll-Avetisyan, N. and Benders, T. (2025). Tongue root harmony cues for speech segmentation in multilingually raised infants learning languages with and without vowel harmony in Ghana (Africa). *Bilingualism: Language and Cognition* 1–11. <https://doi.org/10.1017/S1366728925100485>


Received: 18 July 2024
Revised: 15 May 2025
Accepted: 31 July 2025

Keywords:

multilingual infants; speech segmentation; vowel harmony; Africa; Akan; understudied languages

Corresponding author:

Paul Okyere Omane;
Email: omaneokyere@hotmail.com

 This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Tongue root harmony cues for speech segmentation in multilingually raised infants learning languages with and without vowel harmony in Ghana (Africa)

Paul Okyere Omane^{1,2,3,4,5,6} , Natalie Boll-Avetisyan^{1,7}  and Titia Benders^{8,9}

¹Department of Linguistics, University of Potsdam, Potsdam, Germany; ²School of Psychological Sciences, Macquarie University, Sydney, NSW, Australia; ³International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), Newcastle University, Newcastle upon Tyne, UK; ⁴International Doctorate for Experimental Approaches to Language and Brain, School of Psychological Sciences, Macquarie University, Sydney, NSW, Australia; ⁵International Doctorate for Experimental Approaches to Language and Brain, Department of Linguistics, University of Potsdam, Potsdam, Germany; ⁶International Doctorate for Experimental Approaches to Language and Brain, Center for Language and Cognition Groningen, University of Groningen, Groningen, The Netherlands; ⁷Research Focus Cognitive Sciences, University of Potsdam, Potsdam, Germany; ⁸Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands and ⁹Department of Linguistics, Center for Language Sciences, Macquarie University, Sydney, NSW, Australia

Abstract

This study investigated the hypothesis that 9- to 11-month-old multilingual infants learning Advanced Tongue Root (ATR) harmony languages (such as Akan) alongside other non-vowel harmony languages in Ghana (Africa) can use ATR harmony cues for speech segmentation. Using the central fixation procedure, infants were familiarized with bisyllabic words in two passages, one with ATR cues and one without, and then tested on isolated familiarized and novel bisyllabic words. Results indicate that, as a group, infants segmented words in their native language using ATR harmony cues, showing a familiarity preference. No effect of exposure to ATR harmony language(s) was found. These results provide the first evidence of word segmentation in infants learning between two and five languages, and with infants in Africa. The findings contribute to our understanding of multilingual infants' language processing, suggesting their sensitivity to phonotactic cues for speech processing.

Highlights

- Infants learning ATR harmony languages use ATR cues for speech segmentation.
- Multilingual infants in Ghana are sensitive to native language segmentation cues.
- Little exposure to harmony language may suffice for multilinguals to track harmony cues.

1. Introduction

Word segmentation poses a fundamental challenge for language-learning infants. This task involves extracting individual words from continuous speech, which is intricate due to the absence of clear pauses or boundaries between words in infants' input (Cutler & Otake, 1994; Gonzelez-Gomez & Nazzi, 2013). Segmentation, therefore, presents a challenge for infants who are yet to acquire a lexicon: they must find out where words start and end in the speech stream (e.g., Jusczyk, 1997), as achieving this feat is critical for infants' lexical acquisition (Cristia et al., 2014; Junge et al., 2012; Singh et al., 2012).

Previous research has revealed that infants rely on various cues from their ambient language to facilitate the detection of words from continuous speech (e.g., Höhle & Weissenborn, 2003). Some of the cues explored in infants so far include prosodic regularities, which involve rhythmic units such as the strong/weak stress pattern of words in languages like English (e.g., Echols et al., 1997; Jusczyk et al., 1999; Orena & Polka, 2019), and phonotactic regularities, which refers to the rules or constraints governing the possible sound combinations, phoneme positions and syllable structures in a language (e.g., Friederici & Wessels, 1993; Gonzalez-Gomez & Nazzi, 2013; Mattys & Jusczyk, 2001). The latter (i.e., phonotactic regularities) is the focus of the present study.

Our understanding of infant word segmentation primarily comes from studies conducted with infants growing up in Western contexts learning Indo-European languages, particularly English (Nazzi et al., 2008). This lack of diversity could, however, cause biases in theory building (Aravena-Bravo et al., 2023; Cristia et al., 2023; Kidd & Garcia, 2022; Singh et al., 2023). The current study contributes to diversifying infant language acquisition research and cross-linguistic understanding of phonological acquisition by focusing on infants growing up in Ghana, Sub-

Saharan Africa (an understudied population). Language acquisition in this population is important in particular because infants typically grow up multilingually with several understudied African languages (e.g., Ga, Akan and Ewe) in multi-ethnic and multi-generational homes (see also Omane et al., 2023, 2024, 2025). To this point, very little is known about how multilingual exposure affects infants' speech perception.

The present study investigated infants' use of vowel harmony (VH), a language-specific, phonotactic constraint on the co-occurrence of vowels in a word or syllable as a cue for early speech segmentation. The VH constraint posits that vowels in a di- or multi-syllabic word should share some specific vocalic features. To take the language and the vowel harmony rule under test in the present study as an example: in Akan (a Kwa Niger-Congo language), an Advanced Tongue Root (ATR) harmony language, vowels are grouped into two different sets: +ATR (/i,e,u,o/) and -ATR (/ɪ,ʊ,ɛ,ɔ/), based on the position of the tongue root during the articulation of these vowels. The ATR VH constraint in Akan requires vowels from the same set to occur within a word (with few exceptions), for example, +ATR: *kube* 'coconut' and -ATR: *ɛkɔm* 'hunger' (Dolphyne, 1988).

Previous research has established that monolingual infants initially seem to be biased for preferring VH patterns: four-month-old English-learning infants show a listening preference for harmonic over disharmonic syllable sequences (Solá-Llonch & Sundara, 2025). However, from the age of 6 months onwards, only infants learning a VH language exhibit this listening preference (Altan et al., 2016; Hohenberger et al., 2016, 2017; van Kampen et al., 2008). In contrast, infants learning a language without VH show no such preference (Gonzalez-Gomez et al., 2019; van Kampen et al., 2008; Solá-Llonch & Sundara, 2025), suggesting that the preference at later ages is a result of attunement to the input language. Infants growing up multilingually in Ghana typically acquire languages both with and without VH. Our recent work showed that these infants also exhibit a listening preference for VH patterns at the age of 6 months, and this preference did not depend on their degree of regular exposure to VH languages (Omane et al., 2024), suggesting that minimal exposure to a VH might be sufficient for maintaining the VH bias. This pattern of results leads to the question of whether VH has a function in language acquisition and processing.

Research on both adults and infants has found that VH affects speech segmentation, facilitating listeners' detection of potential word boundaries in continuous speech. Specifically, it was found that speakers of languages with VH, but not those of languages without, use the absence of harmony between adjacent syllables to infer the location of word boundaries (e.g., Kabak et al., 2010; Suomi et al., 1997; Vroomen et al., 1998), suggesting that the use of VH is a language-specific cue for speech segmentation. Moreover, two previous studies show that young infants already exploit VH as a segmentation cue (Mintz et al., 2018; van Kampen et al., 2008). Van Kampen et al. (2008) investigated backness harmony in 9-month-old infants learning Turkish, a VH language. In the head-turn preference paradigm, infants were familiarized with two text passages containing three-syllable nonwords consisting of a bisyllabic stem and a monosyllabic nonsense prefix. In one text passage, the vowel of the prefix harmonized with the vowels of the bisyllabic stem (e.g., *nu-namoll*), and in the other, it did not (e.g., *lo-netiss*). At test, each infant heard isolated bisyllabic stems that were familiarized with a harmonizing prefix (e.g., *namoll*), familiarized with a non-harmonizing prefix (e.g., *netiss*) or novel (e.g., *batull*). Infants showed longer looking times for the familiarized stems that had been presented with a non-harmonizing prefix than those that were presented with a

harmonizing prefix. This led the authors to conclude that infants inferred a word boundary within the three-syllabic string after the non-harmonizing prefix, indicating that infants use the absence of VH between adjacent syllables as a speech segmentation cue. Furthermore, Mintz et al. (2018; experiment 2c) explored whether infants learning English, a language without VH, would rely on VH as a universal cue for speech segmentation using an artificial language learning paradigm. Seven-month-old infants were briefly exposed (for less than a minute) to a continuous speech stream (e.g., *ditepubobidetupo*). After this, at test, when presented with bisyllabic sequences that had been present in the speech stream, infants preferred listening to sequences containing exclusively front or back vowels (e.g., *dite, pubo*) over sequences that combined front and back vowels (e.g., *detu, bodi*). Results from these prior studies suggest that infants are sensitive to VH cues and can use them for segmenting words even after minimal exposure to speech containing vowel harmony patterns. These results give rise to the hypothesis that multilingual infants with some exposure to an ATR harmony language should also be able to use ATR harmony cues for segmentation.

Prior research on bilingual infants' speech segmentation with cues other than phonotactic cues (e.g., stress) suggests that bilingual infants might segment speech in each of their two languages similarly to monolingual infants. For example, while monolingual French- and English-learning infants can only segment words in their native language (i.e., French-learning infants relying on iambic stress, and English-learning infants on trochaic stress as segmentation cues) but not from a non-native language that has a non-native stress pattern, French-English bilingual infants can segment bisyllabic words from both their languages (Polka & Sundara, 2003; Polka et al., 2017). Likewise, in another study with French-English bilingual infants, Orena and Polka (2019) tested infants in an inter-mixed dual language task. These infants heard text passages that mixed French and English, with the language changing after every two sentences within a passage and with French target words embedded in French sentences and English target words embedded in English sentences. The result was that bilingual infants segmented bisyllabic words in both languages. Segmentation of monosyllabic words has also been found in Spanish-Catalan (Bosch et al., 2013) and English-Mandarin (Singh, 2018; Singh & Foong, 2012) bilingual infants. Unlike segmentation, in other domains of perception, e.g., in perceptual preferences and speech discrimination, bilingual infants are often found to perform better in the language they are exposed to more (e.g., Liu & Kager, 2015; Sebastián-Gallés & Bosch, 2002).

While these findings suggest robust segmentation in bilinguals' two languages using language-specific cues, it remains an open question whether multilingual infants rely on *phonotactic* cues for speech segmentation. Moreover, it is not yet known whether segmentation ability is or is not impacted by the degree of exposure to the language(s) with this phonotactic cue. While adults speaking VH, but not adults speaking non-VH languages, use VH cues for segmenting artificial languages (e.g., Kabak et al., 2010; Suomi et al., 1997), infant research suggests that infants may be able to use VH for segmentation even when exposed to a language without this cue (Mintz et al., 2018; Van Kampen et al., 2008). However, although Mintz et al.'s (2018) findings suggest that even infants without any exposure to VH can use this cue for segmentation, their study used a very simple artificial language, possibly overestimating the use of vowel harmony in children with no exposure to it. The study by Van Kampen et al. (2008) provides some further indication that infants' use of VH for segmentation from natural speech is not dependent on exclusive exposure to VH languages: the Turkish-learning infants in that study were living in Germany with some exposure to German (a non-VH language).

Yet, it is still not clear how infants with different levels of simultaneous exposure to languages with VH alongside those without VH will use this cue for speech segmentation. Answering this question will help us understand how multilingual exposure impacts phonotactic processing of the ambient language, and, more precisely, how multilingual exposure may affect their segmentation ability with cues that are only relevant to a subset of their languages.

2. The present study

For the first time, the current study investigated whether multilingually raised infants learning minimally one language with and one without VH can use VH cues when dealing with the task of detecting word boundaries in fluent speech. We focused on a specific type of VH, the Advanced Tongue Root (ATR) harmony in Akan, which statistically is the most common type of VH in the world given its occurrence in genetically diverse languages (Rose, 2018). The acquisition of ATR harmony has only been investigated in our recent study on 6-month-old infants in Ghana, finding sensitivity to ATR harmony irrespective of the degree of exposure to ATR harmony languages (Omane et al., 2024). Studying multilingual infants exposed to an ATR harmony language and non-vowel harmony languages will contribute to our understanding of how the degree of exposure to multiple diverse languages affects infants' ability to use VH, specifically ATR harmony, for segmentation. Research questions, hypotheses and methods, including the analysis plan, were pre-registered (see <https://osf.io/pqk2n>).

The current study aimed to address two questions. The first question was whether multilingual infants exposed to minimally one ATR harmony language and one non-VH language would rely on ATR harmony cues for speech segmentation. To address this, multilingually raised infants performed a word segmentation task with naturally recorded passages in Akan (familiarization) followed by isolated bisyllables (test), which had been embedded in the passages as target words. Within the passages, target words occurred in contexts manipulated for having either a harmony context or a disharmony context, where vowels in the target word either harmonize or disharmonize in ATR with vowels of an attached modifier. We hypothesized that multilingually raised infants exposed to an ATR harmony language (e.g., Akan) will rely on ATR harmony cues for segmentation. Two predictions are possible: (1) If infants segment words based on VH cues, it will be most likely that they display a familiarity preference, i.e., that they listen longer to familiarized target words that appeared in the *disharmony* context passages than to both novel words and words that appeared in the *harmony* context passages. However, infants may also show a novelty preference by listening longer to novel words and to familiarized target words that appeared in the harmony context passage than to words that occurred in a disharmony context passage. Both outcomes would suggest infants' use of VH for segmentation.¹

¹We pre-registered the alternative (below); however, upon reflection, based on Reviewer comments and Jusczyk et al. (1999) study, we no longer consider this an equivalent alternative. Alternatively (though this is not what we predict in light of our hypothesis), (2) infants might segment words irrespective of the harmony cue. In this case, infants might show a familiarity preference, i.e., longer listening time to familiarized target words than to novel words, but we might not observe a significant difference between words familiarized in the harmonic versus disharmonic context. Likewise, infants might also show a novelty preference, i.e., longer listening times to novel words than to familiarized target words, again with no significant difference between words familiarized in the harmonic versus disharmonic context. Both outcomes would suggest infants' ability to segment words from Akan without evidence for using VH cues for segmentation.

The second question was whether the relative amount of exposure to (an) ATR harmony language(s) would modulate the use of harmony cues in speech segmentation. There are reasons to hypothesize that (1) exposure to an ATR harmony language might modulate the use of harmony cues for segmentation but also reasons to hypothesize that (2) exposure to an ATR harmony language might not modulate the use of harmony cues for segmentation. If exposure modulates the use of harmony cues in segmentation, the difference in looking times to target words familiarized in a disharmony context versus those familiarized in a harmony context will be more enhanced the more infants are exposed to (an) ATR harmony language(s).

3. Methods

3.1. Participants

The final sample consisted of 40 full-term (i.e., 37 weeks or more gestation) infants aged between 9 and 11 months (22 boys; *Mean age*: 306 days; *range* = 274–363) being raised multilingually in Accra, Ghana. Infants were recruited from two hospitals (the University of Ghana Hospital, Legon, and a hospital that prefers to remain anonymous) and through personal networks (snowball sampling) in Accra. Previous speech segmentation studies have commonly tested between 14 to 32 infants per group (e.g., Gonzalez-Gomez & Nazzi, 2013; Jusczyk et al., 1999; Orena & Polka, 2019; van Kampen et al., 2008). Since we expected heterogeneity regarding language experience in our sample, we aimed for a larger sample. Consistent with our a-priori estimation, we successfully tested 40 infants during the 6 months of fieldwork available for data collection. An additional six infants were tested, but their data were excluded from the analyses presented here because of parent interference during the experiment ($n = 2$), the child crying ($n = 1$), experimenter errors ($n = 1$) and technical failure – the experiment ending abruptly after a few test trials ($n = 2$). No infant was reported to have any hearing or vision problems. As an inclusion criterion, all infants were learning Akan (an ATR harmony language) and minimally one non-VH language (e.g., Ewe, Ga, Ghanaian English). Infants' language exposure was assessed using the Caregiver Interview about Multilingual Exposure (CIME) tool, an interview assessment protocol, and a logbook (see Omame et al., 2025; the same tools were used in Omame et al., 2024). Infants were exposed to between two and four languages based on the interview assessment (two languages: 30%, three languages: 55% and four languages: 15%), and two to five languages according to the logbook (two languages: 32.5%, three languages: 50%, four languages: 12.5% and five languages: 5%). An overview of the language combinations infants experienced is provided in the [Supplementary Materials, Table S1](#). Participants were compensated with a fee. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study was approved by the Ethics Committee for Humanities at the University of Ghana (Ghana) under the project name 'Multilingual infants' use of vowel harmony for speech perception in Akan' (approval number: ECH 226/ 21–22).

3.2. Auditory stimuli

3.2.1. Nonce target words

We created four harmonic nonce target words (henceforth: target words) that infants needed to segment from spoken text passages. These target words were bisyllabic $C_1V_1C_2V_2$ (C = Consonant, V = Vowel) made up of Akan phonemes (see Dolphyne, 1988). Four vowels were selected from the vowel inventory of Akan, which

Table 1. The Set of +ATR and –ATR vowels used in creating the stimuli with their standard phonetic description

+ ATR Vowels	Standard phonetic description	–ATR Vowels	Standard phonetic description
i	High front unrounded	ɛ	Mid front unrounded
u	High back rounded vowel	ɔ	Mid back rounded

Note: The phonetic descriptions are based on the International Phonetic Alphabet (IPA).

comprised two advanced tongue root vowels (+ATR, one front and one back) and two unadvanced tongue root vowels (–ATR, one front and one back) as given in Table 1.

From the selected vowels, four two-vowel templates were created. For each template, vowels shared the same ATR feature: either both vowels are +ATR (/i_u, _u_i /) or –ATR (/ɛ_ɔ, _ɔ_ɛ/). Vowels in each template also shared vowel height (e.g., /i_u/ are both high vowels; /ɔ_ɛ/ are both mid-vowels). Four stop consonants /p, b, t, d/ from the consonant inventory of Akan were selected to create the stimuli. The consonants were combined with the vowel templates to create the bisyllabic C₁V₁C₂V₂ target words: *tupi*, *dibu*, *dɔbe* and *teɔɔ*. Two constraints, based on voicing and place of articulation, determined where consonants were placed: (1) The initial consonant was always a coronal (/t/ or /d/), and the second was a bilabial stop (/p/ or /b/). (2) Both consonants in each template shared the same voicing feature (/t_p_/ or /d_b_/). These restrictions on the consonants in the stimuli ensured a constant degree of complexity across the conditions.

3.3. Nonce modifiers

To create the harmonic and disharmonic sequences in which the target words would occur, we followed Van Kampen *et al.* (2008) in using four monosyllabic nonce modifiers (henceforth: modifiers) that do not exist in the target language, Akan: *ke*, *kɛ*, *ge* and *gɛ*, which either have a + ATR /e/ or –ATR /ɛ/ mid vowel. We chose to attach the monosyllables after the target words instead of before based on a grammatical property of Akan: we wanted to assign target words to noun positions and be treated as a noun, and the monosyllables to a post-nominal modifier (e.g., adjectives) position, where they must agree in ATR with the preceding vowel.

3.3.1. Target word and modifier combinations

With the bisyllabic target words and the monosyllabic modifiers, we formed three-syllable combinations (i.e., CVCV+CV). For each target word, two different modifiers were attached to manipulate harmonicity: one modifier shares ATR with the vowels in the target word (e.g., *tupi-ke*) while the other does not (e.g., *tupi-kɛ*). Harmonic and disharmonic combinations of target word plus modifier were controlled for changes in vowel height and consonant voicing: The –ATR target words always shared their vowel height with the modifiers (all mid-vowels) but not consonant voicing (e.g., *teɔɔ-ge* or *dɔbe-kɛ*), irrespective of harmony context. Conversely, the +ATR target words share consonant voicing (e.g., *tupi-ke*) but not vowel height with the modifiers, regardless of harmony context (see Table 2 for the complete list).

With this, infants were first familiarized with trisyllabic affixed forms (stem + modifier, CVCV+CV), inserted into text passage templates described in the paragraph below and tested on their recognition of the bisyllabic stem, similar to van Kampen *et al.*'s (2008) design (see also Kim & Sundara, 2015). The noun-modifier

Table 2. Bisyllabic target words plus modifiers (italicized) showing the resulting harmonic and disharmonic three-syllable combinations

	+ATR target word plus modifier CVCVCV	–ATR target word plus modifier CVCVCV
Harmonic	<i>tupike</i>	<i>tɛpɔgɛ</i>
	<i>dibuge</i>	<i>dɔbɛkɛ</i>
Disharmonic	<i>tupigɛ</i>	<i>tɛpɔke</i>
	<i>dibukɛ</i>	<i>dɔbɛgɛ</i>

structure requires that infants use backward cues to segment speech streams into word-sized units, which previous studies suggest they are able to do (Thiessen *et al.*, 2019). Consider below the disharmonic and harmonic context examples in (1) and (2), respectively.

- (1) *tupi-ke tiee maame no* 'tupi kɛ listened to the woman'
- (2) *tupi-ke tiee maame no* 'tupi ke listened to the woman'

In (1), the modifier 'kɛ' is not harmonic with the preceding target word *tupi*. So, if infants rely on backward cues from vowel harmony for segmentation, they should assume a word boundary before the modifier (i.e., after the target word) since the ATR feature of the modifier vowel (–ATR) does not spread regressively, indicating that the syllables cannot belong to a single word (e.g., 'tupi kɛ' would be parsed as 'kɛ' modifying 'tupi'). In (2) by contrast, the modifier harmonizes with the preceding target word, suggesting that the three syllables could belong to a single trisyllabic word (e.g., 'tupike').

3.3.2. Text passages

Two different six-sentence text passage templates, 'A' and 'B,' were created (see Supplementary Materials 2, Table S2). Each sentence had one slot for a target word plus modifier combination: three in the sentence-initial and three in the non-initial position. The structure of the sentences in the passage templates was such that no target word plus modifier combination appeared at the end of a sentence (i.e., within the verb phrase) to avoid regressive spreading of the +ATR harmony feature to nearby –ATR syllables or words, as regressive +ATR vowel harmony across word boundaries within verb phrases is a phonological rule in Akan (Dolphyne, 1988; Kügler, 2015). The ATR features of syllables preceding and following the slots were balanced across the two passage templates. Firstly, the slots in both passage templates were followed by +ATR (*di*, *wei*, *tie*) and –ATR (*maa*, *furɔɔ*, *pira*) words three times each. Secondly, the slots in non-initial position were preceded once by +ATR (*wei*) and twice by –ATR (*biaa*, *ɛnɔra*) in passage template A and twice by +ATR (*afei*, *nti*) and once by –ATR (*ɔkyina*) in passage template B. Each passage template was assigned one +ATR and one –ATR nonword. Specifically, *tupi* and *teɔɔ* were assigned to passage template A, and *dibu* and *dɔbe* to passage template B. Both the harmonic and disharmonic trisyllabic sequences created from each nonword were assigned to the same passage template. This ensured that every passage template was assigned an equal number of harmonic and disharmonic combinations (see Table 3 below). Consequently, each infant encountered harmonic and disharmonic contexts in different passage templates. The assignments of harmonic and disharmonic context passages were counterbalanced across infants. For example, in the illustration in Table 3, infant X receives one harmony context passage from template A and one disharmony context passage from template B, while infant Y, in

Table 3. Summary of how target words plus their modifiers were assigned to the passage templates showing harmonic and disharmonic combinations for each passage, and (between parentheses) illustration of counterbalanced assignment to infants

	Passage template A		Passage template B	
Harmonic	tupi ke (Infant W)	teɔ ge (Infant Y)	dibu ge (Infant Z)	dɔbe ke (Infant X)
Disharmonic	tupi ge (Infant X)	teɔ ke (Infant Z)	dibu ke (Infant Y)	dɔbe ge (Infant W)

contrast, receives a disharmony context passage from template A and a harmony context passage from template B. From here on, we use the term ‘passage(s)’ to refer to passage templates with embedded target word plus modifier combinations.

A female native speaker of Akan recorded all stimuli. All target words were produced with a consistent High–Low (H–L) tone pattern. The isolated CVCV test items were elicited in different sentence types: (1) interrogative: *wonim tupi?* ‘do you know tupi?’ (2) declarative: *yefre no tupi!* ‘s/he is called tupi!’ (3) imperative: *Fa tupi!* ‘take tupi!’. This was meant to facilitate the realization of acoustic variability across the tokens of each target word. The speaker was instructed to read all sentence types and passages in an infant-directed manner, with no pauses between words or at phrase boundaries. Thirty different recordings (tokens) of each target word (10 in each sentence type) and four renditions of each passage were recorded. Further processing of the stimuli, such as annotation and extraction of sentences for the passages and test items, and checking the pitch height of extracted target test items, was done in Praat (version 6.2.06, [Boersma & Weenink, 2022]).

To create the *familiarization stimuli (passages)*, we selected one recorded rendition of each sentence of the eight six-sentence familiarization passages. For each passage, the sentences were re-concatenated with 500 milliseconds of silence between them. The average duration of a passage was about 13.2 seconds (range: 13.05–13.30 s).

For the *test stimuli (isolated words)*, 15 tokens of each target word were selected and concatenated into a single stream trial with 650 milliseconds of silence between the tokens. In each experimental trial, the infant heard 15 different tokens of a single target word sequence (e.g., tupi). The average duration for a test trial was 15.6 seconds (range: 15.33–15.88 s).

3.4. Visual stimuli

There were two visual stimuli: a rotating colorful wheel and a static checkerboard. The rotating wheel was used as an attention-getter before each trial was presented, while the checkerboard served as an unrelated visual attractor when the auditory stimulus was being played.

4. Procedure

4.1. Experiment

Infants were tested individually in a mobile lab setup in one of the three locations: a room at the Department of Linguistics at the University of Ghana or one of two hospitals. We used the single-

screen central fixation paradigm (Cooper & Aslin, 1990). To minimize the influence of the experimenter’s presence in the room and activities such as coding and switching trials during testing, the rooms were partitioned into two with a curtain. The infant and the caregiver sat on one side of the partitioned room. The infant sat on the caregiver’s lap in front of a 17.3-inch Asus ROG Strix XG17 screen about 40–50 cm away. Two Logitech z120 pc speakers through which the auditory stimuli were presented were placed at the two bottom edges of the screen. A webcam was placed on top of the screen to monitor the infant during the experiment. The experimenter sat on the other side of the partitioned room to observe and record the infant’s looking behavior via the webcam. The experimenter manually coded the infants’ looking times during the experiment. The experiment was run by the Habit2 program, version 2.2.1 (Oakes et al., 2019).

At the onset of each trial, the visual attention-getter appeared at the center of the screen to attract the infant’s attention. Once the infant’s attention was captured, the experimenter initiated the auditory trial, which was accompanied by the visually presented checkerboard at the center of the screen.

4.1.1. Familiarization phase

Half of the infants were familiarized with text passages containing the target words *teɔ* and *dibu*, while the other half were familiarized with passages containing *tupi* and *dɔbe* (see text passages with embedded target word plus modifier combinations in the [Supplementary Materials, Table S2](#)). Each infant was familiarized with one harmonic and one disharmonic combination of the target word plus modifier, embedded in the passage templates. Each infant heard either template A passage with a harmonic target word plus a modifier combination or template B passage with a disharmonic combination or vice versa, as illustrated in [Table 3](#). The order of presentation of passages was pseudorandomized across infants. Each passage was repeated two times; hence, every infant heard four passages in total.

4.1.2. Test phase

The auditory trials were presented simultaneously with an unrelated visual stimulus (static checkerboard). Each infant listened to the two target words they heard during familiarization and one novel target word. Novel target words were randomly assigned to infants. Note that a novel target word for one infant was a familiarized word for another. Each target word was presented four times, making a total of 12 trials. A trial consisted of 15 repeated isolated tokens of the same target word. Each infant heard a maximum of 12 test trials, and fewer if the infant became fussy and could not complete all trials. The average duration for a test trial was about 15.6 seconds. The trial ended after the infant looked away for more than 2 consecutive seconds. The infants’ listening time was measured by the length of time that they looked at the screen when an auditory stimulus was being played. Thus, infants’ looking time was used as a proxy for how long they listened to the auditory stimulus.

4.2. Language exposure assessment

4.2.1. Caregiver Interview about Multilingual Exposure (CIME)

Caregivers completed the CIME before the experiment to determine whether infants met the inclusion criteria of gestational age (37 or more weeks of gestation), were learning Akan (an ATR harmony language) and minimally one non-VH language (e.g., Ewe, Ga, Ghanaian English; the same intake criteria as in

Omane et al., 2024). The CIME asked about the languages directed to the infant by a range of speakers (e.g., father, mother, grandparents, older sibling[s] and other caregivers) as well as those that the infant generally overhears (e.g., at home and religious gatherings). Exposure to each language was calculated as a proportion of time an infant heard a given language over the total time of exposure to all languages (collapsing over both speech directed to the infant and overheard speech). Similar calculations were performed to arrive at ATR-harmony exposure (combining all languages in each infant's input with ATR harmony, such as Akan and Dagbani; see [Supplementary Table S1](#)).

4.2.2. Logbook

In addition to the CIME, a logbook method was used to estimate language exposure throughout an infant's entire day, as in Omane et al. (2024, 2025). Caregivers were given a logbook in the form of a small card placed in the pocket of a vest worn by the baby at home after the experiment to complete at home. They were instructed to indicate, by ticking different options, the different languages their infant heard every half an hour for an entire day between 7 am and 7 pm while the baby was awake. A demonstration accompanied the instruction to ensure that caregivers understood the task. Caregivers typically filled out the logbook a day or two after the experiment. The experimenter picked up the logbook from the majority of the caregivers' homes after completion (which usually took around 2 hours due to traffic), with only a few caregivers agreeing to return the logbook to the hospital where the experiment was conducted. Language and ATR-harmony exposure were similarly calculated from the logbook data.

5. Data preprocessing and analyses

5.1. Data exclusion

Following our pre-registration and previous studies (e.g., Bosch et al., 2013; Junge et al., 2020), we excluded 11 test trials (2.3% of total trials) across eight infants from the analyses, as total looking time at these trials was less than 1 second. Six of those trials were with familiarized words (disharmonic context: $n = 2$, harmonic context: $n = 4$), and five with novel words.

5.2. Statistical analysis

All analyses were conducted using R (R Core Team, 2023). Looking times (accumulated looks on screen in seconds per trial) were analyzed with linear mixed-effect models (using the lme4 package v1.1–26; Bates et al., 2015). Normality tests indicated that the raw looking times were more normally distributed than the log-transformed looking times. Hence, raw looking times in seconds were used as the dependent measure.

For statistical analysis, we pre-registered a model that included one categorical predictor variable and three continuous predictor variables. The categorical predictor was *condition* (within-participant, three levels: familiarized disharmonic, familiarized harmonic and novel words). The levels were coded using successive difference contrast coding, an orthogonal contrast (using the MASS package, version 7.3–54, Venables & Ripley, 2002), implemented as follows: familiarized harmonic ($-2/3, -1/3$), familiarized disharmonic ($1/3, -1/3$) and novel ($1/3, 2/3$). As such, the output would present the estimated difference between familiarized disharmonic versus harmonic ($1/3, -2/3$) and novel words versus familiarized disharmonic ($2/3, -1/3$). The continuous predictors were ATR-harmony *exposure* (i.e., amount of exposure to all ATR harmony languages in a % of exposure to all languages, centered), *age* in days (centered) and *trial number* (1–12, within-participant factor, centered; a more fine-grained alternative to the predictor *block* to absorb more variance). The random structure included by-participants and by-target word random intercepts and random slopes by both participant and target word for condition, trial number and their interaction. The R formula for the pre-registered model was as follows:

- (1) Looking time \sim (condition) \times (exposure + age + trial number) + (1 + condition * trial number|participant) + (1 + condition \times trial number|target word)

It was pre-registered that we would obtain two ATR-harmony exposure variables: one from the CIME and one from the logbook. However, this would not be possible if they were correlated. An assessment of both measures showed that they were correlated ($r(38) = .43, p = .005$, see [Figure 1](#)).

To reduce collinearity, we performed a Principal Component Analysis (PCA) on the two measures (as in Omane et al., 2024) and

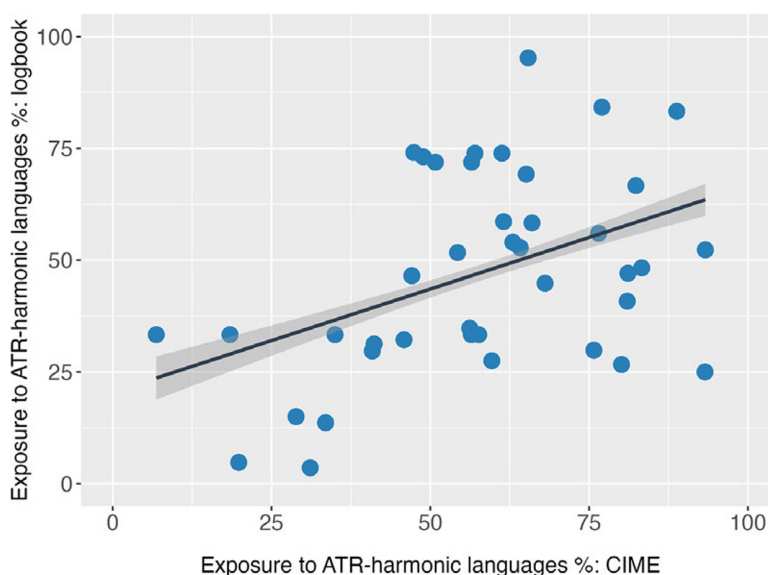


Figure 1. Correlation of ATR harmony exposure measures obtained from logbook versus the CIME.

used the first Principal Component (PC) as the single predictor *exposure* in the statistical analyses. Subsequently, we worked with the pre-registered model above.

The pre-registered model did not converge because of its complexity; hence, the random-effects structure was reduced. Correlations between random effects were either zero or approximately 1; hence, they were removed. The following random slopes had variances (close to) zero and were thus removed from the model: the by-participant random slopes for condition, trial and their interaction, as well as the by-target word random slopes for condition. Furthermore, the fixed-effect component of the model was simplified by removing the condition \times age and condition \times trial number interactions, as including them in the model resulted in a singularity warning. Moreover, these interactions were not directly related to our hypotheses but were included to absorb variance in the data. After this, the model converged with the formula presented in (2):

- (2) Looking time \sim condition \times exposure + age + trial number + (1 | participant) + (1 + trial number ||| target word)

As pre-registered, we continued with model comparisons using the Likelihood Ratio Test to see if the model could be simplified further. This revealed that the model fit was not significantly improved by including either a trial number slope by target word or a random intercept by target word (all p 's = 0.4; see the output of model comparisons in [Supplementary Materials S3a](#) and [S3b](#), respectively). The fixed effect trial number improved the model fit ($p < .001$; see the output of the model comparison in [Supplementary Material S3c](#)) and was, hence, retained. We always retained the fixed effects of condition and exposure, as they were essential for hypothesis testing. The final formula was as follows:

- (3) Looking Time \sim condition \times exposure + age + trial number + (1 | participant)

6. Results

For the descriptive analysis, infants' exposure (in %) to all ATR harmony languages ranged from 6.9% to 93.3% as estimated by the CIME (mean = 57.3%, $SD = 20.1\%$), and from 3.6% to 95.2% as estimated by the logbook (mean = 46.1%, $SD = 23.0\%$). Exposure (in %) to only Akan also ranged from 6.9% to 93.3% as estimated by the CIME (mean = 53.78%, $SD = 20.7\%$), and from 3.6% to 95.2% as

estimated by the logbook (mean = 40.8, $SD = 21.9$). Mean looking times to the familiarized target words from the harmonic and disharmonic contexts and the novel target words were computed for each infant. On average, infants' looking time was highest for familiarized words from the disharmonic context (mean = 8.83 s, $SD = 2.38$), followed by those that appeared in the harmonic context (mean = 7.76 s, $SD = 2.30$) and then novel words (mean = 7.49 s, $SD = 1.99$) (see [Figure 2](#)). Twenty-four (60%) infants had the longest average looking times to familiarized words from a disharmonic context, ten (25%) to familiarized words from a harmonic context and six (15%) to novel words.

The linear mixed-effect model (see [Table 4](#) for results on all fixed-effects parameters) revealed a marginally significant effect of the contrast between familiarized words from disharmonic and harmonic contexts (coefficient: disharmonic versus harmonic, $\beta = 0.71$, $p = .05$, 95% CI $[-0.009, 1.433]$), with infants looking longer to target words from the disharmonic than the harmonic context. Moreover, the contrast between novel words and familiarized words from the disharmonic context (coefficient: novel versus disharmonic) was significant ($\beta = -0.74$, $p < .05$, 95% CI $[-1.472, -0.006]$), with infants looking longer to words from the disharmonic context than novel words. There was also an effect of trial number ($\beta = -0.35$, $p < .001$, 95% CI $[-0.435, -0.261]$), indicating that infants' looking time to trials decreased over the course of the experiment. The other main effects (i.e., exposure and age) were not significant and no significant interaction of condition and exposure was observed (all p 's $> .05$).

In order to assess the difference in looking time to familiarized words from the harmonic context and novel words, we ran an exploratory model ([Table 5](#)) with the same specifications as the planned and fitted model (see model formula 3), except for the following re-coding of levels in the *condition* contrast: familiarized harmonic versus novel (1/3, $-2/3$) and familiarized disharmonic versus familiarized harmonic (2/3, $-1/3$). The model showed no significant effect of the contrast between familiarized words from the harmonic context and novel words (coefficient: harmonic versus novel, $\beta = 0.03$, $p = .94$, 95% CI $[-0.70, 0.75]$).

7. Discussion

The goal of the present study was to investigate whether multilingual infants learning (an) ATR harmony language(s) alongside

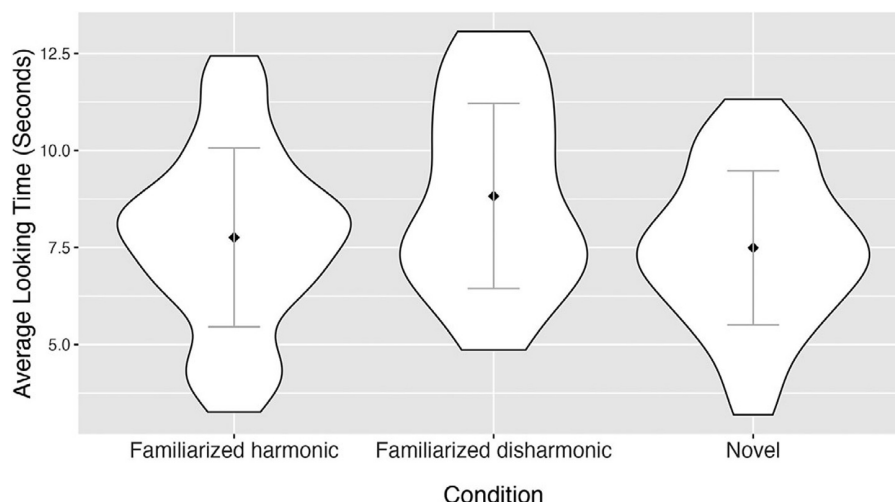


Figure 2. Averages of individual infants' mean looking times (s) to familiarized words from harmonic and disharmonic context passages and novel words. Dots indicate the mean for each condition.

Table 4. Final model output showing parameters of the linear mixed-effect model

Fixed effects	β	SE	t	p	95% CI
(Intercept)	8.04	0.30	27.00	< .001	7.46, 8.62
Disharmonic versus harmonic	0.71	0.37	1.93	.05	−0.01, 1.44
Novel versus disharmonic	−0.74	0.38	−1.97	< .05	−1.48, −0.01
Exposure	0.29	0.25	1.16	.25	−0.30, 0.65
Age	−0.00	0.01	−0.24	.82	−0.02, 0.02
Trial number	−0.35	0.04	−7.81	<.001	−0.43, −0.26
Exposure × disharmonic versus harmonic	0.24	0.31	0.78	.44	−0.49, 0.70
Exposure × novel versus disharmonic	−0.26	0.31	−0.83	.41	−0.55, 0.64

Table 5. Exploratory model output showing parameters of the linear mixed-effect model

Fixed effects	β	SE	t	p	95% CI
(Intercept)	8.04	0.30	27.00	< .001	7.47, 8.62
Harmonic versus novel	0.03	0.37	0.07	.94	−0.70, 0.75
Disharmonic versus harmonic	0.71	0.37	1.93	.05	−0.01, 1.44
Exposure	0.29	0.25	1.16	.25	−0.19, 0.77
Age	−0.00	0.01	−0.22	.82	−0.02, 0.02
Trial number	−0.35	0.04	−7.81	<.001	−0.44, −0.26
Harmonic versus novel × exposure	0.02	0.31	0.06	.95	−0.59, 0.62
Disharmonic versus harmonic × exposure	0.24	0.31	0.78	.44	−0.36, 0.84

non-VH languages segment bisyllables from continuous speech using ATR harmony cues. Nine- to 11-month-old infants learning Akan (an ATR harmony language) in addition to at least one non-VH language were tested in Ghana. Using the central fixation procedure (Cooper & Aslin, 1990), infants were first familiarized with text passages that included trisyllabic sequences, combinations of a bisyllabic target word plus a monosyllabic modifier, which were either harmonic, providing no segmentation cue, or disharmonic, providing a regressive ATR harmony segmentation cue between the target word and the modifier. In a subsequent test phase, infants' looking time to novel words, as well as familiarized bisyllabic words that had occurred in either harmonic or disharmonic context, was recorded. An interview assessment and a diary logbook were used to assess infants' relative exposure to (an) ATR harmony and non-VH language(s). As hypothesized, our findings show that multilingual infants better segment bisyllables from speech when word boundaries are marked by a change in ATR (i.e., disharmony) than when the trisyllabic sequences form a harmonic unit (e.g., *tupike*). The current study thus provides the first evidence for segmentation with ATR harmony cues and in multilingual infants learning (an) ATR harmony language(s) alongside languages without VH. However, we did not detect an effect of ATR harmony language exposure on infants' segmentation with ATR harmony cues. We discuss these findings, study limitations and future research considerations below.

Previous artificial language studies found adult speakers of VH languages, but not speakers of non-VH languages, to use VH cues for segmentation (Kabak *et al.*, 2010; Suomi *et al.*, 1997; Vroomen *et al.*, 1998), suggesting that the use of the cue is language-specific in adulthood. Moreover, segmentation with VH cues has previously been reported for infants learning Turkish, a VH language (with

backness harmony: van Kampen *et al.*, 2008), and infants learning English, a language without VH (Mintz *et al.*, 2018, Experiment 2c).

Our findings extend evidence of infants' speech segmentation with VH cues to ATR harmony and to a genetically unrelated language (Akan). By this, the present study adds to the existing literature on the cues infants cross-linguistically use for segmentation (Höhle & Weissenborn, 2003; Jusczyk *et al.*, 1999; Saffran *et al.*, 1996; van Kampen *et al.*, 2008). While evidence from three studies now converges on the conclusion that VH is a robust segmentation cue for infants, the question about its language-specificity or potentially universal nature remains open. If this study had shown that exposure to ATR harmony relates to infants' use of VH, this would have suggested language-specificity or learning (we provide further discussion on the lack of an effect of VH language exposure later in this section). In the absence of evidence for (or against) this effect, we can speculate that infants are universally sensitive to VH patterns, supporting the results from Mintz *et al.* (2018) with an artificial language, as well as a recent study finding auditory preferences for VH in 4-month-old infants learning non-VH languages (Solá-Llonch & Sundara, 2025). Important to note, however, is that Mintz *et al.* (2018) used an artificial language entirely different from the text passages as those used in the present study and Van Kampen *et al.* (2008). To shed more light on the accessibility of VH for infants' language processing, a next step will be to investigate whether infants learning non-VH languages can use VH cues to segment words from natural language, with all its naturally occurring complexities.

The present results further demonstrate that infants prioritize the VH cue over stimulus-specific transitional probabilities (TPs) between syllables. In our experiment, infants heard trisyllabic sequences with a TP of 1 between all syllables in both the harmonic

(e.g., *tupi ke*) and the disharmonic condition (e.g., *dobe ke*). Our findings, however, suggest that infants segmented at the point of disharmony, before the modifier, thus breaking up units that are statistically coherent in terms of TPs between syllables. This indicates that they weighted the ATR harmony cues as more relevant than domain-general statistical cues to detect word units in their native language. This is consistent with previous studies showing that infants weight language-specific prosodic cues over stimulus-specific statistical cues for word segmentation (e.g., 6-month-old German-learning infants: Marimon et al., 2024; 9-month-old English-learning infants: Thiessen & Saffran, 2003, Experiment 1).

At present, the question remains open of whether infants can or when they will learn to segment words when vowel sequences are harmonic. We found no significant effect of the contrast between novel and familiarized words from the harmonic context, providing no evidence for (or against) infants' recognition of bisyllables from the harmony context, where the harmony cues, as well as TPs of 1 between syllables, are suggestive of a trisyllabic unit. Similarly, Van Kampen et al. (2008) found no evidence that 9-month-old Turkish-learning infants infer a word boundary within a trisyllabic sequence if the first syllable (or prefix) harmonizes with the subsequent bisyllabic string. Speculatively (as this is a null result), this may reflect infants parsing the trisyllabic unit as a single chunk. Akan- and Turkish-learning infants potentially treating trisyllabic units as single words in the absence of harmony cues to word boundaries resemble 7.5-month-old English-learning infants not recognizing subparts from syllable strings (e.g., not recognizing 'king' after familiarization with 'kingdom,' Jusczyk et al., 1999), and 11-month-old French-learning infants segmenting frequent sequences even when they span word boundaries (Ngon et al., 2013). These studies, taken together, thus suggest that infants fail to recognize subparts of words or other frequently occurring units. Future studies may want to explore at what age Akan-learning infants learn to segment words from harmonic contexts, as word boundaries can also occur between adjacent harmonic words in Akan. This research should explore the development of word segmentation in the absence of harmony cues and further explore which other segmentation cues, such as top-down lexical and low-level phonetic cues, may then be used by infants growing up in Ghana.

Regarding the direction of the listening preferences at test, infants demonstrated a longer looking time to familiarized bisyllables from a disharmonic context over those from a harmonic context or (importantly) novel bisyllables. This can be interpreted as a familiarity preference for words that were heard *and* segmented during the familiarization phase. This aligns with reported familiarity preferences in studies that tested the use of other segmentation cues, including phonotactics (VH: van Kampen et al., 2008; consonant clusters: Mattys & Jusczyk, 2001), allophonic cues (Jusczyk et al., 1999) and rhythmic units (stress pattern: Jusczyk & Aslin, 1995; syllable-unit: Bosch et al., 2013; Nazzi et al., 2006), in addition to meta-analytic evidence for a familiarity preference bias in segmentation tasks (Bergmann & Cristia, 2016).

The present study demonstrates that infants use VH cues to segment bisyllables from speech when the segmentation cue is at the end of the target bisyllable. This complements the previous evidence of infants' segmentation with VH cues marking the target word onset (van Kampen et al., 2008). That infants can track harmony cues for identifying word boundaries both at word onset and offset is akin to their processing of other phonotactic cues like consonant clusters in English (e.g., Mattys & Jusczyk, 2001). Adult speech segmentation studies have shown that speakers of VH languages can segment bisyllabic words when the VH cue marks the target word onset (Kabak et al., 2010; Suomi et al., 1997;

Vroomen et al., 1998) but one study with Finnish listeners did not find evidence for the use of VH cues when it marks the target word offset, suggesting a potential positional restriction on using VH information as a segmentation cue (Suomi et al., 1997). In our stimuli, target words were assigned to noun positions, with post-nominal modifiers providing the segmentation cue (at target word offset), as this structure is common in Akan. However, as vowels in prefixes must also harmonize with stem vowels in ATR in Akan, our expectation is that infants should successfully segment target words, too, if target words are preceded by a disharmonizing prefix (i.e., target word onset), as was found with Turkish-learning infants (van Kampen et al., 2008). Whether the position of harmony cues impacts word segmentation in infants learning Akan and other languages remains to be explored, to broaden our understanding of the language- and age-dependent tracking of directionality in fluent speech.

Our findings do not provide evidence for or against the effect of the quantity of exposure to (an) ATR harmony language(s) on infants' use of VH cues for segmentation. This prompts speculation that even limited exposure to an ATR harmony language might suffice for multilingual infants to learn how VH disagreements between adjacent syllables align with word boundaries and use this knowledge to segment words from fluent speech. The absence of a detected exposure effect is consistent with a previous study (Omame et al., 2024), where multilingual infants exposed to (an) ATR harmony language(s) alongside non-VH languages in Ghana demonstrated a familiarity preference for ATR harmonic syllable sequences in Akan, with no evidence for or against the effect of exposure either. The finding aligns with prior research on other domains of speech perception, showing that, for example, simultaneous German–French bilingual adults' chunking of continuous speech (Boll-Avetisyan et al., 2020) and German–French bilingual infants' preference for trochaic patterns was not modulated by language exposure either (Bijeljac-Babic et al., 2016). Furthermore, several prior segmentation studies have also not found evidence for (or against) the amount of input in a language impacting bilingual infants' ability to use native language cues for speech segmentation (Bosch et al., 2013; Orena & Polka, 2019; Polka et al., 2017). Taken together, the absence of a significant exposure effect in the current study leads us to speculate that any input multilingual infants receive in an ATR harmony language allows them to track harmony cues during speech processing and that further input in an ATR harmony language does not further enhance this ability. This suggests the development or maintenance of ATR-based segmentation may require minimal input to an ATR harmony language.

As argued in our previous study (Omame et al., 2024), the absence of an exposure effect could also relate to methodological shortcomings. Infants in our study were exposed to diverse languages (between two and five languages per child) with varying numbers of caregivers (between two and six). Consequently, caregivers may not have accurately reported language input for all caregivers, potentially accounting for our inability to find an exposure effect on segmentation. Future research may want to address this issue by comparing data obtained using our CIME and/or logbook with data from daylong recordings (see Orena et al., 2020, who found correlations between interview assessments and daylong recording data for infants growing up in French–English contexts). Alternatively, more reliable estimates of language exposure in a multilingual context like Ghana could be to survey or interview more than one caregiver per child, although such an approach may be impractical or difficult to realize.

The present study is relevant to multilingual speech segmentation as it presents the first evidence of segmentation abilities in

infants learning between two and five languages. It also provides the first evidence of segmentation with phonotactic cues in infants learning more than one language. In the context of previous research on bilingual infants' speech segmentation, our findings align with studies demonstrating segmentation ability in bilinguals' native languages (Bosch et al., 2013; Mateu & Sundara, 2022; Orena & Polka, 2019; Polka & Sundara, 2003; Polka et al., 2017; Singh & Foong, 2012; Singh, 2018). Unlike prior bilingual studies that tested segmentation in the infant's two native languages (but see Singh, 2018), our study focused on testing infants in only one of their native languages, as Akan, the ATR harmony language that we focused on, was the only language that all infants were exposed to. Our results suggest that multilingual infants can use a phonotactic cue of one of their languages for segmenting speech in that language. It remains to be tested in the future whether infants learning three or more languages have the processing ability to segment words in all their ambient languages, akin to infants with dual language input (e.g., Orena & Polka, 2019; Polka & Sundara, 2003; Polka et al., 2017).

In conclusion, this study investigated word segmentation with VH cues in 9–11-month-old multilingual infants learning languages with and without VH in Ghana. Our results demonstrate that infants use ATR harmony cues from one of their languages to segment bisyllabic words in that language, with no evidence for or against the effect of quantity of exposure to an ATR harmony language on infants' segmentation abilities. These findings prompt further research into multilingual language processing in infants raised in other multilingual contexts. Future studies are encouraged to examine word segmentation in other understudied African languages and multilingual populations to gain deeper insights into their segmentation abilities across all their languages and the cues infants use for such complex tasks.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S1366728925100485>.

Data availability statement. The data, study materials, CIME, logbook and analysis code will be made available in an online repository: [<https://osf.io/h9tvr>].

Acknowledgments. The study received approval from the Ethics Committee for Humanities (ECH) at the University of Ghana, Ghana. This research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317633480 – SFB 1287, Project C07 awarded to NB and a CTiMQRES IDEALAB scholarship from Macquarie University (allocation no. 20201820) awarded to PO. We thank Reginald Akuoko Duah for his numerous contributions, including allowing us to use his office as a lab. We thank the University of Ghana Hospital in Legon and an anonymous Hospital for allowing us to recruit and test infants on their premises. We are grateful to the nurses at the children's ward in both hospitals for their hospitality and for assisting us in the recruitment process. We are thankful to Rebecca Dufie Forson, our research assistant, and Frank Obeng for assisting us in infant recruitment and data collection in Ghana. We also thank Anna Laurinavichyute (SFB 1287, Q-project) for statistical advice. Finally, we thank all parents and infants who participated in the study.

Competing interests. The authors declare none.

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 B. Haznedar & F. N. Ketrez (Eds.), *The acquisition of Turkish in childhood* (pp. 29–56). Amsterdam: John Benjamins. https://www.bu.edu/buclid/files/2016/09/BUCLD_Proceedings_2016_01_25_altan1.pdf.

- Aravena-Bravo, P., Cristia, A., García, R., Nicolas, R. K., Laranjo, R., Arokoyo, B. E., Benavides-Varela, S., Benders, T., Boll-Avetisyan, N., Ben, R. D., Diop, Y., Durán-Urzuá, C., Manalili, M., Narasimhan, B., Omane, P. O., Kolberg, L. S., Ssemata, A. S., Styles, S. J., Cristia, A., ... Early, D. (2023). Towards diversifying early language development research: The first truly global international summer/winter school on language acquisition (/ L + /) 2021. *Journal of Cognition and Development*, 25(2), 242–260. <https://doi.org/10.1080/15248372.2023.2231083>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, 19(6), 901–917. <https://doi.org/10.1111/desc.12341>.
- Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2016). Early prosodic acquisition in bilingual infants: The case of the perceptual trochaic bias. *Frontiers in Psychology*, 7(7), 210. <https://doi.org/10.3389/fpsyg.2016.00210>.
- Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.06, retrieved 27 February 2022 from <http://www.praat.org/>
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2020). Rhythmic grouping biases in simultaneous bilinguals. *Bilingualism: Language and Cognition*, 23(5), 1070–1081. <https://doi.org/10.1017/S1366728920000140>.
- 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), 106. <https://doi.org/10.3389/fpsyg.2013.00106>.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61(5), 1584–1595. <https://doi.org/10.2307/1130766>.
- Cristia, A., Foushee, R., Aravena-Bravo, P., Cychosz, M., Scaff, C., & Casillas, M. (2023). Combining observational and experimental approaches to the development of language and communication in rural samples: Opportunities and challenges. *Journal of Child Language*, 50(3), 495–517. <https://doi.org/10.1017/S0305000922000617>.
- Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, 85(4), 1330–1345. <https://doi.org/10.1111/cdev.12193>.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33(6), 824–844. <https://doi.org/10.1006/jmla.1994.1039>.
- Dolphyne, F. A. (1988). *The Akan (Twi-Fante) language: Its sound system and tonal structure*. Ghana Universities Press.
- Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36(2), 202–225. <https://doi.org/10.1006/jmla.1996.2483>.
- Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, 54(3), 287–295. <https://doi.org/10.3758/bf03205263>.
- Gonzalez-Gomez, N., & Nazzi, T. (2013). Effects of prior phonotactic knowledge on infant word segmentation: The case of nonadjacent dependencies. *Journal of Speech, Language, and Hearing Research*, 56(3), 840–849. [https://doi.org/10.1044/1092-4388\(2012\)12-0138](https://doi.org/10.1044/1092-4388(2012)12-0138).
- Gonzalez-Gomez, N., Schmandt, S., Fazekas, J., Nazzi, T., & Gervain, J. (2019). Infants' sensitivity to nonadjacent vowel dependencies: The case of vowel harmony in Hungarian. *Journal of Experimental Child Psychology*, 178, 170–183. <https://doi.org/10.1016/j.jecp.2018.08.014>.
- Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, 6(2), 122–127. <https://doi.org/10.1111/1467-7687.00261>.
- Hohenberger, A., Altan, A., Kaya, U., Tuncer, O. K., & Avcu, E. (2016). Sensitivity of Turkish infants to vowel harmony: Preference shift from familiarity to novelty. In B. Haznedar & F. N. Ketrez (Eds.), *The acquisition of Turkish in childhood* (pp. 29–56). Amsterdam: John Benjamins. <https://doi.org/10.1075/tilar.20.02hon>.
- 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 [Conference session]. 41st Annual Boston University

- Conference on Language Development, Somerville, MA, United States. <https://www.lingref.com/buclid/41/BUCLD41-25.pdf>.
- Junge, C., Everaert, E., Porto, L., Fikkert, P., de Klerk, M., Keij, B., & Benders, T. (2020). Contrasting behavioral looking procedures: A case study on infant speech segmentation. *Infant Behavior and Development*, *60*, 101448. <https://doi.org/10.1016/j.infbeh.2020.101448>.
- Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, *15* (4), 463–473. <https://doi.org/10.1111/j.1467-7687.2012.1144.x>.
- Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. <https://doi.org/10.1006/cogp.1995.1010>.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. MIT Press.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. <https://doi.org/10.1006/cogp.1999.0716>.
- Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology*, *1*(1), 207–224. <https://doi.org/10.1515/labphon.2010.010>.
- Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. <https://doi.org/10.1177/01427237211066405>.
- Kim, Y. J., & Sundara, M. (2015). Segmentation of vowel-initial words is facilitated by function words. *Journal of Child Language*, *42*(4), 709–733. <https://doi.org/10.1017/S0305000914000269>.
- Kügler, F. (2015). Phonological phrasing and ATR vowel harmony in Akan. *Phonology*, *32*(1), 177–204. <https://doi.org/10.1017/S0952675715000081>.
- Liu, L., & Kager, R. (2015). Bilingual exposure influences infant VOT perception. *Infant Behavior and Development*, *38*, 27–36. <https://doi.org/10.1016/j.infbeh.2014.12.004>.
- Marimon, M., Langus, A., & Höhle, B. (2024). Prosody outweighs statistics in 6-month-old German-learning infants' speech segmentation. *Infancy: The Official Journal of the International Society on Infant Studies*, *29*(5), 750–770. <https://doi.org/10.1111/infa.12593>.
- Mateu, V., & Sundara, M. (2022). Spanish input accelerates bilingual infants' segmentation of English words. *Cognition*, *218*, 104936. <https://doi.org/10.1016/j.cognition.2021.104936>.
- Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91–121. [https://doi.org/10.1016/S0010-0277\(00\)00109-8](https://doi.org/10.1016/S0010-0277(00)00109-8).
- Mintz, T. H., Walker, R. L., Welday, A., & Kidd, C. (2018). Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition*, *171*, 95–107. <https://doi.org/10.1016/j.cognition.2017.10.020>.
- Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54* (3), 283–299. <https://doi.org/10.1016/j.jml.2005.10.004>.
- Nazzi, T., Iakimova, G., Bertoncini, J., Mottet, S., Serres, J., & de Schonen, S. (2008). Behavioral and electrophysiological exploration of early word segmentation in French: Distinguishing the syllabic and lexical levels. In A. Friederici & G. Thierry (Eds.), *Early language development: Bridging brain and behaviour* (pp. 65–89). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.5.05naz>.
- Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (non)words, (non)words: Evidence for a protollexicon during the first year of life. *Developmental Science*, *16*(1), 24–34. <https://doi.org/10.1111/j.1467-7687.2012.01189.x>.
- Oakes, L. M., Sperka, D., DeBolt, M. C., & Cantrell, L. M. (2019). Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behavior Research Methods*, *51*(5), 1943–1952. <https://doi.org/10.3758/s13428-019-01244-y>.
- Omane, P. O., Benders, T., & Boll-Avetisyan, N. (2024). Vowel harmony preferences in infants growing up in multilingual Ghana (Africa). *Developmental Psychology*, *60*(8), 1372–1383. <https://doi.org/10.1037/dev0001776>.
- Omane, P. O., Benders, T., & Boll-Avetisyan, N. (2025). Exploring the nature of multilingual input to infants in multiple caregiver families in an African city: The case of Accra (Ghana). *Cognitive Development*, *74*(101558), 101558. <https://doi.org/10.1016/j.cogdev.2025.101558>.
- Omane, P. O., Benders, T., Duah, R. A., & Boll-Avetisyan, N. (2023). Diversifying language acquisition research can be (partly) achieved in urban societies and with simplified methodologies: Insights from multilingual Ghana. *Journal of Child Language*, *50*(3), 532–536. <https://doi.org/10.1017/S0305000923000090>.
- Orena, A. J., Byers-Heinlein, K., & Polka, L. (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, *23*(2), e12901. <https://doi.org/10.1111/desc.12901>.
- 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. <https://doi.org/10.1111/infa.12296>.
- 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. <https://doi.org/10.1111/desc.12419>.
- Polka, L., & Sundara, M. (2003). Word segmentation in monolingual and bilingual infant learners of English and French. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 1021–1024). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1021.pdf.
- R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rose, S. (2018). ATR vowel harmony: New patterns and diagnostics. *Proceedings of the Annual Meetings on Phonology*, *5*, 1–12.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>.
- Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 974–989. <https://doi.org/10.1037/0096-1523.28.4.974>.
- Singh, L. (2018). He said, she said: Effects of bilingualism on cross-talker word recognition in infancy. *Journal of Child Language*, *45*(2), 498–510. <https://doi.org/10.1017/S0305000917000186>.
- Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*, *28*(4), 708–737. <https://doi.org/10.1111/infa.12545>.
- Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, *124*(2), 128–142. <https://doi.org/10.1016/j.cognition.2012.05.008>.
- Singh, L., Reznick, J. S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*(4), 482–495. <https://doi.org/10.1111/j.1467-7687.2012.01141.x>.
- Solá-Llloch, E., & Sundara, M. (2025). Young infants' sensitivity to precursors of vowel harmony is independent of language experience. *Infant Behavior and Development*, *78*, 102032. <https://doi.org/10.1016/j.infbeh.2025.102032>.
- Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, *36*(3), 422–444. <https://doi.org/10.1006/jmla.1996.2495>.
- Thiessen, E. D., Onnis, L., Hong, S., & Lee, K. (2019). Early developing syntactic knowledge influences sequential statistical learning in infancy. *Journal of Experimental Child Psychology*, *177*, 211–221. <https://doi.org/10.1016/j.jecp.2018.04.009>.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. <https://doi.org/10.1037/0012-1649.39.4.706>.
- van Kampen, A., Parmaksiz, G., van de Vijver, R., & Höhle, B. (2008). Metrical and statistical cues for word segmentation: The use of 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* (pp. 313–324). Newcastle: Cambridge Scholars Publishing.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Vrooomen, J., Tuomainen, J., & De Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, *38*(2), 133–149. <https://doi.org/10.1006/jmla.1997.2548>.