ARTICLE

Voiced aspirates with mixed voicing in Yemba, a Grassfields Bantu language of Cameroon

Matthew Faytak[1,*] and Steffman Jeremy[2]

[1]University at Buffalo and [2]The University of Edinburgh;
*Corresponding author. Email: faytak@buffalo.edu

Abstract

Using electroglottography and acoustic measures, we characterize the strength and quality of voicing in voiced aspirated and unaspirated consonants (stops, fricatives, and approximants) in Yemba (Grassfields Bantu, Cameroon). We show that the Yemba voiced aspirates exhibit MIXED VOICING: modal voicing during the consonant constriction, but voiceless aspiration after release. Breathy or whispery phonation extends slightly into consonant constrictions preceding, and across the entire duration of vowels following, aspiration; this non-modal phonation extends further into prenasalized consonants. Mixed voicing has typically been excluded from the possible range of laryngeal–supralaryngeal coordinative patterns in consonants, and is thought to be unattested in the world's languages; most previous work on this topic assumes that non-modal phonation after voiced consonant release is breathy-voiced. However, we argue that Yemba voiced aspirates differ from more commonly studied breathy-release aspirates only in the settings of some gestural parameters: the late glottal spread gesture is larger in magnitude and more resistant to coarticulation, yielding consistently devoiced aspiration which may even be more perceptually recoverable compared to breathiness.

1 Introduction

1.1 Typology of laryngeal coordination in voiced aspirates

Voiced aspirated consonants, or VOICED ASPIRATES, have long presented difficulties for the study of consonantal laryngeal contrasts. From the introduction of VOICE ONSET TIME (VOT, see Lisker & Abramson 1964) as a measure of laryngeal timing and coordination, it has been observed that VOT and other measures of laryngeal–supralaryngeal coordination alone are insufficient to capture the difference between voiced aspirated and voiced unaspirated consonants (Clements & Khatiwada 2007; Abramson & Whalen 2017; Dmitrieva & Dutta 2020; Schwarz, Sonderegger & Goad 2019). In part, this is because voiced aspirates are not typically realized with discrete periods of prevoicing and post-release aspiration, as their name would suggest. Rather, the primary cue for voiced aspirates has proven to be a period of breathy phonation at and after consonant release and extending into the following vowel, with important secondary roles played by characteristics of the burst spectra and perturbations to the following vowel's f0 (Davis 1994; Rami et al. 1999; Clements & Khatiwada 2007; Dutta 2007; Mikuteit & Reetz 2007; Dmitrieva & Dutta 2020; Schwarz et al. 2019; Schertz &

Khan 2020). These characteristics also cue voiced aspirated or breathy sonorants (Traill & Jackson 1988; Maddieson 1991; Demolin & Delvaux 2001; Berkson 2013, 2019), although their degree of acoustic separation from modally voiced sonorants is typically smaller compared to the separation of voiced aspirated obstruents from modal-voiced obstruents (Berkson 2019).

Because VOT alone cannot distinguish voiced aspirates from voiceless aspirates, their articulatory mechanism has been well investigated, and they have also been extensively compared to their voiceless aspirated counterparts. For voiceless aspiration, a wide glottal opening or spread-glottis gesture is timed such that peak opening occurs at or just after release of a consonant closure (Hirose, Lee, & Ushijima 1974; Kagaya 1974; Kagaya & Hirose 1975; Löfqvist 1980; Löfqvist & Yoshioka 1981; Kingston 1985; Kim, Maeda & Honda 2010). A consensus has emerged that voiced aspirates' post-release breathiness or murmur differs from voiceless aspiration in the magnitude and timing of the associated spread glottis gesture: most voiced aspirates have a spread glottis gesture which is lower in magnitude, resulting in a smaller peak glottal opening area. This gesture also typically peaks later, after release, such that the closure portion typically remains modally voiced (Kagaya 1974; Kagaya & Hirose 1975; Benguerel & Bhatia 1980; Dixit 1989; Keating 1990: 329–330; Davis 1994; Ahn, 2018: 183–186).

In spite of the general agreement that voiced and voiceless aspirates have similar timing of spread-glottis gestures relative to the consonantal constrictions they are coordinated with, there is substantial cross-linguistic and cross-talker variation in the timing and coordination of these gestural complexes (see, e.g., Hoole & Bombien 2017; Pouplier et al. 2022), which is also modulated by prosodic factors such as phrasing and prominence (Bombien, Mooshammer, Hoole & Kuehnert 2008; Hoole, Bombien, Kühnert, & Mooshammer 2009; Bombien 2011; Hoole & Bombien 2017). More generally, it is well documented in the literature that speech rate and prosodic structure modulate the production and acoustics of contrasts involving aspiration (see Beckman, Helgason, McMurray & Ringen 2011; Krivokapić 2014; Kim, Kim & Cho 2018). Such variation is already well-known for voiceless aspirates, the study of which has expanded the field's understanding of timing and inter-articulator coordination more generally. Models of speech programming have been refined by the introduction of speaker- and language-specific target ranges for VOT in voiceless aspirated consonants (Cho & Ladefoged 1999; Cho, Whalen, & Docherty 2019) and structured relationships within a language's phonemic inventory (Chodroff & Wilson 2017). Research into other segment types such as preaspirated stops (Löfqvist & Yoshioka 1981; Engstrand 1987; Karlsson & Svantesson 2012) and voiceless aspirated nasals (Bhaskararao & Ladefoged 1991; Chirkova, Basset & Amelot 2019; Terhijia & Sarmah 2020) has revealed still more inter-language variation in the coordination of the laryngeal and supralaryngeal articulators.

The full range of coordinative possibilities has not been explored to a similar extent for voiced aspirates. Cross-linguistic variation in timing of the spread-glottis gesture is already known from a handful of studies. Languages such as Eastern Armenian (Seyfarth & Garellek 2018) and Owerri Igbo (Ladefoged, Williamson, Elugbe & Owulaka 1976; Henton, Ladefoged & Maddieson 1992: 81–82) appear to time the spread-glottis gesture earlier than the canonical, breathy-release voiced aspirates discussed above, resulting in more non-modal phonation during the stop closure and relatively little breathiness after stop release.

It stands to reason that the magnitude of the spread-glottis gesture could also vary cross-linguistically. While canonical voiced aspirates most often exhibit breathy phonation, a language could have voiced aspirates which are voiced during closure and voiceless at and after release, achieved with a particularly large-magnitude spread-glottis gesture, timed sufficiently late to avoid devoicing the stop closure in anticipation. In fact, such a

phonetic variant, with some voicing during closure and an interval of voiceless aspiration after release, is not an uncommon realization for voiced aspirates in many south Asian languages (Poon & Mateer 1985: 46; Maddieson 1991; Davis 1994: 186–188; Mikuteit & Reetz 2007). Stevens (1998: 476–478) directly attributes this sporadic devoicing of breathy release to wide glottal aperture. Control of the magnitude of glottal spread is less precise compared to its timing (Löfqvist, Baer & Yoshioka 1981), so it stands to reason that this devoicing is not controlled but occasionally occurs due to production noise.

1.2 *Mixed voicing and its omission from the typology*

However, the possibility of a stop voiced during closure and voicelessly aspirated at release is typically omitted from typologies of stop laryngeal contrasts (Ladefoged 1971; Henton, Ladefoged, & Maddieson 1992). In Ladefoged (1971, 1973), it was noted that 'a sound in which the vocal cords are vibrating during the articulation and then come apart into the voiceless position during the release of the stricture . . . has not yet been observed in any language' (1973: 9). More recent global surveys of speech sounds fail to mention the possibility of this stop type and its apparent absence from the typology (Catford 1982: 111–116; Catford 1988: 57–61; Laver 1994: 348–355; Ladefoged & Maddieson 1996: 47–73, 80; Ladefoged & Johnson 2011: 159–164). This omission is not unreasonable, especially since a consonant exhibiting modal voicing during its constriction and wide glottal abduction at its release would be quite effortful for the speaker: voicing during constriction requires glottal adduction, and aspiration requires glottal abduction. The modulation from phonating to spread vocal folds would need to be rapid and precisely timed, an arrangement which ought to be avoided under general principles of effort minimization.

Nonetheless, there seems to be no reason to consider this stop type to be impossible. Few formal arguments have been advanced for the impossibility of a single consonant exhibiting voiced closure and voiceless (aspirated) release, which we will refer to here as MIXED VOICING. Variable devoicing of the breathy release of South Asian voiced aspirates has already been discussed above, and languages of mainland Southeast Asia often exhibit partially devoiced stops under registrogenesis (Brunelle, Tạ, Kirby & Đinh 2020; Brunelle, Brown & Phạm 2022). More significantly, mixed voicing has in fact been proposed for unit segments in some languages. Perhaps the earliest and best-known proposal is for Kelabit, where Blust (1974, 2006, 2016) has argued for a series of voiced aspirate stops on phonetic, distributional, and morphological grounds. Blust describes this series of stops /bʰ dʰ gʰ/ as exhibiting voiceless release and 'heavy' aspiration, such that /dʰ/ is transcribed in its typical realization as [dtʃ] (Blust 2016: 269). Similar apparent mixed-voicing plosives, affricates, and clicks are also attested in the Kx'a and Tuu (so-called 'Khoisan') languages (Snyman 1975; Ladefoged & Traill 1984: 15; Maddieson 1984; Ladefoged & Maddieson 1996: 63, 80–81; Gerlach 2015; Naumann 2016). However, the apparent mixed-voicing structures in these languages have typically been analyzed as clusters, sometimes explicitly to avoid positing mixed voicing in the consonant inventory (Snyman 1975: 83–84; Traill 1985: 208–211, Ladefoged & Maddieson 1996: 80–81; Blevins 2010: 209–210; Güldemann & Nakagawa 2018: 5–15).

Reanalysis of apparent mixed-voice stops as clusters does not solve the puzzle of mixed voicing, however, since the same articulatory events are merely redistributed between multiple consonants in sequence. More problematically for this analysis, formal arguments have been advanced against mixed-voicing clusters. Phoneticians and phonologists alike have argued that such sequences, i.e. a voiced stop followed by a voiceless stop, are impossible within a syllable onset (Lindblom 1983: 24; Fujimura 1990: 338–339; Lombardi 1991: 59; Lombardi 1999: 272). However, once again, counterexamples to this generalization are easily

found: a number of languages are known to permit voiced–voiceless and voiceless–voiced onset clusters (Greenberg 1978: 257; Steriade 1997; Blevins 2004; Kreitman 2010; Kirby 2014), implying this type of sequence is merely unusual rather than impossible.

As noted by Blust (2016: 275), then, there is no obvious restriction against mixed voicing in single segments or clusters. This possibility has nonetheless gone unconsidered in most work on stop laryngeal contrasts. This is perhaps due in part to the frequent conflation of 'voiced aspiration' with breathy or murmured release (Ladefoged 1971: 13; Benguerel & Bhatia 1980: 141–143; Blust 2016: 248) and the arguments advanced (perhaps erroneously) against mixed voicing in tautosyllabic clusters. Furthermore, there is a dearth of instrumental work examining cross-linguistic variation of 'voiced aspirates' in the timing and magnitude of glottal spread or the resulting patterns of the voice source, which would help to clarify whether some subtype exhibits mixed voicing. In this paper, we begin to fill this gap by characterizing the apparent mixed voicing present in the Bamileke languages' voiced aspirated consonants, with particular focus given to Yemba.

1.3 Aspiration in Yemba

Yemba, also known as Dschang or Dschang Bamileke (ISO 639-3: ybb), is a Grassfields Bantu language spoken by 300,000 or more people in the West (Ouest) Region of Cameroon (Figure 1) and by diaspora populations located primarily in Europe and North America (Eberhard, Simons & Fennig 2020). It is one of roughly eleven Bamileke languages, a large subgroup within Grassfields Bantu (Watters 2003; Hammarström et al. 2020). Most speakers use French as a second or third language, as is typical of the region (Kouega 2007; Tsofack 2010). Yemba has been the subject of some phonological work focusing on its complex tonal phonology (see Hyman & Tadadjeu 1976; Hyman 1985), but its unusual system of aspiration contrasts has yet to form the basis of any phonetic study.

Most Bamileke languages, including Yemba, exhibit an aspiration contrast which cross-cuts the voicing contrast typical of wider Grassfields Bantu (Watters 2003). Historically,
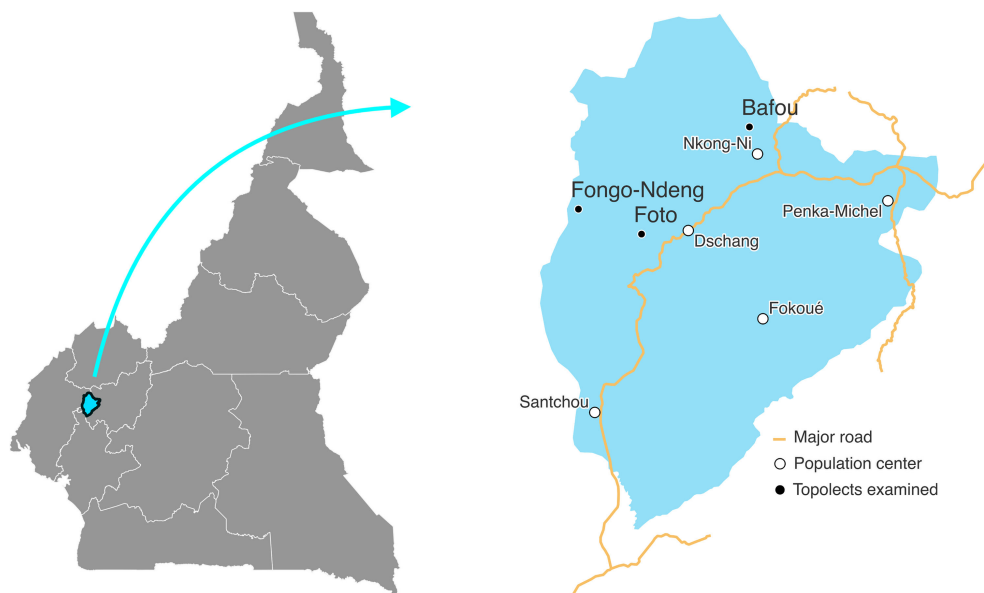

Figure 1. (Colour online) Approximate Yemba-speaking area within Cameroon, adapted from Nanfah (2003).

Table 1. Consonantal inventory of Yemba, adapted from Bird (1999) with post-nasal allophones included. Voiced aspirated-voiced unaspirated pairs examined in the present study are emboldened.

	Bilabial	Labden.	Alveolar	Postalv.	Velar	Glottal
Plosive [-voi]	p p^h		t t^h nt nt^h		k k^h ŋk ŋk^h	ʔ
Plosive [+voi] (allophonic)	**mb mb^h**		**nd nd^h**		**ŋg ŋg^h**	
Affricate		pf pf^h ɱpf ɱpf^h	ts ts^h nts nts^h	tʃ tʃ^h ɲtʃ ɲtʃ^h		
Fricative [-voi]		f f^h ɱf ɱf^h	s s^h	ʃ ʃ^h		(h)
Fricative [+voi]		**v v^h** **ɱv ɱv^h**	**z z^h** **nz nz^h**	**ʒ ʒ^h** **ɲʒ ɲʒ^h**		
Nasal	m m^h		n n^h	ɲ	ŋ	
Approximant			**l l^h**	j	**ɥ ɥ^h** **w w^h**	

aspirated segments developed as allophones of unaspirated segments before high vowels in open syllables (Hyman 1972; Anderson 1982), likely as an outgrowth of the tendency for stops to exhibit longer voice onset time before high vowels (Kagaya 1974: 173; Esposito 2002; Klatt 1975; Stevens 1998; Bang et al. 2018). In some modern Bamileke varieties, aspiration is still allophonic and restricted to this environment (Voorhoeve 1964; Nganmou 1991: 64; Nguendjio 1989: 32). While this development is often discussed as unique to Bamileke within Grassfields Bantu, closely related Noun languages are also reported to have allophonic aspiration of /t k/ and sometimes /g/ before high vowels (Nusi 1986; Tsafack Forku 2000; Ndedje 2003; Njeck 2003).

Subsequent sound changes affecting vowel quality have sometimes led to the development of minimal pairs for aspirated and unaspirated stops (Hyman 1972; Nissim 1981; Ngouagna 1988; Ngueyep 1988; Anderson 2008). In an additional innovative development, several Bamileke varieties such as Ngyembɔɔn and Feʼefeʼe extend the aspiration contrast to voiced and voiceless fricatives (Hyman 1972; Anderson 2008). These aspirated stops and fricatives are most often described as followed by a fricative matching the onset consonant in place, resulting in a system of affricates and lengthened fricatives (Nissim 1981; Anderson 1982). The quality of this frication, and Bamileke aspiration in general, has consistently been described as voiceless (Hyman 1972; Nissim 1981; Anderson 1982).

The aspiration contrast in Yemba has a breadth that is unusual even within Bamileke: contrastive aspiration is observed not only for stops, affricates, and fricatives as in other Bamileke lects, but also for approximants and nasals (Table 1; see also Bird 1999). While aspirated nasals occur only in a small number of words which vary among Yemba dialects (Bird 1999; Nanfah 2003), the aspirated approximants occur frequently.

A number of the aspirated consonants exhibit voicing during the consonantal constriction. The quality of the aspiration for these voiced aspirates is itself unusual: Yemba aspirates consistently give the impression of a period of sustained voiceless phonation after release (Bird 1999: 3–4). This includes the voiced aspirates, which do not give the impression of breathiness or murmur as is typical for phones with this label. Rather, the auditory effect of all voiced aspirates is that voicing is initiated during the onset consonant,
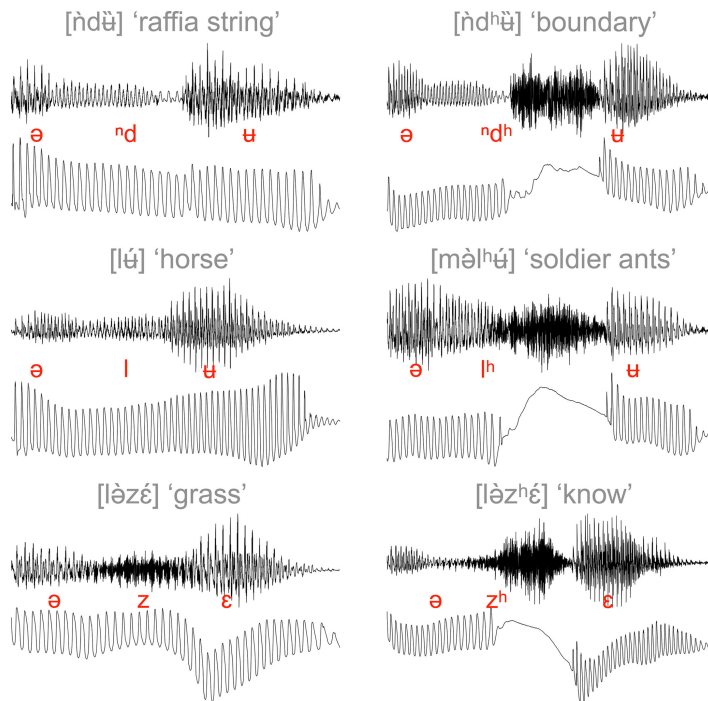
Figure 2. Audio (top) and EGG (bottom) signals from Bird (2003) for six items varying in consonant manner and aspiration. Note sustained lack of vocal fold vibration in the voiced aspirates on the right. The audio/EGG recordings pictured can be found in the Supplementary Materials.

interrupted with voiceless aspiration, then restarted at the onset of the following vowel. Electroglottograph traces published in Bird (2003) suggest that, for the tokens inspected here, this voiceless interruption exhibits wide glottal spread with no discernable vibratory activity (Figure 2), even when the consonant constriction is voiced. Unlike in Ngyembɔɔn and Bandjoun, in Yemba only aspirated fricatives are produced with lengthened frication (e.g., /sʰ/ as [ss] or [sː]; /zʰ/ as [zs], see Haynes 1989); other consonants are produced with aspiration as the term is usually meant (e.g. /pʰ/ as [pʰ], /ɥʰ/ as [ɥʰ], etc.).

Additional voiced (and voiced aspirated) segments occur under the influence of pre-nasalization (see Table 2). When preceded by a nasal prefix, the voiceless bilabial stops /p pʰ/ are realized as [mb mbʰ], the lateral approximants /l lʰ/ as [nd ndʰ], and the (labial-)velar approximants /ɥ ɥʰ w wʰ/ as [ŋg ŋgʰ ŋgw ŋgʰw]. Voiced fricatives and /f/ may occur with or without prenasalization, and the remaining voiceless fricatives may not be prenasalized. Prenasalization in Yemba acts as a tone-bearing unit (Hyman & Tadadjeu 1976; Hyman 1985), suggesting that it is syllabic as in Medɨmba, a closely related Bamileke language (Franich 2018). Regardless of the particular analysis one adopts, prenasalized consonants differ structurally from simple oral consonants, and so we separately analyze prenasalized and oral consonants in this study to account for their longer duration relative to oral consonants (Riehl 2008: 301–302; Franich 2018).

Whether or not supralaryngeal stricture is present in the aspiration after consonant release (i.e. whether fricative-like as in Ngyembɔɔn or merely a change in phonation), it is the apparent voicelessness of this interval after voiced consonants, and the resulting voiced–voiceless–voiced sequence of laryngeal settings, which is the focus of this study. As discussed in Section 1.1, this specific sequence of mixed voicing associated with onset consonant release is not only typologically unusual, but has generally been omitted from

Table 2. Derivation of Yemba voiced stops (aspirated and unaspirated) from approximants and /p/ by addition of a placeless nasal prefix /N/, which assimilates to the place of the following consonant. Morphemes glossed 'NC#' are overt noun class concord marking.

Unaspirated [mb, nd, ŋg, ŋgw]				Aspirated [mbʰ, ndʰ, ŋgʰ, ŋgʰw]			
Prefix	Stem		Inflected form	Prefix	Stem		Inflected form
Ǹ- 'NC6'	-pǐ 'liver'	→	m̀bǐ	Ń- 'INF'	-pʰǐ 'sow'	→	ḿ'bʰí
Ń- 'INF'	-lú 'rain (v.)'	→	ńdú	Ń- 'INF'	-lʰ̆ù 'wrestle'	→	ń'dʰú
Ǹ- 'NC1'	-ɰ̃ɥə̀ 'guest'	→	ŋ̀gɥə̀	Ń- 'INF'	-ɰʰ̆ù 'do'	→	ŋ́'gʰú
Ń- 'INF'	-wɛ̀ 'have'	→	ŋ́'gwɛ́	Ǹ- 'NC9'	-wʰǐ 'deer'	→	ŋ̀gʰwǐ

the typology of laryngeal contrasts. The voiced aspirates in Yemba appear to provide an example of just such a structure, which we investigate here.

1.4 Against two alternative interpretations of aspiration

Given the unusual nature of the phonetic structure at issue, before moving to define our hypotheses, we consider the phonological representation of aspiration in Yemba. There are at least three possibilities: a UNIT CONSONANT analysis, which views aspiration as an attribute of the onset consonant, i.e. /CʰV/; a VOCALIC analysis, which views aspiration as an attribute of the vowel, i.e. /CV̤/; and a CLUSTER analysis, which treats aspiration as a standalone consonant participating in an onset cluster as the second member, i.e. /ChV/. In this study, we follow the unit consonant analysis, which deviates from prior treatments of Yemba aspiration (Haynes 1989; Harro & Haynes 1991; Bird 1999; Nanfah 2003) but which constitutes the usual analysis for aspiration in other Bamileke languages (Hyman 1972; Nissim 1981; Anderson 1982, 2008). Here, we summarize the evidence for the unit consonant analysis and against the two other analyses.

Distributional evidence favors the unit consonant analysis over the cluster analysis. Yemba aspiration, if treated as a separate consonant /h/, would be highly constrained in its distribution: virtually all /h/ would occur in sequence after the release of a stem-initial onset consonant, with only a handful of /h/ acting as unambiguous simple onsets, and then only in loanwords.[1] Furthermore, clusters containing /h/ would be the only morpheme-internal clusters posited for Yemba's native vocabulary; treating this as release-associated aspiration is more parsimonious than positing complex onsets containing /h/ when complex onsets do not otherwise exist. The frequently occurring nasal prefixes are syllabic, as evidenced by their ability to bear tone, and so cannot be taken to participate in stem onset clusters; see Table 2.[2]

Aspiration could also be associated with the following vowel; it is well known that non-modal phonation may be phonologically affiliated with either consonants or vowels

---

[1] To our knowledge, the entire set of Yemba words containing segmental /h/ is: [hép] 'help' (Eng. *help*), [hápàŋ] '(clothes) iron' (Eng. *iron*), and [hâ(k)] 'distilled alcohol' (perhaps from English *alcohol*, French *alcool*, or Arabic /ʕaraq/).

[2] An anonymous reviewer suggests an alternative analysis in which a syllable boundary intervenes between the stricture of the first onset consonant and aspiration, perhaps with the preceding syllabic nasal acting as the nucleus, e.g. /n̩d.hV/, /ɲʒ.hV/. However, all Yemba syllables bear tone, and fricatives and approximants do not appear to bear tone in Yemba (Hyman & Tadadjeu 1976; Hyman 1985), making the suggested syllabification impossible in cases where fricatives and approximants occur without a preceding nasal, e.g. /ẓ.hV/, /ɥ.hV/, /f̩.hV/, /l̩.hV/.

(Esposito & Khan 2012; Esposito, Khan, Berkson & Nelson 2020). Bird argues that Yemba aspiration is moraic and part of the syllable rhyme, identifying its source as an abstract 'palatal mora' identifiable with, and in complementary distribution with, [i] and [i̯V] syllable nuclei (Bird 1999: 16). However, subsequently collected lexical material in Bird (2003) contains minimal pairs for rhymes containing [i], [i̯V], aspiration, and both types of vowel simultaneously *with* aspiration, e.g. [lə̀-zɛ́] 'NC5-grass', [lə̀-zʰɛ̀] 'NMLZ-forbid', [lə̀-zi̯ɛ̀] 'NMLZ-awaken', [lə̀-ˈzʰi̯ɛ́] 'NC5-offspring'. This undermines the argument that aspiration arises from a palatal element in the rhyme. Bird (1999) also notes that including aspiration in the rhyme motivates an apparent co-occurrence restriction under which aspiration does not occur in syllables containing coda consonants: if both aspiration and codas are part of the rhyme, both contribute to phonological weight, and the pattern can be treated as avoidance of superheavy syllables. However, while segmental contributions to weight are typically assumed to be part of the rhyme, in recent years moraic onsets have been convincingly posited in a range of languages (Shinohara & Fujimoto 2011; Topintzi & Davis 2017; Topintzi & Nevins 2017; Myhre 2021), suggesting that an onset-associated mora could contribute to the weight of the Yemba syllable.

Furthermore, historical–comparative evidence demonstrates that Yemba's aspirated consonants developed as positional variants of simple unaspirated onset consonants, making it more likely than not that they continue to constitute simple onsets down to the present. Bamileke aspirated stops, fricatives, and affricates (voiced and voiceless alike) developed from proto-Eastern Grassfields (PEG) unaspirated stops and affricates (Hyman 1972; Anderson 1982). The PEG onsets developed aspiration in Bamileke only before *i̯, *u̯ (the higher of two series of high vowels) and *i, *u (the lower series) in open syllables, possibly as phonologization of the longer VOT typical for stops released into higher vowels (Table 3; see discussion in Section 1.2). Some small developments from PEG must be posited to sort out the distribution of aspiration in Yemba: PEG coda *d, *l codas were lost before aspiration was triggered in Yemba, but coda *n and *m were lost more recently, after aspiration was triggered. The Ngemba languages, a sister group to Bamileke within Eastern Grassfields, did not develop aspiration; comparanda from two Ngemba languages (Awing and Mbili) are provided in Table 3 to confirm the general accuracy of the PEG reconstructions. Reflexes of PEG closed syllables are not associated with aspiration in any daughter languages; under this account, this may be because they were laxed or lowered due to the presence of a coda. The historical data also undermine the 'palatal mora' account described in Bird (1999): there is no obvious requirement that a PEG root with an aspirated reflex in Yemba contain a front vowel *i̯ or *i, and the co-occurrence restrictions on aspiration noted in Bird (1999) can be treated as a blocking effect on aspiration of *historically present* codas, including codas not attested in present-day Yemba such as *n and *m.

On the whole, the evidence provided here suggests that Yemba aspirated consonants, including voiced aspirated consonants, are unit consonants in the syllable onset. However, regardless of how the underlying gestural content is packaged up into segments and associated to different parts of the syllable, the voiced aspirate structure in Yemba merits a detailed comparison to more commonly studied speech sounds termed 'voiced aspirates'. Adopting various phonological representations of aspiration in Yemba does nothing to change the basic observation that Yemba voiced aspirates seem to involve a rapid alternation of voicing and spread-glottis gestures, a configuration which is omitted from typologies of unit consonants (the 'voiced aspirates') and tautosyllabic consonant clusters alike, as discussed above in Section 1.2. Conflating this structure with murmured stops under the term 'voiced aspiration' would further 'overload' the term (Blust 2016: 248) and ensure that mixed voicing remains unexplored as a possible variant on existing speech sounds.

Table 3. Comparative data demonstrating the development of aspiration before Proto-Eastern Grassfields (PEG) high vowels in open syllables (A) and its blocking in closed syllables (B). In PEG reconstructions, <ˊ> and <ˋ> indicate final floating high and low tones, respectively. Affixes are separated from stems with hyphens.

		Yemba (Bird 2003)	Mbili (Ayuninjam 1998)	Awing (Azieshi 1994; van den Berg 2009)	Proto-Eastern Grassfields (Elias et al. 1984)
A.	leaf	à-fʰò	à-fù	à-fù	*-fù´
	dog	m̀-ˈbʰɨ́	m̀-bvú	ŋ̀-gú	*-bɨ́à
	tree	à-tʰɨ́	à-fɨ̀	à-tsɨ̀	*-tí´
	nose	lə̀-ˈzʰó ~ lə̀-ˈzʰwɛ́	m̀-lwí	nə̀-lwí-ə̀	*-dúɨ̰
	mouth	ɲ̀-tʃʰṵ̀	ɲ̀-tʃù	ǹ-tsòːr-ə̀	*-cùɭ̀
	fish	è-ˈsʰó ~ è-ˈsʰwɛ́	ɨ̀-swɛ̀	ə̀-ʃú-ə̀	*-sɨ́ˋ
	do, make	lə̀-ɥʰṵ̀	ɣə̀	–	*-gɨ̀d-
	foot, leg	à-kʰṵ̀	à-kù	à-kòːlə̀	*-kṵ̀(l)ˋ
B.	egg	lə̀-pò	m̀-bùm	nə̀-pùm-ə́	*-bɨ̀m´
	wine	mə̀-lùʔ	m̀-lùʔ	–	*-dùkˋ
	buy	lə̀-ˈzó	zwὲn	mə̀-ndʒún-ə̀	*-jún-
	chicken	ŋ̀-ˈgáp	m̀-bvé	ŋ̀-gə́b-ə̀	*-gɨ́bˋ

1.5 Interim summary and research goals

The goal of this study is to determine the typical sequence of events which characterize the Yemba voiced aspirates. We use electroglottography and acoustic voice quality measures to better characterize the timing and quality of non-modal phonation in the Yemba voiced aspirates, and to relate the latter to better studied voiced aspirates with breathy release. Voiced unaspirated consonants, which are expected to have uninterrupted modal voicing, are the control against which voiced aspirates are compared. We evaluate two specific hypotheses as in (1–2), formulated by taking the more commonly researched breathy-voiced or murmured stops as our point of departure:

(1) **Hypothesis 1:** Yemba voiced aspirates will exhibit a strong voice source throughout consonant stricture and release, comparable to voiced unaspirated consonants.

(2) **Hypothesis 2:** Aspiration will be associated with the production of non-modal phonation on following vowels and preceding consonant strictures.

Hypothesis 1 is evaluated to determine if the Yemba data can be taken as a counterexample to the generalization that mixed voicing does not occur in voiced aspirates. The timing of the spread-glottis gesture and its coarticulatory effects are indirectly tested by Hypothesis 2, to facilitate comparison with South Asian languages, where post-release non-modal phonation is typically breathy and often fades into the following vowel.

Separate from the two hypotheses, the study also considers the data on voice quality in a purely exploratory manner (Research Question 1, RQ1) relating to the modulatory effects of consonant manner, as in (3):

(3) **Research Question 1:** Is the quality and timing of aspiration mediated by the manner of the consonant in any way?

From the few studies of breathiness in languages with breathy consonants in a range of manners, breathiness is known to be weaker (closer to modal) in sonorants (see Berkson

2019). While Yemba also has voiced aspirates in a range of manners, the precise manners involved (approximant, fricative, prenasalized plosive, prenasalized fricative) do not align well with those manners covered by existing studies (nasals, approximants). As such, under Research Question 1, we aim to contribute descriptive generalizations of the effect of manner on voiced aspiration in Yemba rather than to test a specific hypothesis.

2 Procedure

2.1 Materials

Audio data were obtained for four speakers (1F, 3M), and electroglottograph (EGG) data were obtained for three of these speakers (1F, 2M). Simultaneous audio and EGG recordings for two speakers (1M, 1F) were recorded in the UCLA Phonetics Lab in 2019. Material for two more speakers (2M) was drawn from an audio lexicon of Yemba (Bird 2003) which includes simultaneous EGG recordings for one speaker. Materials in Bird (2003) were recorded in 1997 in a recording studio managed by SIL Cameroon in Yaoundé, Cameroon. In total, 2,089 audio tokens are analyzed: 1,703 tokens from the corpus speakers and 386 tokens from the lab-recorded speakers. Because only three speakers (two lab speakers and one speaker in the audio lexicon) have accompanying EGG recordings, a total of 1,234 EGG recordings were analyzed. Of these total token counts, 610 audio tokens were aspirated (1,479 unaspirated) and 413 EGG tokens were aspirated (823 unaspirated). Lab recordings were made with an EG2 electroglottograph (Glottal Enterprises, and a Shure SM10A head-mounted cardioid dynamic microphone recording at a rate of 44.1 kHz and run through an XAudioBox preamplifier and A-D device. The specific audio and EGG recording equipment used in Bird (2003) is not clear from available metadata. All EGG data were processed using EGGWorks (Tehrani 2020), and subsequently read into VoiceSauce (Shue et al. 2011).

Target syllables in both sets of recordings contained aspirated or unaspirated voiced consonants in stem and syllable onset position, followed by a monophthong or diphthong in an open syllable (with two exceptions in the in-lab materials, described below). Phonological prominence is controlled in several important respects across all materials: all C(h)V examined across both data sets are in the first syllable of the stem, which is phonologically prominent in Yemba (Hyman 1985: 48). Targets contained voiced aspirated prenasalized stops [mbʰ ndʰ ŋgʰ], fricatives [vʰ zʰ ʒʰ], prenasalized fricatives [ɱvʰ nzʰ ɲʒʰ], and approximants [ɥʰ wʰ lʰ], along with the unaspirated equivalent for each segment. Aspirated nasals were not included in analysis due to their very low frequency. Note that voiceless unaspirated and aspirated consonants are not analyzed in this study, since the hypotheses relate to the realization of the spread-glottis gesture flanked by voicing on both sides. The choice to use voiced unaspirated consonants as the basis of comparison is reflected in the structure of Hypotheses 1 and 2, and was motivated by the desire to make parallel comparisons across all manners examined: voiceless approximants (aspirated or not) do not occur in Yemba, whereas voiced unaspirated consonants contrast with voiced aspirates in all manners.

The audio lexicon data differ from the lab-recorded data in their overall composition. All available open-syllable lexical items containing the target set of consonants were extracted from the audio lexicon in Bird (2003), resulting in a larger sample of material unbalanced for vowel type. Unaspirated tokens contained any of Yemba's phonemic monophthongs /i u ʉ ə ɛ o ɔ a/ or complex nuclei /ʉ̯ə ʉ̯ɔ i̯a i̯e u̯ɔ u̯i u̯i̯e u̯ʉ/; aspirated segments occur with a restricted set of monophthongs /i u ʉ ɛ ɔ/ and diphthongs /u̯ɛ u̯ɔ ʉ̯ə ʉ̯ɔ i̯e u̯ʉ/. The lower lexical frequency of aspirated tokens in the audio lexicon also resulted in a larger count of unaspirated compared to aspirated items in those materials. In contrast, the in-lab recordings were nearly balanced for aspiration and vowel type, though not balanced for

lexical tone. Stimuli for the in-lab recordings contained all licit combinations of the smaller set of vowels /i u ʉ ɛ u̯e/ and all target initials mentioned above except for /ʒ ʒʰ ɲʒ ɲʒʰ/. Due to natural gaps for monophthongs in open syllables paired with certain consonants, three closed-syllable items ending in /k/ or /ʔ/ were used and two items replace monophthongs with similar dipthongs (/ʉ/ with /u̯ə/ and /i/ with /i̯e/, respectively). The full list of stimuli for the in-lab speakers can be found in Appendix A.

The use of varied materials also introduces substantial dialect differences: the southern dialect spoken by the two in-lab speakers, who hail from the cities of Foto and Fongo-Ndeng, differs slightly from the northern Bafou dialect used in Bird (2003) in the vowel qualities used in particular words. The southern dialect lacks /ɔ u̯ɔ i̯a/ altogether; these phones are merged into /o u̯ə i̯e/, respectively. Some additional vowel shifts and mergers are specific to a post-aspiration context.[3] However, because there are no appreciable differences in the quality of aspiration between the two speaker groups, we pool the dialect groups throughout the analysis that follows.

Although the lab and audio lexicon data are both read speech collected in a studio setting, the lab and audio lexicon data sets also differ in their elicitation methods, and therefore in the nature of the speech task undertaken. The recorded component of Bird (2003) consists of each head word read once in isolation, along with a possessed form (of nouns) and tensed forms (of verbs), also read once in isolation. In contrast, the lab-recorded tokens were embedded in the frame sentence shown in (4):

(4) mə́ŋ   ˈli̯é   u̯ʰə̀   __   m̀bù   ù̀
    1SG.SUBJ say COMP   __   to   2SG.OBL
    'I say __ to you'.

Morphologically related forms were not collected from in-lab speakers, and more tokens of fewer lexical items were collected. The specific procedure used to elicit readings of the materials in Bird (2003) is not clear from associated metadata; it is therefore possible that additional differences exist in procedure between the two data sets.

2.2 Segmentation

Audio was hand-annotated using Praat TextGrids (Boersma & Weenink, 2020) to demarcate the acoustic boundaries of the onset consonant constriction, aspiration (if present), and following vowel (Figure 3). The constriction interval for prenasalized stops and fricatives included the nasal portion. The aspiration–vowel boundary was marked at the onset of clear formant structure and voice pulsing. Criteria for marking the boundary between the consonant constriction interval (*clo*) and the post-release, pre-vowel interval (*asp*) varied slightly according to consonant manner. The boundary between stop closure and aspiration was placed immediately following the release burst. Approximant constrictions were segmented out from aspiration at the release of the approximant constriction. For /l lʰ/, this was clearly indicated by an increase in intensity and an abrupt change in formant frequencies. In the case of the central approximants [w wʰ u̯ u̯ʰ], release was marked at the F2 and F3 transition visible before or during aspiration.

Because voiced fricatives maintain their supralaryngeal constriction through all or part of the aspiration interval, the boundary between *clo* and *asp* for fricatives was placed not according to release but according to a different event, the cessation of voicing, as visually

---

[3] For instance, southern /Cʰwɛ/ corresponds to northern /Cʰo/, as in 'nose' [là ˈzʰɔ́] (southern [là ˈzʰwɛ́]), and /ə/ occurs after some aspirated initials in the northern dialect, as in the word 'goat' [m̥̀ ˈvʰɔ́] (southern [m̥̀ ˈvʰó]).
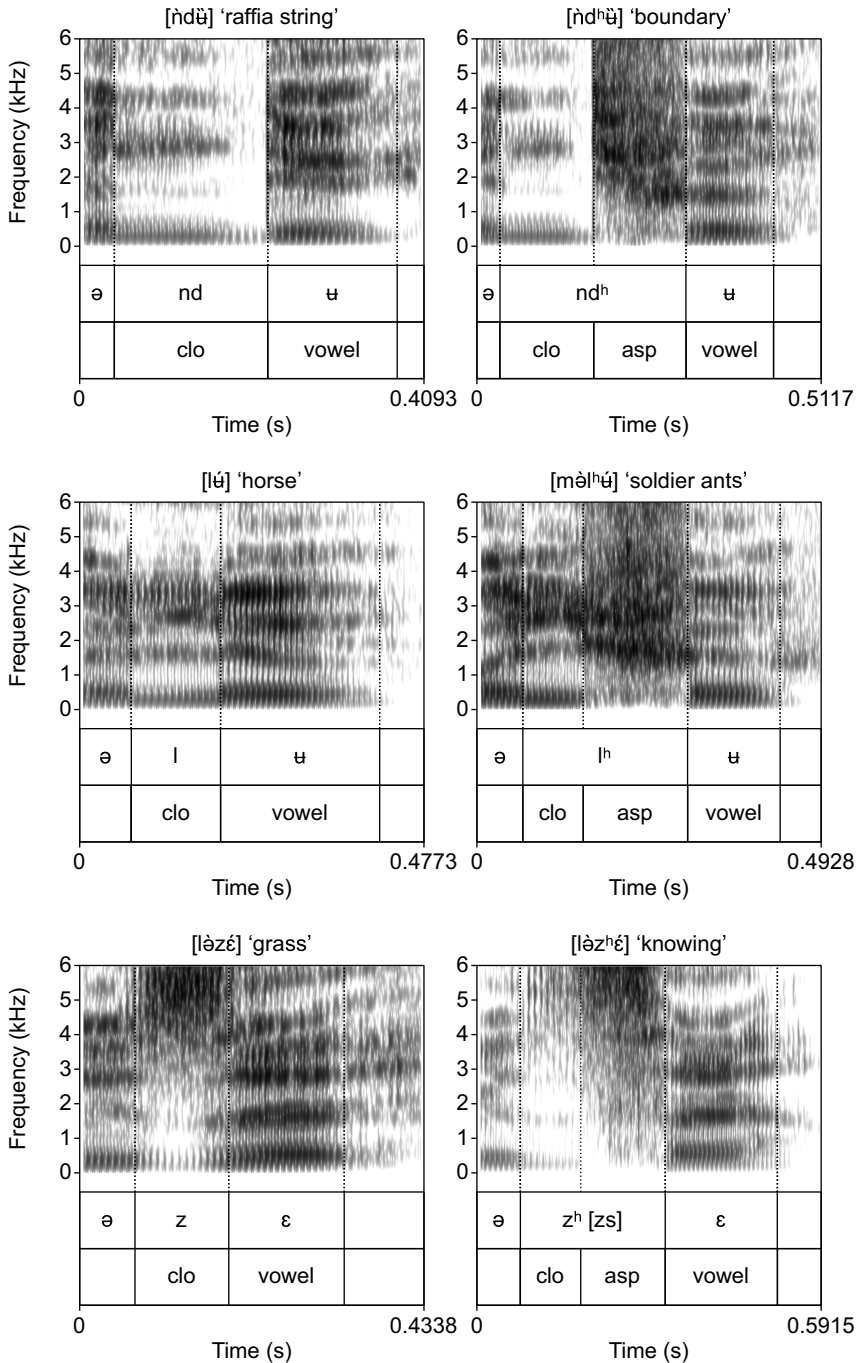
Figure 3. Sample segmentations of aspirated and unaspirated stops (top), approximants (middle), and fricatives (bottom). *Clo* denotes consonantal constriction; *asp* denotes aspiration. The audio recordings pictured can be found in the Supplementary Materials.

identified in the spectrogram by the cessation of periodic voice-source excitation of frication. While this may appear to introduce circularity into our investigation of the timecourse of voicing (as part of Hypothesis 2 and the exploratory research question), we note here that the binary judgment (voiced vs. unvoiced) according to which the events were segmented is not expected to align precisely with strength of excitation (SoE) as a continuous measure of voice source strength. We acknowledge, however, that the use of this segmentation criterion may slightly affect the assignment of transitional material to the constriction and aspiration intervals, with more transitional material marked as *clo* for fricatives than the other manners. This issue is revisited in the discussion in Section 4.1.

2.3 Analysis

All numerical data and scripts are hosted in an Open Science Foundation (OSF) repository which can be accessed at https://osf.io/yx4g6/. We also include in this repository prior specifications for the Bayesian regression models, and numerical summaries of the generalized additive mixed models (GAMMs), with summaries for the Bayesian regression models also included in Appendix B.

2.3.1 Measures

To characterize the presence and strength of the voice source over all intervals (constriction, aspiration, and vowel) and evaluate Hypotheses 1 and 2, we use STRENGTH OF EXCITATION (SoE), which has been used as a measure of voice source strength in a range of recent work on voicing and voice quality (Seyfarth & Garellek 2018; Garellek 2020; Garellek, Chai, Huang & Van Doren 2021; Tabain, Garellek, Hellwig, Gregory & Beare 2022). SoE was calculated at each detectable EPOCH in the acoustic signal using VoiceSauce (Shue et al., 2011). An epoch is a positive to negative zero crossing of the zero frequency-filtered acoustic signal (Murty & Yegnanarayana 2008; Mittal, Yegnanarayana & Bhaskararao 2014). Epochs correspond to glottal closures in the phonatory cycle, and SoE is the slope of the signal at each epoch. The SoE at a given epoch thus corresponds to the amount of energy contributed to the local speech signal by the voice source: higher SoE values indicate a larger contribution.

All models which follow use range-normalized log SoE rather than raw SoE, following Garellek et al. (2021), which factors out setting- and equipment-specific differences in SoE. This is particularly useful for the present study, given that the two pairs of speakers were recorded more than twenty years apart with different equipment. A range-normalized SoE value of 1 represents the strongest (maximum) contribution of voicing to the speech signal for a given speaker, and 0 the weakest (minimum) contribution. We take the minimum range-normalized SoE for each speaker to indicate voicelessness: in the *raw* data, the minimum value for each speaker was in fact an SoE of zero (voicelessness), meaning a zero value in the range normalized data actually corresponds to voicelessness as well. We note each speaker additionally had some near-zero values slightly elevated above zero SoE, due to occasional spurious detection of epochs with associated energy (Dhanajaya & Yegnanarayana 2009). Additional GAMM model code for analysis of raw SoE values is included in the OSF repository.

From prior examination of aspirated tokens (see Section 1.2), we assume that voice quality cannot be reliably characterized on aspiration because phonation is most often absent. However, we expect that phonation *adjacent to* the aspiration interval may be relatively breathy in Yemba, in line with previous experimental findings showing carryover

coarticulation of laryngeal state from onset consonants to following vowels (Löfqvist & McGowan 1992; Ní Chasaide & Gobl 1993; Dutta 2007; Khan 2012; Dmitrieva & Dutta 2020) and the initiation of glottal spread gestures prior to or during associated consonant constriction (Kagaya 1974; Kagaya & Hirose 1975; Benguerel & Bhatia 1980; Dixit 1989). In an effort to characterize the influence of aspiration on its surroundings and evaluate Hypothesis 2, we calculated voice quality measures from the EGG and acoustic signals for each flanking consonantal constriction and vowel. Using VoiceSauce, CONTACT QUOTIENT (CQ) was calculated from the EGG signal, and CEPSTRAL PEAK PROMINENCE (CPP) and H1*–A3* were calculated from the acoustic signal using VoiceSauce's STRAIGHT backend (Kawahara, de Cheveigné & Patterson 1998), this time using VoiceSauce's default 1 ms sampling rate. These measures have been used to characterize breathy phonation (Keating & Esposito 2007; Berkson 2019), and in the signal preceding and following aspiration we expect Yemba speakers to exhibit lower CQ due to reduced vocal fold contact; lower CPP due to weaker harmonic structure; and higher H1*–A3* due to increased spectral tilt.[4]

2.3.2 Statistical modeling

We carried out a set of analyses on the data which modeled the influence of aspiration and consonant manner on voice quality. All analyses were carried out in R (v4.1.2, R Core Team 2021) using RStudio (v2022.2.3.492, RStudio Team 2022).

As a first set of analyses, we modeled SoE, CQ, CPP, and H1*–A3* for constriction and vowel intervals using Bayesian mixed effects modeling. We submitted the mean across all samples in constriction, aspiration, and vowel intervals to a Bayesian mixed-effects linear regression implemented using the R package *brms* (Bürkner 2017, 2021). The SoE mean value model was built to statistically compare constriction, aspiration, and vowel intervals in terms of their strength of voicing, something not done in the GAMM models where the time series of each interval were modeled separately. The dependent measure was the mean SoE for a given interval, predicted as a function of interval type (constriction, aspiration, or vowel). Random effects were specified as random intercepts for speaker, with by-speaker slopes for interval type.

For the other voice quality measures (CQ, CPP, and H1*–A3*), which were not measured during aspiration as noted above, we constructed six models: one for each of the three measures in constriction and vowel intervals. The models predicted each measure as a function of manner (for constriction) or preceding constriction's manner (for vowel), presence of aspiration, and the interaction of these two fixed effects. Random effects were specified as random intercepts for speaker, and random slopes for fixed effects and their interactions. As in the GAMM modeling, oral and prenasalized fricatives were not included in the CPP and H1*–A3* models. In coding aspiration and manner variables in the models, we contrast-coded aspiration, mapping aspirated to –0.5 and unaspirated to 0.5. When there were only two manners in the model (as in the consonant models for CPP and H1*–A3*), we also contrast-coded manner, mapping approximant to –0.5, and prenasalized stop to 0.5. When all four manners were included in a model (CQ for consonants and all vowel models), approximant was set as the reference level.

All Bayesian regression models were fit to draw 4,000 samples, with a burn-in period of 1,000 samples in each of four Markov chains, from the posterior distribution over parameter values. We retained 75 percent of samples for inference. We fitted all models with normally distributed weakly informative priors for fixed effects and the intercept. The mean

---

[4] The particular set of measures we adopt here is commonly used in previous literature on voice quality, but we omit the very commonly used measure H1*–H2*, opting to use H1*–A3* in its place as a measure of spectral tilt. This decision was made in light of recent research showing that H1*–H2* does not directly reflect laryngeal articulation (Chai & Garellek 2022; Gobl & Ní Chasaide 2019).

for the intercept priors was specified as the approximate mean of a variable for the reference level (or grand mean if both factors were contrast coded), with a wide standard deviation. For example, in the consonant data, the mean CPP was observed to be 21 dB. We thus set the prior for the intercept as normal(20,10) (mean of 20, standard deviation of 10): an informed but very wide distribution. Fixed effect priors were set to be equally wide with a mean of zero: that is, a prior expectation of no effect on the measure being modeled.

In reporting results from these models, we focus on the strength of evidence for an effect provided by the posterior distribution. We report the median of the posterior estimate for a given effect and its 95 percent CREDIBLE INTERVAL (CrI), which indicates the range in which 95 percent of the estimates from the posterior fall. When this interval excludes the value of zero, this indicates a reliably estimated non-zero effect, i.e. that an effect is robust in the data. Conversely, intervals including zero indicate non-trivial variation in the estimated directionality of the effect, though we also will consider this in terms of the strength of evidence for an effect. In other words, 95 percent CrIs which only narrowly include zero may be taken to provide some weaker evidence for an effect's existence. We additionally report a metric which captures the same intuition from the posterior distribution for an effect: the PROBABILITY OF DIRECTION, henceforth *pd*, as computed with the *bayestestR* package (Makowski, Ben-Shachar & Lüdecke 2019). This metric indicates the percentage of the posterior which shows a given directionality. The metric can range between 50 (a distribution centered on zero; no evidence for an effect) and 100 (a distribution which totally excludes zero; very strong evidence for an effect), and corresponds intuitively to a frequentist *p*-value. As suggested in Makowski et al. (2019), when $pd > 95$, we can take this as evidence for an effect's existence. We focus on reporting those effects found to be credible in the model.

As an additional set of analyses, in order to model the dynamics of these effects *within* constriction, aspiration and vowel intervals, time series for each measure of interest were converted to percent duration of labeled interval and submitted to an AR1 GAMM. The model was fit to each measure over time as a function of consonant manner and aspiration, which were treated as a single combined variable. Structuring the variable in this way allowed us to model non-linear differences based on both manner and aspiration. For SoE, we built three models, one for the consonant constriction, one for the aspiration interval, if present, and one for the following vowel. For both aspiration and vowel models, we included the manner of the preceding consonant (approximant, prenasalized stop, fricative, prenasalized fricative) as a predictor, with preceding manner and aspiration treated as a single combined variable.

We did not collect voice quality measures during aspiration. Thus, CQ, CPP and H1*–A3* models were only constructed for the C and V intervals, again predicting each measure by aspiration and by manner (for C), or aspiration and preceding C manner (for V). The independent variable was again coded as a combined variable, as with the SoE GAMM models. For the C interval, models for acoustic voice quality measures (CPP and H1*–A3*) were not constructed for fricatives and prenasalized fricatives, given that frication noise present during these manners introduces aperiodic noise into the signal, complicating accurate measurement. CQ, being a non-acoustic measure, is used to analyze all manners. All measures were scaled (z-scored) within speaker to account for possible differences in individual speakers' voice quality.

Generalized additive mixed models for all measures were fitted using the R packages *mgcv* and *itsadug* (Wood 2017; van Rij et al. 2022). The models included a parametric term and smooth terms for the combined aspiration–manner variable. To model possible nonlinear differences across speakers based on aspiration and manner, we coded random effects as reference-difference factor smooths, as described by Sóskuthy (2021). Factor

smooths for speaker and factor smooths for speaker by combined aspiration–manner were coded as an ordered factor. The *k* parameter (knots) for all smooth terms was set to 20, which was assessed to be adequate via the *gam.check()* function. We also set the *m* parameter to 1 for factor smooth terms, following Sóskuthy (2021). In assessing the GAMM fits we focus on how measures differ visually over time, and when confidence intervals (95% CI) from the fits do not overlap, suggesting a significant difference based on aspiration.

2.3.3 Predictions

Given the structure of the models outlined above, we consider the following predictions. First, with respect to Hypothesis 1, we predict that Yemba voiced consonant constrictions, aspiration, and vowels will show high range-normalized SoE. Hypothesis 1 would be rejected if aspiration were to show close to zero range-normalized SoE, suggesting absence of the voice source. We assume that a range-normalized SoE value consistently lower than voiced segments and close to zero would indicate voicelessness in the signal, in line with prior work using the measure (Dhanajaya & Yegnanarayana 2009; Mittal et al. 2014; Garellek et al. 2021). Time series SoE modeling should provide evidence for the same pattern as the mean SoE model, but should also provide a more precise picture of SoE dynamics over time: rejection of the null hypothesis would follow if the GAMMs show a dip in SoE during aspiration reaching a low minimum in the middle of the aspiration interval.

With respect to Hypothesis 2, we predict that segments adjacent to aspiration should be impacted in terms of their voice quality, specifically exhibiting breathier phonation. Expressed in terms of our voice quality measures, lower CQ in segments adjacent to aspiration (compared to those which are not) would indicate less vocal fold contact and indicate breathy voicing. Lower CPP adjacent to aspiration would likewise indicate breathier voicing as evidenced by weaker harmonic structure, and higher H1*–A3* adjacent to aspiration would index greater spectral tilt, consistent with breathy voicing, as another measure of voice quality. To the extent that we find effects on mean measures, we expect this should line up with the time series data, which will clarify whether the overall effect is localized or spread across the segment's duration; this detailed information is additionally expected to aid in interpretation of Research Question 1.

3 Results

3.1 Strength of excitation

3.1.1 Mean measures

To assess how constriction, aspiration, and vowel intervals differ in terms of voice source strength, we submitted the mean log SoE for each interval to a mixed effects model. Vowel interval was set as the reference level in this model. We report the *pd* metric, described above in Section 2.3.3, and the estimate in log SoE. Comparing the three pairwise marginal estimates extracted from the fit, we find that vowels have higher SoE compared to both aspiration ($\beta = 1.20$, $pd = 94$) and constriction intervals ($\beta = 0.11$, $pd = 92$; a smaller effect). We also see evidence for lower SoE in aspiration compared to constriction intervals ($\beta = 1.10$, $pd = 93$). These data (see Figure 4) suggest that as a whole, the aspiration interval shows considerably weaker voicing compared to flanking segments, with many means at or near zero, suggesting voicelessness. Note also that aspiration of constriction and vowel intervals entails a slight reduction of SoE as compared to unaspirated intervals. We examine this difference in more detail in GAMM modeling of the time series data.
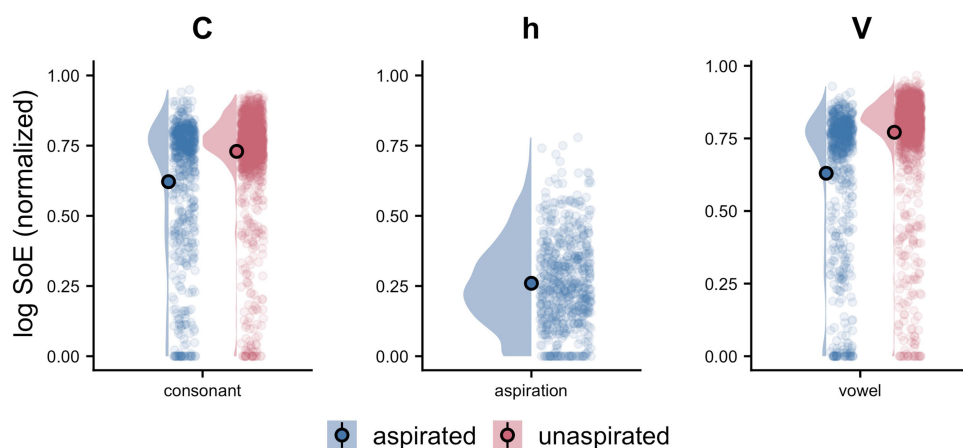
Figure 4. (Colour online) Range-normalized log SoE values for segment means in constrictions (left), aspiration (center), and vowels (right).

3.1.2 Generalized additive mixed models

We turn next to the GAMM results for SoE. Overall, SoE is high during voiced consonant constrictions, regardless of aspiration (C; Figure 5, left). For both vowels and constrictions, there is no overall effect of aspiration on SoE, but there are differences in trajectory. There is a small effect of aspiration on the curvature of the SoE trajectory for aspirated consonants of all manners, but particularly non-prenasalized fricatives: SoE deflects downward in roughly the last 10 percent of normalized segment duration. Vowels (V; Figure 5, right) exhibit no analogous lowering of SoE early in their duration when they follow aspiration.

During aspiration itself (h, Figure 5, center), SoE shows a falling–rising trajectory hitting a minimum value at or near range-normalized zero; that is, the lowest observed values for each speaker. The dynamics of this trajectory depend on consonant manner: oral and prenasalized fricatives start lowest and bottom out the fastest, while approximants and prenasalized stops start higher and have a slower decline. The effect on SoE levels relative to those of vowels and constrictions is large and robust: aspiration consistently has lower SoE than the preceding consonant constriction and following vowel.

This data suggests that aspiration is characterized by a consistent, long voiceless portion occurring near segment midpoint, away from flanking segments. We note this here, in advance of further discussion of results, to justify excluding the aspiration interval from the voice quality measures which follow: in the absence of a consistent voice source, these measures cannot be reliably extracted. It must also be noted that all preceding consonant constrictions have reliably high SoE, suggesting voicing regardless of the presence of prenasalization. Yemba voicing does not appear to rely on nasal venting for maintenance of voicing over most of a constriction's duration, as attested in numerous other languages (Ohala 1997; Solé 2018), even when voiceless aspiration immediately follows. Some local effects on SoE are seen immediately adjacent to aspiration, suggesting the start of the glottal spread gesture during the constriction. We pursue confirmation of this effect in the analyses of voice quality measures which follow.

3.2 Voice quality measures beyond stength of excitation

3.2.1 Mean measures

Next, we consider the effects of aspiration on mean voice quality measures beyond SoE, to assess aspiration's influence on the immediately surrounding phonation. In reporting these
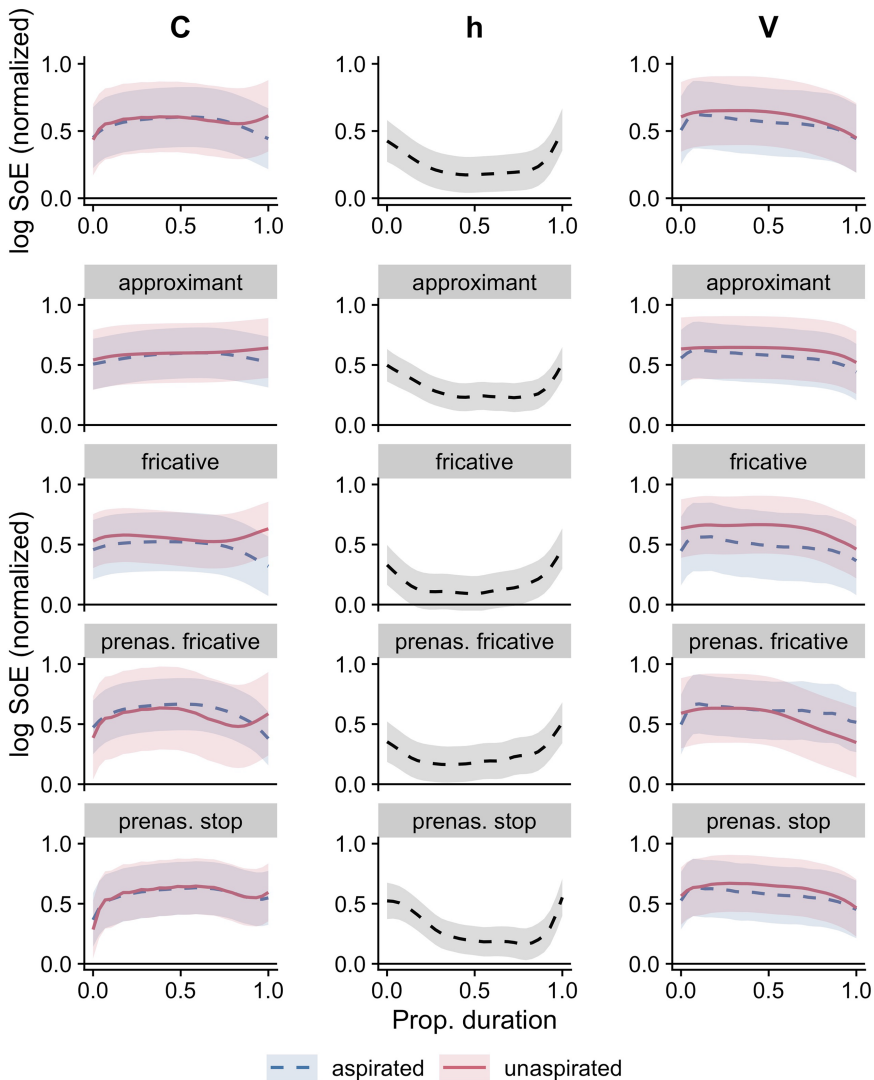
Figure 5. (Colour online) Strength of excitation (SoE) during constriction (left), aspiration (middle), and vowel intervals (right), split by aspiration. Fits pooled by manner shown in top row; bottom four rows show fits split by manner.

results, we again provide the *pd* metric and the estimate's median in the appropriate units (proportion of glottal cycle for CQ; dB for CPP and H1*–A3*). Ninety-five percent credible intervals (CrI) for each measure are included in the full model output in the appendix. Recall that *pd* values greater than 95 suggest especially strong evidence for an effect.

We begin with voice quality measures during consonant constriction. For CQ (Figure 6, top left), we find a credible effect of aspiration in reducing CQ overall during constriction ($\beta = 0.08$, $pd = 96$). We also find an effect of manner, whereby oral fricatives show lower CQ as compared to reference-level approximants ($\beta = -0.06$, $pd = 98$). Pairwise comparisons of all manners, carried out with the R package *emmeans* (Lenth 2021), show that this is the only credible difference between manners in terms of CQ, though there is weaker evidence for
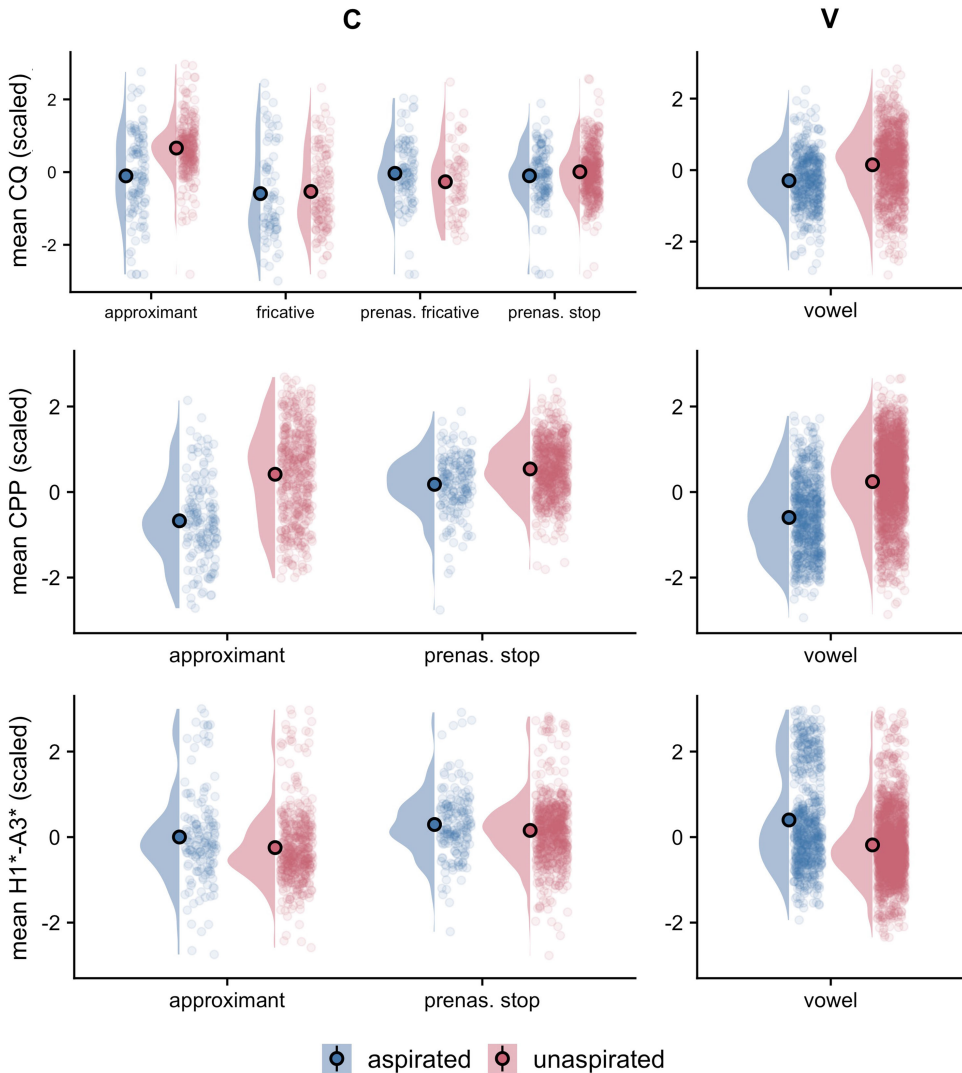
Figure 6. (Colour online) Segment means for constriction (C, left) and vowel (V, right) intervals: CQ (top), CPP (middle), and H1*–A3* (bottom). Consonant data is split by manner, while vowel measures are pooled for preceding consonant manner. Large points indicate grand means.

slightly lower CQ in fricatives compared to both prenasalized fricatives ($\beta = -0.05$, $pd = 88$) and prenasalized stops ($\beta = -0.06$, $pd = 91$).

The two acoustic measures generally align with the CQ observations, though recall that only prenasalized stops and (oral) approximants are examined in these models. As shown in Figure 6, middle left, voiced consonant constrictions preceding aspiration have lower CPP ($\beta = 1.85$, $pd = 100$), consistent with breathy voicing, while no strong evidence for a difference between manners was apparent. There was, however, a credible interaction between manner and aspiration ($\beta = -2.32$, $pd = 100$), whereby approximants show a clearly larger effect of aspiration as compared to prenasalized stops. The presence of aspiration is seen to increase consonants' H1*–A3* (Figure 6, bottom left), again suggesting breathy voicing ($\beta =$

−3.18, $pd = 97$). Consonant manner also exerts an effect on H1*–A3* in this model: prenasalized stops exhibit higher H1*–A3*, suggesting that they have breathier voicing compared to approximants (β = 3.25, $pd = 97$), in line with previously demonstrated effects of consonant manner on voice quality (Mittal et al. 2014; Chong et al. 2020).

Turning to vowels, each of the vowel models for CQ, CPP, or H1*–A3* finds no evidence for an effect of preceding consonant manner on the relevant measure (all $pd < 95$). In further discussion, we therefore pool consonant manners (Figure 6, right). Following aspiration, vowels exhibit mean VQ measurements suggesting slight breathiness: there is robust evidence for lower CPP (β = 3.10, $pd = 98$) and higher H1*–A3* (β = −6.19, $pd = 99$) following aspiration, and weaker evidence for lower CQ (β = 0.06, $pd = 89$).

In summary, the mean model results show that aspiration exerts small influences on mean values for voice quality measures, possibly due to anticipation of a wide spread-glottis gesture in the upcoming aspiration interval. Voice quality measures suggest breathiness to a greater extent in consonant constriction and vowel intervals when adjacent to aspiration. Mean measures additionally appear to be mediated to a small extent by consonant manner, where fricatives show overall lower CQ as compared to other manners, and prenasalized stops show overall higher H1*–A3* as compared to approximants.

3.2.2 Generalized additive mixed models

We next turn to the GAMM fits for CPP, H1*–A3*, and CQ, focusing on the preceding constriction and following vowel to assess how any differences in voice quality due to aspiration manifest over (normalized) time. Recall from Section 2.3 above that aspiration is not analyzed here due to the absence of reliable voicing, and not all voice quality measures were calculated for all consonant manners: cepstral peak prominence (CPP) and H1*–A3* were not calculated for oral and prenasalized fricatives due to potential interference from the supralaryngeal noise source. Contact quotient (CQ), however, was calculated for all manners, since it is an articulatory measure less impacted by the presence of a supralaryngeal noise source.

We first consider CQ, shown in Figure 7. Fricatives (particularly oral fricatives) have somewhat lower overall CQ than other consonant manners and vowels, but CQ is otherwise similar across the intervals compared. There is a substantial difference in trajectory for consonant constrictions, but not for vowels: a clear downward deflection of CQ values occurs for aspirated consonants as the constriction duration elapses, seemingly in anticipation of following aspiration. For prenasalized stops and fricatives, the deflection occurs later in normalized time, possibly due to differences in the overall duration of the nasal–oral consonant sequence (see Section 3.3). The prenasalized stops also have less of an aspiration-induced difference in trajectory compared to the other manners. Vowels show no substantial difference in CQ across their duration, which is peculiar in light of acoustic voice quality differences discussed below.

Next, we consider CPP and H1*–A3*. The results for CPP are shown in Figure 8. There is a small effect of aspiration on the timecourse of CPP for consonants, driven by a small effect on approximants: again, the only sound at issue which is not prenasalized. Unlike the results obtained for CQ, vowels following aspiration also show a tendency to have lower CPP, though the trajectories do not noticeably differ as they do for consonant constrictions. Time series for the second acoustic measure, H1*–A3*, are shown in Figure 9. Much as with CPP, there is a slight effect of aspiration on H1*–A3* trajectory during consonant constrictions, again driven by approximants, the non-prenasalized category included in the model. There is a small but clear effect of aspiration on the H1*–A3* timecourse of following vowels. Like the difference observed for CPP, this difference is not one of overall trajectory,
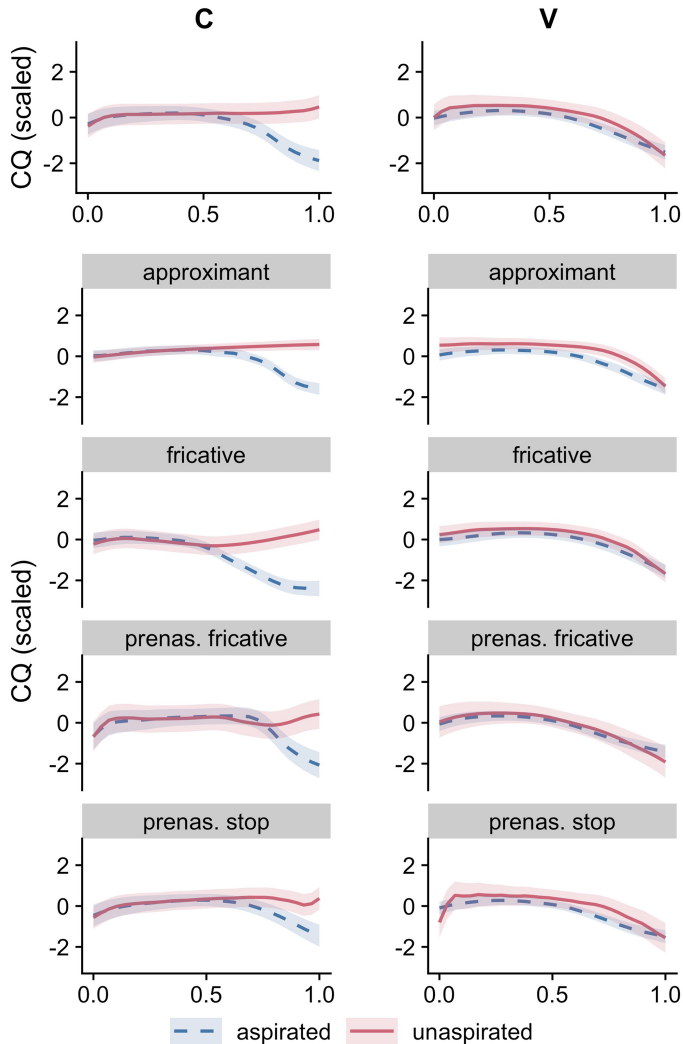
Figure 7. (Colour online) Modeled CQ timecourse during constriction (left) and vowel (right) intervals. The *x* axis shows normalized time. Fits pooled by manner are shown in the top row; the bottom four rows show fits split by manner.

and occurs over most of the vowel's duration and regardless of the preceding consonant's manner.

Overall, GAMM fits suggest that aspiration impacts the voice quality of surrounding segments, but over different timecourses, and in different measures. Recall that effects on mean measures, even when credible, were mostly small in magnitude. For consonant constrictions, this may reflect that they are driven by trajectory changes *local to aspiration* in the GAMM analyses, particularly for CQ: only the second half or last third of voiced aspirated consonants could be described as breathy. Approximants and oral fricatives generally show a larger impact of aspiration on voice quality compared to the prenasalized fricatives and stops, with the modeled timecourses suggesting earlier and breathier breathiness.
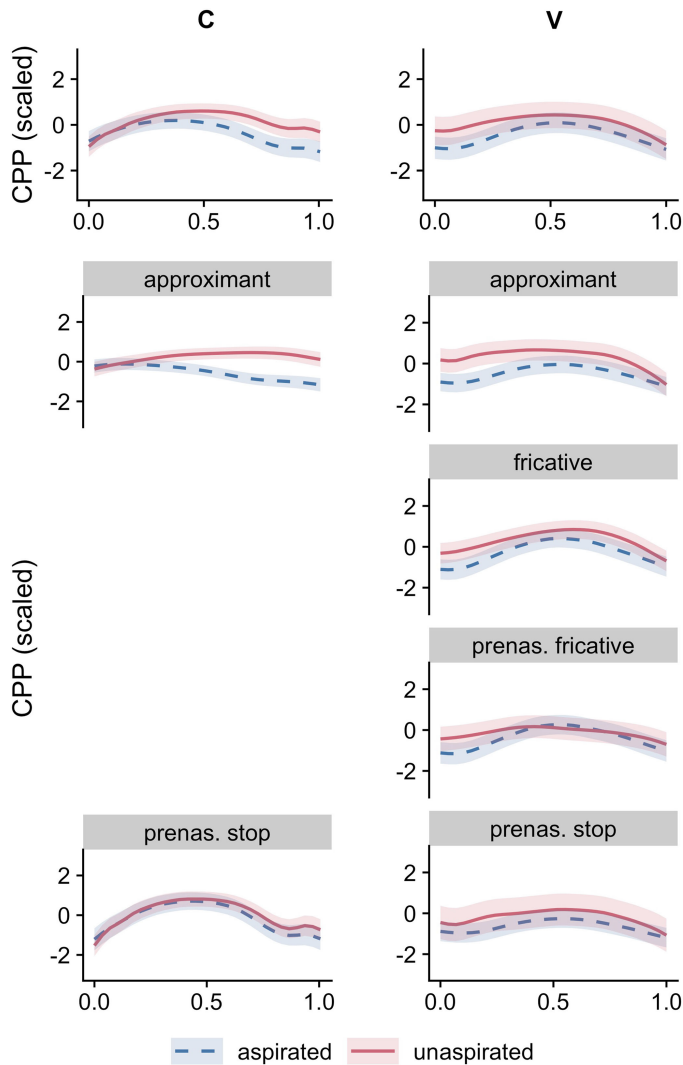
Figure 8. (Colour online) Modeled CPP timecourse during constriction (left) and vowel (right) intervals. The *x* axis shows normalized time. Fits pooled by manner are shown in the top row; the bottom four rows show fits split by manner.

Vowels occurring after aspiration differ in detailed timecourse, exhibiting consistently elevated CPP and (especially) H1*–A3* over their entire duration, but CQ is not consistently affected.

3.3 Interpreting consonant manner effects: duration of C and h intervals

The GAMMs discussed above reveal differences in the timing and trajectory of voice quality measures during constriction and aspiration intervals for prenasalized fricatives and stops, at first glance suggesting that consonant manner mediates the timing of aspiration. These timecourses are modeled over normalized time, which may obscure the nature of the differences in timing: they may be driven by differences in overall consonant duration rather
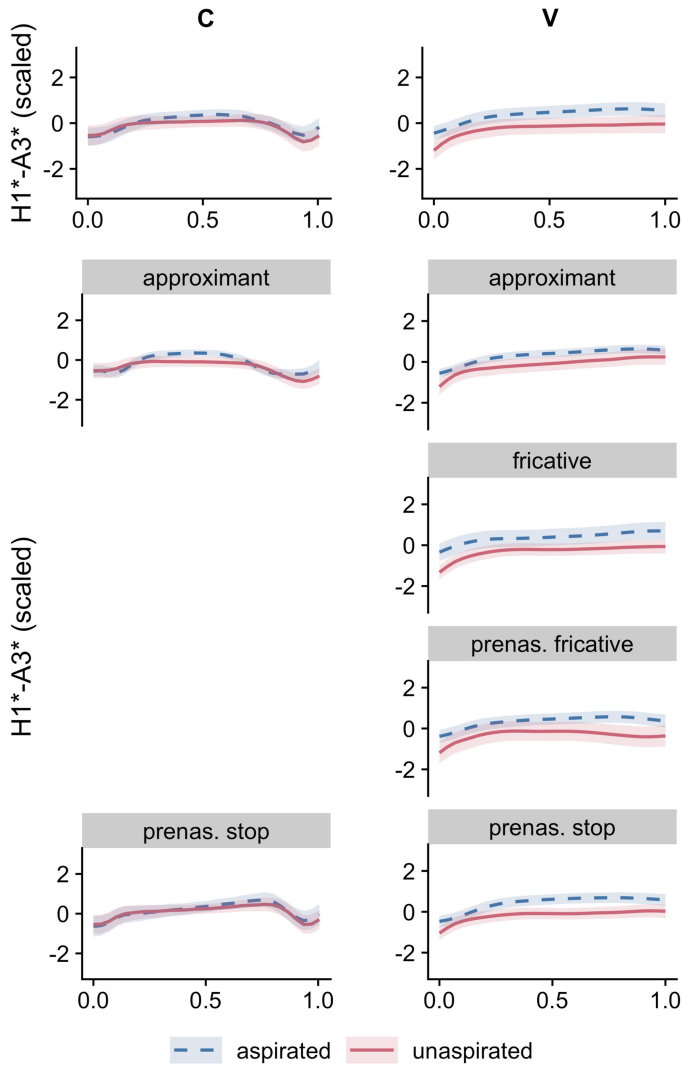
Figure 9. (Colour online) Modeled H1*–A3* time course during constriction (left) and vowel (right) intervals. The *x* axis shows normalized time. Fits pooled by manner are shown in the top row; the bottom four rows show fits split by manner.

than differences in the spread glottis gesture. Because of the complex status of prenasalized fricatives and stops, they are likely to differ in duration compared to the singleton, non-prenasalized consonants.

To aid interpretation of the timecourse data and better address Research Question 1, we provide raw durational figures for the constriction intervals for the different consonant manners here. Prenasalized constriction intervals are generally longer than non-prenasalized constriction intervals. Non-prenasalized approximants ($M = 108$ ms, $SD = 29$ ms) and non-prenasalized fricatives ($M = 132$ ms, $SD = 39$ ms) are somewhat shorter than prenasalized stops ($M = 188$ ms, $SD = 46$ ms) and much shorter than prenasalized fricatives ($M = 212$ ms, $SD = 58$ ms). Splitting by aspiration (Table 4), aspirated consonants have

Table 4. Mean (standard deviation in parentheses) for constriction duration (ms), split by manner and aspiration.

	approximant	fricative	prenasalized stop	prenasalized fricative
aspirated	97(27)	102(22)	180(39)	197(52)
unaspirated	111(29)	147(36)	190(47)	228(60)

slightly shorter constrictions than unaspirated consonants, though the same durational differences across manners hold for both aspirated and unaspirated consonants.

Thus, GAMMs were computed over a longer time frame for prenasalized stops and prenasalized fricatives, compared to approximants and fricatives. Coarticulatory changes to voice quality due to anticipation of the spread-glottis state of aspiration may be expected to extend *less* into this longer segment in terms of normalized duration. This appears to be reflected in the GAMM data for all measures, and is especially clear for CQ (Figure 5), which was calculated for all manners. As alluded to in earlier discussion of the CQ GAMMs, the later downward deflection of CQ in normalized time for prenasalized segments may be driven in part by the overall greater duration of the nasal–oral consonant sequences.

4 Discussion

Below, we review how the study results relate to our hypotheses and research question (see Section 1.3). With regards to Hypothesis 1, we note that the data are consistent with Yemba voiced aspirates exhibiting voicing during consonant constriction and voiceless aspiration after release (Section 4.1), a possibility not entertained in existing typologies of laryngeal contrast. As for Hypothesis 2, we consider the impact of aspiration on the following vowel (Section 4.2). We subsequently consider how Yemba's mixed-voice voiced aspirates compare to breathy-release voiced aspirates in other languages (Section 4.3), and the role of differences in the magnitude and timing of laryngeal gestures in the typology which emerges from considering both types together. Finally, we consider some functional reasons for the success of mixed-voice aspirates in Yemba and the broader Bamileke family (Section 4.4).

4.1 Constriction and post-release (aspiration) intervals

Strength of excitation mean models and GAMMs both suggest that a weak or nonexistent voicing target is generally achieved during aspiration, leading us to reject Hypothesis 1, which held that a strong voice source should be maintained through the entire voiced aspirated consonant. In the middle of the aspiration interval, normalized SoE is lower than all voiced segments examined here and close to zero. We note that the slight elevation above zero of this signal, which is seen for aspiration in all consonant types, is likely due to small contributions of energy associated with epochs spuriously detected during voicelessness (see Dhanajaya & Yegnanarayana 2009). These SoE data are also consistent with other recent studies where known voiceless consonants similarly show a low but non-zero SoE (Seyfarth & Garellek, 2018; Tabain et al. 2022). Weak voicing, likely breathy or whispery, is indicated by moderate SoE away from the midpoint of aspiration in both directions: towards the preceding voiced constriction and the following voiced vowel. This can be regarded as a coarticulatory influence of the adjacent voiced segments on aspiration which is local in

character: the highest SoE measurements during aspiration come in the first and last 25 percent of duration of the interval.

In spite of the strong observed tendency toward voicelessness during aspiration, the strength of excitation data also suggest the consistent presence of a voice source during consonant constrictions. Constrictions exhibit consistently high SoE regardless of the presence of prenasalization, and do not differ in SoE as an effect of aspiration. Voice quality measures do, however, diverge local to consonant release: consonant constrictions exhibit lower CQ and lower CPP immediately before aspiration, partly supporting Hypothesis 2. (There is no effect of aspiration on H1*–A3* in consonants, for reasons which remain unclear.) The data suggest that voiced aspirate consonant constrictions maintain a voice source which is only slightly affected by anticipatory vocal fold spreading ahead of the spread-glottis posture required for voiceless aspiration. Altogether, these SoE data support the claim that Yemba's voiced aspirates sequence a voiced constriction with a *voiceless*, aspirated release, and so present a counterexample to influential typologies of laryngeal contrasts which exclude this possibility (Ladefoged 1971; Henton et al. 1992; Ladefoged & Maddieson 1996; Cho et al. 2019).

Finally, we consider the exploratory research question on the modulating effect of consonant manner. While fricatives (both prenasalized and oral) appear to exhibit somewhat lower SoE and CQ around the constriction–aspiration boundary, this may be due to the different segmentation criteria used for fricatives, which lack a clear release before the onset of aspiration, unlike the other manners. This may have led more transitional material to be assigned to the consonant closure, as suggested in Section 2.2, potentially making this effect spurious. As such, we focus in our discussion on the other apparent influence on voicing's time course which emerges from the data: nasality. Voice quality measures indicate that anticipatory non-modal phonation before aspiration extends less into the consonant constriction for the two prenasalized manners. It remains unclear whether this effect is driven by prenasalized onsets' complexity and longer duration (Riehl 2008; Franich 2018; see Section 3.3) or some property of nasality itself. One plausible source of this effect, which should be considered in future research, is that modal voicing is lost more readily under narrower supralaryngeal constrictions, which encourage a fast buildup of intra-oral pressure (Stevens 1998; Solé 2010; Chong et al. 2020). Nasals and prenasalized consonants, which vent airflow through the open velopharyngeal port, may thus be less prone to anticipatory loss of non-modal voicing, even given a similar extent and timecourse of vocal fold spreading (e.g. Garellek, Ritchart & Kuang 2016).

4.2 Following vowel phonation

Acoustic measures indicate that Yemba vowels are breathier when they are adjacent to aspiration, providing support for Hypothesis 2. SoE does not credibly differ for vowels as a function of preceding aspiration or the manner of the preceding consonant, suggesting that aspiration does not greatly impact the strength of the vowel's voice source. Voice quality measures point to breathiness over most of the vowel's duration: vowels after aspiration are characterized by lowered CPP and elevated H1*–A3*, which respectively signal reduced periodicity and greater spectral tilt, both acoustic signatures of breathier phonation (Esposito 2010; Esposito & Khan 2020, Keating et al. 2023). This pattern is also attested for breathy-release voiced aspirates in the Indo-Aryan languages (Clements & Khatiwada 2007; Dutta 2007; Mikuteit & Reetz 2007; Esposito & Khan 2012; Berkson 2013; Dmitrieva & Dutta 2020) but not for all voiced aspirates universally (cf. Seyfarth & Garellek 2018 on Armenian).

Unlike for consonant constrictions, contact quotient (CQ) was not significantly lower following aspiration, as is common for breathy phonation (Esposito & Khan 2012; Khan 2012; Esposito et al. 2020). Bayesian regression on mean measures found only a trend toward lower mean CQ after aspiration, and GAMMs found no substantial reduction in CQ across vowels' durations. Vowels after aspiration thus seem to achieve greater airflow, and breathiness, through some means other than reduced contact of the vocal folds. This may suggest that Yemba vowels exhibit *whispery* voice after aspiration, which maintains medial compression of the vocal folds but exhibits elevated airflow and aperiodicity (Laver 1980; Rose 1989; Mazaudon & Michaud 2008; Tian & Kuang 2021).

4.3 Expanding the typology of voiced aspirates

Typologies of laryngeal contrast typically admit only one type of voiced aspirate, the breathy-release or "murmured" aspirates common in Indo-Aryan languages (Ladefoged 1971; Henton et al. 1992; Ladefoged & Maddieson 1996; Cho et al. 2019). Yemba appears to present a distinct variant on this type. In this section, we compare the Yemba voiced aspirates to breathy-release voiced aspirates to clarify this distinction. The breathy phonation associated with breathy-release stops has been described as differing from voiceless aspiration chiefly in the timing and magnitude of a spread-glottis gesture (Kagaya & Hirose 1975; Schiefer 1987; Dixit 1989; Davis 1994; Mikuteit & Reetz 2007; Ahn 2018: 183–186). We assume a similar spread-glottis gesture occurs in Yemba voiced aspirates, and our discussion here focuses on the timing and magnitude of this gesture.

Previous work on breathy-release voiced aspirates suggests that they exhibit a late spread-glottis gesture of variable magnitude, timed to peak after stop release (Kagaya 1974; Kagaya & Hirose 1975; Benguerel & Bhatia 1980; Yadav 1984; Traill & Jackson 1988; Dixit 1989; Maddieson 1991; Demolin & Delvaux 2001; Berkson 2019). They are most often phonated through both closure and aspiration (Dixit 1989; Dmitrieva & Dutta 2020; Islam 2019). Upon release, there is most often a discrete interval of voiced noise production (Davis 1994; Mikuteit & Reetz 2007; Berkson 2013). However, voicing is occasionally absent during this interval (Maddieson 1991; Davis 1994: 186–188; Mikuteit & Reetz 2007), and if voicing is present, the aspiration interval may be absent altogether (Schertz & Khan 2020; Davis 1994; Mikuteit & Reetz 2007; Ahn 2018: 184). This spread-glottis gesture can thus be treated as having a wide WINDOW or target region in terms of magnitude (Keating 1990), as depicted in Figure 10a: a speaker's spread glottis is typically open enough to trigger voiced aspiration, but may sporadically be wider, yielding voiceless aspiration, or narrower, yielding no discrete interval of aspiration.

A second type of voiced aspirate is described by Seyfarth & Garellek (2018), who contrast Armenian voiced aspirates with the breathy-release type described above (see also Cho et al. 2019: 58 for additional discussion). In this type (Figure 10b), a lower-magnitude glottal spread gesture begins early and peaks during closure. This results in slight breathiness of the following vowel, but no discrete interval of aspiration noise, and devoicing is not observed. Similar breathy stops with early, low-magnitude, and less variable glottal spread gestures are reported for Owerri Igbo (Ladefoged et al. 1976; Henton et al. 1992: 81–82); a similar gestural configuration may also characterize low register-associated stops in languages of mainland Southeast Asia, which exhibit partial voicing and co-occur with breathiness on the following vowel (Brunelle et al. 2020; Brunelle et al. 2022). These data suggest a spread-glottis gesture which is different from the breathy-release type both in timing and magnitude: initiated earlier and with a lower degree of spread. The variability of this spread is also reduced compared to breathy-release aspirates.
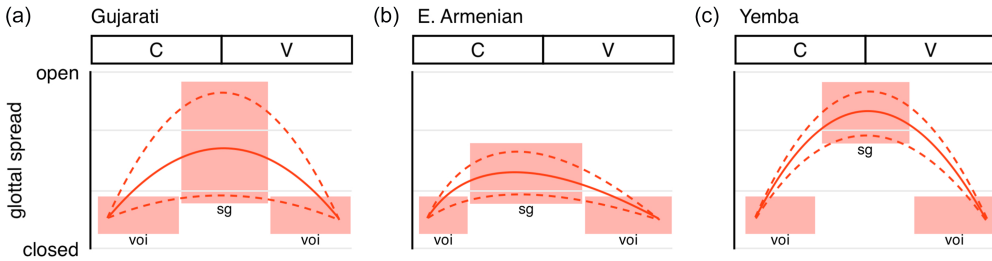
Figure 10. (Colour online) Gestural windows and simplified gestural scores for three types of voiced aspirates exemplified by Gujarati (a), Eastern Armenian (b), and Yemba (c), after Keating (1990). Solid lines indicate central tendency of the spread-glottis gesture; dashed lines indicate extent of variability.

Yemba seems to exhibit a third type of voiced aspirate. The spread-glottis gesture begins late, as in breathy-release voiced aspirates, and is typically large enough in magnitude to trigger voicelessness in post-release aspiration (Figure 10c). The magnitude of this gesture can be regarded as greater than either type discussed above, and it requires preparatory activity far enough in advance to result in vocal fold abduction during the consonant constriction (see Section 4.2). The gesture appears typically to trigger voicelessness in post-release aspiration, unlike breathy-release aspirates, whose aspiration is only sporadically devoiced due to a narrower (or less consistently wide) glottal aperture (Stevens 1998: 476, 478): even when flanked by voicing on both sides, aspiration is reliably voiceless, suggesting a narrow but extreme window for the spread-glottis gesture. Yemba can perhaps be thought of as an outlier in the particularly high magnitude of its spread-glottal gesture and lower tolerance of variance in the magnitude of that gesture, much as Tlingit or Navajo are in the duration of their VOT lag time (Cho & Ladefoged 1999).

We can tentatively view the mixed-voice (Yemba, 10c) and breathy-release (Gujarati, 10a) types of voiced aspirate as language-specific strategies for addressing the antagonism inherent to producing a spread-glottis gesture flanked by intervals of modal or nearly modal voicing. Phonologically, voiced aspirates of this type have been argued to be specified for both privative [voice] and [spread glottis] (Mikuteit & Reetz 2007; Schwarz et al. 2019); Schwarz et al. (2019) note that full voicing in Nepali is more consistently achieved than an aspiration interval. This apparent prioritization of voicing suggests that glottal spread accommodates to full voicing through the breathy-release aspirate. Likewise, in Yemba, the early, large glottal abduction starting during constriction appears to allow for consistent post-release voicelessness, suggesting that constriction voicing accommodates to voiceless aspiration in the Yemba type. This typology is well modeled by existing dynamical accounts under which gesture-like primitives are specified for their COARTICULATORY AGGRESSIVE-NESS, or propensity to influence their neighboring segments while maintaining their own invariance (Saltzman & Munhall 1989; Fowler & Saltzman 1993; Recasens & Espinosa 2009). Cast in this framework, the aggressiveness of [voice] is higher than [spread glottis] in breathy-release aspirates, inclusive of types A–B in Figure 10, and the aggressiveness of [spread glottis] is higher than [voice] for the Yemba type.[5]

---

[5] An anonymous reviewer notes that a large co-intrinsic effect on f0 would be expected under this model. While we acknowledge that f0 perturbation effects from aspiration could be a fruitful area for future inquiry, the data considered in this study (particularly the audio lexicon data) are not balanced for Yemba's lexical tone contrasts. As such, we consider investigation of f0 beyond the scope of the present study.

4.4 Ecological validity of the data and future work

The conclusions discussed above on the magnitude and consistency of the spread-glottis gesture in Yemba must be regarded as tentative, given that they are based on data that present a number of limitations (see Section 2.1 for further details). The degree of inter-speaker variation is difficult to assess given the fragmentary nature of the data: the data were not recorded in the same setting and were likely recorded with differing microphones and electroglottographs, presenting a number of confounds as to the consistency of the data across the speaker sample. The speaker sample is small and heterogenous, consisting of four speakers speaking at least two different dialects of Yemba; two speakers had resided outside of the Yemba-speaking area for some time prior to recording. Because inter-speaker variation in the magnitude and inter-articulator coordination of laryngeal gestures is amply attested in prior work (Chodroff & Wilson 2017; Hoole & Bombien 2017) including work on voiced aspirates specifically (Poon & Mateer 1985: 46; Davis 1994: 186–188), there is the possibility that the present work overstates the extent and consistency of voicelessness in post-release aspiration in Yemba.

Furthermore, all data considered here are read speech collected in a laboratory or recording studio setting, which may limit the ecological validity of the study's findings. It is particularly likely that faster, more reduced speech (and spontaneous speech in particular) would lead to undershoot of the production of laryngeal cues, including the spread glottis gesture discussed here (Beckman et al. 2011; Schwarz et al. 2019; Narayan 2022), again possibly leading to an overstatement of the degree of devoicing during post-release aspiration in the present findings. Because the materials contained only a restricted set of prosodic frames, it also has not been possible to investigate interactions of higher-level prosodic structure and the timing and magnitude of aspiration's spread-glottis gesture in Yemba, as amply attested in case studies of other languages (Krivokapić 2014; Hoole & Bombien 2017; Kim et al. 2018). As such, some aspects of the present study's findings await confirmation in future work, which may leverage more abundant, naturalistic data which has not yet been collected for Yemba and other Bamileke languages.

4.5 Functional motivation

In the present study, Yemba voiced aspirates exhibit a very effortful modulation of laryngeal state. This raises the question of why this effortful articulation, while rare cross-linguistically, has flourished within Bamileke: other Bamileke languages with impressionistically similar voiced aspirates include Fe'efe'e (Hyman 1972), Ghomálá' (Nissim 1981), and Ngyembɔɔn (Anderson 1982, 2008). As discussed above, breathy-release voiced aspirates permit a relatively low-magnitude glottal spread gesture, which is less reliably recoverable in perception: accordingly, merger of voiced aspirated and voiced unaspirated stops is common (Hussain 2018). Non-obstruents are more vulnerable to this loss of contrast: Berkson (2019) notes that breathy sonorants in Marathi are less well differentiated acoustically from modal sonorants, compared to breathy and modal obstruents. The size of glottal opening is known to be relatively poorly controlled in speech production (Löfqvist et al. 1981), which may underpin the fragility of breathy–modal contrasts and particularly of breathy–modal contrasts in sonorants.

Yemba notably does not exhibit the pattern reported in Berkson (2019): its voiced aspirated sonorants are just as strongly (and voicelessly) aspirated as its voiced aspirated obstruents. Functionally, mixed-voicing aspirates may be more resistant to contrast loss due to their especially large modulation of laryngeal state, which should be more consistently perceptually recoverable. The unusually wide range of consonant manners which may be aspirated in Yemba may have encouraged the development of mixed voicing as

an enhancement to aspiration of voiced consonants (Kirby 2013; Wedel, Kaplan & Jackson 2013) to discourage loss of phonemic contrasts with high functional load. Because tools for calculating functional load in Yemba have yet to be developed, and because there is little work on the perceptual recoverability of voiced aspiration, the plausibility of this account remains to be confirmed. Further work on other Bamileke languages may provide a means of doing so, since the nature of phonemic contrasts involving aspiration and frequency of aspirated consonants varies from language to language (see Section 1.2).

5 Conclusion

The articulatory and acoustic evidence analyzed here suggests that Yemba voiced aspirated consonants exhibit both modal voicing during closure and voiceless aspiration after release. This *mixed voicing* is argued to involve a sequence of particularly low-variance gestures for glottal adduction and spread, in that order, with the latter half of the consonant closure tending to become somewhat breathy in anticipation of the upcoming voiceless aspiration. This mixed voicing has generally been excluded from the typology of laryngeal–supralaryngeal coordination, with most previous work following the assumption that non-modal phonation associated with the release of a voiced consonant constriction is invariably breathy.

While the proposed Yemba gestural specification is presumably more effortful than that of breathy-release voiced aspirates, which have more variable and accommodating glottal spread magnitude at the release of the supralaryngeal constriction, this can be seen as a relatively minor difference in gestural magnitude and window size for aspiration itself, rather than a fundamental difference of type which should preclude the existence of mixed-voicing consonants altogether. Because of possible functional differences from breathy-release aspirates, in particular that voiceless aspiration may be more easily recoverable in perception, we urge further work on this consonant type, including a reassessment of the typical voice quality in voiced aspirates in more frequently studied languages.

Appendix A Stimuli for in-lab speakers

Table A1. List of stimuli used for in-lab elicitation (Southern Yemba speakers) with the ad-hoc orthographic representations agreed upon by participants and used as prompts. Nouns and their obligatory noun class prefixes are separated by hyphens

Prompt	Phonetic transcription	Gloss
Mbie	m̀-bì	'CL9-raffia knife'
Mbeuh	m̀-ˈbʉ́	'CL6-breasts'
Mbouh	m̀-ˈbú	'CL6-hands'
mbhi-he	m̀ˈbʰí	'before'

Table A1. Continued

Prompt	Phonetic transcription	Gloss
mbeu-he	m̀-ˈbʰú	'CL9-dog'
Ndie	ǹ-ɗì	'CL1-lord'
Ndeuh	ǹ-ɗ̈	'CL9-raffia string'
ndouh	ǹ-ɗũ̀	'CL3-river'
ndi-he	ǹ-ɗʰì	'NC1-descendant'
ndeu-he	ǹ-ɗʰ̈	'NC3-boundary'
ndou-he	ǹ-ɗʰũ̀	'NC9-distant relative'
ngeuh	ŋ̀-gɰ̈ə̀	'NC1-guest'
ngou	ŋ̀-gù?	'NC9-year'
ngeu-he	ŋ́-ˈgʰú	'INF-do'
ngou-he	ŋ̀-gʰũ̀	'NC9-ritual'
le gi-he	lə̀-ɥʰì	'NMLZ-be intelligent'
le geu-he	lə̀-ɥʰ̈	'NMLZ-do'
lie	lí	'sleep.IMP'
leuh	lú	'horse'
louh	ˈlú	'FUT (post-hodiernal)'
le li-he	lə̀-ˈlʰį̃é	'NMLZ-host'
ma-leu-he	mə̀-ˈlʰú	'NC4-soldier ants'
wi-he	wʰì	'bile'
a-weuh	àwú	'who'
le-weu-he	lə̀-ˈwʰú	'NMLZ-die'
a zie	à-zì	'NC7-shelf'
le zweh-he	lə̀-ˈzʰwɛ́	'NMLZ-kill'
le zeh	lə̀-zɛ́	'NC5-grass'
le zeh-he	lə̀-zʰɛ́	'NMLZ-know'
nzeh-he	ǹ-zʰɛ̀	'NC9-road'
a veu-he	à-ˈvʰú	'NC7-ashes'
le vok	lə̀-ˈvók	'NMLZ-be insufficient'
mvo'	m̀-ˈvó?	'NC9-back part of bed'
mveu-he	m̀-ˈvʰú	'NC9-goat'

Appendix B Model summaries

Table B1. Summary for the SoE model for mean SoE across constriction, aspiration, and vowel: Reference level is the vowel interval

SoE model	Est. (Err)	L- 95%CrI	U-95% CrI	pd
(Intercept)	-4.03(0.62)	-5.24	-2.71	100
interval:constriction	-0.10(0.09)	-0.26	0.09	92
interval:aspiration	-1.13(0.64)	-2.22	0.33	94

Table B2. Summary for the CQ, CPP and H1*–A3*, mixed effects models, for mean measures of consonant intervals

CQ model- consonants	Est. (Err)	L- 95%CrI	U-95% CrI	pd
(Intercept)	0.52(0.04)	0.45	0.59	100
Aspiration	0.08(0.04)	-0.01	0.17	96
manner,fricative	-0.06(0.03)	-0.12	0.00	98
manner, prenas. fricative	-0.02(0.04)	-0.10	0.08	71
manner, prenas. stop	0.00(0.05)	-0.09	0.09	55
manner,fricative: aspiration	-0.02(0.06)	-0.13	0.10	71
manner prenas. fricative,aspiration	-0.08(0.0.5)	-0.17	0.04	94
manner prenas. stop, aspiration	-0.04(0.03)	-0.11	0.02	94
CPP model- consonants	**Est. (Err)**	**L-95%CrI**	**U-95%CrI**	**pd**
(Intercept)	20.42(0.96)	18.48	22.35	100
Aspiration	1.85(0.28)	1.27	2.35	100
Manner	0.86(1.15)	-1.43	3.21	81
aspiration:manner	-2.32(1.26)	-4.89	0.29	97
H1*–A3* model- consonants	**Est. (Err)**	**L-95%CrI**	**U-95%CrI**	**pd**
(Intercept)	8.47(4.71)	-0.94	18.16	96
Aspiration	-3.18(1.65)	-0.30	6.51	97
Manner	3.25(1.69)	-6.33	0.22	97
aspiration:manner	3.44(3.35)	-3.57	10.06	87

Table B3. Summary for mixed-effects models, for mean measures of vowel intervals

CQ model- vowels	Est. (Err)	L- 95%CrI	U-95% CrI	pd
(Intercept)	0.49(0.03)	0.43	0.55	100
Aspiration	0.06(0.06)	-0.06	0.17	89
prec. manner,fricative	0.01(0.04)	-0.07	0.10	59
prec. manner, prenas. fricative	0.00(0.04)	-0.08	0.08	52
prec. manner, prenas. stop	-0.01(0.02)	-0.05	0.02	90
prec. manner,fricative: aspiration	-0.03(0.04)	-0.10	0.05	84
prec. manner prenas. fricative,aspiration	-0.03(0.04)	-0.10	0.06	85
prec. manner prenas. stop, aspiration	-0.01(0.03)	-0.08	0.06	67
CPP model- vowels	**Est. (Err)**	**L-95%CrI**	**U-95%CrI**	**pd**
(Intercept)	21.46(1.47)	18.46	24.37	100
Aspiration	3.10(1.34)	0.26	5.61	98
prec. manner,fricative	0.27(0.50)	-0.66	1.40	77
prec. manner, prenas. fricative	-0.67(0.66)	-2.10	0.67	91
prec. manner, prenas. stop	-1.28(1.07)	-3.43	0.88	92

Table B3. Continued

CQ model- vowels	Est. (Err)	L- 95%Crl	U-95% Crl	pd
prec. manner,fricative: aspiration	-0.90(1.30)	-3.38	1.91	81
prec. manner prenas. fricative,aspiration	-2.25(0.80)	-3.65	-0.58	99
prec. manner prenas. stop, aspiration	-1.14(0.87)	-3.05	0.50	94
H1*–A3* model- vowels	**Est. (Err)**	**L-95%Crl**	**U-95%Crl**	**pd**
(Intercept)	10.98(4.05)	2.71	19.02	99
Aspiration	-6.19(2.13)	-10.17	-1.58	99
prec. manner,fricative	-0.51(2.12)	-5.00	3.74	61
prec. manner, prenas. fricative	-0.12(2.37)	-5.31	4.52	50
prec. manner, prenas. stop	0.71(1.39)	-2.21	3.59	74
prec. manner,fricative: aspiration	-3.12(2.91)	-8.21	3.48	88
prec. manner prenas. fricative,aspiration	-0.82(2.90)	-6.88	4.96	63
prec. manner prenas. stop, aspiration	-0.78(2.58)	-6.13	4.17	64

Appendix C. List of abbreviations

SoE Strength of excitation

CQ Contact quotient

CPP Cepstral peak prominence

H1*–A3* Amplitude of first harmonic (corrected for formant frequencies) minus amplitude of harmonic closest to F3 (corrected for formant frequencies)

CrI Credible interval

pd Probability of direction

NC# Noun class agreement, according to the numbering system used by Harro & Haynes (1991) and Bird (2003).

All other abbreviations in glosses conform to the Leipzig Glossing Rules.

References

Abramson, Arthur S. & Doug H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Ahn, Suzy. 2018. *The role of tongue position in voicing contrasts in cross-linguistic contexts.* PhD dissertation, New York University. https://www.proquest.com/dissertations-theses/role-tongue-position-voicing-contrasts-cross/docview/2133650209/se-2

Anderson, Stephen. 1982. From semivowels to aspiration to long consonants in Ngyembɔɔn-Bamileke. *Journal of West African Languages* 12(2), 58–68.

Anderson, Stephen. 2008. *A phonological sketch of Ngiemboon-Bamileke.* Unpublished manuscript. http://orthographyclearinghouse.org/papers/aPhonologicalSketchofNgiemboonBamileke.pdf

Ayuninjam, Funwi F. 1998. *A reference grammar of Mbili.* Lanham, MD: University Press of America.

Azieshi, Gisele. 1994. *Phonologie structurale de l'awing.* Master's thesis, Université de Yaoundé I. https://pure.mpg.de/rest/items/item_403400/component/file_403399/content

Bang, Hye-young, Morgan Sonderegger, Yoonjung Kang, Meghan Clayards & Tae-Jin Yoon. 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics* 66, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005

Beckman, Jill, Pétur Helgason, Bob McMurray & Catherine Ringen. 2011. Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics* 39(1), 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Benguerel, André-Pierre & Tej K. Bhatia. 1980. Hindi stop consonants: An acoustic and fiberscopic study. *Phonetica* 37, 134–148. https://doi.org/10.1159/000259987

van den Berg, Bianca. 2009. *A phonological sketch of Awing*. Yaoundé: SIL Cameroon. www.silcam.org/resources/archives/32292

Berkson, Kelly H. 2013. *Phonation types in Marathi: An acoustic investigation*. PhD dissertation, University of Kansas. http://hdl.handle.net/1808/12339

Berkson, Kelly H. 2019. Acoustic correlates of breathy sonorants in Marathi. *Journal of Phonetics* 73, 70–90. https://doi.org/10.1016/j.wocn.2018.12.006

Bhaskararao, Peri, & Peter Ladefoged. 1991. Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(2), 80–88. https://doi.org/10.1017/S0025100300004424

Bird, Steven. 1999. Dschang syllable structure. In Harry van der Hulst & Nancy A. Ritter, (Eds.), *The syllable: Views and facts*, 447–476. New York: Walter de Gruyter.

Bird, Steven. 2003. *Grassfields Bantu fieldwork: Dschang lexicon* [*electronic resource*]. Philadelphia, PA: Linguistic Data Consortium. https://doi.org/10.35111/z5x4-4x59

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2010. Phonetically-based sound patterns: Typological tendencies or phonological universals. In Cécile Fougeron, Barbara Kuehnert, Mariapaola d'Imperio & Nathalie Vallée (Eds.), *Laboratory phonology 10: Phonology and phonetics*, 201–224. New York: Walter de Gruyter.

Blust, Robert. 1974. A double counter-universal in Kelabit. *Research on Language & Social Interaction* 7(3–4), 309–324. https://doi.org/10.1080/08351817409370376

Blust, Robert. 2006. The origin of the Kelabit voiced aspirates: A historical hypothesis revisited. *Oceanic Linguistics* 45(2), 311–338. http://www.jstor.org/stable/4499967

Blust, Robert. 2016. Kelabit-Lun Dayeh phonology, with special reference to the voiced aspirates. *Oceanic Linguistics* 55(1), 246–277. https://doi.org/10.1353/ol.2016.0010

Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer (v6.1.16). Computer software. https://praat.org

Bombien, Lasse. 2011. *Segmental and prosodic aspects in the production of consonant clusters: On the goodness of clusters*. PhD dissertation, Universität München.

Bombien, Lasse, Christine Mooshammer, Philip Hoole & Barbara Kuehnert. 2008. Prosodic effects on articulatory coordination in initial consonant clusters in German. *The Journal of the Acoustical Society of America* 123(5), 3331. https://doi.org/10.1121/1.2933848

Brunelle, Marc, Jeanne Brown & Phạm Thị Thu Hà. 2022. Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis. *Phonetica* 79(2), 151–188. https://doi.org/10.1515/phon-2022-2019

Brunelle, Marc, Tạ Thành Tấn, James Kirby & Đinh Lư' Giang. 2020. Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology* 11(1), 15. https://doi.org/10.5334/labphon.278

Bürkner, Paul C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, Paul C. 2021. Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Catford, James C. 1982. *Fundamental problems in phonetics*. Bloomington, IN: Indiana University Press.

Catford, James C. 1988. *A practical introduction to phonetics*. Oxford: Clarendon Press.

Chai, Yuan, & Marc Garellek. 2022. On H1–H2 as an acoustic measure of linguistic phonation type. *The Journal of the Acoustical Society of America* 152(3), 1856–1870. https://doi.org/10.1121/10.0014175

Chirkova, Katia, Patricia Basset & Angélique Amelot. 2019. Voiceless nasal sounds in three Tibeto-Burman languages. *Journal of the International Phonetic Association* 49(1), 1–32. https://doi.org/10.1017/S0025100317000615

Cho, Taehong, & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Cho, Taehong, Doug Whalen & Gerard Docherty. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72, 52–65. https://doi.org/10.1016/j.wocn.2018.11.002

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61, 30–47. https://doi.org/10.1016/j.wocn.2017.01.001

Chong, Adam, Megan Risdal, Ann Aly, Jesse Zymet & Patricia Keating. 2020. Effects of consonantal constrictions on voice quality. *The Journal of the Acoustical Society of America* 148(1), EL65–EL71. https://doi.org/10.1121/10.0001585

Clements, George N. & Rajesh Khatiwada. 2007. Phonetic realization of contrastively aspirated affricates in Nepali. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 629–632. Saarbrücken: International Phonetic Association. www.icphs2007.de/conference/Papers/1650/1650.pdf

Davis, Katherine. 1994. Stop voicing in Hindi. *Journal of Phonetics* 22(2), 177–193. https://doi.org/10.1016/S0095-4470(19)30192-5

Demolin, Didier & Véronique Delvaux. 2001. Whispery voiced nasal stops in Rwanda. In Paul Daalsgard, Børge Lindberg & Henrik Benner (Eds.), *Proceedings of Eurospeech 2001 – Scandinavia*, 651–654. www.researchgate.net/profile/Veronique-Delvaux-2/publication/221484409_Whispery_voiced_nasal_stops_in_rwanda/links/556c118308aeab7772214f59/Whispery-voiced-nasal-stops-in-rwanda.pdf

Dhanajaya, N. & Bayya Yegnanarayana. 2009. Voiced/nonvoiced detection based on robustness of voiced epochs. *IEEE Signal Processing Letters* 17(3), 273–276. https://doi.org/10.1109/LSP.2009.2038507

Dixit, R. Prakash. 1989. Glottal gestures in Hindi plosives. *Journal of Phonetics* 17(3), 213–237. https://doi.org/10.1016/S0095-4470(19)30431-0

Dmitrieva, Olga and Indranil Dutta. 2020. Acoustic correlates of the four-way laryngeal contrast in Marathi. *Phonetica* 77(3), 209–237. https://doi.org/10.1159/000501673

Dutta, Indranil. 2007. *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt.* PhD dissertation, University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/82651

Eberhard, David, Gary Simons & Charles Fennig. 2020. *Ethnologue: Languages of the world.* Dallas, TX: SIL International.

Elias, Philip, Jacqueline Leroy, Jan Voorhoeve, Etienne Sadembouo, Engelbert Domche & Robert Breton. 1984. Mbam-Nkam or Eastern Grassfields. *Afrika und Übersee* 67(1), 31–107.

Engstrand, Olle. 1987. Preaspiration and the voicing contrast in Lule Sami. *Phonetica* 44(2), 103–116. https://doi.org/10.1159/000261784

Esposito, Adrian. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59(4), 197–231. https://doi.org/10.1159/000068347

Esposito, Christina M. 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2), 71–139. https://doi.org/10.1016/j.wocn.2010.02.002.

Esposito, Christina M. & Sameer ud Dowla Khan. 2012. Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association* 42(2), 123–143. https://doi.org/ 10.1017/S0025100312000047

Esposito, Christina M. & Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14(12), e12392. https://doi.org/10.1111/lnc3.12392

Esposito, Christina M., Sameer ud Dowla Khan, Kelly H. Berkson & Max Nelson. 2020. Distinguishing breathy consonants and vowels in Gujarati. *Journal of South Asian Languages and Linguistics* 6(2), 215–243. https://doi.org/10.1515/jsall-2019-2011

Fowler, Carol & Eliot Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36(2–3), 171–195. https://doi.org/10.1177/002383099303600304

Franich, Kathryn. 2018. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* 67, 21–33. https://doi.org/10.1016/j.wocn.2017.11.001

Fujimura, Osamu. 1990. Demisyllables as sets of features: Comments on Clements' paper. In John Kingston & Mary E. Beckman (Eds.), *Papers in Laboratory Phonology: Volume 1, Between the Grammar and Physics of Speech*, 334–340. Cambridge: Cambridge University Press.

Garellek, Marc. 2020. Acoustic discriminability of the complex phonation system in !Xóõ. *Phonetica* 77(2), 131–160. https://doi.org/10.1159/000494301

Garellek, Marc, Yuan Chai, Yaqian Huang & Maxine Van Doren. 2021. Voicing of glottal consonants and non-modal vowels. Journal of the International Phonetic Association First Look, 1–28. https://doi.org/10.1017/S0025100321000116

Garellek, Marc, Amanda Ritchart & Jianjing Kuang. 2016. Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics* 59, 110–121. https://doi.org/10.1016/j.wocn.2016.09.001

Gerlach, Linda. 2015. Phonetic and phonological description of the N!aqriaxe variety of ǂ'Amkoe and the impact of language contact. PhD dissertation, Humboldt University, Berlin.

Gobl, Christer & Ailbhe Ní Chasaide. 2019. Time to frequency domain mapping of the voice source: The influence of open quotient and glottal skew on the low end of the source spectrum. In Gernot Kubrin & Zdravko Kačič (Eds.), *Proceedings of INTERSPEECH 2019*, 1961–1965.

Greenberg, Joseph H. 1978. Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsík (Eds.), *Universals of human language, volume 2: Phonology.* Stanford: Stanford University Press.

Güldemann, Tom, & Hirosi Nakagawa. 2018. Anthony Traill and the holistic approach to Kalahari Basin sound design. *Africana Linguistica* 24, 45–73. https://doi.org/10.2143/AL.24.0.3285491

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2020. Glottolog 4.3. Jena: Max Planck Institute for the Science of Human History. https://doi.org/10.5281/zenodo.4061162

Harro, Gretchen & Nancy Haynes. 1991. *Grammar sketch of Yemba*. Yaoundé: SIL Cameroon. www.sil.org/resources/archives/47892

Haynes, Nancy. 1989. Une esquisse phonologique du yemba. In Daniel Barreteau & Robert Hedinger (Eds.), *Descriptions de langues camerounaises*, 179–236. Paris: Agence de Coopération Culturelle et Technique.

Henton, Caroline, Peter Ladefoged & Ian Maddieson. 1992. Stops in the world's languages. *Phonetica* 49(2), 65–101. https://doi.org/10.1159/000261905

Hirose, Hajime, Charles Y. Lee & Tatsujiro Ushijima. 1974. Laryngeal control in Korean stop production. *Journal of Phonetics* 2(2), 145–152. https://doi.org/10.1016/S0095-4470(19)31189-1

Hoole, Philip & Lasse Bombien. 2017. A cross-language study of laryngeal-oral coordination across varying prosodic and syllable-structure conditions. *Journal of Speech, Language, and Hearing Research* 60(3), 525–539. https://doi.org/10.1044/2016_JSLHR-S-15-0034

Hoole, Philip, Lasse Bombien, Barbara Kühnert & Christine Mooshammer. 2009. Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In Gunnar Fant, Hiroya Fujisaki & Jiaxuan Shen (Eds.), *Frontiers in phonetics and speech science*, 275–287. Beijing: Commercial Press. https://shs.hal.science/halshs-00681953

Hussain, Qandeel. 2018. A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics* 71, 284–305. https://doi.org/10.1016/j.wocn.2018.09.011

Hyman, Larry M. 1972. A phonological study of Fe'fe'-Bamileke. PhD dissertation, UCLA. https://linguistics.ucla.edu/images/stories/Hyman.1972.pdf

Hyman, Larry M. 1985. Word domains and downstep in Bamileke-Dschang. *Phonology Yearbook* 2(1), 47–83. https://doi.org/10.1017/S0952675700000385

Hyman, Larry M. & Maurice Tadadjeu. 1976. Floating tones in Mbam-Nkam. In Larry M. Hyman (ed.), *Southern California Occasional Papers in Linguistics 3: Studies in Bantu Tonology*, 59–1. Los Angeles, CA: University of Southern California.

Islam, Md. Jahurul. 2019. Phonetics and phonology of 'voiced-aspirated' stops: Evidence from production, perception, alternation and learnability. PhD dissertation, Georgetown University. http://hdl.handle.net/10822/1056010

Kagaya, Ryohei. 1974. A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics* 2, 161–180. https://doi.org/10.1016/S0095-4470(19)31191-X

Kagaya, Ryohei & Hajime Hirose. 1975. Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 9, 27–46.

Karlsson, Anastasia M. & Jan-Olof Svantesson. 2012. Aspiration of stops in Altaic languages: An acoustic study. *Altai Hakpo* 22, 205–222. www.researchgate.net/profile/Anastasia-Karlsson-2/publication/262067964_Aspiration_of_stops_in_Altaic_languages_An_acoustic_study/links/0c960536916fe593ef000000/Aspiration-of-stops-in-Altaic-languages-An-acoustic-study.pdf

Kawahara, Hideki, Alain de Cheveigné & Roy Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: Revised TEMPO in the STRAIGHT-suite. In *Proceedings of the Fifth International Conference on Spoken Language Processing* (*ICSLP98*). www.isca-speech.org/archive_v0/archive_papers/icslp_1998/i98_0659.pdf

Keating, Patricia. 1990. The window model of coarticulation: Articulatory evidence. In John Kingston & Mary E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 451–470. Cambridge: Cambridge University Press.

Keating, Patricia & Christina Esposito. 2007. Linguistic Voice Quality. *UCLA Working Papers in Phonetics* 105, 85–91. https://escholarship.org/uc/item/04r5q6qn

Keating, Patricia, Jianjing Kuang, Marc Garellek, Christina M. Esposito & Sameer ud DowlaKhan. (2023). A cross-language acoustic space for vocalic phonation distinctions. *Language* 99(2), 351–389. https://doi.org/10.1353/lan.2023.a900090

Khan, Sameer ud Dowla. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40(6), 780–795. https://doi.org/10.1016/j.wocn.2012.07.001

Kim, Hyunsoon, Shinji Maeda & Kiyoshi Honda. 2010. Invariant articulatory bases of the features [tense] and [spread glottis] in Korean plosives: New stroboscopic cine-MRI data. *Journal of Phonetics* 38(1), 90–108. https://doi.org/10.1016/j.wocn.2009.03.003

Kim, Sahyang, Jiseung Kim & Taehong Cho. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics* 71, 65–80. https://doi.org/10.1016/j.wocn.2018.07.004

Kingston, John. 1985. *The phonetics and phonology of the timing of oral and glottal events.* PhD dissertation, University of California, Berkeley.

Kirby, James. 2014. Acoustic transitions in Khmer word-initial clusters. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia & Doris Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production*, 234–237. www.lel.ed.ac.uk/~jkirby/docs/kirby2014transitions-preprint.pdf

Kirby, James. 2013. The role of probabilistic enhancement in phonologization. In Alan C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization*, 228–246. Oxford: Oxford University Press.

Klatt, Dennis. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18(4), 686–706. https://doi.org/10.1044/jshr.1804.686

Kouega, Jean-Paul. 2007. The language situation in Cameroon. *Current Issues in Language Planning* 8(1), 3–93. https://doi.org/10.2167/cilp110.0

Kreitman, Rina. 2010. Mixed-voicing word-initial onset clusters. In Cécile Fougeron, Barbara Kuehnert, Mariapaola d'Imperio & Nathalie Vallée (Eds.), *Laboratory Phonology 10: Phonology and Phonetics*, 169–200. New York: Walter de Gruyter.

Krivokapić, Jelena. 2014. Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658), 20130397. https://doi.org/10.1098/rstb.2013.0397

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics.* Chicago, IL: University of Chicago Press.

Ladefoged, Peter. 1973. The features of the larynx. *Journal of Phonetics* 1(1):73–83. https://doi.org/10.1016/S0095-4470(19)31376-2

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics: Seventh edition.* Stamford, CT: Cengage.

Ladefoged, Peter & Ian Maddieson 1996. *The sounds of the world's languages.* Malden, MA: Blackwell.

Ladefoged, Peter & Anthony Traill. 1984. Linguistic phonetic descriptions of clicks. *Language* 60(1), 1–20. https://doi.org/10.2307/414188

Ladefoged, Peter, Kay Williamson, Benny Elugbe & S. Owulaka. 1976. The stops of Owerri Igbo. *Studies in African Linguistics* supplement 6, 147–163.

Laver, John. 1980. *The phonetic description of voice quality.* Cambridge: Cambridge University Press.

Laver, John. 1994. *Principles of phonetics.* Cambridge: Cambridge University Press.

Lenth, Russell. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means (R package, version 1.7.1-1). https://CRAN.R-project.org/package=emmeans

Lindblom, Björn. 1983. Economy of speech gestures. In Peter MacNeilage (Ed.), *The production of speech*, 217–245. New York: Springer.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422. https://doi.org/10.1080/00437956.1964.11659830

Löfqvist, Anders. 1980. Interarticulator programming in stop production. *Journal of Phonetics* 8(4), 475–490. https://doi.org/10.1016/S0095-4470(19)31502-5

Löfqvist, Anders & Hirohide Yoshioka. 1981. Laryngeal activity in Icelandic obstruent production. *Nordic Journal of Linguistics* 4(1), 1–18. https://doi.org/10.1017/S0332586500000639

Löfqvist, Anders, Thomas Baer & Hirohide Yoshioka. 1981. Scaling of glottal opening. *Phonetica* 38(5–6), 265–276. https://doi.org/ 10.1159/000260032

Löfqvist, Anders & Richard S. McGowan. 1992. Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* 20, 93–110. https://doi.org/10.1016/S0095-4470(19)30256-6

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization.* PhD dissertation, University of Massachusetts Amherst.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17, 267–302. https://doi.org/10.1023/A:1006182130229

Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge: Cambridge University Press.

Maddieson, Ian. 1991. Articulatory phonology and Sukuma "aspirated nasals". In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 17(2), 145–154. http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/1646/1420

Makowski, Dominique, Mattan Ben-Shachar & Daniel Lüdecke. 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4(40), 1541. https://doi.org/10.21105/joss.01541

Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65(4), 231–256. https://doi.org/10.1159/000192794

Mikuteit, Simone & Henning Reetz. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50(2), 247–277. https://doi.org/10.1177/00238309070500020401

Mittal, Vinay Kumar, Bayya Yegnanarayana & Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America* 136(4), 1932–1941. https://doi.org/10.1121/1.4894789

Murty, Kodukula Sri Rama & Bayya Yegnanarayana. 2008. Epoch extraction from speech signals. In *IEEE Transactions on Audio, Speech, and Language Processing* 16(8), 1602–1613. https://doi.org/ 10.1109/TASL.2008.2004526

Myhre, Matias. 2021. *The weight and representation of Ryukyuan Miyako onsets: Initial geminate moraicity, markedness, and sonority.* PhD dissertation, The Arctic University of Norway. https://munin.uit.no/handle/10037/22926

Nanfah, Gaston. 2003. *Analyse contrastive des parlers Yémba du département de la Ménoua de l'Ouest-Cameroon.* Köln: Rüdiger Köppe Verlag.

Narayan, Chandan R. 2022. Speaking rate, oro-laryngeal timing, and place of articulation effects on burst amplitude: Evidence From English and Tamil. *Language and Speech*, online first. https://doi.org/10.1177/00238309221133836

Naumann, Christfried. 2016. The phoneme inventory of Taa (West !Xoon dialect). In Rainer Voßen & Wilfrid H. G. Haacke (Eds.), *Lone tree: Scholarship in service of the Koon: Essays in memory of Anthony T. Traill*, 341–382. Köln: Rüdiger Köppe Verlag.

Ndedje, René. 2003. *La morphologie nominale du mǝŋgambo.* Master's thesis, Université Yaoundé I. https://pure.mpg.de/rest/items/item_403509/component/file_403508/content

Nganmou, Alise. 1991. *Modalités verbales: temps, aspects et modes en madumba.* PhD dissertation, Université Yaoundé 1. https://pure.mpg.de/rest/items/item_403507_6/component/file_403506/content

Ngouagna, Jean Pierre. 1988. *Esquisse phonologique du ngomba.* Master's thesis, Université de Yaoundé. https://pure.mpg.de/rest/items/item_403547/component/file_403546/content

Nguendjio, Émile-Gille. 1989. *Morphologie nominale et verbale de la langue baŋgwa.* PhD dissertation, Université de Yaoundé. https://pure.mpg.de/rest/items/item_403433_1/component/file_403432/content

Ngueyep, Justin. 1988. *Essai de description phonologique du bamena.* Master's thesis, Université de Yaoundé. https://pure.mpg.de/rest/items/item_403427_3/component/file_403426/content

Ní Chasaide, Ailbhe & Christer Gobl. 1993. Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36(2–3), 303–330.

Nissim, Gabriel M. 1981. *Le Bamileke-Ghomálá' (parler de Bandjoun, Cameroun): Phonologie, morphologie nominale, comparaison avec des parlers voisins.* Paris: Société d'Etude Linguistiques et Anthropologiques de France.

Njeck, Mathaus Mbah. 2003. *A phonology of Bǎŋgulaŋ and a proposed orthography.* Master's thesis, Université Yaoundé I. https://pure.mpg.de/rest/items/item_403431/component/file_403430/content

Nusi, Jean. 1986. *Esquisse phonologique du Ti: Parler des Ti de la province de l'Ouest-Cameroun.* Master's thesis, Université de Yaoundé. https://pure.mpg.de/rest/items/item_403573_4/component/file_403572/content

Ohala, John J. 1997. Aerodynamics of phonology. In *Proceedings of the Seoul International Conference on Linguistics*, 1–6. Seoul: Linguistic Society of Korea.

Poon, Pamela & Catherine Mateer. 1985. A study of VOT in Nepali stop consonants. *Phonetica* 42, 39–47. https://doi.org/10.1159/000261736

Pouplier, Marianne, Manfred Pastätter, Philip Hoole, Stefania Marin, Ioana Chitoran, Tomas O. Lentz & Alexei Kochetov. 2022. Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics* 93, 101153. https://doi.org/10.1016/j.wocn.2022.101153

Rami, Manish, Joseph Kalinowski, Andrew Stuart & Michael Rastatter. 1999. Voice onset times and burst frequencies of four velar stop consonants in Gujarati. *The Journal of the Acoustical Society of America* 106(6), 3736–3738. https://doi.org/10.1121/1.428226

Recasens, Daniel & Aina Espinosa. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America* 125(4), 2288–2298. https://doi.org/ 10.1121/1.3089222

Riehl, Anastasia. 2008. *The phonology and phonetics of nasal-obstruent sequences.* PhD dissertation, Cornell University. https://conf.ling.cornell.edu/pdfs/Riehl-2008.pdf

Rose, Phil. 1989. Phonetics and phonology of the yang tone phonation types in Zhenhai. *Cahiers de Linguistique: Asie Orientale* 18(2), 229–245. https://doi.org/10.1163/19606028-90000316

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/

RStudio Team. 2022. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. www.rstudio.com/

Saltzman, Elliot & Kevin Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2

Schertz, Jessamyn & Sarah Khan. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics* 81, 100979. https://doi.org/10.1016/j.wocn.2020.100979

Schiefer, Leiselotte. 1987. 'Voiced aspirated' or 'breathy voiced' and the case for articulatory phonology. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 27, 257–278.

Schwarz, Martha, Morgan Sonderegger & Heather Goad. 2019. Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics* 73, 113–127. https://doi.org/10.1016/j.wocn.2018.12.007

Seyfarth, Scott & Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71, 425–450. https://doi.org/10.1016/j.wocn.2018.09.001

Shinohara, Shigeko & Masako Fujimoto. 2011. Moraicity of initial geminates in the Tedumuni dialect of Okinawa. In Wai-Sum Lee & Eric Zee (Eds.), *Proceedings of ICPhS 17*, 1826–1829. Hong Kong: City University of Hong Kong.

Shue, Yen-Liang, Patricia Keating, Chad Vicenik & Kristine Yu. 2011. VoiceSauce: A program for voice analysis. In Wai-Sum Lee & Eric Zee (Eds.), *Proceedings of ICPhS 17*, 1846–1849. Hong Kong: City University of Hong Kong.

Snyman, Jan W. 1975. *Žuǀ'hõasi fonologie en woordeboek.* Rotterdam: A. A. Balkema.

Solé, Maria-Josep. 2010. Effects of syllable position on sound change: An aerodynamic study of final fricative weakening. *Journal of Phonetics* 38(2), 285–305. https://doi.org/10.1016/j.wocn.2010.02.001

Solé, Maria-Josep. 2018. Articulatory adjustments in initial voiced stops in Spanish, French and English. *Journal of Phonetics* 66, 217–241. https://doi.org/10.1016/j.wocn.2017.10.002

Sóskuthy, Martín. 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Unpublished ms, UCLA. https://doi.org/10.1.1.16.9312.

Stevens, Kenneth N. 1998. *Acoustic phonetics.* Cambridge, MA: MIT Press.

Tabain, Marija, Marc Garellek, Birgit Hellwig, Adele Gregory & Richard Beare. 2022. Voicing in Qaqet: Prenasalization and language contact. *Journal of Phonetics* 91, 101138. https://doi.org/10.1016/j.wocn.2022.101138

Tehrani, Henry. 2020. EGGWorks, a free program for EGG analysis [software].

Terhijia, Viyazonuo & Priyankoo Sarmah. 2020. Aspiration in voiceless nasals in Angami. In *Proceedings of Meetings on Acoustics* 42(1), 060008. https://doi.org/10.1121/2.0001403

Tian, Jia & Jianjing Kuang. 2021. The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association* 51(2), 202–228. https://doi.org/10.1017/S0025100319000148

Topintzi, Nina & Stuart Davis. 2017. On the weight of edge geminates. In Haruo Kubozono (ed.), *The phonetics and phonology of geminate consonants*, 260–282. Oxford: Oxford University Press.

Topintzi, Nina & Andrew Nevins. 2017. Moraic onsets in Arrernte. *Phonology* 34(3), 615–650. https://doi.org/10.1017/S0952675717000306

Traill, Anthony. 1985. *Phonetic and phonological studies of !Xóõ Bushman.* Quellen zur Khoisan-Forschung, vol. 1. Hamburg: Helmut Buske.

Traill, Anthony & Michel Jackson. 1988. Speaker variation and phonation type in Tsonga nasals. *Journal of Phonetics* 16(4), 385–400. https://doi.org/10.1016/S0095-4470(19)30517-0

Tsafack Forku, Doris. 2000. *A sketch of the phonology of Mamənyan and standardization perspectives.* Master's thesis, Université Yaoundé I. https://pure.mpg.de/rest/items/item_403501_1/component/file_403500/content

Tsofack, Jean-Benoît. 2010. Le français langue pluricentrique: Des aspects dans quelques pratiques à l'Ouest-Cameroun. *Le français en Afrique* 25, 243–258. www.unice.fr/bcl/ofcaf/25/Tsofackə20Jean-Benoit.pdf

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Dirk van Rijn. 2022. Itsadug: Interpreting time series and autocorrelated data using GAMMs (R package, version 2.4.1). https://cran.r-project.org/web/packages/itsadug/index.html

Voorhoeve, Jan. 1964. The structure of the morpheme in Bamileke (Bangangté dialect). *Lingua* 13, 319–334. https://doi.org/10.1016/0024-3841(64)90034-8

Watters, John R. 2003. Grassfields Bantu. In Derek Nurse & Gérard Philippson (Eds.), *The Bantu Languages*, 225–256. London: Routledge.

Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2), 179–186. https://doi.org/10.1016/j.cognition.2013.03.002

Wood, Simon. 2017. *Generalized additive models: An introduction with R.* Boca Raton, FL: Chapman and Hall/CRC.

Yadav, Ramawatar. 1984. Voicing and aspiration in Maithili: A fiberoptic and acoustic study. *Indian Linguistics* 45, 1–30.