Journal of Phonetics 108 (2025) 101378

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Coarticulation and coordination in phonological development: Insights from children’s and adults’ production of complex–simplex stop contrasts in Gã

Felix Kpogo [a], Charles B. Chang [b,*]

[a] Brown University, United States
[b] City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 13 June 2023
Received in revised form 29 September 2024
Accepted 26 November 2024
Available online 19 December 2024

Keywords:
Coarticulation
Coordination
Labio-velars
Voice onset time
Closure duration
Formant transitions
Ghanaian languages

ABSTRACT

Achieving adult-like coarticulation, which relies on precise gestural coordination, is known to be a challenging aspect of phonological development. Unique coordination challenges are posed by doubly articulated stops, typologically uncommon complex consonants that show crosslinguistic variation in their acoustic contrast with simplex (singly articulated) consonants. We examined the acoustics and development of complex–simplex stop contrasts between labio-velars (/k͡p/, /g͡b/) and bilabials (/p/, /b/) in Gã (Niger-Congo, Kwa), with special attention to coarticulation with adjacent sonorants. We found that Gã adults mostly produced differences in voice onset time and closure duration to implement these contrasts, and Gã five-year-olds also produced differences in these dimensions. Crucially, however, five-year-olds also produced significant differences in onset formants, which adults did not. These findings provide evidence of age-graded variation in the implementation of complex–simplex stop contrasts in Gã, suggesting that over the course of development there may be a shift away from production of carryover coarticulatory differences toward greater reliance on durational differences. We argue that children’s initial reliance on carryover coarticulation capitalizes on a tendency toward greater consonant–vowel coarticulation as compared to adults, discussing implications for our understanding of how coarticulation develops.

© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

Achieving adult-like coarticulation requires the precise coordination of multiple articulatory gestures and is therefore known to be a challenging aspect of speech production for young children, whose gestural coordination is still developing (Zharkova, 2018; Cychosz et al., 2021). Most of our knowledge of gestural and coarticulatory development comes from widely-studied, and generally Indo-European, languages such as English, an empirical bias that is found throughout the developmental literature (Kidd & Garcia, 2022; Singh et al., 2023). This bias has limited our understanding of phonological development—in particular, the development of coarticulation and coordination—because it has encouraged a narrow focus on the possibilities for gestural overlap that exist in those widely-studied languages.

Across the world’s languages, however, there are diverse possibilities for gestural overlap. Apart from the “phonetic” types of gestural overlap that have been argued to be universal (Lindblom & MacNeilage, 2011), such as the overlap between consecutive consonants and vowels that leads to vowel formant transitions (i.e., “coarticulation”), there are also “phonological” (i.e., contrastive) types of gestural overlap that are found in specific languages. For example, some languages include consonant phonemes with a contrastive secondary articulation (e.g., palatalization, pharyngealization), while other languages include consonant phonemes that have two primary places of articulation of the same manner (i.e., doubly articulated consonants). Doubly articulated consonants such as complex stops can be understood to represent the most extreme case of gestural overlap, which therefore makes them particularly insightful to consider alongside coarticulatory development. In fact, some researchers have described complex stops as potentially falling under the rubric of coarticulation (see Ohala, 1993, p. 156), consistent with the view of certain theories of speech organization (e.g., Articulatory

* Corresponding author at: Mailing address: City University of Hong Kong, Department of Linguistics and Translation, 83 Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region.
E-mail address: cbchang@post.harvard.edu (C.B. Chang).

https://doi.org/10.1016/j.wocn.2024.101378
0095-4470/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Phonology; Browman & Goldstein, 1986, 1989) that there is no meaningful division between gestural overlap in the coarticulation of different segments and gestural overlap within one segment. In this paper, we use the term “coarticulation” in the narrow sense of gestural overlap between consecutive segments (Menzerath & Lacerda, 1933) and the terms “gestural overlap” and “coordination” in the broad sense of overlap between distinct gestures and the process of achieving this, regardless of whether or not the gestures are affiliated with different segments (see Mildner, 2018, for more on terminology).

In this paper, our focus is the development of labio-velar stops, an example of a complex stop type referred to in the literature as doubly articulated (Connell, 1994; Cahill, 1999). Labio-velar stops are of interest for research on coarticulation given their high intrinsic gestural coordination demands and the relevance of “the specific articulatory demands of the segments under consideration, as maturation of speech motor control may be a considerable factor for the observed developmental changes in coarticulatory processes” (Rubertus, 2024, pp. 7–8; see also Recasens et al., 1997; Recasens, 2014). In the case of labio-velar stops, coordination with adjacent segments such as vowels must ordinarily co-occur with the coordination of dual stop constrictions; this is a unique coarticulatory situation specific to doubly articulated stops. Labio-velar stops are relatively uncommon among the world’s languages, occurring in about 6% of the 317 languages in the UPSID database (Maddieson, 1984), but are widespread in languages of Sub-Saharan Africa and Trans-New Guinea. Previous acoustic investigations into contrasts between complex stops and their simplex (i.e., singly articulated) counterparts (e.g., /k͡p/ vs. /p/ or /k/) have indicated that such contrasts may be signaled in several acoustic dimensions, such as voice onset time (VOT), closure duration, and vowel formant transitions; further, there is crosslinguistic variability in the pattern of their acoustic realization (Connell, 1987, 1991, 1994; Maddieson & Ladefoged, 1989; Maddieson, 1993).

Given this crosslinguistic variability, as well as the paucity of developmental data on complex–simplex stop contrasts, the present study examined how contrasts between complex labio-velar stops (/k͡p, g͡b/) and simplex bilabial stops (/p, b/) are produced by adults and five-year-old children in Gã (Niger-Congo, Kwa), an understudied language of Ghana. By reporting acoustic data on complex–simplex stop contrasts in Gã from adult and child speakers, including on formant transitions into and out of adjacent sonorants, this paper contributes new insights on the nature of coarticulatory development from an understudied language, as well as the first evidence of developmental variation in the production of these Gã stops. In addition, the paper contributes further information about the range of crosslinguistic variation in the realization of complex–simplex contrasts.

The rest of the paper is structured as follows. Section 1 situates the study by summarizing previous research on complex–simplex stop contrasts, their acquisition, and the acquisition of coarticulation, providing an overview of the Gã phonological system, and outlining the research questions of the study. Section 2 describes the methodology of the production experiment. Section 3 presents the results of the experiment. Finally, Section 4 discusses the findings of the study in light of the developmental literature on young children’s speech acquisition and the typological literature on complex–simplex stop contrasts.

1.1. Characteristics of labio-velar stops vis-a-vis simplex counterparts

Previous studies of complex stops have examined different phonetic aspects of their production, and here we focus on labio-velar complex stops (/k͡p, /g͡b/) specifically. Some studies have examined the gestural timing of labio-velars (e.g., Painter, 1978; Connell, 1987, 1991; Maddieson & Ladefoged, 1989), including whether the velar and bilabial closures are synchronous or staggered. In Gã, the velar and bilabial closures were found to be simultaneous for one speaker, on the basis of pharyngeal pressure, oral pressure, and microphone tracings for tokens uttered in isolation and in a frame sentence (Painter, 1978). On the other hand, in Yoruba, Ibibio, Ewe, and Mangbetu, the velar closure was found to precede the bilabial closure (Painter, 1978; Connell, 1987, 1991; Maddieson & Ladefoged, 1989; Demolin, 1991; Maddieson, 1993) or to show variability in its relative timing according to speaker and vowel environment (Connell, 1994). Despite the observed variability in the timing of closure onsets, the above studies were consistent in showing that the velar closure is released before the bilabial one.

Others have contrasted the complex stops with simplex ones on acoustic dimensions such as closure duration, spectral features, and voicing characteristics (e.g., Ladefoged, 1964; Garnes, 1975; Connell, 1994; Grawunder et al., 2011). Although some studies produced conflicting results, others found robust acoustic differences between the two stop types. In Ibibio, for instance, labio-velars were found to have longer closure durations than both simplex bilabials and velars (Connell, 1987, 1991). On the other hand, in Igbo and Obolo, labio-velars and simplex stops had similar durations, whereas in Kalabari, voiceless labio-velars had shorter durations than simplex bilabials and velars ([k͡p]: 147 ms, cf. [p]: 225 ms, [k]: 163 ms; Connell, 1994). As for the voiced series in Kalabari, labio-velars showed numerically longer mean durations than simplex stops ([g͡b]: 211 ms, cf. [b]: 131 ms, [g]: 142 ms), although these differences were not statistically significant. Closure durations for the voiced series in Kalabari were similar to those in Ibibio, Igbo, and Obolo. In Yoruba, as in Igbo and Obolo, labio-velars were found to have similar mean durations as the corresponding simplex stops (e.g., [g͡b] in /ag͡ba/ ‘jaw’: 132 ms, cf. [b] in /aba/: 128 ms; Maddieson & Ladefoged, 1989); this was also the case in Mangbetu (Demolin, 1991).

Unlike closure duration, which shows mixed findings, data on the spectral characteristics of labio-velars paints a more consistent picture. Most studies reported a labial-like second formant (F_2) consonant-to-vowel (CV) transition in the release of labio-velars, which is not present for the simplex velar; this is consistent with the velar closure being released before the bilabial closure, as discussed above. Further, the F_2 transition is steeper and more prominent for labio-velars than for simplex bilabials (see Ladefoged, 1964; Garnes, 1975; Connell, 1987, 1991; Dogil, 1988, cited in Connell, 1994). Cahill (2006) also pointed out that F_2 in a CV transition is a reliable cue for distinguishing /g͡b/ and /b/ among Yoruba speakers. The generally steeper F_2 transition for complex stops vis-a-vis simplex stops may reflect a larger gestural magnitude, which could arise from the multiple constrictions in a complex stop effectively competing with each other to be perceived. If such competition were to result in larger release gestures, the larger gestures could lead to steeper formant transitions.

As for vowel-to-consonant (VC) transitions in a postvocalic context, to our knowledge there is no published acoustic data on either of the complex–simplex contrasts examined in this study (i.e., /k͡p/ vs. /p/; /g͡b/ vs. /b/). However, one study of Ibibio suggested that VC transitions generally decline more steeply for labio-velars than for bilabials, although the differences vary according to vowel (Garnes, 1975). Focusing on the vowel environment included in the current study ([ɔ]), we observe that Garnes' comparison of /k͡p/ and /b/ shows similar offset F_2 values for /k͡p/ and /b/, which are slightly higher for /k͡p/ (980 vs. 940 Hz). A later study of Ibibio, which did not include [ɔ], also showed significant variation both within and across speakers in terms of the degree to which VC transitions differed between /k͡p/ and /b/ (Connell, 1994); this may reflect variability in the relative timing of the two closures in a labio-velar, although generally the velar closure occurs no later than the bilabial one. In short, VC transitions may or may not reliably distinguish labio-velars from bilabials.

In addition to F_2 transitions, f_0 differences in the following vowel are a reliable cue for distinguishing /g͡b/ and /b/, at least in Yoruba (Cahill, 2006). Grawunder et al. (2011) also showed that f_0 is higher for /k͡p/ than for /g͡b/ in Yoruba. Of the stops they examined, /k͡p/ and /k/ showed the highest f_0 in terms of mean f_0 level and onset f_0; however, /k/ showed higher f_0 than /k͡p/. There was also a stop voicing effect on vowel duration: vowels following /k/ and /k͡p/ were shorter than those following /b/, /g/, and /g͡b/, but there was no difference between vowels following /k/ and /k͡p/. In the current study of Gã, we also examine the effect of stop type on the following vowel with respect to spectral properties.

Other acoustic correlates of complex–simplex stop contrasts are voice onset time (VOT) and (pre)voicing. In Ewe (which is genetically related to Gã), VOTs for voiceless labio-velars are shorter than those for the simplex bilabials and velars (Maddieson, 1993). In Yoruba, labio-velars are prevoiced stops (Ladefoged, 1964; Puech, 1989; Grawunder et al., 2011); however, Grawunder et al. (2011) pointed out that prevoicing (i.e., closure voicing duration) in Yoruba /k͡p/ is shorter than in /g͡b/, /b/, or /g/, whereas there is no difference in prevoicing for /b/ vs. /g/, /g/ vs. /g͡b/, or /b/ vs. /g͡b/. Connell (1987, 1991) and Garnes (1975) also reported prevoicing in Ibibio voiced and voiceless complex stops as well as voiced simplex stops. Furthermore, electroglottographic (EGG) data and prevoicing intensity data in Yoruba /g͡b/ showed that /g͡b/ has a higher intensity than /b/ and /g/, which suggests stronger voicing enhancement in the complex than simplex stops (Grawunder et al., 2011). The tendency of complex stops to show more prevoicing than simplex stops may reflect a certain degree of implosion (see Ladefoged, 1964; Connell, 1994, p. 446); in fact, complex stops can alternate with implosives in some languages, such as Bekwel (Puech, 1989) and Yoruba (Cahill, 2006). Because implosion is associated with voicing, the implosive tendency of complex stops favors earlier-onset voicing than in simplex stops.

The crosslinguistic phonetic literature thus points to VOT, closure duration, f_0, and F_2 CV transition as possible markers of complex–simplex stop contrasts. However, it also reveals variability in the implementation of such contrasts, raising the question of how such contrasts are realized acoustically in other under-documented languages. The current study contributes acoustic data on complex–simplex stop contrasts in Gã. In this study, we concentrated on labio-velars and simplex bilabials because the data was collected as part of a project that examined Gã-speaking children's simplification of complex stops, which generally reduces complex stops to bilabials and not to velars (Kpogo et al., 2021).

1.2. Acquisition of gestural coordination, coarticulation, and labio-velar stops

A developmental study of complex stop production in Gã is of interest because of the abundant evidence that gestural coordination presents a challenge for young children. For example, in the case of English, articulatory data on lateral approximants, which involve two lingual constrictions, suggests that these sounds are still not adult-like by age 7 (Lin & Demuth, 2015), while acoustic data on oral–laryngeal coordination indicates that VOT of plosives may not become adult-like until after age 10 (Whiteside et al., 2003). As one aspect of gestural coordination, coarticulation between temporally adjacent consonants and vowels is also known to show divergent patterning in children as compared to adults. The nature of child–adult differences in coarticulation varies across the literature (Barbier et al., 2015; Noiray et al., 2018; Rubertus & Noiray, 2018), especially in connection with phonetic features such as place of articulation (Noiray et al., 2013; Zharkova, 2017, 2018), but recent work on English-, German-, and Quechua-learning children shows a tendency for children to produce greater degrees of carryover (CV) and anticipatory (VC) coarticulation than adults (Cychosz et al., 2021; Howson & Redford, 2021; Abakarova et al., 2022). Such a tendency may be related not only to children's still-developing speech motor control, but also to immature vocal anatomy such as an under-differentiated tongue (Gibbon, 1999; Green et al., 2000). Furthermore, children's coarticulation of consecutive gestures does not appear to be uniformly organized, but rather "sensitive to the underlying articulatory properties of the segments combined" (Noiray et al., 2018, p. 1355).

In light of these facts, examining how complex stops—obstruents with intrinsically high coordination demands for (near-)simultaneous oral gestures—come to be coordinated with adjacent sonorants has the potential to shed light on different approaches to coarticulation. At least four models of coarticulation have been proposed: the feature spreading model (e.g., Daniloff & Hammarberg, 1973), the phonetic window model (Keating, 1990), the Directions Into Velocities of Articulators (DIVA) model (Guenther, 1995), and the coproduction model (Fowler, 1980); see Noiray et al. (2019), Redford (2019), and Rubertus (2024) for detailed reviews of these models. All of these models account for how speakers find a coarticulatory compromise between intelligibility and articulatory efficiency, but they differ in terms of their assumptions concerning phonological primitives and, consequently, their view of where and how coarticulation arises. For example, the feature spreading model sets up a division between "grammatical"

Table 1
Examples of complex–simplex stop contrasts in Gã.

word-initial				word-medial			
[k͡pɛ̀]	‘sew’	[g͡bá]	‘tear’	[àk͡pá]	‘dupe’	[àg͡bà]	‘bridge’
[pɛ́]	‘only’	[bá]	‘come’	[àpám]	‘shackles’	[ábá]	‘similar (appearance)’
[kɛ́]	‘definitely’	[gà]	‘ring’	[àkà]	‘trial’	[àgá]	‘grasshopper’

types of coarticulation based on feature-changing rules and “physical” types that do not involve feature changes, whereas the coproduction model, following from the assumption of articulatory gestures rather than features as the phonological primitives, understands all coarticulation as the outcome of articulatory implementation of gestures that differ in terms of strength and the articulators involved.

In regard to the development of coarticulation, these models have made different, testable predictions. For instance, models that must derive gradient coarticulatory patterns through intensive computation, such as the feature spreading model, predict that coarticulation is computationally demanding and should thus increase over development, whereas models that can derive such patterns via gestural activation and overlap, such as the coproduction model, predict a decrease in coarticulation over development as gestures become increasingly differentiated (see further discussion in Redford, 2019). In addition, the models' diverging views of anticipatory and carryover coarticulation have implications for predicting developmental differences. Some models, such as the DIVA model, understand anticipatory and carryover coarticulation in terms of different sources—planning in the former case, but non-planning factors such as physical constraints in the latter case (see, e.g., Lindblom, 1963); however, other models, such as the coproduction model, explain both directions in the same way. Thus, insofar as development results in a better ability both to plan and to overcome non-planning factors, only models such as the DIVA model predict complementary patterns of development in the two directions of coarticulation: an increase in anticipatory coarticulation, but a decrease in carryover coarticulation. Are these predictions borne out in the case of complex stops, whose intrinsic coordination demands might cause their coarticulation with other segments to develop differently?

While the literature on complex stops has documented their phonetic properties in comparison to simplex stops, few studies have looked at children's acquisition of these stops, much less their coarticulatory development. Those studies that do exist have suggested that, crosslinguistically, complex stops are acquired later than simplex stops, which are generally mastered relatively early in development (before age 4; McLeod & Crowe, 2018).[1] For example, Yoruba-speaking children begin to show some articulatory control of the complex stops at age 4 (Isaiah, 2015), but they, as well as Igbo-speaking children, still show errors on these stops at this age (Ajolore, 1974; Nwokah, 1986; Oyebade, 1990; Orie, 2012). More recently, Kpogo et al. (2021) found that Gã-speaking children between the ages of 5 and 8 show accurate production of these stops in intervocalic position, but make errors on them in word-initial position before a vowel or lateral; however, their production improves with age. In addition, Gã-speaking children show more accurate production of the voiceless than voiced labio-velar. In contrast to the complex stops, the simplex stops are mastered by age 5, a milestone that Isaiah (2015) also reported for Yoruba.

Children's later mastery of complex stops vis-a-vis simplex stops may be related to two challenges posed by complex stops. First, relative to the simplex stops, the complex stops, which involve making two oral closures, require greater motor control and intergestural timing skills for their articulation. Second, given the fact that complex stops (both /k͡p/ and /g͡b/) tend to be prevoiced, young children's incomplete articulatory control of these stops may be related to their struggles with prevoicing. Prevoicing is said to be a later development in production as it requires coordinating laryngeal control and supralaryngeal articulatory gestures; hence, it does not become adult-like until age 5 or later (Kewley-Port & Preston, 1974; Zlatin & Koenigsknecht, 1976; Macken & Barton, 1980).

Crucially, if the above factors make the complex stops relatively difficult to produce, then even children's “accurate” (to adult ears) production of these stops may differ acoustically from adults' production. Moreover, children's production of the voiced labio-velars may diverge from adult norms to an even greater degree than their production of the voiceless labio-velars, given that voiced stops require more articulatory effort to enhance closure voicing (Grawunder et al., 2011). Thus, we examined variation in the development of labio-velar stops in Gã by comparing children's production of multiple stop types, including both voiced and voiceless stops, to that of adults.

1.3. Gã phonology

The phonemic inventory of Gã consists of seven oral and five nasal vowels; two tones, high and low; and 31 consonant phonemes, including several stops (Dakubu, 2002). In its stop series, Gã contrasts complex labio-velar stops (/k͡p, g͡b/) with simplex bilabial and velar stops (/p, b, k, g/) in word-initial and word-medial positions, but not in word-final position (see Table 1).

According to Dakubu (2002), a simple phonological word in Gã has a syllable structure of (C/V)CV(C) (e.g., /bá/ ‘come’; /òbɔ̃́/ ‘full’; /ŋ̀kṹ/ ‘shea butter’) whereas a complex word has an underlying CVCV(C) structure but can surface as CCV(C) (e.g., /k͡pàlá/ →[k͡pl̀á]; see Dakubu, 2002). In coda position, only nasals are permissible, but the velar nasal occurs at a higher rate than other nasal segments (e.g., /m/, /n/). In this study, target items included words with both syllable types (see Table 2).

Most work on the Gã stop system, especially the stops of interest, has relied primarily on impressionistic descriptions (Dakubu, 1968; Kropp, 1968; Kotey, 1974). To the best of our knowledge, the only acoustic study of the stop contrast exam-

[1] We use the terms “acquisition” and “mastery” to indicate the achievement of production that is “accurate” (in the sense of being intelligible to adult listeners) at a high rate. By convention, the threshold accuracy level for “acquisition” in acquisition research is often set at 80% (e.g., Stites et al., 2004; Long et al., 2012).

Table 2
Target words, by contrast and phonological environment.

Contrast	#_V	Meaning	#_l	Meaning
/k͡p/ vs. /p/	[k͡pé]	‘chew’	[k͡pĺɛ́]	‘shiny’
	[péńtà]	‘painter’	[pl̀ɛ̃̂kòó]	‘nail’
	[k͡pìtíóó]	‘short’	[k͡pĺìŋ̀]	‘short and stout’
	[pìlá]	‘get hurt’	[pl̀ízĩ̀ŋ̀]	‘prison’
	[k͡pɛ̀]	‘sew’	[k͡pĺɔ́]	‘pluck’
	[pɛ́]	‘only’	[pl̀ɛ́tè]	‘plate’
	[k͡pók͡pòì]	‘Gã festive food’	[k͡pĺòtòó]	‘pig’
	[pònà]	‘type of yam	[pl̀ótó]	‘underpants’
	[k͡pàtá]	‘kitchen’	[k͡pl̀á]	‘rashes’
	[pàpá]	‘father’	[pl̀ástà]	‘plaster’
			[k͡pl̀é]	‘multitude/mighty’
			[pl̀é]	‘suffer and die’
/g͡b/ vs. /b/	[g͡bèé]	‘dog’	[g͡bl̀àmú]	‘fish-smoking oven/grill’
	[béélá]	‘big bush rat’	[bl̀àmɔ́]	‘tying’
	[g͡bɛ̀]	‘road’	[g͡bl̀ɛ̀]	‘grind’
	[bɛ́lì]	‘mattress’	[bĺɛ́]	‘whistle’
	[g͡bìnè]	‘claw’	[g͡bl̀ã̀]	‘marriage’
	[bíbìóó]	‘small’	[bl̀òdò]	‘bread’
	[g͡bòg͡bò]	‘wall’	[g͡bl̀ú]	‘a stool in the shrine’
	[bò]	‘shout’	[bl̀úkù]	‘leggings’
	[g͡bùdùg͡báŋ̀]	‘metal drum’	[g͡bl̀ìnìì]	‘dull/inactive’
	[bú]	‘hole’	[bĺíbóó]	‘exhausted’
	[g͡bá]	‘tear’	[g͡bl̀àg͡bl̀à]	‘salty’
	[bá]	‘come’	[bl̀ɔ́ʃì]	‘brush’

ined here is Painter (1978), who looked at data from one Gã speaker on the gestural timing of the complex stop and on f_0 in the context of larynx lowering for implosives. Thus, in this study, we provide the first detailed acoustic data from multiple speakers at different stages of linguistic development on complex and simplex (bilabial) stop contrasts in Gã as produced by first-language (L1) speakers.

1.4. Research questions and predictions

In this study, we investigated three research questions about the acoustic implementation, development, and coarticulatory outcomes of complex–simplex stop contrasts in Gã:

Q1: What are the acoustic properties that reliably separate the Gã complex stops from simplex stops in the speech of adult L1 speakers?
Q2: What are the acoustic properties (if any) that reliably separate the Gã complex stops from simplex stops in the speech of child L1 speakers (five-year-olds)?
Q3: Does children’s production of complex–simplex stop contrasts show a greater reliance on differences in coarticulation with flanking sonorants (i.e., formant transitions) than adults’ production?

Q1 and Q2 contribute to the broader literature on the acquisition of gestural coordination as well as the literature on the typology of complex consonants, while Q3 contributes to the literature on the acquisition of coarticulation specifically. To address these questions, we conducted a production experiment testing Gã adults and children on complex–simplex stop contrasts in two contexts, phrase-initial and phrase-medial.

In regard to Q1, we generally expected the complex–simplex stop contrasts in Gã to be implemented in terms of acoustic properties alluded to in the literature as reliably distinguishing complex–simplex stop contrasts in other languages. Thus, we examined several acoustic properties, both temporal and spectral, with special attention to VOT, closure duration, f_0, and F_2 (see Section 1.1). Given the crosslinguistic variability observed in production of complex–simplex stop contrasts, we did not have specific predictions about which properties would distinguish these contrasts in Gã adults. However, in connection with the two contexts tested, we predicted that acoustic differences between complex and simplex stops in the properties that are available in both contexts would be larger phrase-initially than phrase-medially (**Prediction 1, P1**), due to the smaller cohort of cues in the phrase-initial context (e.g., VC transition cues are available phrase-medially but not phrase-initially).

In regard to Q2, given the findings of Kpogo et al. (2021), we expected five-year-old Gã children to produce significant acoustic differences between the complex and simplex stops; that is, we did not expect children to uniformly fail to produce a contrast, especially because the target contexts included intervocalic contexts (see Section 2.3), which appear to facilitate production of the complex stops. Additionally, we predicted that children would, overall, implement the complex–simplex stop contrasts with a similar set of acoustic properties as adults; that is, we predicted that children would approximate what is present in adult speech (**Prediction 2, P2**).

In regard to Q3, we distinguished between carryover coarticulation (i.e., CV formant transitions) and anticipatory coarticulation (i.e., VC formant transitions) and considered three possibilities for how Gã children might compare to adults in terms of differences in formant transitions between complex and simplex stops: (1) more consistent differences, (2) less consistent differences, and (3) statistically similar differences (i.e., patterning as adult-like). Following from the DIVA model’s view of carryover coarticulation as the outcome of non-planning factors such as physical constraints, we predicted that children would be more influenced by these factors than adults and would therefore produce overall larger degrees of carryover coarticulation, leading to more consistent differences in the onset formants of CV transitions than in adults (**Prediction 3, P3**). In other words, we predicted a type of coarticulatory amplification to result in clearer differentiation of onset formants by children.

Our prediction for anticipatory coarticulation was also based on the DIVA model, as well as findings on the roles of morphology and articulator overlap. Based on the DIVA model’s view of anticipatory coarticulation as a reflection of planning, we predicted that children would produce anticipatory coarticulation less reliably than adults, due to less developed planning mechanisms. In addition, there is evidence that a morphological boundary between target segments can result in reduced anticipatory coarticulation in children (Cychosz, 2020, 2021), which is relevant for the current study where a word boundary always intervened in a target VC sequence (see Section 2.3). Finally, we expected anticipatory coarticulation to differ from carryover coarticulation because of differences in articulator overlap, which are known to have consequences for coarticulatory variation and resistance (see, e.g., Graetzer, 2007). In our case, there was articulator overlap in the VC context (where both the lingual gesture for the vowel and the dorsal gesture of a complex stop involved the tongue body), but not in the CV context (where the labial gestures of a complex stop did not involve the tongue body); articulator overlap could thus make anticipatory coarticulation more challenging to produce reliably. Taken together, these considerations led us to predict that, compared to adults, children would produce less consistent differences in the offset formants of VC transitions between complex and simplex stops (**Prediction 4, P4**).

2. Materials and methods

2.1. Participants

Participants in this study consisted of two groups recruited from James Town, Accra: adult L1 Gã speakers (N = 10; 5 female, 5 male; M_{age} = 28.1 yr, SD = 5.0) and five-year-old L1 Gã-speaking children (N = 10; 6 female, 4 male; M_{age} = 5.3-yr, SD = 0.2). Data for two additional adults was collected, but later excluded from analysis due to poor audio quality. The adult participants were not caregivers to the child participants.

Based on a language background questionnaire, accessible via the Open Science Framework at https://osf.io/zhqud/, both the adult and the child participants were representative of Gã-dominant L1 speakers. The questionnaire was adapted from one used in the Welsh Language Board project for bilingual children and bilingual families (Gathercole & Thomas, 2007), based on the first author's knowledge of and direct experience with the community (see, e.g., Esene Agwara, 2020). Data collected via the questionnaire, including other languages that participants knew apart from Gã, is available at https://osf.io/eby5f/. Broadly, this data indicated that the adults used only Gã at home and had used both Gã and English in school (meaning they spoke Gã more than English on average daily). Similarly, the children used Gã 100% of the time at home and used English only in school.

In regard to participants' language exposure and use, there are some relevant community-level details, which may lead to differences between the adult and child groups. First, by "English" above, we refer to Ghanaian English, as opposed to Ghanaian Pidgin English (GPE), although both varieties are likely present in the linguistic milieu of both groups (see Huber, 2008). The variety that children were exposed to primarily is Ghanaian English, but they may additionally have had exposure to GPE; that said, it is not the norm for Ghanaian children to use GPE, and to our knowledge, the only published description of Ghanaian children using GPE is of pre-school children in middle-class families (Huber, 2008, p. 96). By contrast, adults have more exposure to GPE, and thus may use GPE to a greater extent than children. Second, the Gã-speaking community in Ghana has been affected by ongoing demographic and linguistic shifts. The proportion of Gã speakers in Ghana has declined over the last two decades (Ghana Statistical Service, 2013, 2021), and most Gã speakers are shifting from Gã to other languages such as Twi and English (Akpanglo-Nartey & Akpanglo-Nartey, 2012; Bibiebome et al., 2019). All this implies that the language ecology of Gã may differ between when the adults started acquiring Gã and when the children did, a point we return to in Section 4.

2.2. Stimuli

A set of 23 (near-)minimal pairs containing a labio-velar stop or the corresponding simplex bilabial stop (i.e., /k͡p/ vs. /p/; /g͡b/ vs. /b/) were selected as stimuli. All target words were existing lexical items in Gã that contained the critical segment in one of two environments: (1) word-initial before a vowel, and (2) word-initial before a lateral.[2] Selected words were common in everyday speech to children (determined based on a two-way classification—frequent vs. infrequent—provided by the adult Gã-speaking participants) and easily picturable. Furthermore, to the extent possible, they were balanced in terms of the vowel types surrounding the consonants as well as the tones.

Because it was difficult to control for word length within each word pair, we compared the word lengths for the two stop types to ascertain whether they differed statistically. This comparison revealed that the mean word lengths for the two stop types did not differ significantly ($M_{simplex}$ = 2.30 syllables, SD = 0.77; $M_{complex}$ = 2.17 syllables, SD = 0.82; $t(45)$ = -0.60, p = 0.55). Table 2 shows the full list of target words elicited in the study.

Pictures depicting the objects or actions denoted by the target words were adapted from Kpogo et al. (2021). All pictures were uniform in size (height = 4.45 cm; width = 4.05 cm) and designed to catch the attention of the participants and keep them engaged during testing. The full set of pictures is publicly accessible at https://osf.io/zhqud/.

2.3. Procedure

Informed consent was obtained from all participants; in the case of the child participants, this included the consent of both school authorities and children's parents/guardians. All participants were audio-recorded in a quiet room (more specifically, a recording studio in Accra), using a Zoom H4N Pro recorder (at 48 kHz and with 16-bit resolution) and the recorder's internal microphone placed about 15 cm from the participant's mouth. Each participant was asked to sit in front of a computer screen for testing.

After being given instructions regarding the test session, participants completed a set of practice trials to get acquainted with the production task, a picture naming task. The practice trials were identical in structure to the experimental trials but tested items not included among the target words: [g͡békɛ̃́bîː] 'children' and [ák͡pl̀ɔ̀tɔ̀] 'toad'. Participants then proceeded to the experimental trials, which tested production in two contexts: phrase-initial (in isolation) and phrase-medial (in a carrier phrase). Elicitation of target words in the phrase-initial condition was accompanied by short stories in Gã (e.g., for /g͡bá/ 'tear (v.)', the experimenter said something like *Àkwélé bàákɔ́ wòlò lɛ̀ nĩ́___lɛ́* 'Akwele will take a sheet of paper and___it' while showing the picture for /g͡bá/). The full set of story/sentence prompts is available at https://osf.io/zhqud/. After this condition, participants were asked to say the same words in a phrase-medial condition, within the carrier phrase *Kɛ̀ɛ́mɔ́ ___ékóóŋ̀* 'Say___again'.

In general, each item was elicited once within each condition, in order to limit the length of the testing session for the child participants to under half an hour. Between the two options of collecting more repetitions of fewer items or collecting fewer

[2] Whereas either /l/ or /r/ can occur after simplex stops (bilabial or velar) in Gã, only /l/ can occur after labio-velar stops. The phonetic sequence of [k͡pl] or [g͡bl], however, is the result of deletion of a weak vowel between the labio-velar stop and the lateral (this same process also results in sequences of a simplex stop + liquid; Dakubu, 2002). The weak vowel can be deleted only when the vowels in successive syllables of a word are the same. When the weak vowel is deleted, the lateral assumes the tone of the weak vowel.

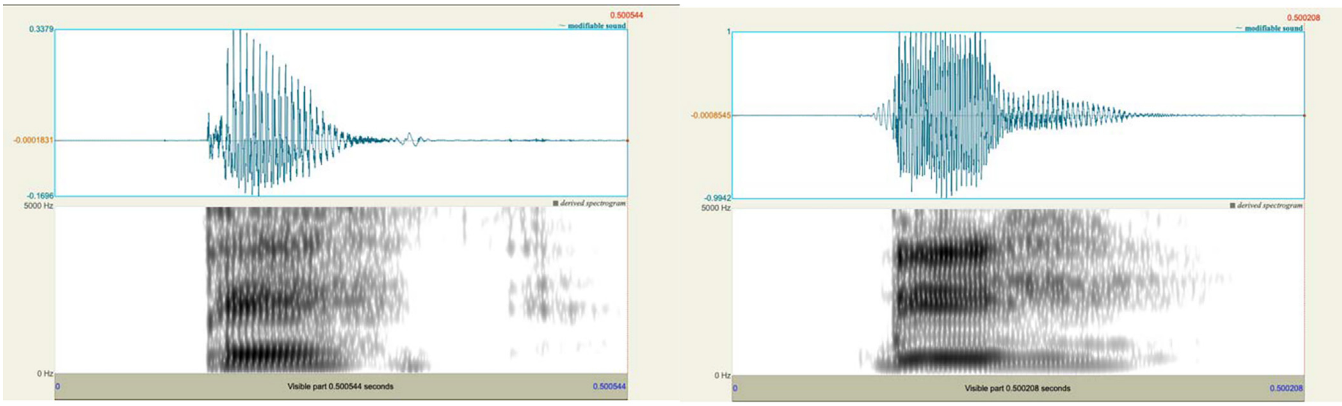

Fig. 1. Waveforms and spectrograms of [pɛ́] 'only' (left) and [k͡pɛ̀] 'sew' (right).

repetitions of more items within this limited time, we opted for the latter, collecting one token each of several different items per item group (i.e., combination of stop type, voicing, and phonological environment) shown in Table 2. This design was preferred because it allowed for the production task to be more natural and engaging, especially for the child participants.

Although the production task collected only one token per item per participant, a few contingencies were included in the testing protocol in order to maximize analyzable productions. First, on any given trial, if a participant's production was inaudible they were asked to rename the picture. If they found it challenging to recognize the target word, they were given definitions or descriptions of the action or the object. In instances where they still faced difficulty in identifying the word depicted, they were played a recorded model of the word's pronunciation; this latter method was used in about half of trials for the child participants (and never for the adults).[3] Audio files for the recorded models (uttered by a male L1 speaker) are available at https://osf.io/zhqud/. Testing took up to 20 min for adult participants and about 25 min for child participants. After the study, participants were given a small reward as a way of showing appreciation.

2.4. Acoustic analysis

The acoustic analysis of the target sounds was carried out using Praat (Boersma and Weenink, 2019) on a wideband spectrogram with a window length of 5 ms, maximum view range up to 5000 Hz, and dynamic range of 50 dB, or on the corresponding waveform. Figs. 1 and 2 show waveforms and spectrograms for four of the recorded models, which together exemplify the /p/ vs. /k͡p/ and the /b/ vs. /g͡b/ contrasts in phrase-initial position. As shown in these figures, both pairs of sounds differ acoustically. For example, the complex voiceless stop shows more prevoicing than the simplex one, while the complex voiced stop shows an increase in the amplitude of voicing going into the vowel, unlike the simplex one.

Given previous findings, we analyzed the acoustic properties of VOT, closure duration, f_0, and formant transitions (see Garnes, 1975; Connell, 1987, 1991, 1994; Maddieson & Ladefoged, 1989), according to protocols described below at the beginning of the respective results sections. The time points for all acoustic measures were extracted using a Praat script. Spectral properties (i.e., f_0 and formants F_1, F_2, and F_3) were measured at three different time points, following previous studies (Garnes, 1975; Connell, 1994; Grawunder et al., 2011; among others); however, only the measurements temporally closest to the stop were submitted to statistical analysis. On some trials, more than one token of the target item was produced; in such cases, the analysis generally targeted the last production.

To account for potential errors in annotating the time points for the acoustic measures, inter-rater reliability was measured for a random subset of 10% of the dataset, which was submitted to another phonetically trained linguist to re-measure all of the acoustic properties of interest. For all measures, reliability was good to excellent: Pearson's R = 0.99 for VOT/closure duration, R = 0.99 for f_0, R = 0.85 for F_1, R = 0.89 for F_2, and R = 0.85 for F_3.

2.5. Statistical modeling

The acoustic data was submitted to statistical analysis using a series of mixed-effects regression models of the dependent variables. All models were based on the raw acoustic values (Hz for spectral measures; ms for temporal measures) and not normalized values.[4] The models excluded outlier values for the dependent variable (according to the 3-*SD* criterion) and were built separately by context and/or group in accordance with the research question(s) they addressed (Q1, Q2, Q3; see Section 1.4). Due to the data gap for closure duration of voiceless stops (see Section 3.1.1), models fully

[3] Given the high rate of production following an auditory model only in the child group, we checked whether being provided an auditory model significantly influenced the children's production. Mixed-effects models built on the children's production data with a fixed-effect predictor coding the presence vs. absence of an auditory model on a given trial showed no significant effect of this predictor for any of the dependent variables discussed below [all $|t|$'s < 1.7, p's > 0.05]. Therefore, trials produced with and without an auditory model were combined in all of the following analyses.

[4] Formant measurements in particular were not normalized for two reasons. First, the vowels included in this study were not representative of all the Gã vowels; hence, normalizing vowels would have been based on skewed means, leading to exaggerated differences between vowels in a normalized vowel space. Second, our primary interest in respect to the vowels was in determining whether those vowels adjacent to the target stops help to realize the complex–simplex stop contrasts. That is, our focus was not on the vowel space per se. Thus, to be consistent across models and to provide effect sizes in acoustic units, we report models built on the raw data for all models. Note that the pattern of results for spectral variables remains the same in alternative models built on semitone values and, therefore, cannot be attributed to our use of raw acoustic data.

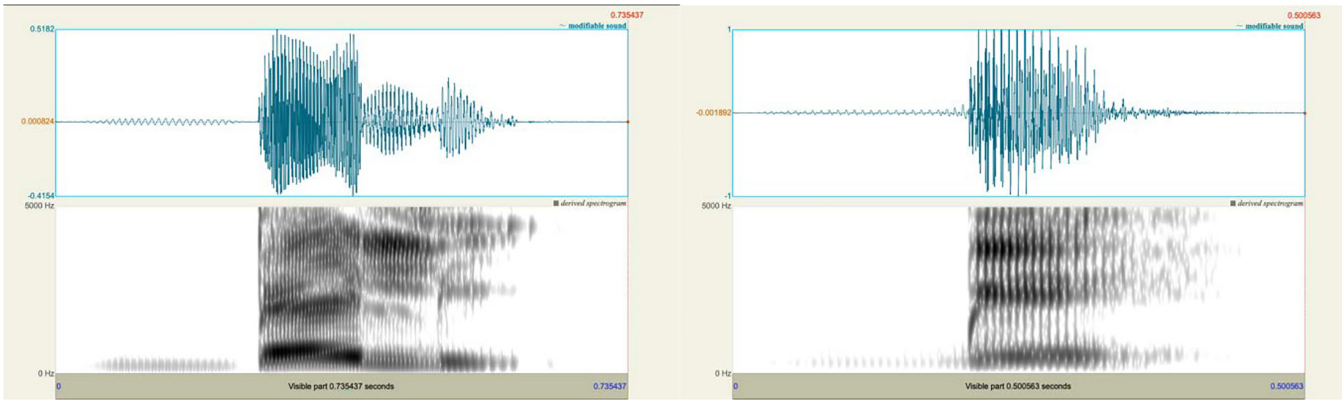

Fig. 2. Waveforms and spectrograms of [bɛ́lì] 'mattress' (left) and [g͡bɛ̀] 'road' (right).

crossing the critical predictors were not built, because such a model was not possible for closure duration. Thus, for the sake of consistency across dependent variables, as well as model interpretability, we built separate models by context and/or group for all of the dependent variables.

Depending on the model, fixed effects for critical predictors, which were all treatment-coded, included one or more of: StopType (complex, simplex; reference level = simplex), Voicing (voiceless, voiced; reference level = voiceless), Context (phrase-initial, phrase-medial; reference level = phrase-initial), and Group (adults, children; reference level = adults). In addition, each model included up to three control predictors as sum-coded fixed effects: a participant's Gender (female, male), PhonEnv (whether the following phonological environment of the target stop comprised a lateral or a vowel), and PostPausal (whether or not a phrase-medial target stop was uttered following an audible pause). Random effects included by-participant and by-item intercepts; random slopes (e.g., for StopType by participant) either did not consistently allow models to converge or resulted in singularity, so were not included in the final models. The full dataset submitted to modeling, along with analysis code, is available open-access at https://osf.io/eby5f/.

All analyses were done in R (R Development Core Team, 2023). Models were built using the *lmerTest* package (Kuznetsova et al., 2017); analysis of variance (ANOVA) was performed on the models using the *car* package (Fox & Weisberg, 2019); and effect sizes for contrasts not directly tested in a given model were obtained in follow-up models with releveled predictors. When comparing adults' and children's production, we focus on the StopType × Group interaction, since we are interested in how the two groups differ (or not) in terms of the effect of stop type (i.e., their differentiation of the complex and simplex stops). We report results from 17 models in all, summarizing all fixed-effect coefficients in the appendix; hence, only the critical coefficients in these models are referenced in the text.

3. Results

3.1. Q1: Adults' production

3.1.1. Adults' VOT and closure duration

VOT and closure duration were measured as follows. VOT was measured by marking the interval between the onset of the release burst and the onset of voicing. Cases of multiple bursts were rare (accounting for less than 5% of tokens), but in these cases the onset of the release burst was identified with the first burst. The time at voicing onset minus the time at stop release was coded as VOT. Thus, in prevoiced tokens of stops such as /g͡b/ and /b/, VOT had a negative value. Closure duration was also measured for phrase-medial voiced stops by inspecting the waveform for amplitude variation and closure voicing.[5] Because the voiced stops (/g͡b/ and /b/) are generally produced with voicing during the whole closure and the phrase-medial context favors voicing, we assumed that the voiced stops would be fully voiced in this context and used stop voicing duration as a proxy for closure duration; therefore, in the case of stops preceded by a pause, our measures may underestimate closure duration if the stop is not fully voiced. The first point in the waveform following the drop in amplitude going from the vowel to the stop constriction, and corresponding to the beginning of a voicing bar indicative of stop voicing in the spectrogram, marked the onset of the voiced closure, while the point corresponding to the beginning of the release burst marked the offset of the voiced closure; in the case of a stop preceded by a pause, the onset of the closure was identified by the beginning of the voicing bar for stop voicing only. The time at offset minus the time at onset was coded as closure duration.

The results of Models 1 and 2 (see Tables A1 and A2 in the appendix) indicated that, for adults, both voiceless and voiced complex stops significantly differed from their simplex counterparts in VOT and/or closure duration. Descriptive statistics on VOT and voiced stop closure duration are summarized in Tables 3 and 4, respectively. Figures plotting the adult data distributions are included in Section 3.2 together with the child data distributions.

Starting with the results of Model 1 for VOT in phrase-initial context (i.e., isolation), compared to /p/, adults produced both /k͡p/ [β = -35.912, p < 0.001] and /b/ [β = -139.264, p < 0.001] with significantly more negative VOT (i.e., longer prevoicing). The stop type effect on VOT was not significantly different for /g͡b/–/b/ as compared to /k͡p/–/p/ [β = -0.797, p = 0.929]; a follow-up model confirmed that /g͡b/ was also produced with significantly more negative VOT than /b/ [β = -36.708, p < 0.001].

[5] Although we attempted to measure closure duration for phrase-medial voiceless stops also, the beginning of the stop constriction for voiceless stops was often not easy to identify, especially for child participants (who often paused before the target word). Therefore, this data was ultimately not analyzed.

Table 3
Mean VOT in ms (with *SD*) of Gã stops in phrase-initial context.

Group	/p/	/k͡p/	/b/	/g͡b/
adults	+16	−20	−124	−160
	(12)	(7)	(43)	(38)
children	+19	−18	−103	−128
	(20)	(17)	(43)	(45)

Table 4
Mean closure duration in ms (with *SD*) of Gã voiced stops in phrase-medial context.

Group	/b/	/g͡b/
adults	104	133
	(30)	(27)
children	107	133
	(39)	(41)

Additionally, there was an effect of phonological environment: VOTs were significantly more positive preceding a lateral as compared to the grand mean including the prevocalic environment [β = 4.252, p = 0.041]. In short, initial complex stops were characterized by more negative VOTs than simplex counterparts, regardless of whether they were phonologically voiceless or voiced.

As for closure duration of phrase-medial voiced stops, Model 2 showed that adults produced the complex stop /g͡b/ with significantly longer closure duration than simplex /b/ [β = 28.790, p < 0.001]. That is, the /g͡b/–/b/ contrast was realized not only with a VOT difference in phrase-initial context, but also with a durational difference in phrase-medial context.

3.1.2. Adults' f_0

Fundamental frequency (f_0) was measured at three time points near vowel onset 10 ms apart. The first time point, subjected to statistical analysis, was 20 ms from stop release. If Praat failed to capture the f_0 track at this time point, manual measurements were taken by marking off the duration of two periods from the start of the f_0 track while maintaining the 10-ms interval between measurements. The pitch settings in Praat were 75 Hz (minimum) and 650 Hz (maximum) for child participants and 75 Hz and 500 Hz for adults.

Model 3 (Table A3) showed no evidence that adults consistently produced differences in onset f_0 to implement complex–simplex stop contrasts. An ANOVA on Model 3 showed a main effect of Context [$\chi^2(1)$ = 36.773, p < 0.001] and Gender [$\chi^2(1)$ = 18.820, p < 0.001], as well as a marginal effect of Voicing [$\chi^2(1)$ = 3.391, p = 0.066], but no main effect of StopType [$\chi^2(1)$ = 0.388, p = 0.533] or PhonEnv [$\chi^2(1)$ = 0.001, p = 0.981]. Context was also involved in two significant interactions: StopType × Context [$\chi^2(1)$ = 10.591, p = 0.001] and StopType × Voicing × Context [$\chi^2(1)$ = 8.599, p = 0.003]. No other interactions were significant. The context effect reflected a tendency for phrase-medial stops to be produced with lower onset f_0 than phrase-initial stops; however, reflecting the above interactions, only two of the contextual contrasts were significant: phrase-medial vs. −initial /k͡p/ [β = -10.049, p < 0.001] and phrase-medial vs. −initial /g͡b/ [β = -15.585, p < 0.001].

Consistent with the StopType × Voicing × Context interaction, Model 3 coefficients and follow-up models indicated a very limited effect of stop type on onset f_0. Voiceless /k͡p/ was produced with virtually the same onset f_0 as /p/ in phrase-initial context [β = -0.032, p = 0.997], and this (lack of) effect did not differ significantly in phrase-medial context [β = -0.547, p = 0.891]. Similarly, /g͡b/ was not produced with significantly different onset f_0 than /b/ in phrase-initial context [β = 2.171, p = 0.783], whereas /g͡b/ was produced with marginally lower onset f_0 than /b/ in phrase-medial context [β = -14.655, p = 0.067]. Together, these results suggest that Gã-speaking adults mostly do not produce differences in onset f_0 to implement complex–simplex stop contrasts.

3.1.3. Adults' formant transitions

To measure CV formant transitions, we first identified the beginning of the target sonorant (vowel or lateral) by the onset of formant structure following stop release, marking the nearest zero-crossing as the sonorant onset. Measurements were taken for each of F_1, F_2, and F_3 at the sonorant onset. Different formant settings were used for adults and children: a maximum formant of 5000 Hz for adult males, 5500 for adult females, and 6000 Hz for children. Although a maximum formant value of 8000 Hz for children has been used by others (e.g., Levy & Hanulíková, 2019; Schweinberger, 2022), we found that 8000 Hz often produced erroneous F_1 values; therefore, we used a lower value of 6000 Hz.

To measure VC formant transitions in the phrase-medial condition, we identified the end of the vowel preceding the target stop (i.e., [ɔ́] in *Kɛ̀ɛ́mɔ́*; see Section 2.3) by the offset of formant structure before the stop constriction. Measurements were taken for each of F_1, F_2, and F_3 at the end of this vowel. These measurements were taken only for tokens where the target item was uttered within the carrier phrase with no audible pause.

3.1.3.1. Adults' F_1. Model 4 (Table A4) showed no evidence that adults produced differences in onset F_1 to implement complex–simplex stop contrasts. An ANOVA on Model 4 showed a main effect of the control predictors Gender [$\chi^2(1)$ = 14.871, p < 0.001] and PhonEnv [$\chi^2(1)$ = 6.039, p = 0.014], but not of any of the critical predictors: StopType [$\chi^2(1)$ = 0.171, p = 0.679], Voicing [$\chi^2(1)$ = 1.273, p = 0.259], or Context [$\chi^2(1)$ = 0.044, p = 0.835]. No interactions were significant. Crucially, no Model 4 coefficients showed a stop type effect on onset F_1.

On the other hand, with regard to offset F_1 in the phrase-medial context (i.e., the VC transition within the precursor vowel [ɔ́]), Model 10b (Table A11) indicated that adults did produce differences in offset F_1 for complex–simplex stop contrasts. In particular, they produced significantly lower offset F_1 values preceding complex as compared to simplex voiceless stops [β = -35.237, p = 0.011], a difference that was non-significantly enhanced for the voiced stops [β = -13.861, p = 0.459].

3.1.3.2. Adults' F_2. Model 5 (Table A5) also showed no evidence that adults produced differences in onset F_2 to implement complex–simplex stop contrasts. An ANOVA on Model 5 showed a main effect of Context [$\chi^2(1)$ = 11.278, p = 0.001] and Gender [$\chi^2(1)$ = 6.886, p = 0.009], but no main effect of StopType [$\chi^2(1)$ < 0.001, p = 0.987], Voicing [$\chi^2(1)$ = 0.482, p = 0.488], or PhonEnv [$\chi^2(1)$ = 0.673, p = 0.412] and no significant interactions [all $\chi^2(1) < 1$, p's > 0.1]. The context effect was reflected in a tendency for stops to be produced with lower onset F_2 phrase-medially than phrase-initially, which was significant for /k͡p/ [β = -91.112, p = 0.004] and for /g͡b/ [β = -59.435, p = 0.047]. No Model 5 coefficients showed a stop type effect on onset F_2.

With regard to offset F_2 in the phrase-medial context, Model 12b (Table A14) showed that adults also produced differences in offset F_2 for complex–simplex stop contrasts. More specifically, they produced significantly higher offset F_2 values preceding complex as compared to simplex voiceless stops [β = 231.123, p < 0.001]. This difference was consistent with the formant transitions expected for velar as compared to bilabial constrictions, and was not significantly different in magnitude for the voiced stops [β = 12.197, p = 0.734].

3.1.3.3. Adults' F_3. Model 6 (Table A6) resembled Models 4–5 in showing no evidence that adults produced differences in onset F_3 to implement complex–simplex stop contrasts. An ANOVA on Model 6 showed a main effect of the control predictors Gender [$\chi^2(1)$ = 25.990, p < 0.001] and PhonEnv [$\chi^2(1)$ = 9.909, p = 0.002], but no main effects of the critical predictors (StopType, Voicing, Context) and no significant interactions [all $\chi^2(1) < 2$, p's > 0.1]. No Model 6 coefficients showed a stop type effect on onset F_3.

As for offset F_3 in the phrase-medial context, Model 14b (Table A17) showed that adults also produced differences in offset F_3 for complex–simplex stop contrasts. The pattern for offset F_3 was the same as for offset F_2: significantly higher offset F_3 values preceding complex as compared to simplex voiceless stops [β = 133.732, p = 0.003], with the difference between complex and simplex being non-significantly reduced for the voiced stops [β = -53.377, p = 0.392]. A follow-up model indicated that the offset F_3 difference for the voiced stops was still marginal [β = 80.355, p = 0.063].

In sum, the results so far suggest that Gã-speaking adults implement complex–simplex stop contrasts mostly in terms of temporal dimensions (VOT, closure duration). In regard to spectral dimensions (f_0, F_1, F_2, F_3), we found evidence of only limited production of f_0 differences and no clear differentiation of the stop types in terms of onset formants. However, in phrase-medial position specifically, there was clear differentiation of complex and simplex bilabial stops in terms of offset formants. The offset formant differences were consistent with the staggered nature of the two closures previously reported for labio-velar stops, in which the velar closure precedes the bilabial closure.

3.2. Q2 & Q3: Children's production vs. adults' production

In this section, we examine what acoustic differences Gã children produce to signal the complex–simplex contrast and compare their results to the adults' results discussed above. In order to directly test for between-group differences (Q3) and also avoid higher-order (in particular, four-way) interactions, which are difficult to interpret, we included Group as a fixed effect in the models discussed below, and built separate models by context. The results are reported in the same order as the adults': first, temporal dimensions (VOT, closure duration), then spectral dimensions (f_0, formants).

3.2.1. Children's VOT and closure duration

As can be seen in Fig. 3, Gã-speaking children, like adults, implemented complex–simplex stop contrasts by producing differences in VOT; however, children differed from adults in their specific pattern of VOT differences. An ANOVA on Model

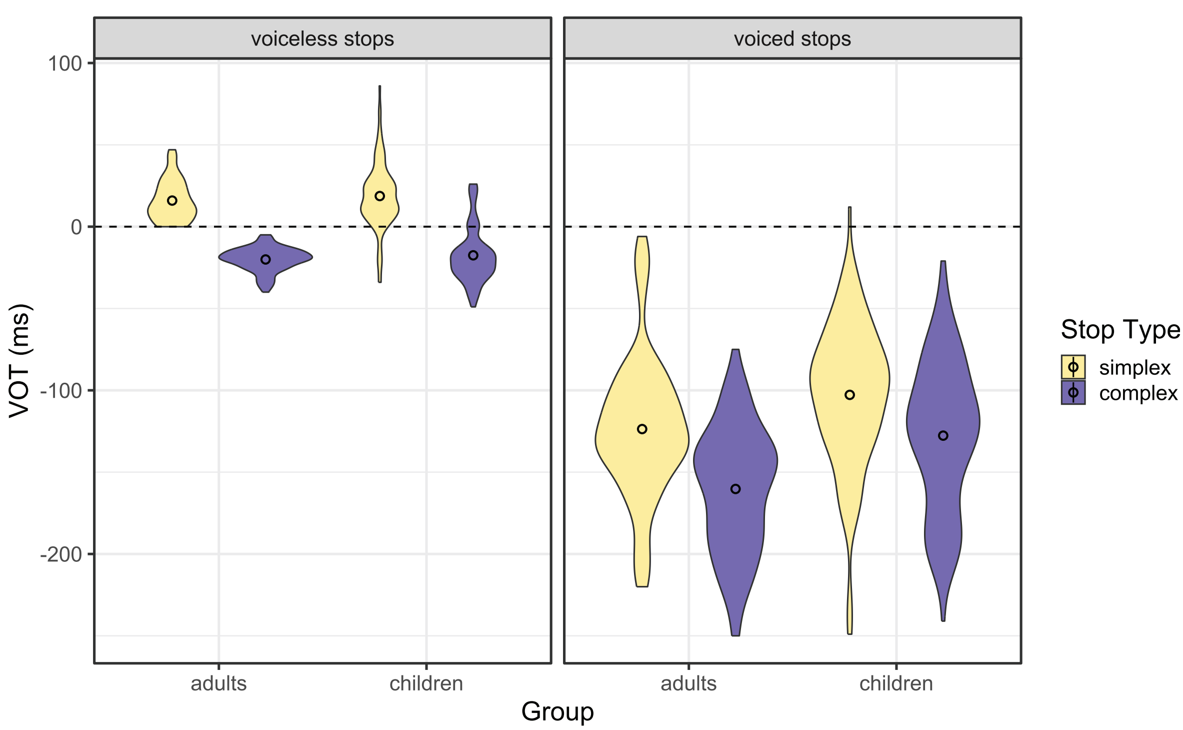

Fig. 3. Violin plot of VOT of Gã stops in phrase-initial context, by voicing, group, and stop type. Hollow circles represent mean values. The dotted line marks zero. (color online only).

1 (Table A1) showed main effects of StopType [$\chi^2(1) = 68.010$, $p < 0.001$], Voicing [$\chi^2(1) = 999.971$, $p < 0.001$], Group [$\chi^2(1) = 7.312$, $p = 0.007$], and PhonEnv [$\chi^2(1) = 4.441$, $p = 0.035$], but no main effect of Gender [$\chi^2(1) = 0.326$, $p = 0.568$]. There were significant or marginal interactions for StopType × Group [$\chi^2(1) = 2.872$, $p = 0.090$], Voicing × Group [$\chi^2(1) = 42.594$, $p < 0.001$], and StopType × Voicing × Group [$\chi^2(1) = 2.850$, $p = 0.091$]; the StopType × Voicing interaction was not significant [$\chi^2(1) = 0.453$, $p = 0.501$]. The interactions with Group reflected children's tendency to produce smaller differences in VOT between simplex and complex stops, and between voiceless and voiced stops, than adults (see Table 3), especially for the /b/–/g͡b/ contrast.

Model 1 coefficients and follow-up models supported the view that, overall, children resembled adults in their production of VOT differences for complex–simplex stop contrasts. Children did not significantly differ from adults in VOT of /p/ [$\beta = 3.053$, $p = 0.646$] or in the decrease of VOT for /k͡p/ [$\beta = -0.253$, $p = 0.962$] (i.e., the stop type effect). However, they did differ from adults in the voicing effect on VOT: relative to adults, children produced a smaller decrease in VOT for /b/ vis-à-vis /p/ [$\beta = 17.805$, $p = 0.001$], a between-group difference that was marginally stronger for /g͡b/ vis-à-vis /k͡p/ [$\beta = 12.414$, $p = 0.092$]. Nevertheless, follow-up models confirmed that children produced significant differences in VOT for all voicing and stop type contrasts (see Table 3): /p/–/b/ [$\beta = -121.459$, $p < 0.001$], /p/–/k͡p/ [$\beta = -36.165$, $p < 0.001$], /b/–/g͡b/ [$\beta = -24.547$, $p < 0.001$], and /k͡p/–/g͡b/ [$\beta = -109.841$, $p < 0.001$].

The results for closure duration of voiced stops in phrase-medial context also revealed similarities between children and adults, as shown in Fig. 4. An ANOVA on Model 2 (Table A2) showed a main effect of StopType [$\chi^2(1) = 27.770$, $p < 0.001$] and PostPausal [$\chi^2(1) = 7.675$, $p = 0.006$], but not of Group, Gender, or PhonEnv [all $\chi^2(1) < 2.5$, p's > 0.1]. Crucially, there was no significant StopType × Group interaction [$\chi^2(1) = 0.868$, $p = 0.352$]. Model 2 coefficients showed that children did not significantly differ from adults in closure duration of /b/ [$\beta = -3.201$, $p = 0.732$] or in the stop type effect on closure duration of /g͡b/ [$\beta = -4.423$, $p = 0.352$]. Moreover, a follow-up model confirmed that, like adults, children produced a significant difference in closure duration between /b/ and /g͡b/ [$\beta = 24.368$, $p < 0.001$]. These results imply that children produce similarly robust differences in closure duration for the /b/–/g͡b/ contrast as adults.

3.2.2. Children's f_0

As shown in Fig. 5, although adults did not produce f_0 differences for complex–simplex stop contrasts in phrase-initial context, children showed a trend toward doing so, but only for the /p/–/k͡p/ contrast. An ANOVA on Model 7 (Table A7) revealed a main effect of Group [$\chi^2(1) = 119.293$, $p < 0.001$] as expected, but no main effects of StopType [$\chi^2(1) = 1.764$, $p = 0.184$] or Voicing [$\chi^2(1) = 1.555$, $p = 0.212$]; there were also no main effects of the control predictors Gender and PhonEnv [$\chi^2(1) < 2.2$, p's > 0.1]. Crucially, there was a significant StopType × Group interaction [$\chi^2(1) = 9.588$, $p = 0.002$], suggesting that children differed from adults in their production of f_0 differences for complex–simplex stop contrasts in phrase-initial context. There were also significant or marginal interactions for Voicing × Group [$\chi^2(1) = 2.787$, $p = 0.095$] and StopType × Voicing × Group [$\chi^2(1) = 4.283$, $p = 0.038$], which reflected the fact that children, unlike adults, tended not to produce a significant f_0 difference between initial voiceless and voiced stops and, moreover, produced f_0 differences between

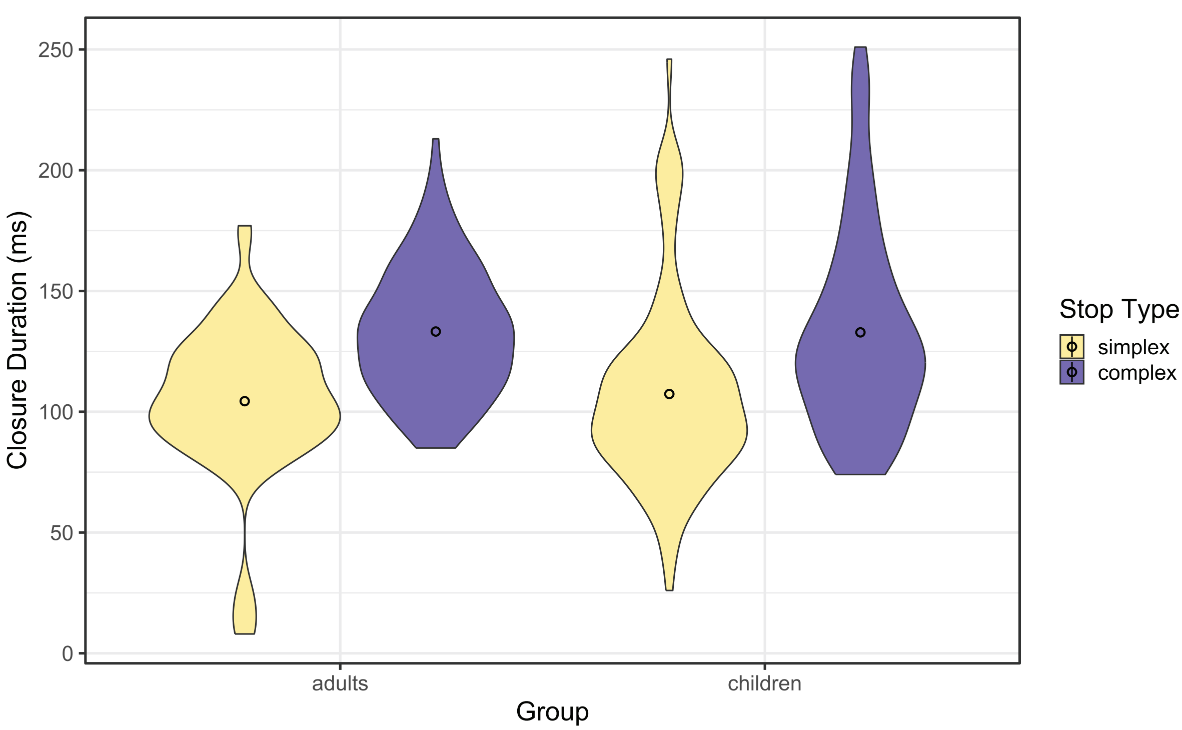

Fig. 4. Violin plot of closure duration of Gã voiced stops in phrase-medial context, by group and stop type. Hollow circles represent mean values. (color online only).

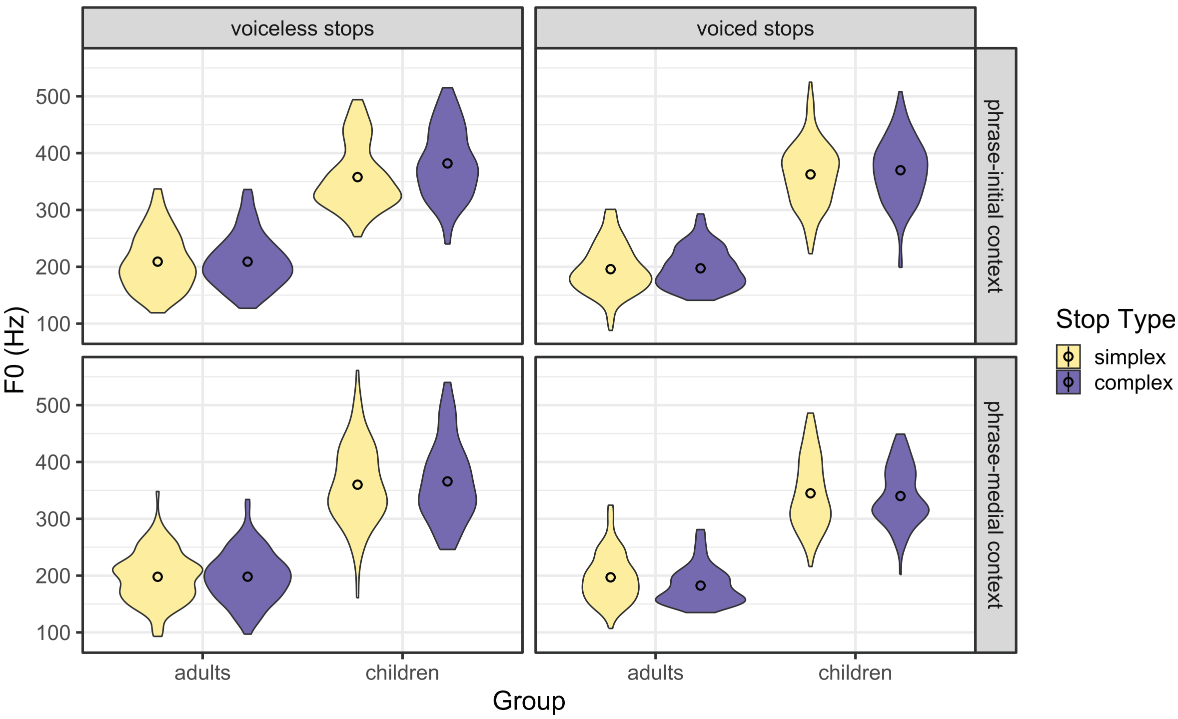

Fig. 5. Violin plot of onset f_0 following Gã stops, by voicing, context, group, and stop type. Hollow circles represent mean values. (color online only).

voiceless and voiced stops that went in different directions for the two stop types. The StopType × Voicing interaction was not significant [$\chi^2(1) = 0.339$, $p = 0.560$].

Model 7 coefficients and follow-up models revealed that children indeed differed from adults in their production of f_0 differences. As expected due to their vocal fold anatomy (see, e.g., Zhang, 2021), children's onset f_0 was much higher than adults' on /p/ [$\beta = 146.719$, $p < 0.001$]. Additionally, in comparison to the non-significant stop type effect observed in adults (see Section 3.1.2), children produced a significantly larger difference in onset f_0 between /p/ and /k͡p/ [$\beta = 24.035$, $p < 0.001$], which a follow-up model indicated was reliable [$\beta = 24.004$, $p = 0.015$]. Children also differed from adults in producing higher, rather than lower, onset f_0 for /b/ relative to /p/ [$\beta = 17.123$, $p = 0.008$]. Thus, the results support the view that children produce onset f_0 differences to a greater degree than adults in phrase-initial context, particularly for the voiceless /p/–/k͡p/ contrast.

Results for the phrase-medial context suggested that, like adults, children did not consistently produce differences in onset f_0 to implement the complex–simplex stop contrasts in this context. An ANOVA on Model 8 (Table A8) showed a main effect of Group [$\chi^2(1) = 86.083$, $p < 0.001$] and Voicing [$\chi^2(1) = 4.582$, $p = 0.032$], as well as PostPausal [$\chi^2(1) = 16.247$, $p < 0.001$], but no main effects of StopType [$\chi^2(1) = 0.451$, $p = 0.502$], Gender [$\chi^2(1) = 0.442$, $p = 0.506$], or PhonEnv [$\chi^2(1) < 0.001$, $p = 0.998$]. The PostPausal effect reflected lower onset f_0 values following a pause [$\beta = -15.521$, $p < 0.001$]. In addition, there was a significant Voicing × Group interaction [$\chi^2(1) = 11.069$, $p = 0.001$], as well as a marginal StopType × Group interaction [$\chi^2(1) = 2.716$, $p = 0.099$]; no other interactions were significant. Unlike the phrase-initial context, the Voicing × Group interaction in the phrase-medial context reflected the fact that, compared to adults, children produced *more* of a decrease in onset f_0 for phrase-medial /b/ relative to phrase-medial /p/ [$\beta = -17.349$, $p = 0.003$]. The marginal StopType × Group interaction was reflected in a follow-up model targeting the phrase-medial /b/–/g͡b/ contrast, the one place where adults produced a significant difference in onset f_0: unlike adults, children did not produce /g͡b/ with reliably different onset f_0 than /b/ [$\beta = -4.397$, $p = 0.675$]. Overall, these results suggest that children, like adults, do not consistently produce onset f_0 differences for complex–simplex stop contrasts in phrase-medial context.

3.2.3. Children's formant transitions

3.2.3.1. Children's F_1. Results for F_1 revealed that children, in contrast to adults, produced significant differences in onset F_1 for complex–simplex stop contrasts in both phrase-initial and phrase-medial contexts—specifically, higher onset F_1 following complex stops (see Fig. 6). Starting with the results for the phrase-initial context, an ANOVA on Model 9 (Table A9) revealed a significant main effect of Group [$\chi^2(1) = 291.931$, $p < 0.001$] as expected due to children's shorter vocal tracts, but only marginal effects of StopType [$\chi^2(1) = 2.997$, $p = 0.083$] and Voicing [$\chi^2(1) = 3.413$, $p = 0.065$]. Additionally, there were significant main effects of the control predictors Gender [$\chi^2(1) = 6.854$, $p = 0.009$] and PhonEnv [$\chi^2(1) = 7.800$, $p = 0.005$]. Crucially, there was a significant StopType × Group interaction [$\chi^2(1) = 26.350$, $p < 0.001$], as well as a marginal Voicing × Group interaction [$\chi^2(1) = 3.804$, $p = 0.051$]. No other interactions were significant [all $\chi^2(1) < 0.5$, p's > 0.1].

Model 9 coefficients and follow-up models indicated that children consistently produced onset F_1 differences for complex–simplex stop contrasts in phrase-initial context. Relative to the non-significant stop type effect in adults (see Section 3.1.3.1), children produced a significantly larger onset F_1

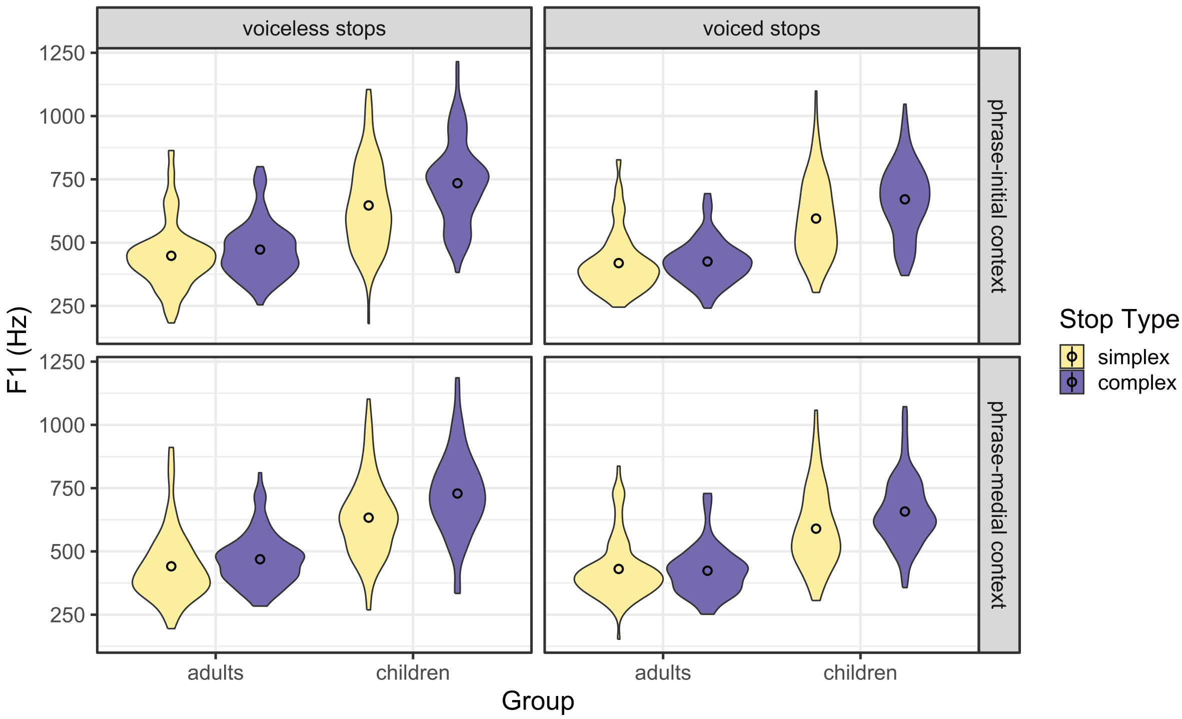

Fig. 6. Violin plot of onset F_1 following Gã stops, by voicing, context, group, and stop type. Hollow circles represent mean values. (color online only).

difference for the /p/–/k͡p/ contrast [β = 59.326, p = 0.002], and this stop type effect on onset F_1 was not significantly different for /b/–/g͡b/ [β = 15.756, p = 0.550]. Follow-up models confirmed that children produced significantly higher onset F_1 for /k͡p/ than /p/ [β = 84.814, p = 0.041] as well as marginally higher onset F_1 for /g͡b/ than /b/ [β = 76.042, p = 0.055].

Results for the phrase-medial context were similar to those for the phrase-initial context. An ANOVA on Model 10a (Table A10) showed a main effect of Group [$\chi^2(1)$ = 164.083, p < 0.001] but not of StopType [$\chi^2(1)$ = 2.460, p = 0.117] or Voicing [$\chi^2(1)$ = 2.412, p = 0.120]. There was an effect of PhonEnv [$\chi^2(1)$ = 4.614, p = 0.032] and PostPausal [$\chi^2(1)$ = 5.554, p = 0.018], but not of Gender [$\chi^2(1)$ = 0.872, p = 0.350]. Again, there was a significant StopType × Group interaction [$\chi^2(1)$ = 27.266, p < 0.001]. There was also a significant Voicing × Group interaction [$\chi^2(1)$ = 5.520, p = 0.019], which reflected children's tendency to produce larger onset F_1 differences between voiceless and voiced stops than adults. No other interactions were significant.

Model 10a coefficients and follow-up models showed, overall, the same pattern as in Model 9. Relative to adults' stop type effect for the phrase-medial context, children produced a significantly larger onset F_1 difference for the phrase-medial /p/–/k͡p/ contrast [β = 63.570, p = 0.001]; this stop type effect on onset F_1 was not significantly different for the phrase-medial /b/–/g͡b/ contrast [β = 8.770, p = 0.737]. Follow-up models confirmed that children produced significantly higher onset F_1 for phrase-medial /k͡p/ than /p/ [β = 95.174, p = 0.026]; however, their onset F_1 difference for the phrase-medial /b/–/g͡b/ contrast, while going in the same direction, did not reach significance [β = 64.676, p = 0.108].

Results for offset F_1 in the phrase-medial context differed from those for onset F_1, but still showed children producing F_1 differences between the two stop types that were similar to the differences that adults produced (see Fig. 7). An ANOVA on Model 10b (Table A11) showed main effects of Group [$\chi^2(1)$ = 82.209, p < 0.001], StopType [$\chi^2(1)$ = 44.396, p < 0.001], and Voicing [$\chi^2(1)$ = 27.665, p < 0.001]. There was no main effect of Gender or PhonEnv [all $\chi^2(1)$ < 1, p > 0.1]. There were no significant interactions, although two interactions were marginal: StopType × Group [$\chi^2(1)$ = 3.805, p = 0.051] and StopType × Voicing × Group [$\chi^2(1)$ = 2.793, p = 0.095], which were reflected in model coefficients.

Model 10b coefficients and follow-up models provided evidence that children produced offset F_1 differences that went in the same direction as those produced by adults and were, moreover, larger than adult differences. Relative to adults' stop type effect for /p/–/k͡p/, in which complex stops were produced with lower offset F_1 than simplex ones (see Fig. 7), children showed an enhanced stop type effect, meaning a larger offset F_1 difference between complex and simplex stops [β = −45.520, p = 0.011]. This enhancement of the stop type effect for children was marginally reduced for the /b/–/g͡b/ contrast [β = 40.995, p = 0.095]. Nevertheless, follow-up models indicated that children produced significantly lower offset F_1 for complex than simplex stops for both the /p/–/k͡p/ contrast [β = −80.757, p < 0.001; cf. adults' β = −35.237] and the /b/–/g͡b/ contrast [β = −53.623, p < 0.001; cf. adults' β = −49.098].

3.2.3.2. Children's F_2. Results for F_2 suggested that children produced F_2 differences to signal complex–simplex stop contrasts, but much less consistently than they did for F_1. In the case of onset F_2, children produced lower values for complex than simplex stops (see Fig. 8). For the phrase-initial context, an ANOVA on Model 11 (Table A12) showed a main effect of

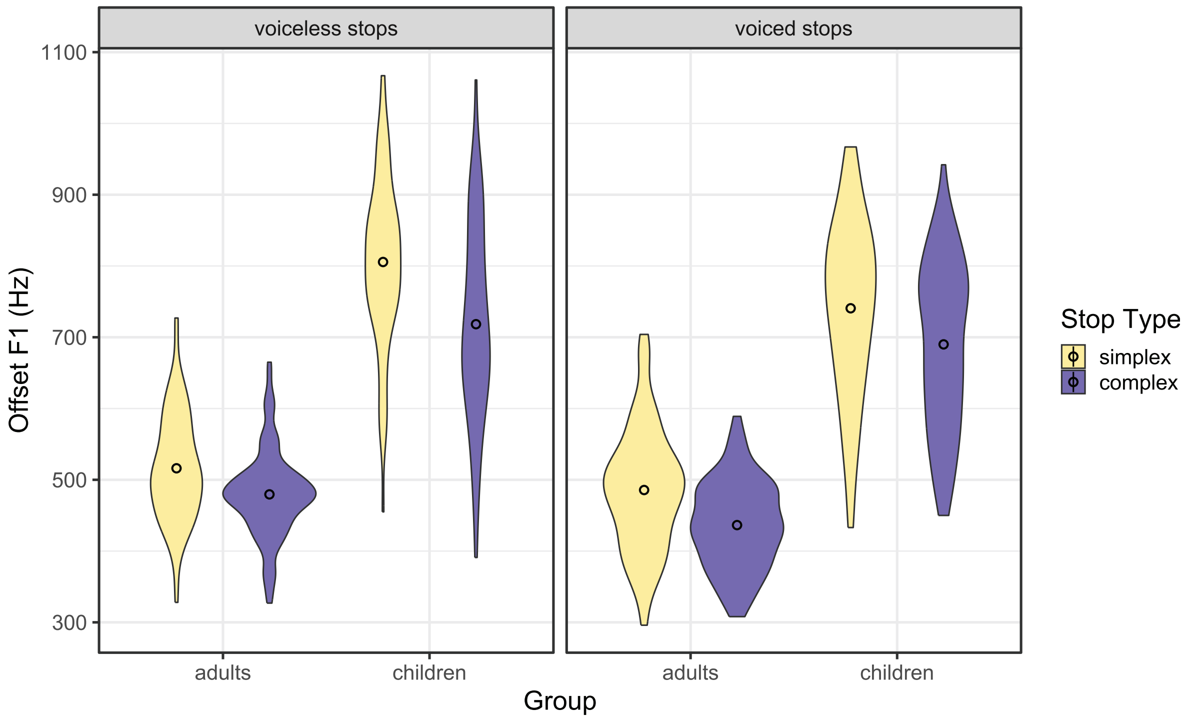

Fig. 7. Violin plot of offset F_1 preceding Gã stops in phrase-medial context, by voicing, group, and stop type. Hollow circles represent mean values. (color online only).

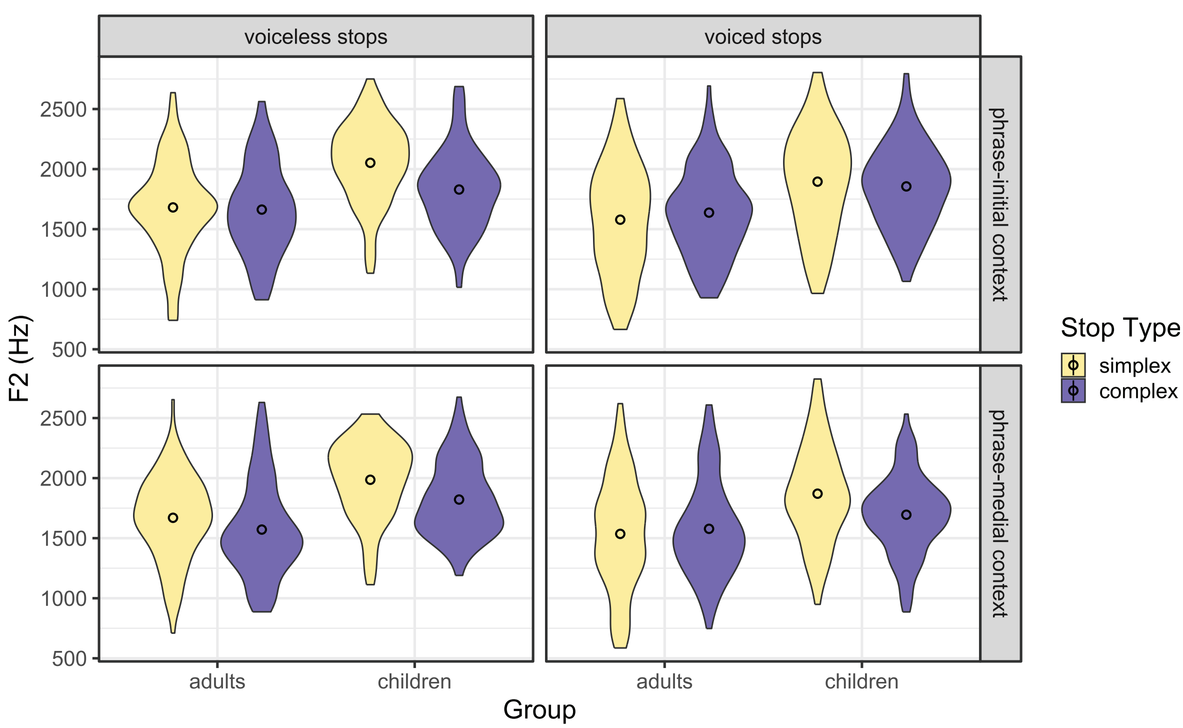

Fig. 8. Violin plot of onset F_2 following Gã stops, by voicing, context, group, and stop type. Hollow circles represent mean values. (color online only).

Group [$\chi^2(1) = 29.876$, $p < 0.001$] but not of StopType [$\chi^2(1) = 0.398$, $p = 0.528$] or Voicing [$\chi^2(1) = 0.628$, $p = 0.428$]. There was also an effect of Gender [$\chi^2(1) = 4.148$, $p = 0.042$], but not of PhonEnv [$\chi^2(1) = 0.026$, $p = 0.872$]. The only significant interaction was the StopType × Group interaction [$\chi^2(1) = 18.957$, $p < 0.001$].

Model 11 coefficients and follow-up models suggested that children differed from adults in their production of onset F_2 differences for complex–simplex stop contrasts in phrase-initial context. Relative to the non-significant stop type effect in adults, children produced a significantly larger F_2 difference for /p/–/$\widehat{\text{kp}}$/ [$\beta = -201.788$, $p < 0.001$]; however, this stop type effect was numerically smaller for /b/–/g͡b/ [β = 103.636, p = 0.127]. Accordingly, follow-up models indicated that while children's onset F_2 difference for the /p/–/k͡p/ contrast approached significance [β = −219.973, p = 0.076], their onset F_2 difference for the /b/–/g͡b/ contrast did not [β = −39.600, p = 0.734].

Results for the phrase-medial context also suggested that children produced larger differences in onset F_2 than adults, although these differences did not reach significance. An ANOVA on Model 12a (Table A13) showed a main effect of Group [$\chi^2(1)$ = 31.898, p < 0.001] but no main effects of StopType [$\chi^2(1)$ = 1.391, p = 0.238] or Voicing [$\chi^2(1)$ = 1.376, p = 0.241]. There was also an effect of Gender [$\chi^2(1)$ = 8.006, p = 0.005], but not of PhonEnv or PostPausal [all $\chi^2(1)$ < 1.5, p's > 0.1]. There was, again, a significant StopType × Group interaction [$\chi^2(1)$ = 16.053, p < 0.001], as well as a StopType × Voicing × Group interaction [$\chi^2(1)$ = 4.618, p = 0.032], but no other significant interactions. Model 12a coefficients revealed that, compared to adults, children produced a numerically, but non-significantly, larger onset F_2 difference for /p/–/k͡p/ [β = −62.808, p = 0.225], and this stop type effect was significantly larger for /b/–/g͡b/ [β = −153.650, p = 0.032]. However, follow-up models showed that children's onset F_2 distinction was not significant for either the /p/–/k͡p/ contrast [β = −161.326, p = 0.175] or the /b/–/g͡b/ contrast [β = −174.370, p = 0.127].

Results for offset F_2 in the phrase-medial context differed from those for onset F_2, in that children produced offset F_2 differences between the two stop types that were smaller than the differences that adults produced (see Fig. 9). An ANOVA on Model 12b (Table A14) showed main effects of Group [$\chi^2(1)$ = 212.026, p < 0.001] and StopType [$\chi^2(1)$ = 151.453, p < 0.001] but not of Voicing [$\chi^2(1)$ = 0.130, p = 0.718]. There was an effect of Gender [$\chi^2(1)$ = 6.003, p = 0.014] but not of PhonEnv [$\chi^2(1)$ = 0.078, p = 0.780]. There were also significant interactions for StopType × Group [$\chi^2(1)$ = 29.551, p < 0.001], StopType × Voicing [$\chi^2(1)$ = 4.894, p = 0.027], and StopType × Voicing × Group [$\chi^2(1)$ = 8.259, p = 0.004]. The Voicing × Group interaction was not significant [$\chi^2(1)$ = 0.052, p = 0.819].

Similar to the results for offset F_1, Model 12b coefficients and follow-up models provided evidence that children produced offset F_2 differences that went in the same direction as those produced by adults; however, child differences were generally smaller than adult differences. In contrast to adults' stop type effect for /p/–/k͡p/, in which complex stops were produced with higher offset F_2 than simplex ones (see Fig. 9), children showed hardly any stop type effect for /p/–/k͡p/, reflected in an interaction coefficient that largely reversed the stop type effect observed in adults [β = -196.129, p < 0.001]. This reversal was significantly weakened for /b/–/g͡b/, however [β = 133.538, p = 0.004]. Thus, follow-up models indicated that children's offset F_2 distinction was not significant for the /p/–/k͡p/ contrast [β = 34.994, p = 0.240; cf. adults' β = 231.123], but was significant for the /b/–/g͡b/ contrast [β = 180.729, p < 0.001; cf. adults' β = 243.320].

3.2.3.3. Children's F_3. Results for F_3 suggested that, to a limited degree, children also produced F_3 differences to signal complex–simplex stop contrasts. Across the board, the pattern of F_3 differences involved higher F_3 for complex than simplex stops (see Figs. 10 and 11). For onset F_3 in the phrase-initial context, an ANOVA on Model 13 (Table A15) showed a main effect of StopType [$\chi^2(1)$ = 5.701, p = 0.017] and Group [$\chi^2(1)$ = 161.256, p < 0.001] but not of Voicing [$\chi^2(1)$ = 0.049, p = 0.825]. There was no significant effect of Gender [$\chi^2(1)$ = 1.589, p = 0.207] but a marginal effect of PhonEnv

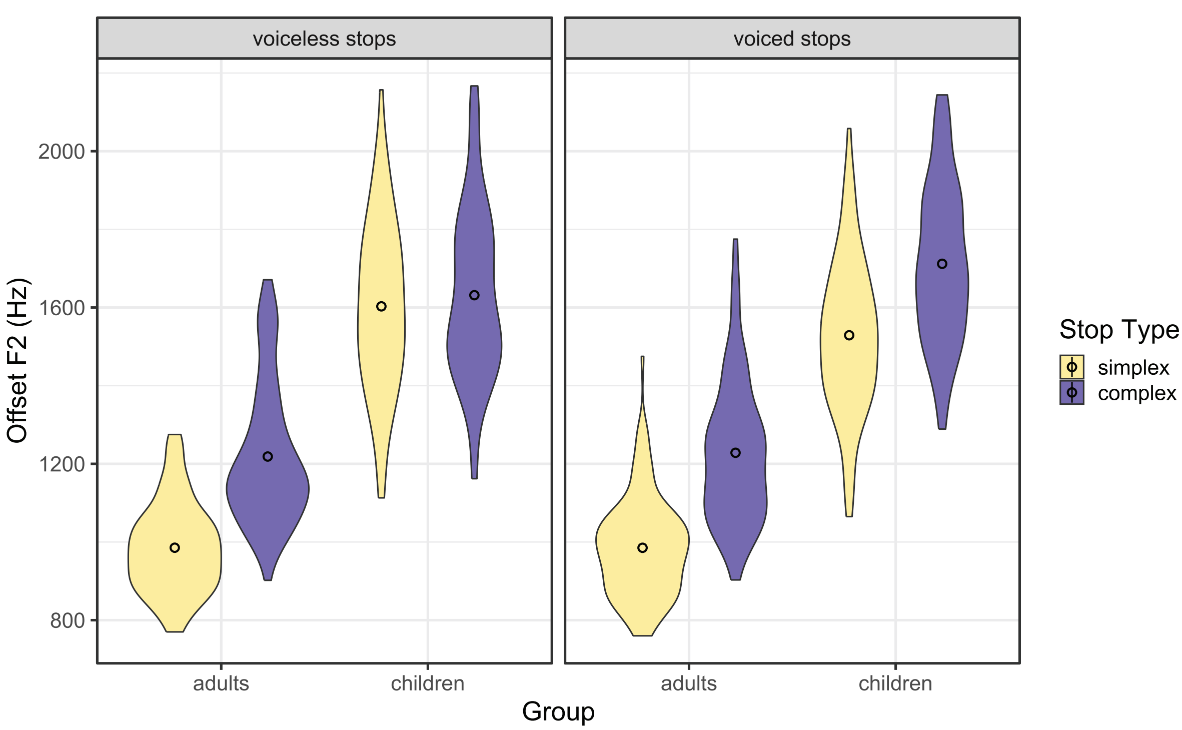

Fig. 9. Violin plot of offset F_2 preceding Gã stops in phrase-medial context, by voicing, group, and stop type. Hollow circles represent mean values. (color online only).

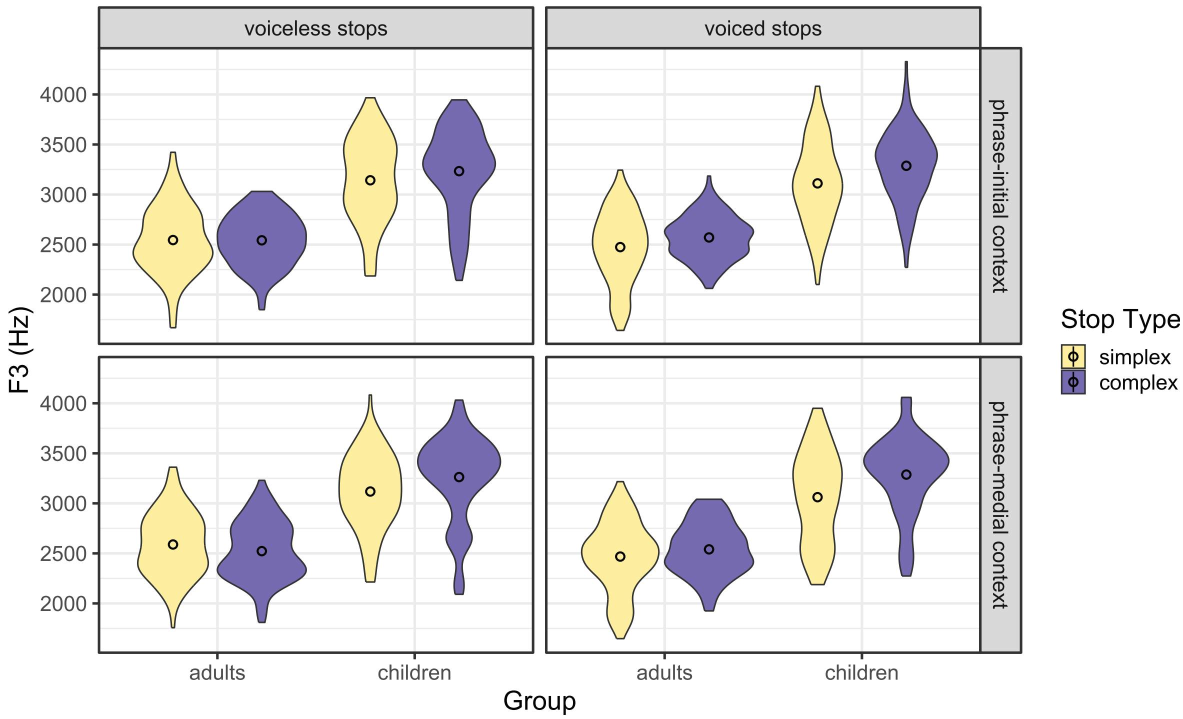

Fig. 10. Violin plot of onset F_3 following Gã stops, by voicing, context, group, and stop type. Hollow circles represent mean values. (color online only).

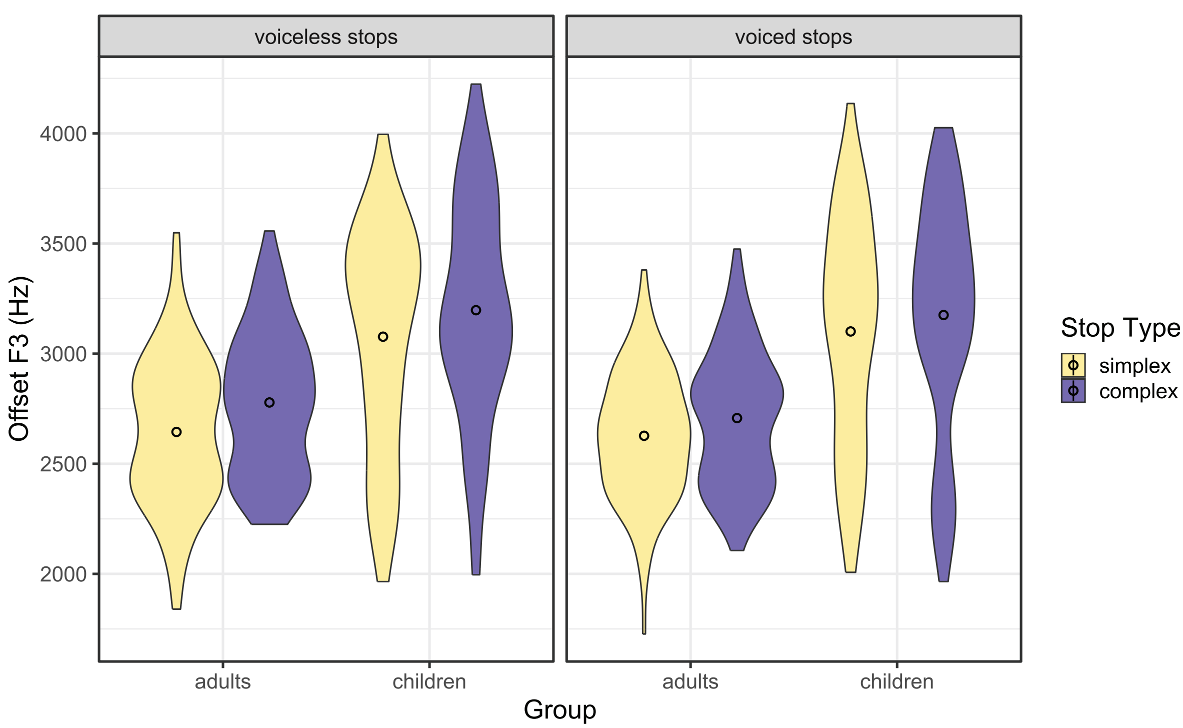

Fig. 11. Violin plot of offset F_3 preceding Gã stops in phrase-medial context, by voicing, group, and stop type. Hollow circles represent mean values. (color online only).

[$\chi^2(1)$ = 3.811, p = 0.051]. As above, there was a significant StopType × Group interaction [$\chi^2(1)$ = 4.173, p = 0.041]. No other interactions were significant.

Model 13 coefficients and follow-up models suggested that children differed from adults in their production of onset F_3 differences for complex–simplex stop contrasts in phrase-initial context. Whereas adults produced onset F_3 values for /k͡p/ that were slightly lower than those for /p/ [β = −2.633, p = 0.967], children produced higher onset F_3 values for /k͡p/ compared to /p/ [β = 94.087, p = 0.122], although this reversal of the stop type effect for children was not significant. Nevertheless, follow-up models showed that while children's onset F_3 distinction was not significant for /p/–/k͡p/ [β = 91.455, p = 0.154], it was significant for /b/–/g͡b/ [β = 174.653, p = 0.005].

Results for the phrase-medial context provided additional evidence of children's production of onset F_3 differences. An ANOVA on Model 14a (Table A16) showed a significant main effect of Group [$\chi^2(1)$ = 110.536, p < 0.001] and StopType [$\chi^2(1)$ = 5.894, p = 0.015] and no main effect of Voicing [$\chi^2(1)$ = 1.154, p = 0.283]. There was also an effect of Gender [$\chi^2(1)$ = 4.800, p = 0.028] and PhonEnv [$\chi^2(1)$ = 13.584, p < 0.001], but not of PostPausal [$\chi^2(1)$ = 0.005, p = 0.944]. The only significant interaction was the StopType $\times$ Group interaction [$\chi^2(1)$ = 18.163, p < 0.001]. Crucially, Model 14a coefficients indicated that, in contrast to adults, children produced significantly higher onset F_3 for phrase-medial /$\widehat{\text{kp}}$/ vis-à-vis /p/ [β = 210.622, p = 0.001]. Further, follow-up models confirmed that children's onset F_3 difference was significant for both the /p/–/$\widehat{\text{kp}}$/ contrast [β = 144.448, p = 0.026] and the /b/–/$\widehat{\text{gb}}$/ contrast [β = 226.417, p < 0.001].

Results for offset F_3 in the phrase-medial context again differed from those for onset F_3, but showed children producing F_3 differences that were similar to adults' (see Fig. 11). An ANOVA on Model 14b (Table A17) showed main effects of Group [$\chi^2(1)$ = 13.706, p < 0.001] and StopType [$\chi^2(1)$ = 15.979, p < 0.001], but not of Voicing [$\chi^2(1)$ = 1.246, p = 0.264]. There were no main effects of Gender [$\chi^2(1)$ = 0.029, p = 0.864] or PhonEnv [$\chi^2(1)$ = 2.292, p = 0.130]. No interactions were significant, including the StopType $\times$ Group interaction [$\chi^2(1)$ = 0.258, p = 0.611].

Model 14b coefficients and follow-up models suggested that children produced offset F_3 differences that went in the same direction as adults' but tended to be smaller in magnitude. Relative to adults' stop type effect for /p/–/$\widehat{\text{kp}}$/, children's stop type effect was not significantly different, but numerically smaller [β = -44.924, p = 0.526]. This weakening of the stop type effect for children was effectively cancelled for /b/–/$\widehat{\text{gb}}$/, however [β = 38.401, p = 0.694]. Regardless, follow-up models indicated that children's offset F_3 differences were not significant for the /p/–/$\widehat{\text{kp}}$/ contrast [β = 88.808, p = 0.104; cf. adults' β = 133.732] or the /b/–/$\widehat{\text{gb}}$/ contrast [β = 73.832, p = 0.152; cf. adults' β = 80.355].

4. Discussion

This study examined the acoustic implementation of complex–simplex stop contrasts in Gã in order to understand how children differ from adults in producing, as well as coarticulating, these typologically uncommon and articulatorily challenging consonants. Our findings suggest that, apart from limited f_0 differences, adult Gã speakers differentiated complex–simplex stop contrasts in phrase-initial context primarily in terms of differences in a temporal dimension (i.e., VOT). By contrast, child Gã speakers did so in terms of differences in both temporal (VOT) and spectral dimensions (onset f_0, onset formants). In phrase-medial context, adults differentiated the stop types in terms of temporal (closure duration) and spectral dimensions (offset formants and, to a limited degree, onset f_0). Children also did so, but with a different pattern of acoustic differences: robust differences in closure duration and onset formants and inconsistent differences in offset formants. Thus, in general, children implement these stop type contrasts across contexts in terms of a broader swath of acoustic dimensions than adults do, including clear differences in CV formant transitions. The results for all dependent variables are summarized in Table 5, including effect sizes in acoustic units.

Returning to our four predictions about the acoustic implementation of these stop type contrasts in different contexts (Section 1.4), we observe that these findings provide support for some, but not all, of these predictions. Recall that we predicted a context effect to augment certain acoustic differences between complex and simplex stops in the phrase-initial context (P1). In the end, we did not find support for P1: acoustic differences were not clearly enhanced in the phrase-initial context (see Figs. 5, 6, 8, and 10), and in the one instance of a significant StopType $\times$ Context interaction for adults, the interaction arose because of a larger difference occurring in the phrase-medial context (see Section 3.1.2).

On the other hand, our findings did provide evidence of children approximating the acoustic differences present in adult speech (P2), which was often reflected in the lack of a significant StopType $\times$ Group interaction in our modeling. Children largely reproduced the differences adults produced between the stop types in VOT and closure duration; in fact, the only aspect of adult differentiation of the complex–simplex stop contrasts that children did not replicate was offset formant differences (discussed further below). That said, we also observed variation in children's approximation of adult-like differences. For example, for voiced stops, children tended to be closer to adult-like on closure duration in phrase-medial context than on VOT in phrase-initial context. This context effect may reflect more difficulty with initiating and maintaining stop voicing in phrase-initial position as compared to phrase-medial position, where voiced stops can benefit from the voicing that is underway for a preceding vowel (see Westbury & Keating, 1986).

Table 5
Summary of effect sizes for complex–simplex contrasts (i.e., β for the fixed predictor StopType: complex), by group, context, contrast, and dependent variable. Units are ms for VOT and closure duration, Hz for f_0, F_1, F_2, and F_3. Significance codes: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.

Group	Context	Contrast	Dependent variable								
			VOT	Closure duration	Onset f_0	Onset F_1	Offset F_1	Onset F_2	Offset F_2	Onset F_3	Offset F_3
adults	initial	/$\widehat{\text{kp}}$/ – /p/	-36^{***}		0	25		−18		−3	
		/$\widehat{\text{gb}}$/ – /b/	-37^{***}		2	0		59		96	
	medial	/$\widehat{\text{kp}}$/ – /p/			−1	33	−35*	−98	231^{***}	−66	134^{**}
		/$\widehat{\text{gb}}$/ – /b/		29^{***}	$-15^{\dagger}$	−8	-49^{***}	42	243^{***}	72	$80^{\dagger}$
children	initial	/$\widehat{\text{kp}}$/ – /p/	-36^{***}		24*	85*		$-220^{\dagger}$		91	
		/$\widehat{\text{gb}}$/ – /b/	-25^{***}		7	$76^{\dagger}$		−40		175^{**}	
	medial	/$\widehat{\text{kp}}$/ – /p/			2	95*	-81^{***}	−161	35	144*	89
		/$\widehat{\text{gb}}$/ – /b/		24^{***}	−4	65	-54^{***}	−174	181^{***}	226^{***}	74

As for adult–child disparities in formant transitions, our findings supported the prediction of children producing more consistent differentiation of the stop types in onset formants as compared to adults (P3), as well as the prediction of less consistent differentiation in offset formants (P4). Although we did not expect the group disparity in onset formant differentiation to involve adults failing to produce significant onset formant differences altogether, this aspect of our results is not crucial to supporting P3 and, further, is unlikely to be due to statistical factors such as low power, as the same sample of adults was, conversely, observed to produce significant or marginal differences in all offset formants. Rather, we interpret the group disparity in onset formant differentiation to be the byproduct of children's tendency toward greater coarticulation of consonants with vowels, as compared to adults'.

Given that mature Gã speakers do not appear to implement complex–simplex stop contrasts in terms of CV formant transitions but Gã-dominant five-year-olds do, the current results are consistent with the view that the implementation of complex–simplex stop contrasts in Gã follows a developmental trajectory that begins with higher degrees of carryover coarticulation and proceeds to lower degrees of carryover coarticulation. Because CV formant transitions arise due to coarticulation, children's tendency toward greater coarticulation may play an important role in Gã children's greater differentiation of onset formants for complex–simplex stop contrasts as compared to Gã adults. This still leaves the question of why Gã children produce onset formant *differentiation* between the complex and simplex stops even though Gã adults do not. After all, an alternative outcome of coarticulatory amplification would be for children to produce large amounts of carryover coarticulation indiscriminately, resulting in no differentiation in onset formants between complex and simplex stops. Given this, as well as the likelihood that this pattern of coarticulatory differentiation is not in the adult input, where does it come from?

We believe there are two possible contributors to the children's coarticulatory patterns, which are not mutually exclusive: (1) strategy and (2) immature gestural coordination. In regard to strategy, given the articulatory challenges of initiating and maintaining prevoicing (see Section 1.2), a characteristic of both voiceless and voiced complex stops in Gã, Gã children may be initially predisposed toward exploring spectral methods of implementing complex–simplex stop contrasts, and their already-high degree of coarticulation may help establish a foothold on the way to doing so. Recall, however, that Gã five-year-olds were close to adult-like in terms of producing VOT and closure duration differences; therefore, we speculate that the point in development we have glimpsed at age 5 represents a period of transition, during which Gã children are mastering voicing-related durational dimensions even while still producing differentiation along spectral dimensions. This is speculative, however, and ultimately much more research examining Gã speakers of different ages is needed to understand when, and how, young Gã speakers begin to move away from spectral dimensions to focus on durational dimensions of complex–simplex stop contrasts. The crucial point is that, at the age of 5, Gã children do not produce carryover coarticulation passively, but rather have harnessed it to help implement complex–simplex stop contrasts.

As for immature gestural coordination, this explanation follows from our results on VC transitions (i.e., offset formant differentiation of the stop types in phrase-medial context). Recall that offset formant data indicated that Gã adults differentiate complex and simplex bilabial stops in terms of offset formants consistently, whereas Gã children do so less consistently. Notably, this data complements previous findings on Ibibio (Garnes, 1975; Connell, 1994) by providing evidence that, at least in the mid back vowel environment, these stop types in Gã can be robustly differentiated in terms of VC transitions. The offset formant data also rules out a passive, "overlap only" account of Gã children's onset formant differentiation that is purely based on children's aforementioned tendency toward greater coarticulation between consonants and vowels, because such an account incorrectly predicts that, compared to adults, children should more consistently produce differences for complex–simplex stop contrasts in all formant transitions, both onset and offset (cf. our P3 + P4). Contrary to this account, children in this study produced offset formant differentiation for complex–simplex stop contrasts *less* consistently than adults did, which may reflect immature gestural coordination of stops with vowels at word junctures or other morphological boundaries (see also Cychosz, 2020, 2021). Allowing morphological boundaries to unduly affect anticipatory coarticulation may in turn be related to immature speech planning, as discussed in Section 1.4 and reflected in the children's high rate of pausing in the phrase-medial context. Although we interpret these results cautiously (because children's results for offset formants are based on a smaller dataset than adults', due to exclusions for pausing), they reveal that the children's production of coarticulation is, indeed, not fully adult-like.

Consequently, we argue that Gã children's more consistent differentiation of the complex–simplex stop contrasts in terms of onset formants, but less consistent differentiation of these contrasts in terms of offset formants, is the outcome of children capitalizing on their typically ample coarticulation between consonants and vowels (i.e., strategy) while still grappling with the limitations of their current articulatory and/or planning abilities (i.e., immature gestural coordination). Crucially, this account presents a dual picture of Gã five-year-olds as developing producers of speech: sophisticated enough to find ways of producing difficult contrasts that align well with their articulatory predispositions, yet not quite skilled enough to implement a given production strategy consistently across contexts. If this picture is on the right track, then we expect that future research may uncover a later developmental stage in which Gã children of older ages, equipped with more refined gestural coordination, continue to produce onset formant differentiation and additionally produce consistent offset formant differentiation. Thus, there are clear avenues for future work to delve deeper into the developmental variation found here.

Our findings of developmental variation in both anticipatory and carryover coarticulation of Gã stop contrasts dovetail with previous findings (e.g., Rubertus, 2024) and have a number of implications for models of coarticulation. First, our finding of less carryover coarticulation in adults than children lends support to models of coarticulation, such as the coproduction model, that predict a developmental decrease in coarticulation with progressive gestural refinement, but is inconsistent with other models, such as the feature spreading model, the window model, and the DIVA model, that predict a developmental increase in coarticulation due to the computation needed to generate gradient gestural overlap. On the other hand, our finding of more anticipatory coarticulation in adults than children has the complementary implication for these models. This finding also converges with the results of Graetzer (2007), which suggest that the presence of conflict between the lingual gestures of an adjacent consonant–vowel pair may lead to greater variation in coarticulation; in our case, such a conflict occurred in the anticipatory context, but not in the carryover context, and could possibly interact with children's immature gestural coordination to lead to the less consistent anticipatory coarticulation observed in children as compared to adults. In addition, our finding of disparities between carryover and anticipatory coarticulation, both among adults and among children, are more in line with certain models, such as the DIVA model, that view these different directions of coarticulation in terms of different sources. Thus, taken together, our findings are not fully consistent with any one existing model, but rather suggest that a comprehensive model will need to be able to predict both developmental decreases and increases in coarticulation—or, more precisely, coarticulatory differentiation of gesturally distinct segments—as well as disparities between carryover and anticipatory coarticulation.

In addition to implications for models of coarticulation, the current findings have implications for the phonetic typology of complex–simplex stop contrasts, converging with previous results in some ways and diverging in others. As shown in Table 3, /k͡p/ has a more negative VOT than /p/ in Gã, which is consistent with Maddieson's (1993) results for /k͡p/ and /p/ in Ewe, a related language. As for the voiced complex–simplex contrast, /g͡b/ also has a more negative VOT than /b/ in Gã. In other words, both /k͡p/ and /g͡b/ are prevoiced (i.e., characterized by negative VOT), but prevoicing is longer for /g͡b/ than /k͡p/. Overall, these voicing characteristics are consistent with those reported for complex stops in Yoruba (Ladefoged, 1964; Puech, 1989; Grawunder et al., 2011) and Ibibio (Garnes, 1975; Connell, 1987, 1991). Our results for closure duration, which showed that /g͡b/ has a longer closure duration than /b/ in Gã (Table 4), converge with results on Ibibio, Igbo, and Obolo (Connell, 1994) but diverge from results on Ewe (Maddieson & Ladefoged, 1989), where complex and simplex stops do not show a significant closure duration difference.

Our findings on spectral dimensions of complex–simplex stop contrasts in Gã also differ in some ways from results on other languages. First, our results for f_0 in Gã, which showed adults using onset f_0 only for the /b/-/g͡b/ contrast in phrase-medial context, are partly inconsistent with results on Yoruba (Cahill, 2006), where the /b/-/g͡b/ contrast is realized in part through f_0 differences in phrase-initial context. Notably, other studies have also documented a voicing effect on f_0 (i.e., higher f_0 after voiceless than voiced stops; see House & Fairbanks, 1953; Ohde, 1984; Dmitrieva et al., 2015), but we observed few, and inconsistent, f_0 differences between voicing categories in Gã. Second, our results for onset formants in Gã, which provided no evidence of adults' differentiation of complex–simplex stop contrasts in these dimensions, contrast with results on other languages, in which complex stops show a lower F_2 locus and steeper F_2 transition (implying a lower onset F_2) than simplex bilabial stops (e.g., Ladefoged, 1964; Garnes, 1975; Connell, 1994; Cahill, 2006). Thus, complex–simplex stop contrasts in Gã appear to differ from those in other languages in not clearly being signaled through CV formant transitions. However, we are cautious to point out that, in respect to CV formant transitions, we measured only onset frequencies in this study, leaving open the possibility that mature speakers use other aspects of CV formant transitions such as slope to distinguish complex from simplex stops in Gã.

In closing, we would like to acknowledge some limitations of this study, and point out a few promising directions for future research. In regard to limitations, first, we were not able to perfectly control the items used to elicit complex and simplex stops. For example, the post-stop sonorant in the items was left to vary, because there were not enough common, picturable words in the lexicon to be able to limit the items to only one post-stop environment; instead, we attempted to balance the items representing the two stop types in terms of the variety of following vowel types and tones. Although phonological variation among the items is largely accounted for via the fixed effect of PhonEnv and the random effect of Item included in our statistical modeling, these two effects may not fully account for the potential effect of phonological differences between the item sets for the two stop types. Thus, in future work, it would be useful to explore the potential effects of phonological variation (e.g., following vowel quality) that the current study was not designed to examine. In addition, as mentioned in Section 2.1, the Gã community in Ghana is undergoing demographic and linguistic changes, with the result that there may be some differences in language experience between adults and children that cannot be fully controlled, such as in exposure and use of other languages and in the vitality of Gã at the time of acquisition. We cannot say for sure whether the developmental differences found in this study are influenced by such experiential differences, but we can say that any such effects are unlikely to be simple. For example, one Gã speaker more exposed to Akan than another may or may not be adversely affected in their acquisition of complex stops, given the presence of contrastively labialized velars in Akan. Thus, the role of language ecology, including multilingualism and language vitality, is an important topic for future research on the development of Gã.

Besides looking at phonological effects and language experience effects, there are several other directions for future research on the development of complex–simplex stop contrasts. First, as mentioned above, examining Gã children of different ages will help shed light on the plausibility of our account of Gã children's pattern of onset formant differentiation. Second, given the limitation of the current study in respect to closure duration in the phrase-medial context, we have data on Gã children's closure durations only for the voiced stops and not for the voiceless stops /p/ and /k͡p/; therefore, it would be useful in further work to elicit the voiceless stops in phrase-medial context, using alternative methods that may be more successful at encouraging production of connected speech by children. Third, as alluded to in Section 1.2, children's production may not only diverge from adults' acoustically, but also auditorily. Collecting auditory-based judgment data on Gã children's production will therefore provide a complementary perspective on the quality of their developing speech abilities, including whether their acoustically adult-like productions (in terms of the properties examined here) are actually perceived by adult Gã listeners as target-like or, instead, as reduced (i.e., sounding like simplex stops; see Kpogo et al., 2021). Fourth, above we made the assumption that children's input would resemble the adult speech analyzed here, but it is possible that Gã adults produce complex–simplex stop contrasts differently in child-directed speech. Thus, examining acoustic differentiation of the complex–simplex stop contrasts in child-directed speech would help sharpen our understanding of the degree to which children's production patterns are converging with or diverging from their input. Finally, it is worth pointing out that we have provided developmental acoustic data on complex–simplex stop contrasts in only one language, and there remain many other understudied languages for which such developmental data is completely lacking. Thus, it is our hope that the current study will be part of a trend of future investigations that contribute data on child speech in other African languages, strengthening the empirical basis of claims about both phonetic typology as well as phonetic and phonological development.

Declarations of interest

None.

CRediT authorship contribution statement

Felix Kpogo: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Charles B. Chang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Formal analysis, Data curation.

Acknowledgments

The authors gratefully acknowledge financial support from the Department of Linguistics at Boston University, logistical assistance from St. Mary's Anglican School (Accra), research assistance from Kevin Samejon and Jonathan Nsiah Tetteh, and helpful feedback from Jonathan Barnes, Kate Lindsey, four anonymous reviewers, and audience members at Harvard's Language and Cognition Group and the 95th Annual Meeting of the Linguistic Society of America. Any errors, however, are those of the authors.

Appendix. Mixed-effects regression model output

Significance codes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A1
Fixed-effect coefficients in Model 1 of VOT in phrase-initial context [N = 912]. Model formula: VOT (ms) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	15.597	5.917	2.636	0.011*
StopType: complex	−35.912	6.407	−5.605	<0.001***
Voicing: voiced	−139.264	6.264	−22.231	<0.001***
Group: children	3.053	6.584	0.464	0.646
Gender: female (vs. grand mean)	−1.639	2.873	−0.571	0.576
PhonEnv: lateral (vs. grand mean)	4.252	2.018	2.107	0.041*
StopType: complex × Voicing: voiced	−0.797	8.866	−0.090	0.929
StopType: complex × Group: children	−0.253	5.313	−0.048	0.962
Voicing: voiceless × Group: children	17.805	5.194	3.428	0.001**
StopType: complex × Voicing: voiced × Group: children	12.414	7.354	1.688	0.092†

Table A2
Fixed-effect coefficients in Model 2 of closure duration for voiced stops in phrase-medial context [N = 452]. Model formula: Duration (ms) ~ StopType * Group + Gender + PhonEnv + PostPausal + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	118.722	8.600	13.805	<0.001***
StopType: complex	28.790	5.539	5.198	<0.001***
Group: children	−3.201	9.192	−0.348	0.732
Gender: female (vs. grand mean)	3.902	4.429	0.881	0.393
PhonEnv: lateral (vs. grand mean)	−3.970	2.534	−1.567	0.132
PostPausal: yes (vs. grand mean)	14.338	5.176	2.770	0.007**
StopType: complex × Group: children	−4.423	4.747	−0.932	0.352

Table A3
Fixed-effect coefficients in Model 3 of adults' f_0 [N = 912]. Model formula: F0 (Hz) ~ StopType * Voicing * Context + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	209.021	8.621	24.245	<0.001***
StopType: complex	−0.032	8.164	−0.004	0.997
Voicing: voiced	−12.494	7.998	−1.562	0.125
Context: medial	−9.502	2.837	−3.349	0.001**
Gender: female (vs. grand mean)	27.920	6.436	4.338	0.002**
PhonEnv: lateral (vs. grand mean)	0.067	2.743	0.024	0.981
StopType: complex × Voicing: voiced	2.202	11.307	0.195	0.846
StopType: complex × Context: medial	−0.547	4.007	−0.137	0.891
Voicing: voiced × Context: medial	10.743	3.926	2.736	0.006**
StopType: complex × Voicing: voiced × Context: medial	−16.279	5.552	−2.932	0.003**

Table A4
Fixed-effect coefficients in Model 4 of adults' onset F_1 [N = 904]. Model formula: F1 (Hz) ~ StopType * Voicing * Context + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	452.347	29.805	15.177	<0.001***
StopType: complex	25.409	41.321	0.615	0.542
Voicing: voiced	–22.515	40.481	−0.556	0.581
Context: medial	−8.022	8.406	−0.954	0.340
Gender: female (vs. grand mean)	23.534	6.103	3.856	0.005**
PhonEnv: lateral (vs. grand mean)	−34.814	14.167	−2.457	0.018*
StopType: complex × Voicing: voiced	−25.775	57.213	−0.451	0.655
StopType: complex × Context: medial	7.282	11.916	0.611	0.541
Voicing: voiced × Context: medial	14.285	11.685	1.223	0.222
StopType: complex × Voicing: voiced × Context: medial	−14.965	16.526	−0.906	0.365

Table A5
Fixed-effect coefficients in Model 5 of adults' onset F_2 [N = 920]. Model formula: F2 (Hz) ~ StopType * Voicing * Context + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	1684.531	103.040	16.348	<0.001***
StopType: complex	−18.185	141.867	−0.128	0.899
Voicing: voiced	−105.783	138.950	−0.761	0.451
Context: medial	−11.271	31.176	−0.362	0.718
Gender: female (vs. grand mean)	63.889	24.346	2.624	0.030*
PhonEnv: lateral (vs. grand mean)	−39.841	48.555	−0.821	0.417
StopType: complex × Voicing: voiced	76.736	196.406	0.391	0.698
StopType: complex × Context: medial	−79.841	44.089	−1.811	0.071†
Voicing: voiced × Context: medial	−31.702	43.161	−0.734	0.463
StopType: complex × Voicing: voiced × Context: medial	63.379	61.039	1.038	0.299

Table A6
Fixed-effect coefficients in Model 6 of adults' onset F_3 [N = 920]. Model formula: F3 (Hz) ~ StopType * Voicing * Context + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	2551.061	50.579	50.437	<0.001***
StopType: complex	−2.633	65.249	−0.040	0.968
Voicing: voiced	−76.495	63.904	−1.197	0.237
Context: medial	43.803	31.189	1.404	0.161
Gender: female (vs. grand mean)	112.155	22.000	5.098	0.001**
PhonEnv: lateral (vs. grand mean)	−66.975	21.276	−3.148	0.003**
StopType: complex × Voicing: voiced	98.912	90.333	1.095	0.279
StopType: complex × Context: medial	−63.584	44.107	−1.442	0.150
Voicing: voiced × Context: medial	−50.230	43.179	−1.163	0.245
StopType: complex × Voicing: voiced × Context: medial	39.493	61.064	0.647	0.518

Table A7
Fixed-effect coefficients in Model 7 of children's and adults' f_0 in phrase-initial context [N = 916]. Model formula: F0 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	209.257	12.362	16.927	<0.001***
StopType: complex	−0.032	9.540	−0.003	0.997
Voicing: voiced	−12.694	9.348	−1.358	0.180
Group: children	146.719	15.447	9.498	<0.001***
Gender: female (vs. grand mean)	10.934	7.487	1.460	0.162
PhonEnv: lateral (vs. grand mean)	−2.526	3.102	−0.814	0.420
StopType: complex × Voicing: voiced	2.275	13.217	0.172	0.864
StopType: complex × Group: children	24.035	6.608	3.637	<0.001***
Voicing: voiced × Group: children	17.123	6.476	2.644	0.008**
StopType: complex × Voicing: voiced × Group: children	−18.953	9.158	−2.070	0.039*

Table A8
Fixed-effect coefficients in Model 8 comparing children's and adults' f_0 in phrase-medial context [N = 864]. Model formula: F0 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + PostPausal + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	184.380	14.555	12.668	<0.001***
StopType: complex	−0.833	10.808	−0.077	0.939
Voicing: voiced	−2.118	10.585	−0.200	0.842
Group: children	169.082	17.979	9.404	<0.001***
Gender: female (vs. grand mean)	5.897	8.867	0.665	0.516
PhonEnv: lateral (vs. grand mean)	0.011	3.616	0.003	0.998
PostPausal: yes (vs. grand mean)	−15.521	3.850	−4.031	<0.001***
StopType: complex × Voicing: voiced	−13.807	14.960	−0.923	0.361
StopType: complex × Group: children	2.903	5.944	0.488	0.625
Voicing: voiced × Group: children	−17.349	5.824	−2.979	0.003**
StopType: complex × Voicing: voiced × Group: children	7.340	8.208	0.894	0.371

Table A9
Fixed-effect coefficients in Model 9 comparing children's and adults' onset F_1 in phrase-initial context [N = 909]. Model formula: F1 (Hz) $\sim$ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	452.776	29.480	15.359	<0.001***
StopType: complex	25.489	40.294	0.633	0.530
Voicing: voiced	−24.840	39.487	−0.629	0.532
Group: children	197.289	17.151	11.503	<0.001***
Gender: female (vs. grand mean)	16.454	6.285	2.618	0.018*
PhonEnv: lateral (vs. grand mean)	−37.876	13.562	−2.793	0.008**
StopType: complex × Voicing: voiced	−24.529	55.801	−0.440	0.662
StopType: complex × Group: children	59.326	19.008	3.121	0.002**
Voicing: voiced × Group: children	–33.591	18.659	−1.800	0.072†
StopType: complex × Voicing: voiced × Group: children	15.756	26.336	0.598	0.550

Table A10
Fixed-effect coefficients in Model 10a comparing children's and adults' onset F_1 in phrase-medial context [N = 863]. Model formula: F1 (Hz) $\sim$ StopType * Voicing * Group + Gender + PhonEnv + PostPausal + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	421.004	32.233	13.061	<0.001***
StopType: complex	31.604	41.016	0.771	0.445
Voicing: voiced	−8.937	40.169	−0.222	0.825
Group: children	200.926	20.504	9.799	<0.001***
Gender: female (vs. grand mean)	7.819	8.374	0.934	0.364
PhonEnv: lateral (vs. grand mean)	−29.799	13.873	−2.148	0.038*
PostPausal: yes (vs. grand mean)	–23.288	9.881	−2.357	0.020*
StopType: complex × Voicing: voiced	−39.268	56.795	−0.691	0.493
StopType: complex × Group: children	63.570	18.837	3.375	0.001**
Voicing: voiced × Group: children	−35.069	18.462	−1.899	0.058†
StopType: complex × Voicing: voiced × Group: children	8.770	26.123	0.336	0.737

Table A11
Fixed-effect coefficients in Model 10b comparing children's and adults' offset F_1 in phrase-medial context [N = 771]. Model formula: F1 (Hz) $\sim$ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	516.271	19.588	26.357	<0.001***
StopType: complex	−35.237	13.462	−2.618	0.011*
Voicing: voiced	−30.027	13.159	−2.282	0.025*
Group: children	271.387	28.974	9.367	<0.001***
Gender: female (vs. grand mean)	8.585	13.339	0.644	0.531
PhonEnv: lateral (vs. grand mean)	2.793	3.933	0.710	0.482
StopType: complex × Voicing: voiced	−13.861	18.618	−0.745	0.459
StopType: complex × Group: children	−45.520	17.823	−2.554	0.011*
Voicing: voiced × Group: children	−31.226	17.307	−1.804	0.072†
StopType: complex × Voicing: voiced × Group: children	40.995	24.529	1.671	0.095†

Table A12

Fixed-effect coefficients in Model 11 comparing children's and adults' onset F_2 in phrase-initial context [N = 918]. Model formula: F2 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	1680.307	91.211	18.422	<0.001***
StopType: complex	−18.185	121.099	−0.150	0.881
Voicing: voiced	−101.559	118.608	−0.856	0.396
Group: children	362.531	56.331	6.436	<0.001***
Gender: female (vs. grand mean)	48.612	23.868	2.037	0.058†
PhonEnv: lateral (vs. grand mean)	6.625	41.093	0.161	0.873
StopType: complex × Voicing: voiced	76.736	167.654	0.458	0.649
StopType: complex × Group: children	−201.788	49.079	−4.111	<0.001***
Voicing: voiced × Group: children	−55.668	47.990	−1.160	0.246
StopType: complex × Voicing: voiced × Group: children	103.636	67.868	1.527	0.127

Table A13

Fixed-effect coefficients in Model 12a comparing children's and adults' onset F_2 in phrase-medial context [N = 873]. Model formula: F2 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + PostPausal + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	1645.074	90.107	18.257	<0.001***
StopType: complex	−98.518	116.419	−0.846	0.402
Voicing: voiced	−136.319	114.024	−1.196	0.238
Group: children	305.035	53.981	5.651	<0.001***
Gender: female (vs. grand mean)	60.756	21.472	2.829	0.012*
PhonEnv: lateral (vs. grand mean)	−21.611	39.439	−0.548	0.587
PostPausal: yes (vs. grand mean)	−27.020	26.420	−1.023	0.309
StopType: complex × Voicing: voiced	140.607	161.174	0.872	0.388
StopType: complex × Group: children	−62.808	51.674	−1.215	0.225
Voicing: voiced × Group: children	20.928	50.610	0.414	0.679
StopType: complex × Voicing: voiced × Group: children	−153.650	71.503	−2.149	0.032*

Table A14

Fixed-effect coefficients in Model 12b comparing children's and adults' offset F_2 in phrase-medial context [N = 771]. Model formula: F2 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	986.965	27.817	35.480	<0.001***
StopType: complex	231.123	25.788	8.963	<0.001***
Voicing: voiced	1.539	25.285	0.061	0.952
Group: children	598.923	40.048	14.955	<0.001***
Gender: female (vs. grand mean)	41.576	16.970	2.450	0.028*
PhonEnv: lateral (vs. grand mean)	2.121	7.576	0.280	0.781
StopType: complex × Voicing: voiced	12.197	35.747	0.341	0.734
StopType: complex × Group: children	−196.129	33.660	−5.827	<0.001***
Voicing: voiced × Group: children	−72.016	32.823	−2.194	0.029*
StopType: complex × Voicing: voiced × Group: children	133.538	46.468	2.874	0.004**

Table A15
Fixed-effect coefficients in Model 13 comparing children's and adults' onset F_3 in phrase-initial context [N = 919]. Model formula: F3 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	2548.393	55.720	45.736	<0.001***
StopType: complex	−2.633	63.377	−0.042	0.967
Voicing: voiced	−73.827	62.068	−1.189	0.238
Group: children	590.827	63.717	9.273	<0.001***
Gender: female (vs. grand mean)	32.641	25.895	1.261	0.225
PhonEnv: lateral (vs. grand mean)	−37.626	19.273	−1.952	0.058†
StopType: complex × Voicing: voiced	98.912	87.742	1.127	0.264
StopType: complex × Group: children	94.087	60.770	1.548	0.122
Voicing: voiced × Group: children	39.595	59.490	0.666	0.506
StopType: complex × Voicing: voiced × Group: children	−15.713	84.176	−0.187	0.852

Table A16
Fixed-effect coefficients in Model 14a comparing children's and adults' onset F_3 in phrase-medial context [N = 872]. Model formula: F3 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + PostPausal + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	2597.405	65.801	39.474	<0.001***
StopType: complex	−66.174	61.823	−1.070	0.288
Voicing: voiced	−126.929	60.547	−2.096	0.040*
Group: children	506.522	71.136	7.121	<0.001***
Gender: female (vs. grand mean)	65.332	29.821	2.191	0.043*
PhonEnv: lateral (vs. grand mean)	−69.695	18.910	−3.686	0.001**
PostPausal: yes (vs. grand mean)	2.336	33.221	0.070	0.944
StopType: complex × Voicing: voiced	138.363	85.589	1.617	0.111
StopType: complex × Group: children	210.622	61.426	3.429	0.001**
Voicing: voiced × Group: children	64.464	60.162	1.072	0.284
StopType: complex × Voicing: voiced × Group: children	−56.393	85.160	−0.662	0.508

Table A17
Fixed-effect coefficients in Model 14b comparing children's and adults' offset F_3 in phrase-medial context [N = 774]. Model formula: F3 (Hz) ~ StopType * Voicing * Group + Gender + PhonEnv + (1 | Participant) + (1 | Item).

Predictor	β	*SE*	*t*	*Pr(>|t|)*
(Intercept)	2646.249	84.274	31.401	<0.001***
StopType: complex	133.732	45.024	2.970	0.003**
Voicing: voiced	−19.121	44.181	−0.433	0.665
Group: children	446.668	128.792	3.468	0.002**
Gender: female (vs. grand mean)	10.303	60.337	0.171	0.867
PhonEnv: lateral (vs. grand mean)	−18.192	12.016	−1.514	0.130
StopType: complex × Voicing: voiced	−53.377	62.325	−0.856	0.392
StopType: complex × Group: children	−44.924	70.754	−0.635	0.526
Voicing: voiced × Group: children	27.519	68.989	0.399	0.690
StopType: complex × Voicing: voiced × Group: children	38.401	97.637	0.393	0.694

References

Abakarova, D., Fuchs, S., & Noiray, A. (2022). Developmental changes in coarticulation degree relate to differences in articulatory patterns: An empirically grounded modeling approach. *Journal of Speech, Language, and Hearing Research, 65*, 3276–3299.

Ajolore, O. (1974). *Learning to use Yoruba focus sentences in a multilingual setting*. University of Illinois at Urbana-Champaign. Doctoral dissertation.

Akpanglo-Nartey, J. N., & Akpanglo-Nartey, R. A. (2012). Some endangered languages of Ghana. *American Journal of Linguistics, 1*(2), 10–18.

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K., & Perkell, J. S. (2015). Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses. In *Proceedings of the 16th International Conference on Spoken Language Processing (INTERSPEECH-2015)* (pp. 374–378). Dresden, Germany: International Speech Communication Association.

Bibiebome, E. Z., Anderson, J., & Jones-Mensah, I. (2019). Language shift and maintenance of Ga in Accra. *Journal of Education and Practice, 10*(15), 87–94.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer.* Version 6.0.37, retrieved 14 September 2019, from http://www.praat.org/.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*(1986), 219–252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(2), 201–251.

Cahill, M. (1999). Aspects of the phonology of labial-velar stops. *Studies in African Linguistics, 28*, 155–187.

Cahill, M. (2006). Perception of Yoruba word-initial [gb͡] and [b]. *Selected Proceedings of the 36th Annual Conference on African Linguistics*, Vol. 1, 37–41.

Connell, B. (1987). Temporal aspects of labiovelar stops. *Works in Progress from the Dept. of Linguistics. Univ. of Edinburgh, 20*, 53–60.

Connell, B. (1991). Accounting for the reflexes of labial-velar stops. In M. Rossi (Ed.), *Proceedings of the 12th International Congress of Phonetic Sciences (ICPhS XII)*, Vol. 3, 110–113.

Connell, B. (1994). The structure of labial-velar stops. *Journal of Phonetics, 22*(4), 441–476.

Cychosz, M. (2020). *Phonetic development in an agglutinating language*. Berkeley: University of California. Doctoral dissertation.

Cychosz, M. (2021). The coarticulation-duration relationship in early Quechua speech. *Journal of Phonetics, 87* 101052.

Cychosz, M., Munson, B., & Edwards, J. R. (2021). Practice and experience predict coarticulation in child speech. *Language Learning and Development, 17*, 366–396.

Dakubu, M. E. K. (1968). *A comparative study of Ga and Adangme with special reference to the verb*. SOAS University of London. Doctoral dissertation.

Dakubu, M. E. K. (2002). *Ga phonology. Monograph 6. Legon: Institute of African Studies*. University of Ghana.

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics, 1*(3), 239–248.

Demolin, D. (1991). Les consonnes labio-vélaires du mangbétu. *Pholia, 6*, 85–105.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49*, 77–95.

Dogil, G. (1988). On the acoustic structure of multiply articulated stop consonants (labio-velars). *Wiener Linguistische Gazette, 42–43*, 3–55.

Esene Agwara, A. D. (2020). What an ethnographically informed questionnaire can contribute to the understanding of traditional multilingualism research: Lessons from Lower Fungom. In P. Di Carlo & J. Good (Eds.), *African multilingualisms: Rural linguistic and cultural diversity* (pp. 181–203). Lanham, MD: Lexington Books.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*(1), 113–133.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd edition). Thousand Oaks: Sage.

Garnes, S. (1975). An acoustic analysis of double articulations in Ibibio. *Ohio State University Working Papers Linguistics, 20*, 44–55.

Gathercole, V. C. M., & Thomas, E. M. (2007). Factors contributing to language transmission in bilingual families: The core study – adult interviews. In V. C. Gathercole (Ed.), *Language transmission in bilingual families in Wales* (pp. 59–181). Cardiff: Welsh Language Board.

Ghana Statistical Service. (2013). *2010 population and housing census: National analytical report*, Chapter 4: Population size, composition and age-sex structure. Retrieved 18 September 2024, from https://www2.statsghana.gov.gh/docfiles/2010phc/National_Analytical_Report.pdf.

Ghana Statistical Service. (2021). *Ghana 2021 population and housing census: General report*, Vol. 3C: Background characteristics. Retrieved 18 September 2024, from https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/2021%20PHC%20General%20Report%203C_revised%20print_281121a.pdf.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*, 382–397.

Graetzer, S. (2007). Consonantal coarticulation resistance in vowel-consonant-vowel sequences. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 893–896). Dudweiler, Germany: Pirrot.

Grawunder, S., Winter, B., & Atoyebia, J. (2011). Voicing of labiovelar stops in Yoruba. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 767–770.

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43*, 239–255.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*(3), 594–621.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105–113.

Howson, P. J., & Redford, M. A. (2021). The acquisition of articulatory timing for liquids: Evidence from child and adult speech. *Journal of Speech, Language, and Hearing Research, 64*, 734–753.

Huber, M. (2008). Ghanaian Pidgin English: Phonology. In R. Mesthrie (Ed.), *Varieties of English 4: Africa, South and Southeast Asia* (pp. 93–101). Berlin: Mouton de Gruyter.

Isaiah A. A. (2015). Segmental substitution patterns in Yoruba phonological acquisition. Paper presented at the 46th Annual Conference on African Linguistics (ACAL 46), March 26–28, University of Oregon, USA.

Keating, P. A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 451–470). Cambridge: Cambridge University Press.

Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics, 2*, 195–210.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language, 42*, 703–735.

Kotey, P. F. A. (1974). Consonant labialization and consonant clusters in Ga. *Journal of West African Languages, 9*, 49–56.

Kpogo, F., Gathercole, V. C. M., & Tetteh, J. N. (2021). Acquisition of doubly articulated stops among Ga-speaking children. *Journal of African Languages and Linguistics, 42*, 101–146.

Kropp, M. E. (1968). An analysis of the consonant system of Gã. *Journal of West African Languages, 5*, 59–61.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.

Ladefoged, P. (1964). *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge: Cambridge University Press.

Levy, H., & Hanulí´ková, A. (2019). Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology, 10*, 9.

Lin, S., & Demuth, K. (2015). Children's acquisition of English onset and coda /l/: Articulatory evidence. *Journal of Speech, Language, and Hearing Research, 58*, 13–27.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*(11), 1773–1781.

Lindblom, B., & MacNeilage, P. (2011). Coarticulation: A universal phonetic phenomenon with roots in deep time. *TMH – QPSR (Speech, Music, and Hearing – Quarterly Progress and Status Report), 51*, 41–44.

Long, M. H., Gor, K., & Jackson, S. (2012). Linguistic correlates of second language proficiency: Proof of concept with ILR 2–3 in Russian. *Studies in Second Language Acquisition, 34*, 99–126.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word initial stop consonants. *Journal of Child Language, 7*, 41–74.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Maddieson, I. (1993). Investigating Ewe articulations with electromagnetic articulography. *UCLA Working Papers in Phonetics, 85*, 22–53.

Maddieson, I., & Ladefoged, P. (1989). Multiply articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics, 72*, 116–138.

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*, 1546–1571.

Menzerath, P., & Lacerda, A. (1933). *Koartikulation, Steuerung und Lautabgrenzung: Eine experimentelle Studie*. Berlin: Ferdinand Dümmlers Verlag.

Mildner, V. (2018). Aspects of coarticulation. In M. Gósy & T. E. Gráczi (Eds.), *Challenges in analysis and processing of spontaneous speech* (pp. 27–48). Budapest: Research Institute for Linguistics of Hungarian Academy of Sciences.

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research, 61*, 1355–1368.

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America, 133*, 444–452.

Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019). Back from the future: Non-linear anticipation in adults' and children's speech. *Journal of Speech, Language, and Hearing Research, 62*(8S), 3033–3054.

Nwokah, E. (1986). Consonantal substitution patterns in Igbo phonological acquisition. *Language and Speech, 29*, 159–176.

Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech, 36*(2–3), 155–170.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*, 224–230.

Orie, O. (2012). *Acquisition reversal: The effects of postlingual deafness in Yoruba*. Boston: Mouton de Gruyter.

Oyebade, F. (1990). Language acquisition: The phonology of a Yoruba child. *Research in African Languages and Linguistics, 1*, 17–34.

Painter, C. (1978). Implosives, inherent pitch, tonogenesis and laryngeal mechanisms. *Journal of Phonetics, 6*, 249–274.

Puech, G. (1989). L'Opposition implosive/mi-voisées en Bekwel. Paper presented at the 19th Colloquium on African Languages and Linguistics, September 1989, Leiden, The Netherlands.

R Development Core Team. (2023). *R: A language and environment for statistical computing*. Version 4.3.0, retrieved 7 June 2023, from http://www.r-project.org/.

Recasens, D. (2014). *Coarticulation and sound change in Romance*. Amsterdam: John Benjamins.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America, 102* (1), 544–561.

Redford, M. A. (2019). Speech production from a developmental perspective. *Journal of Speech, Language, and Hearing Research, 62*(8S), 2946–2962.

Rubertus, E. (2024). *Coarticulatory changes across childhood: Implications for speech motor and phonological development*. University of Potsdam. Doctoral dissertation.

Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PLoS ONE, 13* e0203562.

Schweinberger, M. (2022). *Creating vowel charts in R*. Version 2022.10.10, retrieved 5 June 2023, from https://slcladal.github.io/vc.html.

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*, 708–737.

Stites, J., Demuth, K., & Kirk, C. (2004). Markedness vs. frequency effects in coda acquisition. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Proceedings of the 28th Annual Boston University Conference on Language Development* (pp. 565–576). Somerville, MA: Cascadilla Press.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics, 22*, 145–166.

Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: A developmental study of motor speech skills in humans. *Neuroscience Letters, 347*, 29–32.

Zhang, Z. (2021). Contribution of laryngeal size to differences between male and female voice production. *Journal of the Acoustical Society of America, 150*, 4511–4521.

Zharkova, N. (2017). Voiceless alveolar stop coarticulation in typically developing 5-year-olds and 13-year-olds. *Clinical Linguistics & Phonetics, 31*, 503–513.

Zharkova, N. (2018). An ultrasound study of the development of lingual coarticulation during childhood. *Phonetica, 75*, 245–271.

Zlatin, M., & Koenigsknecht, R. (1976). Development of the voicing contrast: A comparison of voice onset time in perception and production. *Journal of Speech and Hearing Research, 19*, 93–111.