Chapter 1

Supralaryngeal articulation across voicing and aspiration in Yemba vowels

Jae Weller
University of Wisconsin

Matthew Faytak
University at Buffalo

Jeremy Steffman
Northwestern University

Connor Mayer
University of California, Irvine

G. Teixeira
University of California, Los Angeles

Rolain Tankou
California Bamileke Associations

Both prevoicing and aspiration are thought to affect supraglottal cavity size during and following the production of stop consonants. However, many languages implement stop laryngeal contrasts using both prevoicing and aspiration, making it difficult to independently link observed tongue position differences to one effect or the other. Yemba (Grassfields Bantu) offers a unique opportunity to separately study these two effects, as the language contains stop consonants which independently vary in prevoicing and aspiration. The current study examines tongue position in vowels following Yemba stop consonants using acoustic and ultrasound data. For-

Change author. Change title. To appear in: Change Volume Editor. Change volume title. Berlin: Language Science Press. [preliminary page numbering]

mant frequencies of vowels following stops suggest that stop prevoicing conditions tongue root advancement and tongue body lowering, resulting in an expansion of the supraglottal cavity; the same data suggest that stop aspiration conditions tongue root retraction, resulting in an opposing contraction of the supraglottal cavity. Ultrasound data for a subset of acoustic study participants was not entirely consistent with the acoustic data, showing tongue body lowering and root retraction after aspirated stops, but no consistent lingual adjustments after prevoiced stops. These findings, from the first phonetic study of a Bamileke language's laryngeal contrasts, lend support to active cavity constriction as a supporting mechanism for aspiration.

1 Introduction

1.1 Stop laryngeal contrast and tongue position

Sustaining voicing during stop closure presents a well-documented challenge in speech production: in order to sustain voicing, there must be a sufficient negative pressure gradient across the glottis, with a lower supraglottal pressure compared to subglottal pressure. During a stop, a complete supraglottal closure must be made, blocking airflow out of the vocal tract and rapidly causing the pressure differential critical to maintaining voicing to dissipate (Ohala 1983; 2011; Westbury 1983). While a sufficient pressure gradient can be achieved for a short time during stop closure with no articulatory adjustments, voicing cannot be sustained long enough to account for the extent of prevoicing in most voiced stops (Ohala 1983).

To maintain a sufficiently large negative pressure gradient in support of voicing production, a wide range of adjustments to the articulators can be used to enlarge the supraglottal cavity, thereby decreasing supraglottal pressure. Venting into the nasal cavity through an incompletely sealed velum (Rothenberg 1968; Solé 2018) and lowering the larynx (Hombert et al. 1979; Kirby & Ladd 2016; Ohala & Riordan 1979) also act to maintain this pressure gradient. A range of adjustments to the pharyngeal walls and tongue have also been observed. One long-noted strategy is increasing tongue or vocal tract wall tissue compliance so that the oral or pharyngeal cavities can passively expand during voiced stop closure (Bell-Berti 1975; Kent & Moll 1969; Rothenberg 1968; Sprouse et al. 2008). The stiffness of lingual constrictions is also known to be reduced to increase lingual tissue compliance: voiced stops tend to have lighter constrictions with a smaller area of tongue-palate contact (Dixit 1990; Fletcher 1989; Fujimoto et al. 2021; Kochetov & Kang 2017).

An additional strategy for supraglottal cavity expansion is the primary focus of this study: active expansion of the supraglottal cavity through tongue body lowering and/or tongue root advancement. Such modifications to lingual articulation have been observed during phonologically voiced stops in a range of languages including English, Ikema Miyako, Oromo, Portugese, and Russian (Ahn 2015; 2018; Fujimoto et al. 2021; Matsui & Kochetov 2018; Percival et al. 2018; Westbury 1983). Constriction location itself is also thought to be slightly more anterior for prevoiced stops compared to their voiceless counterparts, particularly in prevoiced geminate stops (Krishnaswamy et al. 2018; Percival et al. 2018).

The production of aspiration, on the other hand, has been associated with a smaller supraglottal cavity, though there has been relatively little research on this topic. Reduced supraglottal cavity size may facilitate the production of aspiration because it allows for more pressure buildup, and thus a more intense burst of air on release (Ahn 2018). A potentially analogous result was observed for voiceless fricatives, which may exhibit an active constriction of the pharyngeal cavity (Proctor et al. 2010). This has been speculated to aid in pressure regulation, with a more constricted vocal tract allowing more turbulence to be produced with less airflow. This finding may extend to aspirated stops: a more constricted supraglottal cavity might enable the speaker to more reliably produce noisier aspiration.

The two phonetic features at issue here, aspiration and prevoicing, are not independently distributed across the stop inventories of many languages. That is, although some phonological "voicing" contrasts are implemented solely using consistent prevoicing in "voiced" stops (Benguerel et al. 1978; Kirby & Ladd 2016), and others solely using aspiration in "voiceless" stops (Deterding & Nolan 2007), many of the most commonly studied languages employ both, realizing "voiced" stops with prevoicing and "voiceless" stops with aspiration (Beckman et al. 2013; Keating 1984). This is the case in English and German, for example, where voiced stops may lack prevoicing but are never aspirated, and where voiceless stops are realized with aspiration in many contexts. The literature on laryngeal contrast and tongue position discussed above primarily concerns English and a series of "true voicing" languages with consistent prevoicing in voiced stops.

The lack of independence between the occurrence of prevoicing and the occurrence of aspiration in such languages makes it difficult to isolate their respective effects on the vocal tract: as discussed in Ahn (2018), if a relatively retracted tongue position is associated with voiceless (aspirated) stops in such a

Table 1: Example words for alveolar stops varying orthogonally by voicing and aspiration. The top item in each cell contains an onset preceded by a nasal consonant.

	voiceless		voiced	
aspirated	ń-tí	‘INF-write’	ǹ-d̀ì	‘CL1-lord’
	tí-í	‘write-IMP’	*[−nas]-di	
unaspirated	ń-$^{!}$t^{h}í	‘INF-host’	ǹ-d^{h}ì	‘CL1-descendant’
	mə̀-t^{h}í	‘CL6-saliva’	*[−nas]-d^{h}i	

language, then it is not clear if the cause of this difference is retraction due to the presence of stop aspiration, or the absence of a tongue-advancing effect caused by stop prevoicing. In the next section we will introduce Yemba, a language in which prevoicing and aspiration are unusually independent, which facilitates research addressing the influence of these features on vocal tract state.

1.2 Yemba laryngeal contrasts

Yemba (ISO 369-3 [ybb]), also known as Dschang or Dschang Bamileke, is a Bamileke language (in the larger Grassfields Bantu family) spoken in the West Province of Cameroon (Hammarström et al. 2021) and in diaspora communities primarily located in North America and Europe. It is spoken by at least 300,000 to 400,000 speakers, most of whom are bi- or trilingual in French, Cameroonian Pidgin English, English, and one or more other languages of non-colonial origin such as other Bamileke varieties (Eberhard et al. 2021).

Yemba is of general interest to phonetics and phonology for the complexity of its laryngeal contrasts. Like most Grassfields Bantu languages, Yemba has a surface contrast between stops with and without prevoicing. A phonemic aspiration contrast cross-cuts the voicing contrast (Bird 1999). Aspiration may be associated with both voiceless and prevoiced segments, including most sonorants and fricatives. Unusually, aspiration is always voiceless, even when associated with a prevoiced segment. The Yemba surface inventory thus includes a four-way laryngeal contrast in stops between voiceless unaspirated, prevoiced unaspirated, voiceless aspirated, and prevoiced aspirated (Table 1).

This four-way laryngeal contrast also occurs in several other Bamileke languages, including varieties of Ngyemboon (Anderson, 2008), Fe’fe’ (Hyman

Table 2: Example alternations in the presence of a prefixal nasal.

Oral preceding		Nasal preceding	
lə-lʰʉ̀	‘CL5-wrestle’ (‘wrestling’)	ń-ꜜdʰʉ́	‘INF-wrestle’
lə-ꜜlí	‘CL5-sleep’ (‘sleeping’)	ń-dí	‘INF-sleep’
à-ꜜpú	‘CL7-hand’ (sg.)	m̀-ꜜbú	‘CL6-hand’ (pl.)
lə̀-ꜜpʰʉ́	‘CL5-hole’ (sg.)	m̀-ꜜbʰʉ́	‘cl6-hole’ (pl.)

1972), Nda’nda’ (Ngueyep 1988), Ngomba (Ngouagna 1988), and Ghomálá’ (Nissim 1981). While superficially similar to the four-way contrast attested in many Indo-Aryan languages, voiced aspirates in Indo-Aryan languages typically have breathy-phonated release (Berkson 2013; Dmitrieva & Dutta 2020; Schertz & Khan 2020; Schwarz et al. 2019). Bamileke voiced aspirates are crucially distinct from this stop type in that they are produced with a sequence of prevoicing followed by voiceless aspiration, a characteristic noted in impressionistic descriptions of Bamileke languages (Anderson 1982; Bird 1999; Hyman 1972; Nissim 1981) and supported by more recent instrumental work (Faytak & Steffman 2021).

Two additional considerations must be mentioned here. First, as Table 1 shows, the Yemba voiced stops [b(ʰ) d(ʰ) g(ʰ)] are always preceded by a nasal, though fricatives exhibit voicing contrasts whether or not they are preceded by a nasal. Second, standard phonological analyses of Yemba have treated the prevoiced stops [b(ʰ) d(ʰ) g(ʰ)] as allophones of /p(ʰ) l(ʰ) ɰ(ʰ)/, respectively (Bird 1999). This has been motivated by alternations triggered by the addition of a tone-bearing, syllabic nasal prefix as shown in Table 2. We return to the connection between voicing and preceding nasality in more detail in the discussion section.

In a sense, then, voicing is not an underlying contrast in stops, but rather one that is contingent on the presence of a preceding nasal. This is not a significant concern for the present study. First, many forms, such as those in the voiced column of Table 1, do not exhibit these alternations at all, which casts doubt on whether they can be analyzed as underlying a prefixal nasal followed by /p l ɰ/. In addition, voicing alone indicates a meaningful phonological contrast elsewhere in the language, with fricatives contrasting for voicing in the absence of a preceding nasal (e.g. /s/ vs. /z/). Finally, the surface phonetic contrast between voiceless stops preceded by a nasal (e.g. [nt]) and voiced stops preceded by a nasal (e.g. [nd]) is robust. Thus, whether or not they correspond to a true phonological voicing contrast at a more abstract level, the dimensions of this surface contrast are aspiration and prevoicing. The relative independence between these dimensions in Yemba provides a unique test case for isolating the effects of prevoicing and aspiration on supraglottal cavity size.

1.3 The present study

In languages whose two-way phonological voicing contrasts are based on some combination of both prevoicing and aspiration, such as English, the causality of differences in tongue position related to laryngeal activity is difficult to work out. It is difficult to specifically attribute observed effects of phonological "voicing" on tongue position to either aspiration or prevoicing, since each phonological category – "voiced" and "voiceless" – can be characterized by the presence of one, and the absence of the other, phonetic event. The laryngeal contrasts present in Yemba provide a unique opportunity to disentangle the effects of prevoicing and aspiration on tongue position, since the surface inventory of the language contains prevoiced stops, aspirated stops, and stops with *both* associated prevoicing *and* voiceless aspiration.

Given these facts, the present study aims to use acoustic and ultrasound data to establish the effects of prevoicing and aspiration on tongue position in vowels following Yemba stops. Based on prior studies, we expect voicing to condition cavity expansion via tongue root advancement or tongue body lowering of the following vowel. Conversely, we expect aspiration to condition cavity contraction via tongue root retraction or tongue body raising. Some aspects of tongue position can be inferred from formant frequency measurements: advancement/expansion due to voicing may be reflected in *raised* F1 or F2, and retraction/contraction due to aspiration may be reflected in *lowered* F1 or F2. Analysis of ultrasound video subsequent to acoustic analysis is used to confirm or deny the existence of the lingual adjustments suggested by the acoustic analysis, not all of which may be straightforwardly reflected in the acoustic signal (Atal et al. 1978; Stevens 1989; Stevens & Keyser 2010).

2 Methods

2.1 Acoustic study

2.1.1 Participants and materials

Four native Yemba speakers' speech was analyzed. Two speakers (1M, 1F) were recorded in the UCLA Phonetics Lab. Audio recordings of two additional Yemba speakers (2M) were taken from a multimedia lexicon which had previously been recorded in Cameroon (Bird 2003). Speakers were between the ages of 31 and 45 at the time of recording. We describe the materials collected separately for each group below since the circumstances of data collection differed.

2.1.2 In-lab recordings

Speakers 1 and 2 were recorded with a head-mounted Shure SM10A microphone (32-bit audio, 44.1 kHz sampling rate) in a sound-attenuated booth at the UCLA Phonetics Lab. Stimuli for in-lab recordings were selected in collaboration with the last author, a native speaker and member of the Yemba speech community. Stimuli consisted of near-minimal sets contrasting minimally in aspiration and voicing (but sometimes differing in tone and prefixal morphology). Open-syllable words with the following onsets were selected: voiceless stops [p], [t], [k], voiced stops [b], [d] , and their aspirated counterparts [pʰ], [tʰ], [kʰ], [bʰ], [dʰ]. Words containing the voiced velar onsets [g] or [gʰ] were omitted from the data due to difficulties in finding words with a balanced set of following vowels. Target onset consonants were followed by one of three high vowels /i/, /u/, or /ʉ/, since aspirated stops only precede high vowels (Bird 1999).

Stimuli were produced in the frame sentence in (1), which was presented on a laptop screen using an ad hoc orthography developed in collaboration with the last author.

(1) máŋ ꜜlíɛ́ ɰʰə̀ ____ ǹ-bù ȕ
1SG.SBJ say COMP ____ INF-give 2SG.OBJ
'I say ____ to you'

A total of 35 word types were recorded per speaker (see Appendix for full list); each word containing a voiced stop (aspirated or unaspirated) was repeated 6 times. The number of repetitions of voiceless unaspirated stop items varied by speaker: each word with a voiceless unaspirated stop was repeated 3 times by Speaker 1, and 2 times by Speaker 2. After excluding speech errors and misreadings, this resulted in a total of 227 tokens for both speakers combined: 62 voiced and aspirated, 77 voiced and unaspirated, 44 voiceless unaspirated, and 54 voiceless unaspirated.

2.1.3 Data selection from Bird (2003)

To augment the lab recordings, a total of 277 stop tokens were taken from the audio lexicon in Bird (2003). The lexicon was reviewed for words that contained open syllables with the onsets used to select lab stimuli of interest. Since Bird (2003) was not constructed with the current analysis in mind, the balance of stop types is not even between the four voicing and aspiration categories: there are more unaspirated than aspirated tokens, and more voiceless than voiced tokens. Specifically, 14 prevoiced aspirated stop tokens, 56 voiceless aspirated stop tokens, 51 prevoiced unaspirated stop tokens, and 156 voiceless unaspirated stop tokens were collected.

Some minor differences exist between the lab and lexicon materials. The lexicon recordings are all read in isolation due to the format of the material, as opposed to the lab-recorded material, which was recorded in a frame sentence. The lexicon recordings contain more word types (139) compared to the lab recordings, with fewer tokens for each type: both speakers produce a large variety of words one or two times. Several word types selected for even coverage of onset types had a glottal stop coda. Tokens were not excluded on the basis of the number of syllables in the word; tokens were taken from words of varying lengths.

2.1.4 Analysis

All recordings, regardless of origin, were segmented manually in Praat. The beginning of modal phonation after stop release was used to determine the beginning of each token's vowel, as signaled by the appearance of periodic F1 and F2 . Likewise, the end of each token's vowel was determined by the cessation of periodic F1 and F2 structure. Example segmentations for an aspirated and unaspirated token are shown in Figure 1. F1 and F2 measures at the acoustic midpoint of the resulting vowel interval (e.g., the midpoint of both tokens' [ʉ] in Figure 1) were automatically extracted in Praat (Boersma & Weenink 2021) using a custom script (Lennes 2003).

Formant measures were modeled using Bayesian mixed effects regressions in the BRMS package in R (Bürkner 2017) predicting F1 and F2 as a function

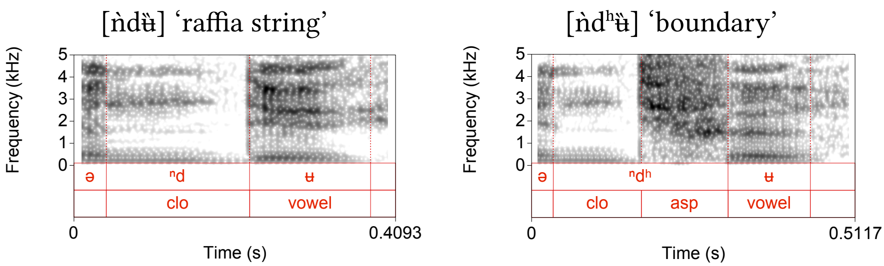

Figure 1: Waveform and spectrogram of prevoiced unaspirated (left) and prevoiced aspirated (right) items showing segmentation of onset and following vowel events for acoustic and ultrasound analysis.

of vowel type, preceding aspiration, voicing, and their interaction, with random intercepts for speaker, and for vowel. A separate regression analysis was carried out for each formant. We employed weak uninformative priors for both the model intercept and fixed effects, with a normal distribution. Intercept priors were based on the mean and standard deviation of F1 and F2, while fixed effect priors were set as Normal(0,100), that is, with no prior expectation of an effect on either F1 or F2, and an SD of 100, i.e., the expectation that effects should be fairly small. Fixed effects were deviation coded (voiced mapped to 0.5, voiceless mapped to −0.5; aspirated mapped to 0.5, unaspirated mapped to −0.5).

2.2 Ultrasound study

2.2.1 Participants and materials

One of the participants in acoustic data collection (1M, participant S1) also provided ultrasound data. Midsagittal ultrasound tongue imaging was recorded for this speaker at a frame rate of 83 Hz using a Telemed Micro ultrasound device. A Telemed convex MC-4 probe was used, held in place submentally by an Articulate Instruments UltraFit stabilization headset (Spreafico et al. 2018). An example of the raw data is shown in Figure 2. Stimuli contained the labial and coronal stops analyzed in the acoustic study, namely [p(^{h}) b(^{h}) t(^{h}) d(^{h})], followed by the same high vowels /i/, /u/, and /ʉ/, and were read in the frame sentence indicated above in (1). A total of 120 tokens from this recording are analyzed here, with 5 tokens per stimulus type. The full list of ultrasound stimuli is provided in the Appendix.

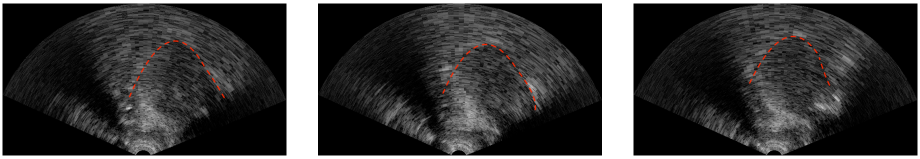

Figure 2: Raw midsagittal ultrasound frames from the acoustic midpoint of a token of each vowel. Approximate tongue surface contours are highlighted with red dashed lines. Right is anterior.

2.2.2 Analysis

Tongue surface contours for each token, extending roughly from tongue blade to tongue root, were extracted using EdgeTrak software (Li et al. 2005) using default settings. While the entire time series of contours was extracted, contours occurring closest to the acoustic midpoint of the vowel were chosen for analysis. The selected contours were submitted to smoothing-spline ANOVA (SSANOVA) to model the typical tongue position for groups of stimulus items (Davidson 2006) using the gss package in R (Gu 2014). The Cartesian coordinates of the extracted tongue surface contours were converted to polar coordinates for analysis and converted back to Cartesian coordinates for display in order to reduce distortion of modeled tongue positions in the area of the tongue root (Mielke 2015).

An initial SSANOVA was first carried out to confirm tongue position differences associated with each of the vowels /i ʉ u/. Two further SSANOVA comparisons were carried out within the data collected for each vowel; a within-vowel design is used here since the size of any prevoicing- or aspiration-related effect on lingual articulation is expected to be much smaller than the difference in tongue position among vowels. The first analysis compares all prevoiced and all voiceless tokens (pooling across aspiration), and the second compares all aspirated and unaspirated tokens (pooling across prevoicing). SSANOVA splines are generated with 95% confidence intervals; confidence intervals for two splines which fail to overlap indicate a statistically significant difference in tongue position at that point for $\alpha = 0.05$. Overlap or non-overlap of splines for prevoiced versus voiceless and aspirated versus unaspirated subsets of the data thus indicate whether tongue position differences beyond those expected by chance are associated with prevoicing or aspiration.

Table 3: Model summary for fixed effects modeling the effect of voicing and aspiration on F1 (left) and F2 (right). Credible estimates are marked with an asterisk (*).

	F1 model			F2 model		
	β	error	95%CrI	β	error	95%CrI
Intercept	347.2	45.3	[256.2, 438.1] *	1464.3	291.4	[889.0, 2031.4] *
Voicing	25.5	9.2	[7.8, 43.6] *	66.6	20.5	[26.12, 106.7] *
Aspiration	-3.1	8.7	[-19.9, 13.8]	-61.3	11.4	[-101.9, -20.5] *
Voi:Asp	7.3	16.9	[-25.6, 40.4]	-60.1	37.1	[-133.0, 11.4]

2.3 Results

2.3.1 Acoustic results

The Bayesian analysis used for the formant data allows us to estimate the size of an effect and our certainty about that estimate, which we present here as 95% credible intervals (CrI). In a Bayesian model, each estimate is drawn from a distribution, and CrIs represent the interval of that distribution in which a certain proportion of the estimates fall (most often 95%). To assess whether a certain factor has a reliable or *credible* impact on formants, we examine whether the 95% CrI includes or excludes zero. Inclusion of zero in the CrI would indicate substantial variation in the directionality of the estimate of the effect, leading to uncertainty. For example, we would be unable to say with much certainty that voicing lowers F1 if the CrI for the effect of voicing on F1 included zero. On the other hand, a CrI which excludes zero gives us clear evidence for the directionality of an effect. Using the same example, a CrI that is entirely negative (excluding zero) suggests that the presence of prevoicing in the preceding stop lowers F1. Thus, when we state an effect is credible we mean that it has a consistently estimated directionality, making us confident that it is robust. This can be considered similar to, though conceptually different from, assessing an effect as *significant* in frequentist models.

First, in Figure 3, we visualize F1 and F2 for all tokens in the analysis, split by vowel and by speaker. As can be seen, each speaker differentiates the vowel categories in F1/F2 space in the expected way. The vowels exhibit a relatively low F1 as expected for their height, though S1 and S2 show more F1 variation than S3 and S4. F2 frequency varies as expected with the frontness of the three vowels.

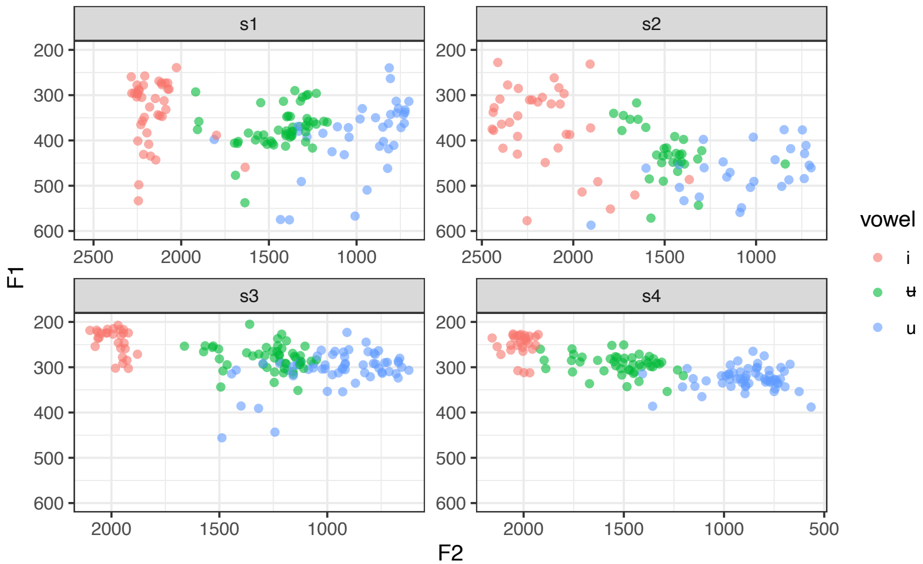

Figure 3: Scatterplot of formant measurements by speaker and vowel.

With this, we turn to the effect of voicing and aspiration on F1 and F2, as shown in Figure 4. Figure 4 plots both F1 and F2, with each measure split by aspiration and voicing. Table 3 additionally contains the output of the fixed effects for the F1 and F2 model, summarizing the model estimate, and lower and upper 95% CrI. Looking first at the effect of voicing on both F1 and F2, we see that voicing credibly raised both F1 ($\beta = 25.5$, $95\%CrI = [9.2, 43.6]$) and F2 ($\beta = 66.6$, $95\%CrI = [26.1, 106.7]$). These two effects are consistent with a lowered tongue body and tongue root advancement, respectively, giving us acoustic evidence for cavity expansion associated with voicing. Aspiration, in contrast, had no credible effect on F1, but did credibly lower F2 ($\beta = -61.3$, $95\%CrI = [-101.9, -20.5]$), consistent with tongue root retraction. The model did not find any credible evidence of an interaction of voicing with aspiration for either formant. Together, these results show acoustic effects consistent with known effects in voiced stops, while also showing an influence of aspiration on F2. We take these results to suggest independent effects of voicing and aspiration on tongue position: root advancement for voicing and root retraction for aspiration.

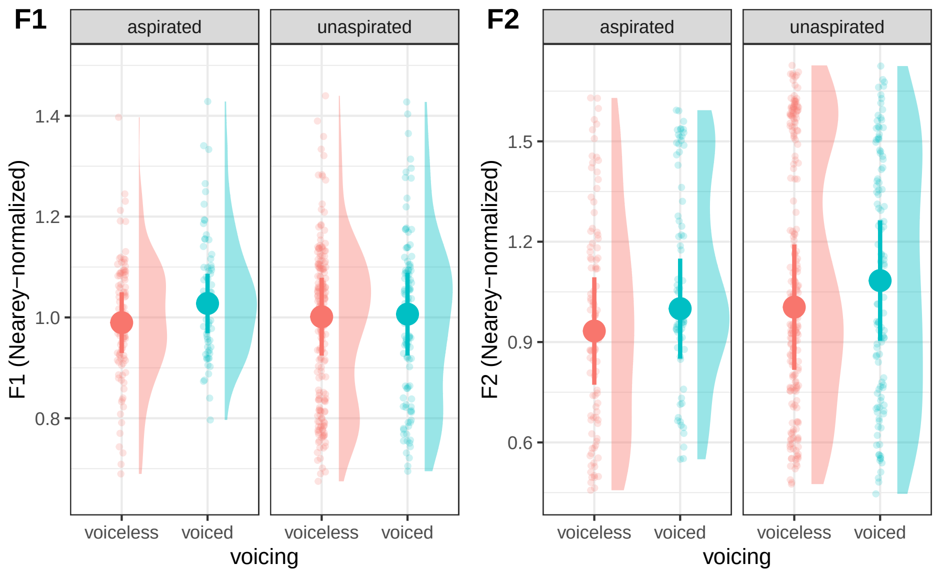

Figure 4: Nearey-normalized vowel midpoint F1 (left) and F2 (right) frequencies by voicing and aspiration of onset consonant.

2.4 Ultrasound results

Next, we turn to the ultrasound data to determine whether the acoustic differences among vowels following different stop types in the full four-speaker set are reflected in the lingual articulation of one speaker. The results of the by-vowel SSANOVA are shown in Figure 5. As expected, there are large differences among the three vowels in the position of the tongue root and dorsum relating to the vowels' backness. As described in §2.2.2, the SSANOVA analyses which follow are within-vowel owing to the larger expected size of this effect compared to the effect of aspiration or prevoicing on tongue position.

SSANOVAs for aspiration within vowel are shown in Figure 6. Recall that in the acoustic results, aspiration has a credible F2-lowering effect on the following vowel. Data for all three vowels /u/, /ʉ/, and /i/ shows some tongue body lowering following aspirated stops. Data for /i/ also has clear tongue root retraction following aspirated items, while /u/ and /ʉ/ do not obviously show this effect. In fact, the superior part of the tongue root for /u/ is slightly advanced following aspirated items. These results are partly consistent with the acoustic study: backing of the tongue is expected from the acoustics, but not tongue dorsum lowering.

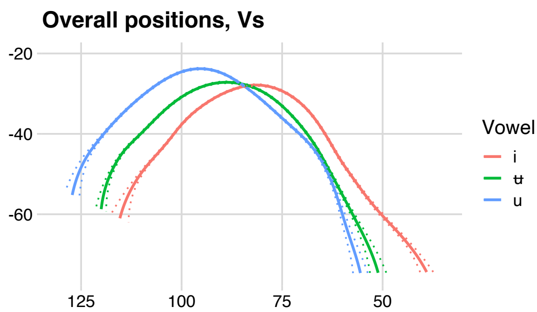

Figure 5: SSANOVA by vowel, pooling across all preceding laryngeal states. Right is anterior.

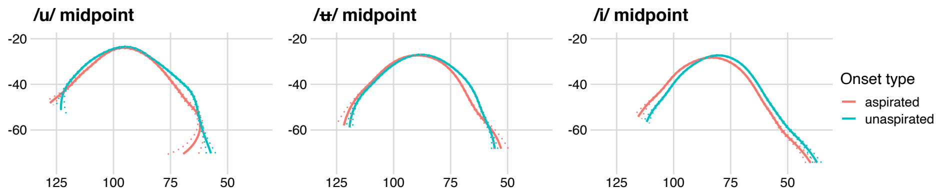

Figure 6: SSANOVA for aspirated versus unaspirated data by vowel. Right is anterior.

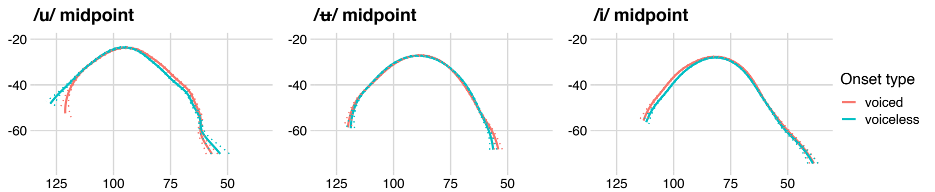

Figure 7: SSANOVA for prevoiced versus voiceless data by vowel. Right is anterior.

SSANOVAs for prevoicing within vowel are shown in Figure 7. Recall that the acoustic results suggest that voicing slightly raises both F1 and F2, leading us to expect a fronter, lower tongue position which could be interpreted as a consequence of cavity expansion. However, cavity expansion is not obviously reflected in the ultrasound data. Only /u/ shows limited voicing-associated tongue root advancement in the most posterior portions of the extracted contours; /i/ unexpectedly shows slight root retraction, counter to predictions. /u/ also shows some degree of tongue body raising. No other significant differences are observed based on stop prevoicing in Figure 7. To summarize the findings from the two studies, while there is a consistent dorsum-lowering (and, for /i/, root-retracting) effect associated with preceding aspiration for the speaker analyzed here, there is no single obvious effect of voicing on lingual articulation.

3 Discussion

The present study examined acoustic and articulatory data from four speakers of Yemba to determine the effects of prevoicing and (consistently voiceless) aspiration on tongue position during the production of the following vowel. The acoustic study with the full set of four speakers provided evidence that prevoicing is associated with an increase in F1 and F2 in the following vowel, suggesting tongue body lowering and advancement, with the latter possibly connected to tongue root advancement. The acoustic study also showed that aspiration is associated with lowered F2 in the following vowel, suggesting tongue body retraction which could relate to root retraction. This generally appeared to support the predictions that voiced stops are associated with an increase in oral cavity volume and aspirated stops are associated with a decrease in oral cavity volume.

The results of the ultrasound study, with data from a single speaker, partially contradicted the findings from the acoustic study. Midsagittal ultrasound frames at the midpoint of target vowels showed that aspiration is associated with tongue root retraction and tongue dorsum lowering during the following vowel, in line with the observed acoustic differences and suggesting a role for the tongue root in producing this difference. Aspiration may coincide with tongue root retraction because a reduced cavity volume facilitates the buildup of pressure in the oral cavity, increasing the reliability and intensity of aspiration at stop release, as suggested directly by Ahn (2018). This is analogous to Proctor et al. (2010)'s finding that voiceless fricatives involve active constriction of the pharyngeal cavity. Further research on the effects of aspiration on tongue position independent of phonetic voicing contrasts is needed to establish if this might be the case.

Prevoicing was not found to have any consistent effect on the lingual articulation of the following vowel, contradicting the acoustic results where small but clear effects on F1 and F2 frequencies were observed, and at odds with previous findings, e.g. (Ahn 2015; 2018; Westbury 1983). The ultrasound study involved a single participant, so this finding should be treated with caution: further study of more speakers is needed to establish if similar changes in tongue position are generally present in Yemba beyond the study's limited population. This is particularly true since there is known to be considerable idiosyncrasy in the extent to which individual speakers modify the volume of the supralaryngeal cavity during the production of voiced obstruents, and which muscular mechanisms they use (Bell-Berti 1975; Proctor et al. 2010; Westbury 1983). The particular speaker investigated in the ultrasound study may undergo cavity expansion but implement it mainly in the lateral portions of the pharyngeal cavity, away from the portion of the vocal tract accessible to midsagittal ultrasound imaging. We note, however, that prevoicing has been found not to affect tongue root position in some other languages, such as in Matsui & Kochetov (2018)'s study of Russian voiced stops, where contrastive secondary palatalization is argued to take priority over reinforcement of voicing in dictating the movements of the tongue root. So, there remains the possibility that Yemba lacks prevoicing-related cavity expansion altogether for language-specific structural reasons.

One possible structural explanation for the apparent lack of prevoicing-related cavity expansion in the stops examined here is that all prevoiced stops in Yemba are also immediately preceded by a nasal consonant (see §1.2). Since nasal venting also facilitates the production of voicing during stop closure (Rothenberg 1968; Solé 2018), it may not be necessary for Yemba speakers to facilitate the production of prevoicing by adjusting the position of the tongue during and after stop articulation. As alluded to in §1.2, Yemba's aspiration contrast extends to the voiced fricative pairs /v/-/vʰ/, /z/-/zʰ/, and /ʒ/-/ʒʰ/ and voiced continuant pairs /l/-/lʰ/, /ɰ/-/ɰʰ/, and /w/-/wʰ/; all of these pairs of phones are voiced but not obligatorily preceded by a nasal, and all three fricatives contrast with voiceless counterparts. Future studies of the voicing contrast in this set of continuant phones may shed light on what lingual adjustments might occur in the absence of a preceding nasal.

4 Conclusion

In this study we aimed to examine how prevoicing and aspiration in stops impact the lingual articulation of following vowels in Yemba, a Grassfields Bantu language with a typologically unusual orthogonal distinction in voicing and aspiration. In an acoustic analysis, formant frequency data suggested tongue root advancement and tongue body lowering following voiced stops, consistent with expansion of the supraglottal cavity. Conversely, formant frequency data indicated that aspiration has the effect of tongue root retraction, consistent with contraction of the supraglottal cavity. In an ultrasound analysis of one participant's data, we found that while there is a consistent dorsum-lowering effect in aspirated stops, and some signs of root retraction, there is no clear effect of voicing on lingual articulation.

Importantly, because Yemba stops (and Bamileke stops more generally) vary independently in voicing and aspiration, we can see that aspiration has a distinct connection to cavity contraction, though evidence for an independent voicing effect is less clear. Taken as a whole, our results suggest that even for "true-voicing" languages with phonetically prevoiced stops, aspiration can coincide with a modification of tongue position. offering an extension of past work in which aspiration and prevoicing do not pattern independently. The present study thus contributes to our cross-linguistic understanding of the effects of laryngeal contrast on supralaryngeal articulation, and represents the first extension of this method to the Bamileke languages, a potentially fruitful area for research on phonation contrasts.

Abbreviations

CL#	noun class # agreement
CrI	credible interval
F*n*	*n*th formant frequency
β	model estimate
!	downstep
´	high tone
`	low level tone
̏	low falling tone

Acknowledgements

We thank Henry Tehrani of the UCLA Phonetics Lab for technical help, as well as Harold Torrence and all the other members of the UCLA field methods course on Yemba (2018–2019), where the preliminary observations that inspired this paper were made. Thanks also go to Steven Bird for providing a copy of Bird (2003) gratis to the research group.

Appendix

Stimuli for lab-recorded speakers are shown in Table 4 in both the ad hoc orthographical versions used for display and corresponding IPA values. All stimuli were used for the acoustic study; if also used in the ultrasound study, this is indicated in the rightmost column.

References

Ahn, Suzy. 2015. Tongue root contributions to voicing in utterance-initial stops in American English. *Proceedings of Meetings on Acoustics* 25(1). 060008. DOI: 10.1121/2.0000428.

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71. 451–467. DOI: 10.1016/j.wocn.2018.10.003.

Anderson, Stephen. 1982. From semivowels to aspiration to long consonants in Ngyembɔɔn-Bamileke. *Journal of West African Languages* 12(2). 58–68. https://main.journalofwestafricanlanguages.org/index.php/downloads/summary/34-volume1202/145-from-semivowels-to-aspiration-to-long-consonants-in-ngyemboon-bamileke.

Atal, Bishnu S., Jih Jie Chang, Max V. Mathews & John W. Tukey. 1978. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *The Journal of the Acoustical Society of America* 63(5). 1535–1555. DOI: 10.1121/1.381848.

Beckman, Jill, Michael Jessen & Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics* 49(2). 259–284. DOI: 10.1017/S0022226712000424.

Bell-Berti, Fredericka. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *The Journal of the Acoustical Society of America* 57(2). 456–461. DOI: 10.1121/1.380468.

Benguerel, André-Pierre, Hajime Hirose, Masayuki Sawashima & Tatsujiro Ushijima. 1978. Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica et Logopaedica* 30(3). 175–198. DOI: 10.1159/000264124.

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

Bird, Steven. 1999. Dschang syllable structure. In H. van der Hulst & N. Ritter (eds.), *The syllable: Views and facts*, 447–476. Berlin: Walter de Gruyter.

Bird, Steven. 2003. *Grassfields Bantu Fieldwork: Dschang Lexicon*. Philadelphia: Linguistic Data Consortium.

Boersma, Paul & David Weenink. 2021. *Praat, a system for doing phonetics by computer [computer program]*. Version 6.1. https://praat.org.

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

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120(1). 407–415. DOI: 10.1121/1.2205133.

Deterding, David & Francis Nolan. 2007. Aspiration and voicing of Chinese and English plosives. In Juergen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 385–388. http://www.icphs2007.de/conference/Papers/1049/1049.pdf.

Dixit, R. Prakash. 1990. Linguotectal contact patterns in the dental and retroflex stops of Hindi. *Journal of Phonetics* 18(2). 189–201. DOI: 10.1016/S0095-4470(19)30401-2.

Dmitrieva, Olga & Indranil Dutta. 2020. Acoustic correlates of the four-way laryngeal contrast in marathi. *Phonetica* 77(3). 209–237. DOI: 10.1159/000501673.

Eberhard, David, Gary Simons & Charles Fennig. 2021. *Ethnologue: languages of the world*. Dallas: SIL International.

Faytak, Matthew & Jeremy Steffman. 2021. Mixed voice in Yemba voiced aspirates. Paper presented at the 95th Annual Meeting of the Linguistic Society of America.

Fletcher, Samuel G. 1989. Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech, Language, and Hearing Research* 32(4). 736–748. DOI: 10.1044/jshr.3204.736.

Fujimoto, Masako, Shigeko Shinohara & Daichi Mochihashi. 2021. Articulation of geminate obstruents in the Ikema dialect of Miyako Ryukyuan: A real-time MRI analysis. *Journal of the International Phonetic Association*. 1–25. DOI: 10.1017/S0025100321000013.

Gu, Chong. 2014. Smoothing spline ANOVA models: R package gss. *Journal of Statistical Software* 58(5). 1–25. http://hdl.handle.net/10.18637/jss.v058.i05.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2021. *Glottolog 4.4*. https://glottolog.org.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language*. 37–58. DOI: 10.2307/412518.

Hyman, Larry M. 1972. *A phonological study of Fe'fe'-Bamileke*. UCLA. (Doctoral dissertation). https://www.proquest.com/docview/302676212.

Keating, Patricia. 1984. Phonetic and phonological representation of stop consonant voicing. *Language*. 286–319. DOI: 10.2307/413642.

Kent, Raymond D. & Kenneth L. Moll. 1969. Vocal-tract characteristics of the stop cognates. *The Journal of the Acoustical Society of America* 46(6B). 1549–1555. DOI: 10.1121/1.1911902.

Kirby, James P. & D. Robert Ladd. 2016. Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America* 140(4). 2400–2411. DOI: 10.1121/1.4962445.

Kochetov, Alexei & Yoonjung Kang. 2017. Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62(1). 18–55. DOI: 10.1017/cnj.2016.39.

Krishnaswamy, Meghavarshini, Indranil Dutta & Ushasi Banerjee. 2018. Active cavity expansion through lingual adjustments to place of constriction in voiced geminates. In *Proceedings of Meetings on Acoustics*, vol. 33, 060002. DOI: 10.1121/2.0001024.

Lennes, M. 2003. *Formant logger [Praat script]*. http://phonetics.linguistics.ucla.edu/facilities/acoustic/collect_formant_data_from_files.txt.

Li, Min, Chandra Kambhamettu & Maureen Stone. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics* 19(6-7). 545–554. DOI: 10.1080/02699200500113616.

Matsui, Mayuki & Alexei Kochetov. 2018. Tongue root positioning for voicing vs. contrastive palatalization: An ultrasound study of Russian word-initial coronal stops. *Journal of the Phonetic Society of Japan* 22(2). 81–94. DOI: 10.24467/onseikenkyu.22.2_81.

Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America* 137(5). 2858–2869. DOI: 10.1121/1.4919346.

Ngouagna, Jean Pierre. 1988. *Esquisse phonologique du ngomba*. Université de Yaoundé. (MA thesis).

Ngueyep, Justin. 1988. *Essai de description phonologique du bamena*. Université de Yaoundé. (MA thesis).

Nissim, Gabriel. 1981. *Le Bamileke-Ghomala': Phonologie, morphologie nominale, comparaison avec des parlers voisins*. Paris: SELAF.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The Production of Speech*, 189–216. New York: Springer-Verlag.

Ohala, John J. 2011. Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th international congress of phonetic sciences*, 64–67. Hong Kong. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session1/Ohala/Ohala.pdf.

Ohala, John J. & Carol J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. *The Journal of the Acoustical Society of America* 65(S1). S23–S23. DOI: 10.1121/1.2017164.

Percival, Maida, Alexei Kochetov & Yoonjung Kang. 2018. An Ultrasound Study of Gemination in Coronal Stops in Eastern Oromo. In *Proceedings of interspeech 2018*, 1531–1535. Hyderabad. https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/2512.pdf.

Proctor, Michael I., Christine H. Shadle & Khalil Iskarous. 2010. Pharyngeal articulation in the production of voiced and voiceless fricatives. *The Journal of the Acoustical Society of America* 127(3). 1507–1518. DOI: 10.1121/1.3299199.

Rothenberg, Martin. 1968. *The breath-stream dynamics of simple-released plosive production*. New York: Karger.

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. DOI: 10.1016/j.wocn.2020.100979.

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. DOI: 10.1016/j.wocn.2018.12.007.

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

Spreafico, Lorenzo, Michael Pucher & Anna Matosova. 2018. Ultrafit: a speaker-friendly headset for ultrasound recordings in speech science. In *Proceedings of interspeech 2018*, 1517–1520. Hyderabad. DOI: 10.21437/Interspeech.2018-995.

Sprouse, Ronald L., Maria-Josep Solé & John J. Ohala. 2008. Oral cavity enlargement in retroflex stops. *UC Berkeley PhonLab Annual Report* 4(4). https://escholarship.org/uc/item/92v0m4mj.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17(1-2). 3–45. DOI: 10.1016/S0095-4470(19)31520-7.

Stevens, Kenneth N. & Samuel J. Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19. DOI: 10.1016/j.wocn.2008.10.004.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America* 73(4). 1322–1336. DOI: 10.1121/1.389236.

Table 4: Stimuli for lab recording.

Orthography	IPA	Gloss	Ultrasound?
le pie	lə̀pȉ	liver	yes
le peuh	lə̀ꜝpʉ́	breast	yes
le pi-he	lə̀pʰì	kola nut	yes
le peu-he	lə̀ꜝpʰʉ́	hole	yes
mbie	m̀bȉ	knife	yes
mbeuh	m̀ꜝbʉ́	breasts	yes
mbouh	m̀ꜝbú	hands	yes
mbie-he	m̀bʰ̏ì	before	yes
mbeu-he	m̀ꜝbʰʉ́	dog	yes
tie	tíí	write! (imp.)	yes
teuh-a	tʉ̀ꜝə́	be strong! (imp.)	yes
ma-touh	mə̀ꜝtú	intestines	yes
ma-ti-he	mə̀tʰí	saliva	yes
a teu-he	àtʰʉ́	tree	yes
a tou-he	àꜝtʰú	head	yes
ntie	ńtí	write (inf.)	
nteuh	ǹꜝtʉ́	heart	
ntouh	ńꜝtú	scoop water (inf.)	
nti-he	ntʰi	host (inf.)	
nteu-he	ǹtʰʉ́	compensation	
ntou-he	ńtʰú	insult (inf.)	
ndie	ǹdȉ	lord	yes
ndeuh	ǹdʉ̏	raffia string	yes
ndouh	ǹdȕ	river	yes
ndi-he	ǹdʰ̏ì	descendant	yes
ndeu-he	ǹdʰʉ̏	boundary	yes
ndou-he	ǹdʰȕ	distant relative	yes
keuh	kʉ̀ꜝʉ́	run! (imp.)	
kouh	kùꜝú	snore! (imp.)	
le keu-he	lə̀kʰʉ̏	trap! (imp.)	
kou-he	àkʰȕ	leg	
nkeuh	ńꜝkʉ́	run (inf.)	
nkouh	ńꜝkú	snore (inf.)	
nkeu-he	ǹkʰʉ̀	rope	
nkou-he	ńꜝkʰú	purge	