ISSN 2436-9179

Crossroads for Phonetic Typology

Edited by
Seunghun J. Lee, Daisuke Shinagawa & Keita Kurabe

JOURNAL OF ASIAN AND AFRICAN STUDIES
SUPPLEMENT

アジア・アフリカ言語文化研究　別冊

No.04

東京外国語大学アジア・アフリカ言語文化研究所
Research Institute for Languages and Cultures of Asia and Africa (ILCAA)

Crossroads for Phonetic Typology

Edited by
Seunghun J. Lee, Daisuke Shinagawa & Keita Kurabe

JOURNAL OF ASIAN AND AFRICAN STUDIES
SUPPLEMENT

アジア・アフリカ言語文化研究　別冊

No.04

東京外国語大学アジア・アフリカ言語文化研究所
Research Institute for Languages and Cultures of Asia and Africa (ILCAA)

Published by the RESEARCH INSTITUTE
for LANGUAGES and CULTURES of ASIA and AFRICA
TOKYO UNIVERSITY of FOREIGN STUDIES
3-11-1 Asahi-cho, Fuchu-shi, Tokyo 183-8534 JAPAN
URL: http://www.aa.tufs.ac.jp
Tel: 81-42-330-5600

◉ Editors: Lee, Seunghun J.
Shinagawa, Daisuke
Kurabe, Keita

◉ Contributors: Abe, Yuko
Aoi, Hayato
Guillemot, Céleste
Kamano, Shigeto
Kurabe, Keita
Lalhminghlui, Wendy
Lee, Seunghun J.
Miyazaki, Kumiko
Patin, Cédric
Sarmah, Priyankoo
Shinagawa, Daisuke
Suzuki, Michinori
Terhiija, Viyazonuo
Uchihara, Hiroto
Ueta, Naoki
Yamamoto, Kyosuke

Printed by Nakanishi Printing Co., Ltd in Japan on 31 July, 2024

Photo on the front cover page:
Photo by Jaakko Kemppainen on Unsplash
https://unsplash.com/photos/green-ferns-k_AXt3VPSrw?utm_content=creditShareLink&utm_medium=referral&utm_source=unsplash

CONTENTS

Journal of Asian and African Studies, Supplement, No. 4, 2024

Introduction:

The Phonetic Typology (PhonTyp) Project

LEE, Seunghun J.
SHINAGAWA, Daisuke
KURABE, Keita

1. Background

This edited volume contains a collection of papers that explore various aspects of phonetic typology. Typological studies on morphosyntactic structures (Chelliah & De Reuse 2010; Croft 2022) or phonological systems (Maddieson 2010; Gordon 2016; Moran & Easterday & Grossman 2023) are relatively common, but phonetic typology remains a field that needs further exploration. From April 2021, a group of researchers specializing in the phonetics or phonology of diverse languages has been meeting regularly over the past three years to discuss and explore issues related to phonetic typology. These meetings are part of the ILCAA joint research project "Phonetic Typology from Cross-Linguistic Perspectives (PhonTyp)."

Exploration of phonetic typology is challenging because it requires the availability of quality recordings that are recorded in a manner that enables cross-linguistic comparison. Two studies examined data from the Illustrations of the International Phonetic Alphabet (IPA), which is published as part of the *Journal of the International Phonetic Association*. Baird *et al.* (2022) address the question regarding the amount of phonetic materials to represent the sound system of a language. Using recordings of the story "North Wind and the Sun (NWS)" from Illustrations of 156 languages, they report that most versions of the NWS story do not include all the allophones present in a language. Perhaps, unsurprisingly, the rarer a phoneme is, the less frequently it was observed in the Illustrations, suggesting a different type of text or a longer text is needed to capture a full phonetic typology. Whalen *et al.* (2022) assess the degree of phonetic documentation from three sources: Illustrations, articles in *Journal of Phonetics (JPhon)*, and papers from Ladefoged & Maddieson (1996) Sounds of the World's Languages (SOWL) documentation project. Based on 23 categories for phonetic documentation (10 for consonants, 7 for vowels, 6 for suprasegmentals), they

DOI: https://doi.org/10.15026/0002000623

found that Illustrations and JPhon articles report measurements pertaining to 12% of the categories, while the SOWL studies covered on average about 41% of the categories. These results show that phonetic documentation still requires much work, and so does the study of phonetic typology that needs such documentation data. Creating resources for conducting phonetic typology work is challenging, as it often requires computation skills. Corpora that are built to investigate phonetic typology are being developed; for example, an openly available corpus such as VoxClamantis v1.0 (Salesky *et al.* 2020), which includes data from 635 languages.

Phonetic typology studies have been conducted in two different ways. First, some studies investigate phonetic aspects of sounds in a language or a language family: preaspiration (Clayton n.d.) and rhotics (Nance & Kirkham 2022) in Scottish Gaelic, glottalized sonorants in Gitksan (Um 1999), and post-velar consonants in Interior Salish languages (Bessell 1992). Second, fewer studies focus on single classes of sounds and explore phonetic aspects across multiple languages: voice onset time in 18 languages (Cho & Ladefoged 1999), the relationship between voicing and fricatives in 4 languages (Bjorndahl 2022), and nasalized fricatives in 3 languages (Shosted 2006). Uncommon or rare sounds tend to be understudied in phonetic research, creating a bias towards common sounds. Addressing this concern calls for the study of the phonetics of under-documented languages (cf. McDonough & Whalen 2008; Maddieson 2016; Gordon 2017, all of them about native American languages).

In this context, the current volume presents twelve papers that predominantly explore uncommon sounds of under-documented languages, shedding light on previously overlooked aspects of phonetic research. In Section 2, the 12 papers will be briefly introduced, and in Section 3, future directions of this phonetic typology research will be discussed.

2. Papers in This Volume

The topics of the papers in the volume target three areas: (i) laryngealized plosives and fricatives, (ii) atypical sonorants such as lateral fricatives, voiceless nasals, and interdental approximants, and (iii) prosodic patterns. A wide range of languages from various language families is included such as Mongolic (Mongolian), Sino-Tibetan (Jinghpaw, Burmese, Mizo, Drenjongke (Bhutia), Angami), Bantu (Swahili, Shingazidja, Tshivenda, Northern Sotho, IsiXhosa, IsiZulu, IsiNdebele, Xitsonga), Austronesian (Kagayanen), Iroquoian (Cherokee), and Japonic (Northern Ryukyuan).

In the first four papers, laryngeal characteristics in obstruents are examined. In Khalkha Mongolian, word-medial aspiration in stops and affricates was described as preaspiration. In his paper (Ueta), preaspiration appears after a vowel, but after another obstruent, the aspiration is realized after the burst, suggesting variable realization of aspiration depending on the phonetic context. The preaspiration in Cherokee (Uchihara) is observed exclusively in /s/ and not in other obstruents. Moreover, the preaspiration pattern interacts with phonological alternation such as vowel deletion and laryngeal alternation, which raises an interesting question about whether or how to represent phonetic information in the phonological representation.

The Northern Ryukyuan paper (Aoi) reports how glottalized consonants are realized in the Ie dialect of Okinawan. He reports that glottalized consonants are realized with a shorter duration and a more abrupt onset than non-glottalized consonants. Examining phonetic characteristics in merging sounds reveals interesting patterns. A phonetic study of aspirated fricatives (Kurabe & Lee) reveals that aspiration duration in Jinghpaw and Burmese aspirated fricatives is much shorter compared to Korean. These cross-linguistic results suggest that aspirated fricatives may not be realized with identical aspiration duration.

The next four papers concern sounds that may display both obstruent-like and sonorant-like features: fricatives that are lateral, and nasals that are voiceless. A typological overview of lateral fricatives (Shinagawa & Lee) in Southern Bantu languages reports the distribution of these fricatives based on phonetic data from six Bantu languages (Tshivenda, Northern Sotho, IsiXhosa, IsiZulu, IsiNdebele, Xitsonga); voiceless lateral fricatives show more restrictive distribution than voiced lateral fricatives in the post-nasal environment.

The acoustic characteristics of voiceless nasals in Mizo, Angami, and Drenjongke (Bhutia) are explored in detail. Mizo and Angami are two Tibeto-Burman languages that Bhaskararao & Ladefoged (1991) mention as examples of two types of phonetic realization of voiceless nasals, and the two studies in this volume add new phonetic characteristics of these nasals. The Mizo study (Sarmah & Lalhminghlui) reports acoustic parameters that correlate with different places of articulation of voiceless nasals. In the Angami paper (Terhija & Sarmah), nasality in the vowel following voiceless nasals is shown to be less present than that in the vowel after voiced nasals. In Drenjongke (Bhuthia), voiceless nasals are reported to be an innovative sound pattern (Guillemot & Lee). Individual variation in the realization of voiceless nasals suggests that nasality gesture must precede the laryngeal gesture, but not the other way around.

The interdental approximant in Kagayanen (Yamamoto) is a typologically rare speech sound although it has been reported in several Austronesian languages. Compared to the alveolar lateral approximant, the interdental approximant shows lower F2 and smaller F2-F1 values.

The next two papers concern the prosody of particles in several Bantu languages. In Zanzibar Swahili, *tu* is an exclusive particle meaning ‘only’. This particle is realized with a high pitch, but interestingly the paper (Abe, Lee, Kamano & Miyazaki) shows that the overall sentence prosody is identical whether the particle *tu* is present or not. A comparative study on the prosody of question elements in Shingazidja and Xitsonga (Patin & Lee) shows that the prosody of the question elements varies by the distribution of the question elements in each language; Shingazidja prefers the clause-final position for the question elements, whereas question elements in Xitsonga are realized mostly non-clause-final position.

The last paper (Lee & Suzuki) in this volume reports a phonetic typology of laryngeal contrasts, using data from the Illustrations of the IPA. After creating a corpus of recordings of plosives from 103 languages, the paper presents VOT data in languages with 2-way, 3-way or 4-way laryngeal contrast. Although the corpus is small in size, the results show that VOT is more constrained in the voiceless unaspirated category in languages with a 2-way or a 3-way laryngeal contrast.

3. Future Directions for the PhonTyp Project

The first phase of this phonetic typology project mainly focused on plosives, fricatives, and nasals. The twelve papers in this volume embody the research results and show that we now understand a little bit more about the phonetic aspects of various languages. From April 2024, the second phase of the project will begin to continue to explore further aspects of phonetic typology focusing on various types of liquids and vowels, which are not always fully examined from cross-linguistic perspectives. The research project will not only explore the phonetics of these sounds in individual languages or language families, but also in the wider context of phonetic typology by reporting on the commonalities and differences of the sounds. Last but not least, we thank the anonymous reviewers who shared their valuable time and provided constructive suggestions during the review process.

References

Baird, Louise, Nicholas Evans, and Simon J. Greenhill. 2022. “Blowing in the Wind: Using ‘North Wind and the Sun’ Texts to Sample Phoneme Inventories.” *Journal of the International Phonetic Association*, 52(3): 453–494.

Bessell, Nicola Jane. 1992. Towards a Phonetic and Phonological Typology of Post-velar Articulation. PhD dissertation, The University of British Columbia.

Bhaskararao, Peri and Peter Ladefoged. 1991. “Two Types of Voiceless Nasals.” *Journal of the International Phonetic Association*, 21(2): 80–88.

Bjorndahl, Christina. 2022. “Voicing and Frication at the Phonetics-phonology Interface: An Acoustic Study of Greek, Serbian, Russian, and English.” *Journal of Phonetics*, 92: 101–136.

Chelliah, Shobhana L. and Willem J. De Reuse. 2010. “What to Expect in Morphosyntactic Typology and Terminology.” *Handbook of Descriptive Linguistic Fieldwork* (S. L. Chelliah and W. J. De Reuse, eds.), 279–356, Dordrecht: Springer Netherlands.

Cho, Taehong and Peter Ladefoged. 1999. “Variation and Universals in VOT: Evidence from 18 Languages.” *Journal of Phonetics*, 27(2): 207–229.

Clayton, Ian. n.d. *Phonetics in Phonology: Evidence from Scottish Gaelic Preaspiration*. ms, University of North Carolina, Chapel Hill.

Croft, William. 2022. *Morphosyntax: Constructions of the World’s Languages*. Cambridge: Cambridge University Press.

Gordon, Matthew K. 2016. *Phonological Typology*. Oxford: Oxford University Press.

Gordon, Matthew K. 2017. “Phonetic and Phonological Research on Native American Languages: Past, Present, and Future.” *International Journal of American Linguistics*, 83(1): 79–110.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World’s Languages*. Oxford: Blackwell Publishers.

Maddieson, Ian. 2010. “Typology of Phonological Systems.” *The Oxford Handbook of Linguistic Typology* (J. J. Song, ed.), 534–548, Oxford University Press.

Maddieson, Ian. 2016. “Under-documented Languages Expand Phonetic Typology.” *Journal of the Acoustical Society of America*, 139(4_Supplement): 2212–2212.

McDonough, Joyce and Doug H. Whalen. 2008. “The Phonetics of Native North American Languages.” *Journal of Phonetics*, 36(3): 423–426.

Moran, Steven, Shelece Easterday and Eitan Grossman. 2023. “Current Research in Phonological Typology.” *Linguistic Typology*, 27(2): 223–243.

Nance, Claire and Sam Kirkham. 2022. “Phonetic Typology and Articulatory Constraints: The Realization of Secondary Articulations in Scottish Gaelic Rhotics.” *Language*, 98(3): 419–460.

Salesky, Elizabeth, Eleanor Chodroff, Tiago Pimentel, Matthew Wiesner, Ryan Cotterell, Alan W. Black, and Jason Eisner. 2020. “A Corpus for Large-scale Phonetic Typology (Version 1)”.

Shosted, Ryan K. 2006. The Aeroacoustics of Nasalized Fricatives: An Instrumental Study in Phonetic Typology. PhD dissertation. University of Illinois Urbana-Champaign.

Um, Hye-Young. 1999. “Glottalized Sonorants in Gitksan.” *Studies in Phonetics, Phonology and Morphology*, 5(2): 94–109.

Whalen, Douglas H., Christian DiCanio, and Rikker Dockum. 2022. “Phonetic Documentation in Three Collections: Topics and Evolution.” *Journal of the International Phonetic Association*, 52(1): 95–121.

Journal of Asian and African Studies, Supplement, No. 4, 2024

The Phonetic Realization of Aspiration in Khalkha Mongolian:

The Phonetic Contrast between Aspirated and Unaspirated Consonants Preceded by an Obstruent*

UETA, Naoki

Mongolian has an aspiration contrast in stops and affricates. Some previous studies claim that the aspiration of word-initial aspirated consonants is realized as post-aspiration, which is characterized by a long voice onset time (VOT), whereas that of word-medial and final aspirated consonants is realized as pre-aspiration, which is typologically rare. Pre-aspiration is normally characterized by a clearly audible aspiration noise and at least partial devoicing of the preceding segment. Pre-aspiration is clearly observed when the preceding segment, including a vowel, is voiced. However, it is unclear whether pre-aspiration is explicitly realized when aspirated consonants are preceded by a voiceless obstruent that cannot be devoiced.

This study addresses the phonetic realization of word-medial aspiration in Mongolian, more specifically, the phonetic difference between aspirated /t^h/ and unaspirated /t/ preceded by an obstruent. A production experiment and acoustic analysis were conducted to clarify whether aspiration is realized when an aspirated stop is preceded by an obstruent. Acoustic analyses revealed that the duration of obstruents, intensity during obstruents, and duration of stop closures did not differ depending on whether the following segment was aspirated or unaspirated, suggesting that preaspiration is not realized in obstruents. In contrast, VOT differs by aspiration type: it tends to be longer for aspirated than for unaspirated consonants. This means that aspiration can be realized as post-aspiration even in a word-medial position under conditions where pre-aspiration is difficult to realize, which can be theoretically explained by the concept of "Licensing by Cue."

Keywords: Mongolian, aspiration contrast, pre-aspiration, VOT, acoustic analysis

DOI: https://doi.org/10.15026/0002000624

* I thank the reviewer for giving helpful and constructive comments. My thanks also go to the students and teachers at the Mongolian University of Science and Technology for their cooperation during the experiment. This research was supported by the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023) and JSPS KAKENHI Grant Numbers 21K20015 and 23K12168.

1. Introduction

Many languages have a laryngeal contrast in obstruents; for example, /t/ and /d/ are contrastive in Japanese based on the difference in laryngeal control between voiceless and voiced.

Khalkha Mongolian, widely spoken in Mongolia and officially considered “Standard Mongolian” (henceforth just “Mongolian”), also has a laryngeal contrast in stops and affricates1). This contrast has traditionally been described using the terms “strong–weak,” “fortis–lenis,” or “tense–lax.” Recent acoustic analyses have revealed that this contrast is based on the presence or absence of aspiration (Svantesson *et al.* 2005, Karlsson and Svantesson 2011, 2012, Svantesson and Karlsson 2012, Ueta 2018a, 2020a). Svantesson and Karlsson 2012 demonstrated that the aspiration of word-initial aspirated consonants is realized as post-aspiration, whereas that of word-medial and word-final aspirated consonants is realized as pre-aspiration. They also stated that pre-aspiration is normally characterized by clearly audible aspiration noise and at least partial devoicing of the preceding segment. Preaspiration is clearly observed when the preceding segment, including vowels, is voiced. However, it remains unclear whether pre-aspiration is explicitly realized when aspirated consonants are preceded by voiceless obstruents, which cannot be devoiced.

This study addresses the realization of aspiration in Mongolian aspirated stop /tʰ/. More specifically, this study conducted a production experiment and acoustic analyses to examine whether the aspiration of /tʰ/ is realized, and whether acoustic differences are observed between /tʰ/ and its unaspirated counterpart /t/ when they are preceded by an obstruent. In conclusion, this study claims that pre-aspiration is not realized in obstruents, whereas aspiration can be realized as post-aspiration, that is, as a longer voice onset time (VOT), even in word-medial positions under conditions where pre-aspiration is difficult to realize.

2. Aspiration Contrast in Mongolian

As noted in Section 1, Mongolian has an aspiration contrast in stops and affricates. The phonological systems of stops and affricates in Mongolian are presented in Table 1. The phonemes in parentheses occur only in loanwords and onomatopoeic words.

The realization of aspiration varies by position. Svantesson and Karlsson 2012 claimed the following.

---

1) There is no laryngeal contrast in fricatives in Mongolian.

Table 1 The phonological system of stops and affricates in Mongolian[2]

		stop	affricate
strong	aspirated	(/pʰ/) /tʰ/ (/kʰ/)	/ʦʰ/ /ʧʰ/
weak	voiceless unaspirated	/t/	/ʦ/ /ʧ/
	voiced unaspirated	/b/ /g/	

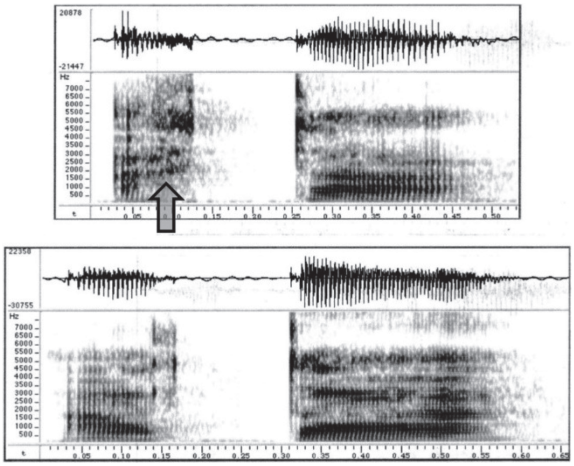


Figure 1 Word-medial /tʰ/ [ʰt] in *atʰaa* 'camel gelding-REFLEXIVE' (top) and /t/ [t] in *ataa* 'demon-REFLEXIVE' (bottom) (Svantesson and Karlsson 2012: 459 Figure 2, arrow added)

(1) In word-initial position, aspiration is realized as postaspiration; if an aspirated consonant is preceded by a voiced segment in the same utterance, however, there is preaspiration as well. In word-medial and final position, aspiration is realized as preaspiration. (Svantesson and Karlsson 2012: 463)

In word-initial position, post-aspiration is characterized by a long VOT, and aspirated and unaspirated consonants can be sufficiently differentiated by VOT: longer VOT for aspirated and shorter VOT for unaspirated consonants (Ueta 2018a). In contrast, in word-medial position, the difference in VOT between aspirated and unaspirated consonants is small and not significant (Svantesson and Karlsson 2012). Aspiration in word-medial position is realized instead as pre-aspiration, which is characterized by clearly audible aspiration noise and at least partial devoicing of the preceding segment (Svantesson *et al.* 2005: 14, Svantesson and Karlsson 2012: 455). Figure 1 illustrates the waveforms and spectrograms of word-medial /tʰ/ [ʰt] and /t/ [t]. The arrow in the figure shows pre-aspiration.

Ueta 2020a examined the acoustic realization of word-medial aspirated obstruents in

---

2) /g/ is regarded as voiced, although it can be realized as voiceless, because it functions phonologically as a voiced consonant from the perspectives of phonotactics and sonority (see Svantesson *et al.* 2005: 65–68). /b/ is also regarded as voiced, although it has characteristics intermediate between voiceless /t/ and voiced /g/ (see Ueta 2020b for details). /kʰ/ is not regarded as a phoneme in Svantesson and Karlsson's studies, but it should be included in the system because it occurs in loanwords and is consistently realized as [kʰ] (Ueta 2020b).

Mongolian and indicated that pre-aspiration appears in various forms according to the phonological environment. The results for Ueta are summarized as follows.

(2) a. When an aspirated obstruent is preceded by a short vowel, preaspiration is explicitly realized as breathy voice and/or devoicing of the vowel. Short vowels are often fully devoiced.
b. When the preceding segment is a long vowel, preaspiration is mainly realized as partial breathy voice on the last part of the vowel. Vowel devoicing rarely occurs and the realization of preaspiration is less salient than on short vowels. In some cases, neither breathiness nor devoicing is observed. This is probably because the [spread glottis] feature conflicts with the phonological heaviness of long vowels.
c. In nasal-aspirated obstruent sequences, preaspiration is in many cases not salient. Instead, aspirated obstruents are characterized as being preceded by shorter nasals and having longer stop closures than their unaspirated counterparts.
(Ueta 2020a: 113 (3))

Previous studies have revealed to a large extent how aspiration in Mongolian is realized in various phonological environments. However, the phonetic realization of aspirated consonants after an obstruent remains unclear. Mongolian stops and affricates can occur after another consonant, as shown in (3)–(6).

(3) a. gar-t^hal 'go out-TERMINATIVE'
b. gar-tag 'go out-HABITUAL'
(4) a. t^hat^h-t^hal 'pull-TERMINATIVE'
b. t^hat^h-tag 'pull-HABITUAL'
(5) a. xat-t^hal 'mow-TERMINATIVE'
b. xat-tag 'mow-HABITUAL'
(6) a. tʊs-t^hal 'drip-TERMINATIVE'
b. tʊs-tag 'drip-HABITUAL'

In (3a), the aspirated stop /t^h/ in the suffix *-t^hal* is positioned after /r/. In this case, the aspiration of /t^h/ is realized as preaspiration in the preceding consonant /r/, resulting in the voiceless liquid [r̥]. In contrast, /r/ followed by an unaspirated consonant, as in (3b), is pronounced as the voiced liquid [r]. The contrast between the aspirated /t^h/ and unaspirated /t/ is realized as the absence or presence of voicing in the preceding consonant /r/.

In (4a), (5a), and (6a), /t^h/ occurs after the obstruents /t^h/, /t/, and /s/, respectively. In these cases, the aspiration of /t^h/ cannot be realized as devoicing of the preceding segments because they are originally voiceless. It is unclear whether the aspiration of /t^h/ is realized as pre-aspiration in the preceding obstruent in some way other than devoicing, realized in another way such as post-aspiration, or is not explicitly realized. This issue is also related to

Table 2 The target words

final consonant	words
/tʰ/	butʰ- ‘consist of’, tʊtʰ- ‘be short of’, tʰatʰ- ‘pull’, ʊtʰ- ‘smoke’, xatʰ- ‘dry up’
/t/	bɔt- ‘think’, but- ‘become dim’, tat- ‘get used to’, ʊt- ‘tarry’, xat- ‘mow’
/ʦʰ/	bʊʦʰ- ‘go back’, gaʦʰ-‘jam’, ʦʰaʦʰ- ‘scatter’
/ʦ/	baʦ- ‘press’, uʦ- ‘see’, xaʦ- ‘bite’
/ʧʰ/	buʧʰ- ‘surround’, ɔʧʰ- ‘go’
/ʧ/	buʧ- ‘dance’, xɔʧ- ‘win’
/s/	nis- ‘fly’, tʰʊs- ‘hit’, tʊs- ‘drip’
/ʃ/	tʰaʃ- ‘slap’, tʰuʃ- ‘support’, xaʃ- ‘enclose’
/x/	nʊx- ‘mash’, tʰɔx- ‘put on’, ux- ‘die’

the contrast between aspirated and unaspirated in word-medial position. If the aspiration is realized as post-aspiration, the realization of aspiration in word-medial position varies by phonological environment, whereas if aspiration is not realized after an obstruent, the contrast between aspirated and unaspirated consonants is neutralized in this environment.

3. Survey

3.1 Production Experiment

A production experiment was conducted to examine the acoustic realization of aspirated and unaspirated consonants after an obstruent.

The target words were 29 verbs whose stems end with an obstruent: /tʰ/, /t/, /ʦʰ/, /ʦ/, /ʧʰ/, /ʧ/, /s/, /ʃ/, or /x/, as shown in Table 2. All the target verbs had mono-syllabic stems with a short vowel, which is pronounced with low pitch followed by high pitch in the next syllable, that is, the first syllable in the following suffix[3]. Each verb was displayed individually on a computer screen in the form of the future participle suffix *-(V)x*[4]; the suffixes *-san*^{4}, *-tag*^{4}, *-tʰal*^{4}, *-ʧʰixsan*^{4}, and *-ʧee* were also presented on the screen[5]. All items were written in Cyrillic orthography.

The production experiment was conducted as follows: One of the verbs and the five suffixes were displayed on a computer. The participants were asked to add each suffix to the verb and read the suffixed form. An example of this experiment is shown in Equation (7), which was conducted for 29 verbs and 1 dummy word. The order of the stimuli was

---

[3] It is known that the degree of aspiration depends on accent (Klatt 1975). Mongolian has a fixed accent system, and it can be generalized that the first two morae are tied with LH pitch, though the entire accent system is not straightforward (see Karlsson 2005 for details). In this experiment, all target words have the same syllable structure and the same accent pattern, and thus the realization of aspiration is not affected by them.

[4] In Mongolian, the verb form with the future participle suffix *-(V)x* is the citation form. Verbs are listed in Mongolian dictionaries in this form.

[5] The superscript ^{4} means that the suffix has four allomorphs according to vowel harmony. For example, *-san*^{4} is realized as *-san*, *-sɔn*, *-sen*, or *-søn*, depending on the preceding vowel. See Svantesson *et al.* 2005: 46–57 for details.

randomized except that the dummy word was arranged to appear first.

(7)							
	Cyrillic orthography	бүтэх:	-сан^{4},	-даг^{4},	-тал^{4},	-чихсан^{4},	-жээ
	transliteration	but^{h}ex:	-san^{4},	-tag^{4},	-t^{h}al^{4},	-ʧ^{h}ixsan^{4},	-ʧee
	read out	*(but^{h}ex)*[6],	*but^{h}sen*,	*but^{h}teg*,	*but^{h}t^{h}el*,	*but^{h}ʧ^{h}ixsen*,	*but^{h}ʧee*[7]

The participants in this experiment were 12 Mongolian native speakers, 6 male and 6 female, ranging in age from 17 to 23 years. All data were recorded using a digital recorder (ZOOM H4n [WAV, 44.1 kHz / 16 bit]) and a head-mounted condenser microphone (AKG C520). Recordings were conducted in a quiet room.

3.2 Acoustic Analysis

The recorded material was analyzed using Praat ver. 6.1.50 (Boersma and Weenink 2021). The forms with *-t^{h}al^{4}* and *-tag^{4}* were the targets of the analysis. Although it is necessary to examine whether pre-aspiration is realized in the obstruent preceding /t^{h}/, it is difficult to express the presence or absence of aspiration directly and numerically because pre-aspiration is normally realized as a breathy voice and/or devoicing of the preceding segment, as shown in (2); however, obstruents are originally voiceless. Therefore, in the present analysis, the following four acoustic elements were measured to estimate whether the aspiration contrast was preserved after an obstruent.

The first element is the duration of the obstruents preceding /t^{h}/ and /t/. Aspiration noise may be quite long compared with the burst noise during the release of stops (Johnson 2012: 174). If the pre-aspiration of word-medial /t^{h}/ is realized in the preceding obstruent, it is possible that the duration of the obstruent followed by /t^{h}/ is longer than that followed by /t/, especially when the obstruent is a stop.

The second element is the mean value of intensity during the turbulent noise of the obstruents preceding /t^{h}/ and /t/. According to Wayland and Jongman 2003, acoustic intensity tends to decrease in breathy phonation. Although they reported this with respect to breathy vowels, we may expect that the intensity would be lower in obstruents where breathiness occurs as the realization of preaspiration. On the other hand, aspirated consonants are generally regarded as “strong.” The strength of consonants may be reflected in their intensity.

Third, the duration of the stop closure of /t^{h}/ and /t/ was measured. As shown in (2c), aspirated obstruents are characterized by longer stop closures than their unaspirated counterparts in nasal-aspirated obstruent sequences (Ueta 2020a). This implies that stop closure duration can affect the distinction between aspirated and unaspirated consonants.

Finally, the VOT of /t^{h}/ and /t/ is an important element. VOT is the time interval

---

[6] The experimenter did not instruct them to read the headword (*but^{h}ex*), but some participants read it.

[7] According to vowel harmony, *-san^{4}*, *-tag^{4}*, *-t^{h}al^{4}*, and *-ʧ^{h}ixsan^{4}* are realized as *-sen*, *-teg*, *-t^{h}el*, and *-ʧhixsen*, respectively, after the vowel *u* in the stem.

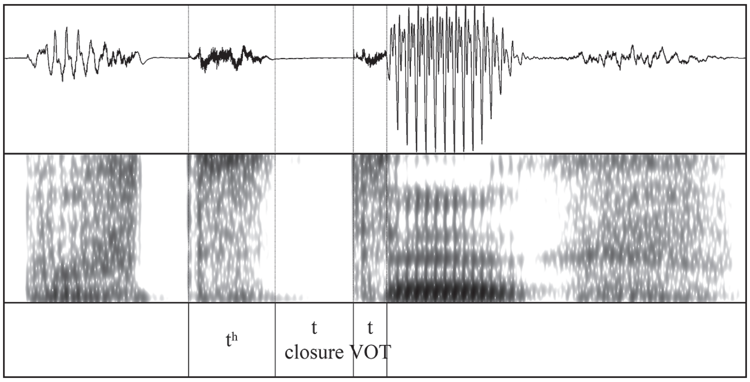


Figure 2 An example of a waveform, spectrogram, and segmentation of *butʰteg*

between the articulatory release of the stop and the onset of vocal fold vibration (Lisker and Abramson 1964), and thus VOT is not a realization of pre-aspiration but post-aspiration. Even if pre-aspiration is not realized during the preceding obstruent, the aspiration of /tʰ/ might be realized as post-aspiration or a longer VOT, as in word-initial position.

To measure these four acoustic elements, segmentation was performed by visually observing the waveforms and spectrograms of each token. The following points were identified:

(8) a. the starting point of the turbulent noise of every obstruent preceding /tʰ/ in *-tʰal*^{4} or /t/ in *-tag*^{4}.
  b. the end point of every obstruent preceding /tʰ/ in *-tʰal*^{4} or /t/ in *-tag*^{4}.
  c. the starting point of /tʰ/ in *-tʰal*^{4} and /t/ in *-tag4*.
  d. the beginning of vocal fold vibrations following /tʰ/ in *-tʰal*^{4} or /t/ in *-tag*^{4}.

The interval between (8a) and (8b) was defined as the duration of the obstruents, and the mean intensity values were measured in this section. Similarly, the interval between (8b) and (8c) is the duration of the stop closures of /tʰ/ and /t/, and VOT is the interval between (8c) and (8d). An example of segmentation is shown in Figure 2.

4. Result

4.1 The Duration of Obstruents

This section addresses the duration of obstruents preceding /tʰ/ and /t/. When the obstruent is a stop (/tʰ/ or /t/), namely, the word has a /tʰ-tʰ/, /t-tʰ/, /tʰ-t/, or /t-t/ sequence, the first stop can be pronounced without a stop release or with only a very short burst noise, such as [tɑt˺tɑg] for /tattag/. In these cases, the duration of the obstruents could not be measured because their endpoints could not be defined. In this experiment, 17 of the 240 tokens were pronounced in this way. A breakdown of the 17 tokens is presented in Table 3. Although the frequency of pronunciations without a stop release was slightly higher when the second stop was /t/ than when it was /tʰ/, the difference was small

Table 3 Stop-stop sequences without a stop release

sequence	$/t^h$-$t^h/$	$/t$-$t^h/$	$/t^h$-$t/$	/t-t/
frequency	4	2	4	7
example	*but^{h}t^{h}el* *t^{h}at^{h}t^{h}al* (twice) *xat^{h}t^{h}al*	*butt^{h}el* (twice)	*but^{h}teg* *tʊt^{h}tag* (twice) *t^{h}at^{h}tag*	*butteg* (4 times) *tattag* (twice) *xattag*

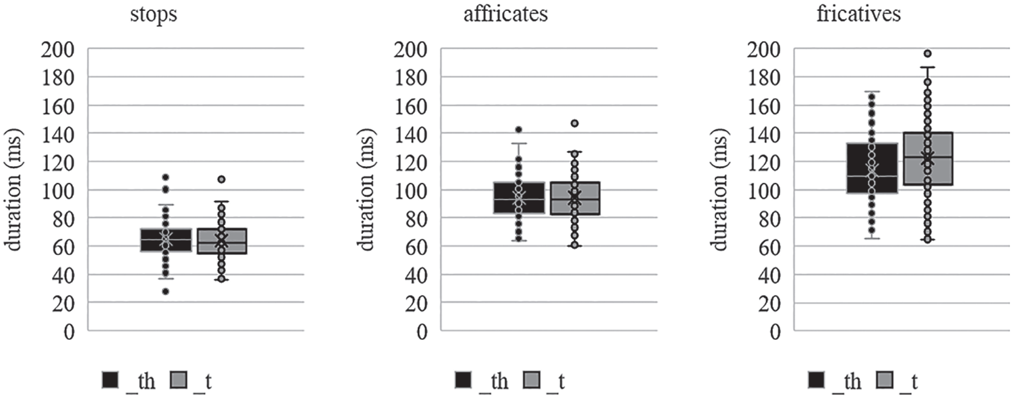


Figure 3 Distributions of stop (left), affricate (center), and fricative (right) durations preceding $/t^h/$ and /t/

and unlikely to be significant.

When the obstruent is an affricate (either $/ʦ^h/$, /ʦ/, $/ʧ^h/$, or /ʧ/) or a fricative (either /s/, /ʃ/, or /x/), the duration of these segments can be consistently measured because frication noise certainly occurs. However, 3 tokens were excluded from the analysis in this experiment because of the obscure pronunciation of the target consonants; thus, the analysis targeted a total of 453 affricates and fricatives.

Figure 3 shows the distribution of the duration of stops, affricates, and fricatives followed by $/t^h/$ and /t/. This figure shows that the duration of the obstruents does not differ depending on whether the following stop is $/t^h/$ or /t/. The mean durations of stops before $/t^h/$ and before /t/ are almost the same, 64.9 ms and 63.8 ms, respectively, and the difference is not statistically significant (paired t-test: $p > .05$). The same holds for affricates: 94.3 ms before $/t^h/$ and 94.2 ms before /t/ (paired t-test: $p > .05$). However, the mean duration of fricatives was slightly different before $/t^h/$ (113.8 ms) and /t/ (122.2 ms), and this difference was statistically significant (paired t-test: $p < .01$). However, it is unlikely that this difference can be attributed to the contrast between $/t^h/$ and /t/.

Duration is influenced by speech rate and thus differs across words and speakers. Next, the intra-speaker durations of the same verbs were compared. For example, the durations of *bu$\underline{t^h}$-t^{h}el* and *bu$\underline{t^h}$-teg* uttered by the same speaker were compared. The duration ratio was calculated by dividing the duration before $/t^h/$ (e.g., 64.3 ms for *bu$\underline{t^h}$-t^{h}el*) by that before /t/ (e.g., 42.7 ms for *bu$\underline{t^h}$-teg*). The result (64.3/42.7 = 1.51) indicates that the duration of *bu$\underline{t^h}$-t^{h}el* is 1.51 times as long as that of *bu$\underline{t^h}$-teg*. Duration ratios were calculated for each pair.

Figure 4 shows a histogram of the duration ratios of obstruents, indicating that the

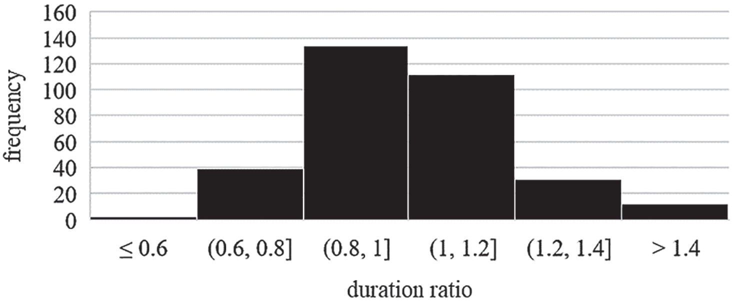
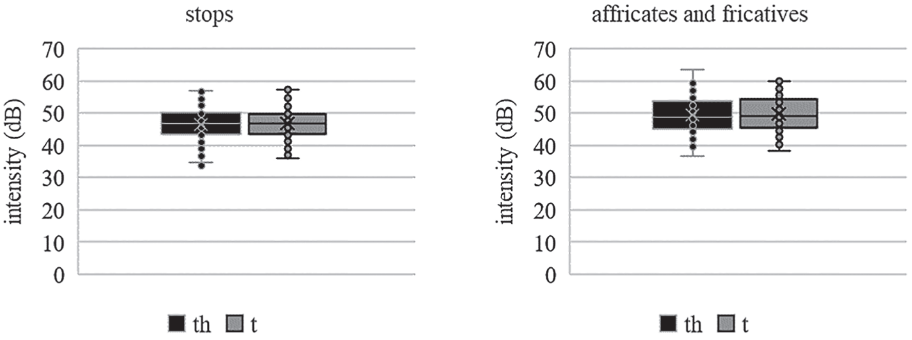


Figure 4 Histogram of duration ratios for obstruents



Figures 5 Distributions of stop (left) and affricate and fricative (right) intensities

duration ratio was approximately 1.00. The mean value of the duration ratio is 1.01, which means, again, that the duration of obstruents before /t/ is almost the same as that before /tʰ/.

4.2 The Mean Value of Intensity

This section reports the mean values of intensity during the obstruents preceding /tʰ/ and /t/. Figure 5 shows the intensity distributions of stops, affricates, and fricatives. Because affricates and fricatives have similar acoustic characteristics, they are summarized on one sheet. This figure demonstrates that the mean intensity of obstruents is almost the same regardless of whether the following stop is /tʰ/ or /t/. The mean intensities of stops before /tʰ/ and /t/ are 46.5 dB and 46.8 dB, respectively, and of affricates and fricatives before /tʰ/ and /t/ are 49.4 dB and 49.8 dB, respectively. Although both these differences were statistically significant (paired *t*-test: $p < .01$), they were extremely small and unlikely to be attributable to the contrast between /tʰ/ and /t/.

Similar to duration, intensity is compared for the same verbs. The intensity ratio is calculated by dividing the intensity before /tʰ/ (e.g., 44.5 dB for *but̲ʰ-tʰel*) by that before /t/ (e.g., 39.0 dB for *but̲ʰ-teg*). The result (44.5/39.0 = 1.14) indicates that the intensity of *but̲ʰ-tʰel* is 1.14 times higher than that of *but̲ʰ-teg*. The intensity ratios were calculated for each pair.

Figure 6 shows a histogram of the intensity ratios. This figure illustrates that the intensity ratio was concentrated at approximately 1.00. The mean value of intensity ratio is 1.00, which means, again, that the intensity during obstruents before /t/ is almost the same as that before /tʰ/.

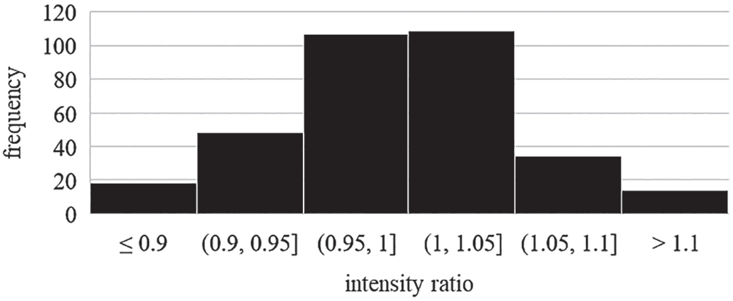


Figure 6 Histogram of intensity ratios

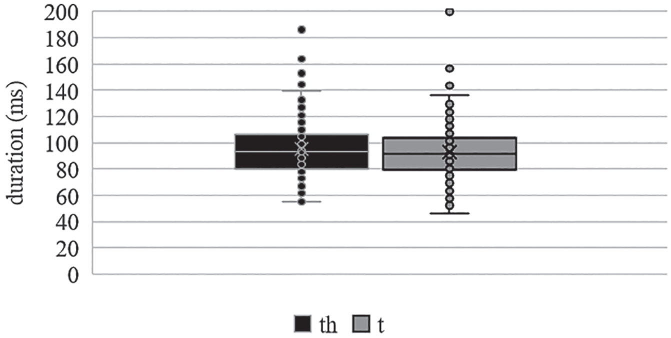


Figure 7 Distribution of stop closure durations

4.3 The Duration of Stop Closure

This section addresses the duration of stop closures of /t^{h}/ and /t/. When the first stop of /t^{h}-t^{h}/, /t-t^{h}/, /t^{h}-t/, or /t-t/ is pronounced without a stop release or with only a very short burst noise—that is, in the 17 tokens presented in Table 3—the duration of the stop closure cannot be measured. In addition, two tokens were excluded from the analysis because of disfluency during stops.

Figure 7 shows the distribution of the duration of stop closures. Again, this figure suggests that the duration of stop closures does not differ between /t^{h}/ and /t/. The mean duration of stop closures of /t^{h}/ and /t/ are almost the same: 95.3 ms and 92.8 ms, respectively. Although the difference was statistically significant (paired *t*-test: $p < .01$), it was extremely small and it is unlikely that the contrast between /t^{h}/ and /t/ is reflected in the duration of stop closures.

The duration ratio was also calculated for the stop closures, a histogram of which is shown in Figure 8. This figure illustrates that the duration ratio of stop closures is concentrated at approximately 1.00. The mean value of the duration ratio is 1.04, which means that the duration of the stop closure for /t/ is almost the same as for /t^{h}/.

4.4 Voice Onset Time

Finally, the voice onset times (VOT) of /t^{h}/ and /t/ were measured. In this analysis, two tokens were discarded because the suffix $-t^{h}al^{4}$ was pronounced [tɬ], and thus no vocal fold vibration was observed after /t^{h}/ in these tokens.

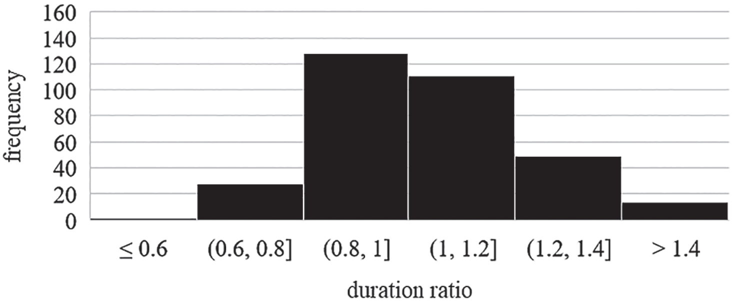


Figure 8 Histogram of stop closure duration ratios

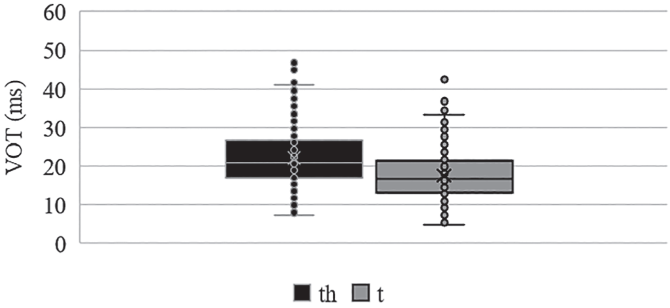


Figure 9 Distribution of VOTs

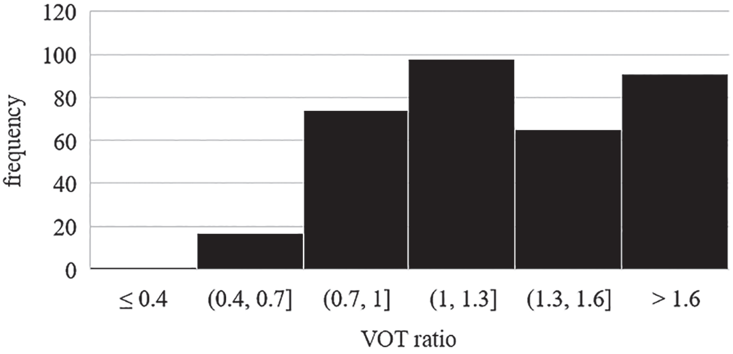


Figure 10 Histogram of VOT ratios

Figure 9 shows the distribution of the VOT for /t^h/ and /t/. The figure demonstrates that the distributions of VOT differ between /t^h/ and /t/; the VOT of /t^h/ tends to be slightly longer than that of /t/. The mean VOT values for /t^h/ and /t/ are 22.1 ms and 17.6 ms, respectively, and the difference is statistically significant (paired *t*-test: $p < .01$). The difference is not very great but certainly exists, and it is natural that the VOT of the aspirated stop /t^h/ is larger than that of its unaspirated counterpart /t/.

The VOT ratio is calculated by dividing the VOT of /t^h/ (e.g., 18.8 ms for but^h-$\underline{t}^hel$) by that of /t/ (e.g., 10.7 ms for but^h-$\underline{t}eg$). The result (18.8/10.7 = 1.76) means that the VOT of but^h-$\underline{t}^hel$ is 1.76 times as long as that of but^h-$\underline{t}eg$. VOT ratios were calculated for each pair.

Figure 10 presents a histogram of the VOT ratios, indicating that more than 70% of the VOT ratios were greater than 1.00. The mean value of the VOT ratio is 1.37, indicating

Table 4 VOTs for all participants

	AN	BB	BO	ER	ES	MM	NE	OT	OY	TN	UN	XL
/t^h/	26.5	20.7	23.7	17.1	19.9	24.5	23.8	24.1	23.8	22.0	27.2	12.4
/t/	24.3	13.3	19.2	19.0	14.8	16.6	14.1	22.4	19.6	14.4	23.2	9.7
t-test	n.s.	**	**	n.s.	**	**	**	n.s.	**	**	*	*

**: $p < .01$, *: $p < .05$, n.s.: $p > .05$

that the VOT of the aspirated stop /t^h/ tends to be larger than that of the corresponding unaspirated stop /t/.

It is clear from Figures 9 and 10 that the VOT of /t^h/ tends to be larger than that of /t/. However, the difference is relatively small, and this tendency is inconsistent: Approximately 30% of the VOT ratios are less than 1.00, that is, there are several pairs in which the VOT of /t/ is larger than that of /t^h/. The characteristics of VOT for each participant must be examined in more detail.

Table 4 shows the mean VOT values of /t^h/ and /t/ for each participant (AN-XL). It is clear from the table that the VOT of /t^h/ was consistently larger than that of /t/ for all participants except ER, which can be interpreted as indicating that the aspiration of /t^h/ was partially realized as post-aspiration after the stop, resulting in a longer VOT. However, the VOT difference may be too small to distinguish /t^h/ and /t/ sufficiently.

5. Discussion

In Section 4, four acoustic quantities were measured: the duration of obstruents, intensity during obstruents, duration of stop closures, and VOT. As shown in Sections 4.1 and 4.2, both the duration of and intensity during obstruents do not differ, regardless of whether the obstruent is followed by aspirated /t^h/ or unaspirated /t/. These facts suggest that the aspiration of /t^h/ is not explicitly realized as pre-aspiration of the preceding obstruent. In addition, as discussed in Section 4.3, the duration of the stop closure of /t^h/ is almost the same as that of /t/. This suggests that /t^h/ and /t/ after an obstruent are not distinguished by differences in stop closure duration.

In contrast, the VOT of /t^h/ tends to be larger than that of /t/ even in word-medial position, although the difference is small, as discussed in Section 4.4. In other words, the contrast between aspirated and unaspirated consonants can be, although not necessarily, realized as a difference in VOT even in word-medial position. According to Svantesson and Karlsson 2012, aspiration in word-initial aspirated consonants is realized as post-aspiration, which is characterized by a long VOT, while aspiration in word-medial aspirated consonants is realized as pre-aspiration. However, the present study revealed that aspiration is not realized as pre-aspiration but as post-aspiration, though not saliently so, when an aspirated consonant occurs after an obstruent. This suggests that Mongolian aspiration is essentially one feature irrespective of position; aspiration can occur either before or after the aspirated consonant, and its realization depends on the phonological environment.

These phenomena can be theoretically explained by the concept of “Licensing by Cue” (Steriade 1997), rather than “Licensing by Prosody” (Itô 1989). In the Licensing by Prosody conception, the distribution of features is controlled by their prosodic position such as syllable onset or coda. However, in the case of Mongolian aspiration, the syllable-based generalization is difficult because aspiration is realized as pre-aspiration in the V.$\underline{C}$V condition versus post-aspiration in the VC$_i$.$\underline{C}$V (C$_i$ = obstruent) condition, in both of which $\underline{C}$ occurs in the onset position of a syllable. In contrast, in the Licensing by Cue conception, phonemic contrasts are permitted (or licensed) based on a scale of perceptibility. According to Steriade 1997: 2, the main factor involved in licensing and neutralization is the distribution of cues to the relevant contrasts. This perception-based account is plausible for the Mongolian aspiration contrast; pre-aspiration (devoicing of the preceding segment) is one of the perceptual cues of aspiration in Mongolian; however, this cue is not available when an obstruent precedes it. In this condition, another perceptual cue, VOT, is utilized to maintain the aspiration contrast, or the contrast is ultimately neutralized because of the weakness of the perceptual cue.

6. Conclusion and Future Research

This study addressed the phonetic realization of word-medial aspiration in Mongolian, more specifically, the phonetic contrast between aspirated /t^h/ and unaspirated /t/ preceded by an obstruent. A production experiment and acoustic analysis were conducted to clarify whether aspiration is realized when an aspirated stop is preceded by an obstruent. Acoustic analyses revealed that the duration of obstruents, intensity during obstruents, and duration of stop closures did not differ depending on whether the following segment was aspirated or unaspirated, suggesting that pre-aspiration is not realized in obstruents. In contrast, VOT differs by aspiration type: It tends to be longer for aspirated than for unaspirated consonants. This implies that aspiration can be realized as post-aspiration, even in word-medial position under conditions where pre-aspiration is difficult to realize.

However, it remains unclear whether aspirated and unaspirated consonants preceded by obstruents can be sufficiently differentiated from the perspectives of both production and perception. It is well known that several acoustic features, such as the fundamental frequency (F0) and first formant (F1) of the adjacent vowel, are related to the distinction of laryngeal contrast (Haggard *et al.* 1970, Kent and Read 1992, Kluender and Lotto 1994). Ueta 2018b pointed out that F1 values and transitions in the following vowels differ between word-initial aspirated and unaspirated obstruents in Mongolian. Ueta 2021 further clarified that the following vowels significantly affect the perception of the distinction between word-initial aspirated and unaspirated consonants in Mongolian: It is possible that the contrast in aspiration is realized in the following vowel section, even in word-medial position, especially after an obstruent. Further research is needed to clarify whether the contrast between aspirated and unaspirated is sufficiently distinct following an obstruent.

References

Boersma, Paul and David Weenink. 2021. Praat: Doing Phonetics by Computer (Version6.1.50). http://www.praat.org/

Haggard, Mark, Stephen Ambler, and Mo Callow. 1970. "Pitch as a Voicing Cue." *The Journal of the Acoustical Society of America*, 47 (2): 613–617.

Itô, Junko. 1989. "A Prosodic Theory of Epenthesis." *Natural Language and Linguistic Theory*, 7 (2): 217–259.

Johnson, Keith. 2012. *Acoustic and Auditory Phonetics* (3rd edition). Malden, MA: Wiley-Blackwell.

Karlsson, Anastasia M. 2005. *Rhythm and Intonation in Halh Mongolian*. Lund: Lund University.

Karlsson, Anastasia and Jan-Olof Svantesson. 2011. "Preaspiration in Mongolian Dialects: Acoustic Properties of Contrastive Stops." *Paper presented at the 10th Seoul International Altaistic Conference*: 125–140.

Karlsson, Anastasia M. and Jan-Olof Svantesson. 2012. "Aspiration of Stops in Altaic Languages: An Acoustic Study." *Altai Hakpo*, 22: 205–222.

Kent, Ray D. and Charles Read. 1992. *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group.

Klatt, Dennis H. 1975. "Voice Onset Time, Frication, and Aspiration in Word-initial Consonant Clusters." *Journal of Speech and Hearing Research*, 18: 686–706.

Kluender, Keith R. and Andrew J. Lotto. 1994. "Effects of First Formant Onset Frequency on [−voice] Judgements Result from Auditory Processes not Specific to Humans." *The Journal of the Acoustical Society of America*, 95 (2): 1044–1052.

Lisker, Leigh and Arthur S. Abramson. 1964. "A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements." *Word*, 20: 384–422.

Steriade, Donca. 1997. "Phonetics in Phonology: The Case of Laryngeal Neutralization." UCLA ms.

Svantesson, Jan-Olof, Anna Tsendina, Anastasia M. Karlsson, and Vivan Franzén. 2005. *The Phonology of Mongolian*. Oxford: Oxford University Press.

Svantesson, Jan-Olof and Anastasia Karlsson. 2012. "Preaspiration in Modern and Old Mongolian." *Suomalais-Ugrilaisen Seuran Toimituksia*, 264: 453–464.

Ueta, Naoki. 2018a. "Voice Onset Time of Word-initial Stops and Affricates in Khalkha Mongolian." *Journal of the Phonetic Society of Japan*, 22 (2): 131–140.

Ueta, Naoki. 2018b. "Mongorugo haruha hoogen no gotoo sogaion no tairitsu ni okeru F0 to F1 no tokuchoo." [The Characteristics of F0 and F1 in the Contrast of Word-initial Obstruents in Khalkha Mongolian]. *Journal of Kijutsuken*, 10: 81–95.

Ueta, Naoki. 2020a. "The Acoustic Realization of Preaspiration in Khalkha Mongolian." *Northern Language Studies*, 10: 99–115.

Ueta, Naoki. 2020b. "Mongorugo haruha hoogen no gochuu heesaon no onseeteki barieeshon to on-in kaishaku." [Phonetic Variations and Phonological Interpretations of Word-medial Stops in Khalkha Mongolian]. *Bulletin of the Japan Association for Mongolian Studies*, 50: 1–18.

Ueta, Naoki. 2021. "The Perception of Word-initial Aspiration Contrasts in Mongolian: The Effect of Voice Onset Time and Following Vowels." *Altai Hakpo*, 31: 155–175.

Wayland, Ratree and Allard Jongman. 2003. "Acoustic Correlates of Breathy and Clear Vowels: The Case of Khmer." *Journal of Phonetics*, 31: 181–201.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Cherokee *s*-preaspiration at the Phonetics-phonology Interface*

UCHIHARA, Hiroto

Cherokee /s/ is realized with preaspiration after a short vowel. This *s*-preaspiration seems like a simple phonetic process, but a categorical phonological process of Vowel Deletion and a morphophonological process of Laryngeal Alternation are sensitive to this phonetic [^{h}] before /s/, thereby posing a problem of whether this *s*-preaspiration is a phonetic or a phonological process. I explore whether enriching the underlying representation by incorporating some phonetic details that are not contrastive can capture such a situation.

1. Introduction

In Oklahoma Cherokee, /s/ is realized phonetically with pre-aspiration (“*s*-preaspiration”) after a short vowel (Feeling 1975: x). This is illustrated in Figure 1. Here, the [s] in the second syllable is preceded by [h] which is not in the underlying representation.

Keywords: phonetics-phonology interface, preaspiration, Cherokee

DOI: https://doi.org/10.15026/0002000625

* I’m grateful for an anonymous reviewer for their insightful suggestions, and the editors of the special volume of the journal for their careful editorial work. A previous version of this paper was presented at *Phonetic typology from cross-linguistic perspectives*, ILCAA on February 20, 2023. I would like to thank the participants of the meeting for their valuable comments. I’m indebted to the speakers of Cherokee, especially David Crawler, Durbin Feeling, Russel Feeling, Ed Jumper, DJ McCarter, Junior Scraper, Eda Scraper, Anna Sixkiller, and Dennis Sixkiller. This work has been funded by the University at Buffalo, State University of New York (2011); the Phillips Fund, American Philosophical Society (2012, 2020); Jacobs Research Fund, University of Washington (2020); and Grant-in-Aid for Research Activity Start-up, Japan Society for the Promotion of Science for the project JSPS KAKENHI Grant Number 22K19988, ‘A cross-dialectal study of Tlapanec verb morphology’ (2022–2024). This paper is part of the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021–2023).

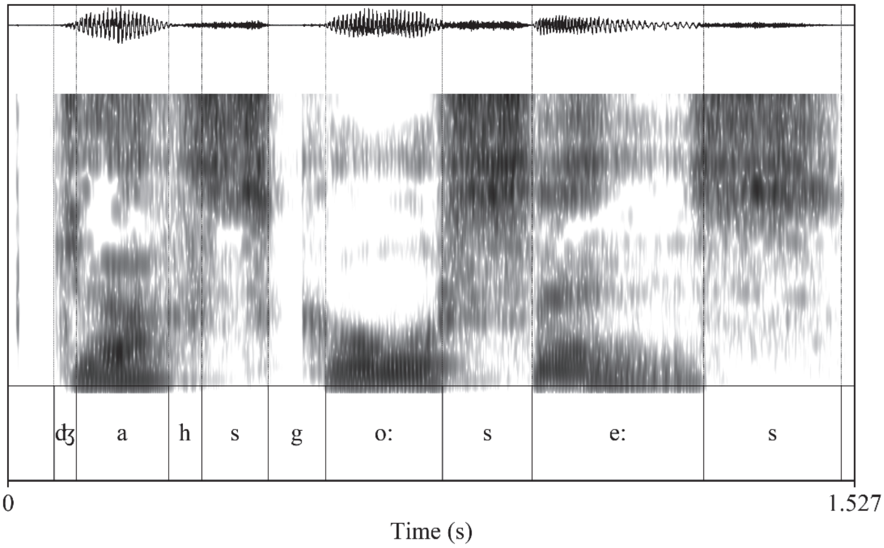


Figure 1 [ʤa^{h}˧sgo:˦se:s˧˦] /ca-skoó-s-éʔi=s/ [2SG.B-dig-PFT-REP=Q]$^{1)}$ ‘Did you dig it?’

Due to this apparent phonetic *s*-preaspiration, /s/ and /hs/ do not contrast in Cherokee. However, this phonetic [^{h}] behaves as if it is there phonologically with respect to a categorical phonological process (Vowel Deletion) as well as to a morpheme-specific morphophonological rule (Laryngeal Alternation), as will be discussed in detail in §4.

In general, phonological processes and morphophonological rules are considered not to refer to subcategorical phonetic details, but in the case of Cherokee *s*-preaspiration, they appear to do so. In this paper, I will examine whether *s*-preaspiration is a phonetic or phonological process by examining the two alternations mentioned above, and explore whether we can handle such a situation by enriching the representation, allowing categorical phonological processes and morphophonological rules to refer to phonetic information (Kingston & Diehl 1994, Hayes & Steriade 2004, Lionnet 2017, Bennett *et al.* 2023).

This paper is organized as follows. First §2 provides some background information on Oklahoma Cherokee and the theories of phonetics-phonology interface. Cherokee /s/ manifests some peculiarities, both in its phonetic characteristics and distribution; this is discussed in §3. §4 examines in detail the two processes that target preaspiration, and §5 explores the alternatives to capture such behavior, followed by conclusion in §6.

2. Background

2.1 Oklahoma Cherokee

Cherokee is the sole representative of the Southern Iroquoian branch of the Iroquoian language family. Cherokee is spoken by ~1,850 remaining first-language speakers in

---

1) List of abbreviations: A: set A; ASR: assertive; B: set B; DIST: distributive; DU: dual; FUT.IMP: future imperative; I: prothetic-*i*; IND: indicative; MID: middle; PCT: punctual; PFT: perfective; PL: plural; PRS: present; Q: interrogative; REP: reportative; SG: singular.

Table 1 Cherokee Consonants.

		Bilabial	Alveolar	Palatal	Velar	Labialized velar	Glottal
Stops			t		k	kw	ʔ
Affricates	Central		c				
	Lateral		tl				
Fricatives			s				h
Nasals		m	n				
Liquid			l				
Glides				y		w	

northeastern Oklahoma and ~250 in western North Carolina (Roy Boney, Jr., Manager of the Cherokee Nation Language Program, p.c). Speakers are mostly over the age of 50.

Cherokee has six vowels (*a, e, i, o, u, ʌ*; the last phoneme is represented as <v> in the orthography), and vowel length is contrastive. Cherokee has six tones (low (*a*), high (*á*), lowfall (*à*), high-low (*â*), low-high (*ǎ*) and superhigh (*a̋*)) that can occur on a syllable with a long vowel (Uchihara 2016). Consonant phonemes are shown in Table 1.

2.2 The Theory of Phonetics-phonology Interface

According to the traditional view, phonology has to do with the discrete, abstract units, while phonetics has to do with the realization in continuous time and physical space. Research in recent years has focused on how they are linked. At least two theories have been proposed on the phonetics-phonology interface (Zsiga 2021, Bennett *et al.* 2023). The first is a modular, substance-free theory (Keating 1990, Zsiga 2000, Boersma 2011, among others), where phonetic representations are considered distinct from phonological processes, and where phonological representations are abstract and categorical, and are mapped to distinct phonetic representations, which are defined in continuous, physical terms. The other is a nonmodular, integrated, phonetically-grounded theory (Kingston & Diehl 1994, Steriade 2001, Flemming 2001, Hayes & Steriade 2004, Pierrehumbert 2016, Lionnet 2017, Bennett *et al.* 2023), where there is no meaningful distinction between phonetic and phonological representations. In this paper, I will examine which theory accounts for Cherokee *s*-preaspiration better.

3. Peculiarity of Cherokee /s/

Cherokee /s/, which undergoes-preaspiration, is peculiar both in terms of its phonetic characteristics (§3.1) and its distribution (§3.2). This section looks at these peculiarities.

3.1 Phonetic Characteristics

Cherokee /s/ has four phonetic characteristics that are not shared by other consonants in Cherokee. First, as mentioned at the beginning of this paper, /s/ is preaspirated when preceded by a short vowel (Feeling 1975: x). This process can be formulated as in (1), and illustrated by an example in (2)

(1) Ø → h /V_s

(2) nisdi:weʔa [niʰ˧sdi:˧we˧ʔa˥˧][2]
/ni-stii-weʔ-a/
PART-2DU-say:PRS-IND
'You two are saying it.' (EJ 2011)

Preaspiration is not applied when the preceding vowel is long, as in (3)a below. In general, a tautosyllabic sequence of a long vowel followed by an /h/ (*V:h]σ) is prohibited in Oklahoma Cherokee (Scancarelli 1987: 27).

Secondly, other obstruents are realized as voiced when not followed by an /h/, either at the initial position as in (3), between vowels as in (4), or after a consonant as in (5), while /s/ never realized as voiced [z], whether at the initial or medial position, as in (6):[3]

(3) #_V
a. do:s [do:s˧]
/toos(a)/
'mosquito'(JRS, Aug 2012)
b. go:k [go:kʰ˧]
/kook(i)/
'summer' (JRS, Aug 2012)

(4) V_V
a. ada [a˧da˥˧]
/ata/
'wood'(JRS, Aug 2012)
b. i:ga [i:˧ga˥˧]
/iika/
'day' (JRS, Aug 2012)

(5) C_
sgǒ:hi [sgo:˦˥hi˥]
/skoóhi/
'ten' (CED-EJ, 2010)

(6) sa:sa [sa:˧sa˥]
/saasa/
'goose' (CED-EJ, 2010)

---

[2] In the examples, the first line shows the surface forms (accompanied by phonetic representations in []), followed by the segmented forms in the underlying representations in the second line. The third and fourth lines provide gloss and free translation.

[3] This could have an aerodynamic motivation: fricatives require a pressure drop across the constriction to generate turbulence, while voicing requires a phonation threshold pressure across the vocal folds. Having both sources operating at the same time requires skillful manipulation of the amount of the airflow (Stevens 1971; Ladefoged & Maddieson 1996: 176ff.). Thus, many languages in the world have voicing contrast for plosives but not for fricatives, as in Spanish.

The third peculiarity is that /s/ is in free variation with [h] in some forms (Scancarelli 2005: 363), as in (7). On the other hand, underlying /h/ does not alternate with [s]; thus, the /h/ in /skoóhi/ in (5) above is never realized as [s].

(7) sé:hnv [se:˦n̥ʌ˦] ~ [he:˦n̥ʌ˦]
/sé:hnv /
‘but’

Finally, As can be seen from Table 1, the place of articulation is not contrastive for sibilants. Thus, [s] is not contrastive with the postalveolar sibilant [ʃ]. Since there is no contrast, the place of articulation is subject to dialectal variation; in North Carolina Cherokee /s/ is usually realized as a voiceless postalveolar [ʃ] or retroflex [ʂ] (Uchihara 2016: 42), while in Oklahoma Cherokee the alveolar place of articulation is more common. Moreover, the lack of contrast in place of articulation has resulted in the following polysemy.

(8) dà:lâ:su:hlv́sga [da:˧˩la:˦˨su:˧l̥ʌ˦ʰsga˦˧]
/t-Ø-alaʔsuul-hv́sk-a/
DIST-3SG.A-put.on.shoe-PRS-IND
‘he is putting on shoes; he is suing (him)’

Here, the same verb means both ‘put on shoes’ and ‘sue’. They have nothing in common semantically, but taking into consideration that [s] and [ʃ] are not contrastive in Cherokee, one could imagine that Cherokee monolinguals categorized [s] and [ʃ] as the same phoneme, thus ‘shoe’ and ‘sue’ may have sounded the same to their ears.

3.2 Distribution

Cherokee /s/ is also peculiar in terms of its distribution. First, /s/ is the only consonant which has its own “syllabary” as a segment. This is possibly because only *s, h* and glottal stop can form clusters with other consonants at the underlying level (Uchihara 2016: 67); other consonant clusters result from the application of various phonological processes, one of which is Vowel Deletion to be discussed in §4.1. The other two consonants that can form underlying clusters, *h* and glottal stop, are unlike *s*, because a glottal stop is never represented in the syllabary, while *h* is only represented when it occurs as the only consonant in the onset position (as in the ‘h’ column in Table 2), and not when it is before another consonant.

Secondly, unlike other obstruents, /s/ cannot be preceded or followed by /h/ phonologically. In other words, that /hs/ vs /s/ vs /sh/ cannot be contrastive, while other obstruents can be preceded (/hT/), followed (/Th/), or flanked (/hTh/) by /h/. First, (9) is an example which contains an /hT/ sequence.

Table 2 Cherokee Syllabary

	a	e	i	O	u	v
V	Ꭰ a	Ꭱ e	Ꭲ i	Ꭳ o	Ꭴ u	Ꭵ v
g/k	Ꭶ ga Ꭷ ka	Ꭸ ge	Ꭹ gi	Ꭺ go	Ꭻ gu	Ꭼ gv
h	Ꭽ ha	Ꭾ he	Ꭿ hi	Ꮀ ho	Ꮁ hu	Ꮂ hv
l	Ꮃ la	Ꮄ le	Ꮅ li	Ꮆ lo	Ꮇ lu	Ꮈ lv
m	Ꮉ ma	Ꮊ me	Ꮋ mi	Ꮌ mo	Ꮍ mu	
n	Ꮎ na, Ꮏ hna, Ꮐ nah	Ꮑ ne	Ꮒ ni	Ꮓ no	Ꮔ nu	Ꮕ nv
gw	Ꮖ gwa	Ꮗ gwe	Ꮘ gwi	Ꮙ gwo	Ꮚ gwu	Ꮛ gwv
s	Ꮜ sa Ꮝ s	Ꮞ se	Ꮟ si	Ꮠ so	Ꮡ su	Ꮢ sv
d/t	Ꮣ da Ꮤ ta	Ꮥ de Ꮦ te	Ꮧ di Ꮨ ti	Ꮩ do	Ꮪ du	Ꮫ dv
dl	Ꮬ dla Ꮭ tla	Ꮮ dle	Ꮯ dli	Ꮰ dlo	Ꮱ dlu	Ꮲ dlv
j	Ꮳ ja	Ꮴ je	Ꮵ ji	Ꮶ jo	Ꮷ ju	Ꮸ jv
w	Ꮹ wa	Ꮺ we	Ꮻ wi	Ꮼ wo	Ꮽ wu	Ꮾ wv
y	Ꮿ ya	Ᏸ ye	Ᏹ yi	Ᏺ yo	Ᏻ yu	Ᏼ yv

(9) hvhda [hʌʰ˧da˥˦]
/h-vht-Ø-a/
2SG.A-use-PCT-IND
'Use it!' (Feeling 1975: 143)

In (10), /t/ is followed by an /h/ and is thus realized as [t] instead of [d] as would have been the case had it not been followed by /h/.

(10) gatv́sga [ga˧tʌʰ˦sga˥˦]
/k-athv́sk-a/
3SG.A-hang.up:PRS-IND
'He is hanging it up.' (Feeling 1975: 116)

Finally, (11)a contains an underlying /hth/ sequence. Although the /h/ preceding the /t/ is deleted in this example, some speakers have this [ʰ] in the same form. In addition, (11)b shows the form which undergoes Laryngeal Alternation to be discussed in §4.2, where the first underlying /h/ alternates with a glottal stop or triggers lowfall tone on the preceding vowel, which is the case here. If there was no underlying /h/ before /t/, we would expect the surface [aʔd], instead of [à:t] as found in here (Uchihara 2013: Ch. 4).

(11) a. [ha(ʰ)˧tʌ:˧da:˥˧sta˧nʌ:˧ʔi˥˦]
hatv:dâ:stanv:ʔi
/h-a(h)thvvta(ʔ)st-ahn-vvʔi/
2SG.A-listen-PFT-FUT.IMP
'Listen later!' (Feeling *et al.* 2003: 164)

b. [ga:˧˩tʌ:˧da:˥˧sdi˧ha˥˦]
gà:tv:dâ:sdiha
/k-ahthvvta(ʔ)st-ih-a/
1SG.A-listen-PRS-IND
'I am listening to it.' (Feeling 1975: 61)

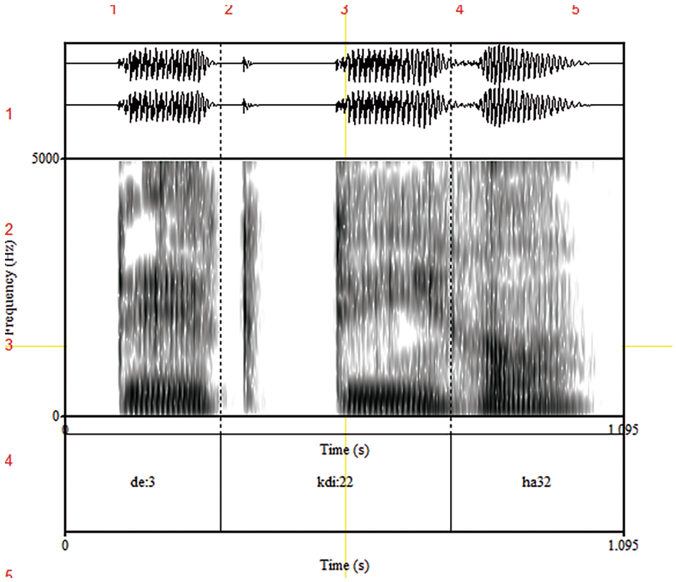


Figure 2 **Spectrogram of *dé:kdi:ha***

4. Behavior of Cherokee *S*-preaspiration

Up to here we have treated the *s*-preaspiration as a mere phonetic process. However, this-preaspiration behaves as if it is preceded by an /h/ phonologically in terms of Vowel Deletion, which is a categorical phonological process, and Laryngeal Alternation, which is a morphophonological rule. This section examines in detail each of these processes.

4.1 Vowel Deletion

Vowel Deletion is one of the phonological processes motivated by the avoidance of a *CVh sequence, which deletes the vowel from this sequence when this sequence is followed by a plosive or an affricate (Flemming 1996, Uchihara 2013: Ch. 3), formulated as in (12). Figure 2 is a spectrogram of (13), which shows that the deletion of the vowel is complete.

(12) CVhT → ChT (T = plosive, affricate)
(13) dé:kdi:ha [de:˦kdi:˧ha˧˨]
/tee-k-(v)ht-iih-a/
DIST-3SG.A-use-PRS-IND
‘He is using it.’

The following examples illustrate additional instances of Vowel Deletion. The deleted vowels are in parentheses in the second lines. The forms in (b) are in contexts where Vowel Deletion is blocked, thereby justifying the underlying forms with the deleted vowels. In (14)

Table 3 Deleted vowels in the syllabary

MEANING	SURFACE	UNDERLYING	Feeling (1975)	EJ (2011)
1. dance	à:lsgí:ʔa	-al(i)skiiʔ-	ᎠᎵᏍᎩᎠ (ali̲sgia)	ᎠᎵᏍᎩᎠ (ali̲sgia)
2. eat meal	à:lsdáayv:hv́sga	-al(i)staʔyvvhvsk-	ᎠᎵᏍᏓᏰᎲᏍᎦ (ali̲sdayvhvsga)	ᎠᎵᏍᏓᏰᎲᏍᎦ (ali̲sdayvhvsga)
3. sew	ga:ye:wsga	-:yeew(i)sk-	ᎦᏰᏫᏍᎦ (gayewi̲sga)	ᎦᏰᏫᏍᎦ (gayewi̲sga)
4. smoke	go:ksga	-ook(i)sk-	ᎪᎦᏍᎦ (goga̲sga)	ᎪᎩᏍᎦ (gogi̲sga)
5. understand	go:lhga	-ool(i)hk-	ᎪᎵᎦ (goli̲ga)	ᎪᎸᎦ (golv̲ga)

b and (15)b the *Vh* sequence is precede by an /h/, which blocks the application of Vowel Deletion. The form in (16) undergoes Laryngeal Alternation (§4.2) where the first /h/ of the stem alternates with a glottal stop or induces a lowfall tone on the preceding vowel. The forms in (17) are in free variation between one which undergoes Vowel Deletion and the other which does not.

	Vowel Deletion		No Vowel Deletion
(14)	a. à:nhdlí [a:n̥˧˩ʤi˦] /an-(v)htl-í(h-a)/ 3PL.A-sharpen-PRS-IND 'They are sharpening it.' (JRS, Aug 2011)		b. hvhdla [hʌʰ˧ʤa˥˧] /h-vhtl-Ø-a/ 2SG.A-sharpen-PCT-IND 'Sharpen it!' (Feeling 1975: 143)
(15)	a. ùwhtánv̌:ʔi [uw̥˧˩ta˦nʌ˦˥:ʔi˥˧] /uw-(v)ht-áhn-vv́ʔi/ 3SG.B-use-PFT-ASR 'He used it.' (Feeling 1975: 143)		b. hvhda [hʌʰ˧da˥˧] /h-vht-Ø-a/ 2SG.A-use-PCT-IND 'Use it!' (ibid.)
(16)	a. de:gú:kdíha [de:˧gu:˦kdi˦ha˥˧] /tee-k-uuk(o)ht-íh-a/ DIST-3SG.A-decide-PRS-IND 'He is deciding.' (Feeling 1975: 78)		b. de:gú:gò:díha [de:˧gu:˦go:˧˩di˦ha˥˧] /tee-k-uukoht-íh-a/ DIST-1SG.A-decide-PRS-IND 'I am deciding.' (ibid.)
(17)	a. itlgv̋ [i˧tɬgʌ˦˥] /i-tl(u)hkvv̋(ʔi)/ I-tree 'tree' (CED-EJ, 2010)	~	b. dluhgv̋ [ʤuʰ˧gʌ˦˥] /tluhkvv̋(ʔi)/ tree 'tree' (Holmes & Smith 1976: 100)

Vowel Deletion is a categorical process, in the sense that the vowel is completely deleted as can be seen in Figure 2, and thus neutralizes with zero. Moreover, when speakers write in syllabary, they can't always recover the underlying vowels (Uchihara 2013: 87ff.). This is shown in Table 3; all of these forms have a deleted /i/, which is justified in other conjugated forms. The final two columns show the syllabary representations by two speakers. As can be seen, the two speakers use the syllabaries with <i> for the sequences which contain a deleted /i/ for the first three verbs, but Feeling (1975) writes with the syllabary for <ga> for the fourth verb 'smoke', while EJ writes with the syllabary for <lv> for the fifth verb 'understand'.

On the other hand, Vowel Deletion is a variable process, since its application is optional as was seen in (17), and also depends on the speech rate.

Here, what is crucial is that not only does the *CVhT* sequence undergo Vowel Deletion, but also the *CVs* sequence. This is formulated as in (18), and examples are given in (19) - (21). The forms in (b) again shows forms which do not undergo Vowel Deletion: in (19)b and (20)b because the vowel is preceded by an /h/, and (21)b because it undergoes Laryngeal Alternation (§4.2).

(18) CVs → Cs

Vowel Deletion	No Vowel Deletion
(19) a. à:ksgo:lĭ:yê:ʔv́ [a:˧˩ksgo:˧li:˦je:˧˨ʔʌ˦] /ak-(a)skoolii-éeʔ-v́v́(ʔi)/ 1SG.B-rub-PRS-IND 'I rubbed it.' (JRS, Aug 2012)	b. hasgo:lĭ:ya [ha^{h}˧sgo:˧li:˦ja˧˨] /h-askooliíy-Ø-a/ 2SG.B-rub-PCT-IND 'Rub it!' (Feeling 1975: 50)
(20) a. à:nsgŏ:sga [a:n̥˧˩sgo:˦sga˧˨] /an-(a)skoó-sk-a/ 3PL.A-dig-PRS-IND 'They are digging it.' (JRS, Aug 2012)	b. hasgŏ:la [ha^{h}˧sgo:˦la˧˨] /h-askoó-l-a/ 2SG.A-dig-PCT-IND 'Dig it!' (Feeling 1975: 51)
(21) a. à:ksósga [a:˧˩kso^{h}˦sga˧˨] /a-k(a)só-sk-a/ 3SG.A-go.downhill-PRS-IND 'He is going downhill.' (Feeling 1975: 34)	b. jigà:sósga [ʤi˧ga:˧˩so^{h}˦sga˧˨] /ci-kasó-sk-a/ 1SG.A-go.downhill-PRS-IND 'I am going downhill.' (ibid.)

Thus, without the pre-*s* /h/ in the underlying representation, the formulation of Vowel Deletion will be as follows. As can be observed, we need two separate rules.

(22) Formulation without pre-*s* /h/
a. CVhT → ChT
b. CVs → Cs

Such an analysis fails to capture the generalization that in both cases there is a preconsonantal [^{h}] phonetically and that in both cases the alternation is motivated by the avoidance of a *CVh sequence.

An alternative is to postulate an underlying /h/ before an /s/, so that we can formulate Vowel Deletion as one rule, as in (23), subdivided into (a) and (b).

(23) Formulation with pre-s /h/
CVhO → ChO (O = obstruent)
a. CVhT → ChT (T = plosive, affricate)
b. CVhs → Chs

Under such an analysis, Vowel Deletion would be captured as follows (repeated from (21)), where the underlying /kahs/ sequence undergoes vowel deletion, just like when the consonant after *h* is a plosive:

(24) [a:˨˩kso^h˧sga˦˧]
/a-k(a)hsó-hsk-a/ (instead of /a-kasó-sk-a/)
3SG.A-go.downhill-PRS-IND
'He is going downhill.'

However, such an analysis is problematic, since /s/ cannot be preceded by an /h/ phonologically as was mentioned above and thus /hs/ vs /s/ cannot be contrastive, even though phonetically [s] is preaspirated (preceded by an [^{h}]). In other words, such an analysis implies the existence of the sequence /hs/ without the singleton /s/ and would also represent a case of absolute neutralization (Kiparsky 1968) which is undesirable.

4.2 Laryngeal Alternation

Laryngeal Alternation (Cook 1979: 40, Munro 1996) is a morpheme-specific morphophonological rule (or morphomic stem alternation) which is triggered by certain pronominal prefixes (such as the 1st person singular agentive). Most pronominal prefixes select the *h*-grade, where the first laryngeal consonant of the stem is *h*, while certain pronominal prefixes such as 1st person singular agentive selects glottal grade where the first laryngeal consonant is a glottal stop, as can be observed in (25).4)

	h-grade	glottal grade
(25)	a. à:de:loho:sga [a:˨˩de:˧lo˧ho:˧sga˦˧]	b. gade:loʔo:sga [ga˧de:˧lo˧ʔo:˧sga˦˧]
	/Ø-ateelohoo-sk-a/	/k-ateelohoo-sk-a/
	3SG.A-find.out-PRS-IND	1SG.A-find.out-PRS-IND
	'He is finding it out.' (Feeling 1975: 9)	'I am finding it out.' (ibid.)

When the laryngeal is followed by a consonant, the glottal grade is realized with a lowfall tone, [V̀:C] instead of a glottal stop as in the following example. Possibly the lowfall tone was induced by creakiness that accompanies a glottal stop. In the following example the *e:hl* sequence in the h-grade in (a) corresponds to the *è:l* sequence in the glottal grade in (b):

---

4) The lowfall tone on the first syllable in (25)a is due to Pronominal Tonic Lowering (Lindsey 1985: 136; Uchihara 2016: 103) which is orthogonal to the discussion here.

h-grade / glottal grade

(26) a. [a:˨˩de:˧l̥oh˧gʷa˦ʔa˦˨]
à:de:hlohgwáʔa
/Ø-ateehlohkw-áʔ-a/
3SG.A-learn-PRS-IND
'He is learning it.' (Feeling 1975: 8)

b. [ga˧de:˨˩loh˧gʷa˦ʔa˦˨]
gadè:lohgwáʔa
/k-ateehlohkw-áʔ-a/
1SG.A-learn-PRS-IND
'I am learning it.' (ibid.)

Table 4 summarizes the incidence of Laryngeal Alternation:

Crucially, Laryngeal Alternation is also triggered by /s/, as if /s/ is preceded by an /h/, as can be seen in the (27)b, where the vowel *i* is assigned lowfall tone before an /s/. The (a) form is in the *h*-grade where the same /i/ undergoes vowel deletion in the CVs context. (28) shows a parallel example where the vowel is followed by an /h/ + plosive (in this case, /k/) sequence, showing that *CVs* and *CVhT* sequences behave the same.

h-grade / glottal-grade

(27) a. à:lsgû:sga [a:l̥˨˩sgu:˥˧sga˦˨]
/Ø-al(i)-skuʔ-sk-a/
3SG.A-MID-nod-PRS-IND
'He is nodding.' (Feeling 1975: 42)

b. galì:sgû:sga [ga˧li:˨˩sgu:˥˧sga˦˨]
/k-ali-skuʔ-sk-a/
1SG.A-MID-nod-PRS-IND
'I'm nodding.' (ibid.)

(28) a. à:lhko:tdíha [a:l̥˨˩ko:˧tdi˦ha˦˨]
/Ø-al(i)hkhootht-íh-a/
3SG.A-shatter-PRS-IND
'He is shattering it.' (Feeling 1975: 23)

b. galì:ko:tdíha [ga˧li:˨˩ko:˧tdi˦ha˦˨]
/k-alihkhootht-íh-a/
1SG.A-shatter-PRS-IND
'I am shattering it.' (Feeling 1975: 23)

/s/ also triggers Laryngeal Alternation even when it is not preaspirated. Recall that *s*-preaspiration is not applied when /s/ is preceded by a long vowel. This is shown in the following pair. (29)a is in the *h*-grade, where we can confirm that /s/ is preceded by a long vowel and thus there is no preaspiration, but Laryngeal Alternation is still applied before this /s/, resulting in the lowfall tone on the preceding vowel, as shown in (29)b.

h-grade / glottal-grade

(29) a. à:gv:sgwó:ʔa [a:˨˩gʌ:˧sgʷo:˦ʔa˦˨]
/Ø-akvvskoo-ʔ-a/
3SG.A-wash.face-PRS-IND
'He is washing his face.' (Feeling 1975: 19)

b. gagv̀:sgwó:ʔa [ga˧gʌ:˨˩sgʷo:˦ʔa˦˨]
/k-akvvskoo-ʔ-a/
1SG.A-wash.face-PRS-IND
'I am washing my face.' (ibid.)

Table 4 Incidence of Laryngeal Alternation

h-grade	glottal grade
VhV	VʔV
VhC	V̀:C

Table 5 Formulation without pre-s /h/

h-grade	glottal grade
VhV	VʔV
V^{h}C	V̀:C
VsX	V̀:sX

Table 6 Formulation with pre-s /h/

h-grade	glottal grade
VhV	VʔV
V^{h}C (C includes /s/)	V̀:C

Thus, formulation of Laryngeal Alternation will be as in Table 5, with separate rows for the forms with an /s/ (X = any segment). Again, such an analysis fails to capture the generalization that a pre-consonantal *h*, whether phonological or phonetic, triggers Laryngeal Alternation.

Alternatively, one can postulate that /s/ is preceded by an /h/ phonologically (which is deleted after a long vowel, as in (29)), as in Table 6; this way we can generalize Laryngeal Alternation.

Under such an analysis, the underlying /h/ regularly triggers lowfall tone on the preceding vowel /a/ in (30), just like an /h/ does before a plosive, as in the following example. Here, Laryngeal Alternation is applied to the underlying /ah/ sequence, resulting in [à:].

(30) [ʤi˧ga:˨˩soʰ˦sga˥˧]
/ci-kahsó-hsk-a/ (instead of /ci-kasó-sk-a/)
1sg.a-go.downhill-prs-ind
‘I am going downhill.’ (Feeling 1975: 34)

Forms where an /s/ is preceded by a long vowel, where Laryngeal Alternation is applied even though it is not preaspirated, can also be handled by such a representation. This is shown in (31), repeated from (29). Here, the underlying /h/ is deleted after the long vowel /vv/ due to the constraint *V:h]σ (Scancarelli 1987: 27) in the *h*-grade form in (a). In (b) Laryngeal Alternation is applied, thereby the vowel *v:* is assigned a lowfall tone before the underlying /hs/ sequence.

	h-grade		glottal-grade
(31) a.	à:gv:sgwó:ʔa [a:˨˩gʌ:˧sgʷo:˥ʔa˥˧]	b.	gagv̀:sgwó:ʔa [ga˧gʌ:˨˩sgʷo:˥ʔa˥˧]
	/Ø-akvv(h)skoo-ʔ-a/		/k-akvvhskoo-ʔ-a/
	3sg.a-wash.face-prs-ind		1sg.a-wash.face-prs-ind
	‘He is washing his face.’		‘I am washing my face.’ (Feeling 1975)

However, recall again that /s/ cannot be preceded by an /h/ phonologically; /hs/ vs /

Table 7 Diagnostics

	Phonological	Phonetic	*s*-preaspiration
(i) Show physical gradience	No	Yes	Maybe
(ii) Dependent on speech rate	No	Yes	Maybe
(iii) Sensitive to morphological structure	Yes	Possibly, but only gradiently	Yes (Laryngeal Alternation)
(iv) Sensitive to phonotactic restrictions	Yes	No	Yes (not after a long vowel, etc.)
(v) Feed/bleed phonological processes	Yes	No	Yes (Vowel Deletion)

s/ cannot be contrastive, even though phonetically [s] is preaspirated (preceded by an [h]). In other words, such an analysis implies the existence of the sequence /hs/ without the singleton /s/.

4.3 Summary

In this section, we have examined how *s*-preaspiration interacts with a categorical phonological process of Vowel Deletion, and a morphophonemic rule of Laryngeal Alternation. In both cases the [h] that precedes an /s/ behaves as if it is there phonologically.

To see whether Cherokee *s*-preaspiration has to do with the domain of phonetics or phonology, we can apply the diagnostics proposed by Bennett *et al.* 2023 to distinguish the two, shown in Table 7. We need a fine-grained acoustic study to determine whether *s*-preaspiration shows physical gradience (i) or dependence on speech rate (ii). In general, phonological processes can be sensitive to morphological structure (iii), such as the type of morphemes, while phonetic processes can be but only gradiently. Since Laryngeal Alternation is sensitive to the morphosyntactic category (such as the 1st person singular agentive prefix), *s*-preaspiration is sensitive to the morphological structure, thus more phonological. *s*-preaspiration is phonotactically conditioned (iv), in the sense that it is not applied after a long vowel as we saw above, thus more phonological. As we saw above in §4.1, *s*-preaspiration feeds (i.e. applies before) Vowel Deletion which is a phonological process and Laryngeal Alternation which is a morphophonological rule (v). Again, in this sense, *s*-preaspiration is more phonological.

As can be seen, most diagnostics indicate that *s*-preaspiration is a phonological process, although this is an automatic, non-neutralizing process (typical of a phonetic process).

5. Discussion: Integrating Phonetic Information in the Underlying Forms?

In the previous section we were faced with the dilemma that *s*-preaspiration is automatic and non-neutralizing, which is typical of a phonetic process but it looks more like a phonological process, when we apply the diagnostics proposed in the literature.

One alternative would be to consider that preaspiration of /s/ is not only phonetic but also phonological, and postulate an underlying /h/ before every /s/ as we have

Table 8 Laryngeal Alternation

h-grade	glottal grade
[VhV]	[VʔV]
[V^{h}C]	[V̀:C]
[V^{h}sX]	[V̀:sX]

intended above. Such an analysis can explain alternations and distribution, but as we saw the problem is that such an analysis implies the existence of the sequence /hs/ without the singleton /s/, which is undesirable. This is because all post-vocalic /s/ triggers Vowel Deletion (when the vowel is short) and Laryngeal Alternation, and thus under such an analysis /s/ would always be preceded by an /h/.

An alternative is to allow categorical phonological processes and morphophonological rules to refer to phonetic information (Kingston & Diehl 1994, Hayes & Steriade 2004:1, Lionnet 2017, Bennett *et al.* 2023). Under such an analysis, Vowel Deletion can be reformulated as follows: it applies to a vowel preceded by a consonant, followed by a *phonetic* (not phonological) [h]:

(32) Vowel Deletion
a. [CV^{h}T] → ChT
b. [CV^{h}s] → Chs

Similarly, Laryngeal Alternation can be reformulated to target phonetic (not phonological) [h], as in Table 8. However, a problem with such an analysis is that /s/ triggers Laryngeal Alternation even when there is no preaspiration, that is when it is preceded by a long vowel. In Table 8, [V:sX] in the *h*-grade would correspond to [V̀:sX] in the glottal grade. Such an issue would not emerge with the analysis postulating an underlying /h/ before an /s/.

6. Conclusion

In this paper, I have shown how *s*-preaspiration, which is automatic and non-neutralizing thus more phonetic, interacts with Vowel Deletion, a phonological process and Laryngeal Alternation, a morphophonological rule. Their complex interactions defy either simple postulation of underlying /h/ before every instance of /s/, or integrating phonetic information in the underlying forms.

It is curious why only /s/ has such peculiarity of triggering pre-aspiration. Especially, it stands out among other obstruents which do not trigger preaspiration and where /hC/ and /C/ are contrastive. Could it be the case that it is related to the fact that /s/ is the only obstruent which is not realized as voiced between vowels (§3.2)? That is, /s/ is preaspirated to avoid voicing between segments that could trigger voicing? But then, why does this preaspiration behave as if it is phonologically there?

References

Bennett, Ryan, Robert Henderson, and Meg Harvey. 2023. "Vowel Deletion as Grammatically Controlled Overlap in Uspanteko." *Language*, 99(3): 399–456.

Boersma, Paul. 2011. "A Programme for Bidirectional Phonology and Phonetics and their Acquisition and Evolution." Anton Benz and Jason Mattausch, eds., *Bidirectional Optimality Theory*, 33–72. Amsterdam: John Benjamins.

Cook, William. 1979. A Grammar of North Carolina Cherokee. Ph.D. dissertation, Yale University.

Feeling, Durbin. 1975. *Cherokee-English Dictionary*. William Pulte, ed. Tahlequah, OK: Cherokee Nation of Oklahoma.

Flemming, Edward. 1996. "Laryngeal Metathesis and Vowel Deletion in Cherokee." In Munro, Pamela, ed. 1996. *Cherokee Papers from UCLA*. U. of California Occasional Papers in Linguistics 16, 23–44. Los Angeles: U. of California, Los Angeles.

Flemming, Edward. 2001. "Scalar and Categorical Phenomena in a Unified Model of Phonetics and Phonology." *Phonology*, 18: 7–44.

Hayes, Bruce and Donca Steriade. 2004. "Introduction: The Phonetic Bases of Phonological Markedness." In Hayes, Bruce, Robert Kirchner, and Donca Steriade, eds. 2004. *Phonetically Based Phonology*. Cambridge: Cambridge University Press.

Holmes, Ruth and Betty Smith. 1976. *Beginning Cherokee*. Norman: U. of Oklahoma.

Keating, Patricia. 1990. "Phonetic Representations in a Generative Grammar." *Journal of Phonetics*, 18: 321–334.

Kingston, John and Randy Diehl. 1994. "Phonetic Knowledge." *Language*, 70: 419–454.

Kiparsky, Paul. 1968. *How Abstract is Phonology?* Bloomington: Indiana University Linguistics Club.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lionnet, Florian. 2017. "A Theory of Subfeatural Representations: The Case of Rounding Harmony in Laal." *Phonology*, 34: 523–564.

Munro, Pamela. 1996. "The Cherokee Laryngeal Alternation Rule. In Munro, Pamela, ed. 1996." *Cherokee Papers from UCLA*. U. of California Occasional Papers in Linguistics 16, 45–60. Los Angeles: U. of California, Los Angeles.

Pierrehumbert, Janet. 2016. "Phonological Representation: Beyond Abstract Versus Episodic." *Annual Review of Linguistics*, 2: 33–52.

Scancarelli, Janine. 1987. Grammatical Relations and Verb Agreement in Cherokee. Ph.D. dissertation, U. of California. Los Angeles.

Steriade, Donca. 2000. "Paradigm Uniformity and the Phonetics-phonology Boundary." In Michael Broe and Janet Pierrehumbert, eds., *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 313–334. Cambridge, UK: Cambridge University Press.

Stevens, Kenneth N. 1971. "Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations." *Journal of the Acoustical Society of America*, 50: 1180–92.

Uchihara, Hiroto. 2013. Tone and Accent in Oklahoma Cherokee. Ph.D. dissertation, University at Buffalo, SUNY.

Uchihara, Hiroto. 2016. *Tone & Accent in Oklahoma Cherokee*. Oxford Studies of Endangered Languages. Oxford University Press.

Zsiga, Elizabeth. 2000. "Phonetic Alignment Constraints: Consonant Overlap and Palatalization in English and Russian." *Journal of Phonetics*, 28: 69–102.

Zsiga, Elizabeth. 2021. *The Phonology/phonetics Interface*. Cambridge, MA: MIT Press.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Acoustic Properties of Glottalized Consonants:
A Pilot Study on the Ie Dialect of the Northern Ryukyuan Languages*

AOI, Hayato

This study explores the phonetic characteristics of glottalized consonants in Ryukyuan languages, with a particular focus on the Ie dialect of Northern Ryukyuan. Despite previous research, the detailed phonetic aspects of these consonants remain largely unexplored. My acoustic phonetic analysis reveals distinctive features, such as short durations and abrupt onsets, that differentiate glottalized from non-glottalized resonants. However, contrary to traditional expectations, voiceless unaspirated stops show no signs of laryngeal constriction. The findings align with historical processes and phonological patterns, shedding new light on the phonological structure of Ryukyuan languages.

1. Introduction

This paper discusses the issues related to the phonetic properties of the glottalized consonants in the Ryukyuan languages. It is said that the glottalized consonants are noticeable phonetic characteristics of the Northern Ryukyuan languages, as in Uemura (1993, 2000). Although studies have been made on the glottalized consonants in Ryukyuan,

---

Keywords: glottalized resonants, glottalized stops, acoustic analysis, the Ryukyuan languages, the Ie dialect of Okinawan

DOI: https://doi.org/10.15026/0002000626

* This paper is a product of the following project: the ILCAA joint research project "Phonetic typology from cross-linguistic perspective (PhonTyp)" (2021–2023). This study is funded by a grant from JSPS (17H06666, 20K13001).

their phonetic details are still unknown. In this paper, I report the results of an acoustic phonetic study conducted on the Ie dialect of Okinawan (henceforth Ie), the Northern Ryukyuan languages.

The Ryukyuan phonetics and phonology has recognized glottal contrasts in the following two classes: sonorants (nasals, liquids, and semi-vowels) and stops (plosives and affricates). From the result, it was found that glottalized and non-glottalized resonants can be distinguished by a short duration and/or an abrupt onset. On the other hand, there are no acoustic traits expected from traditional descriptions that voiceless unaspirated stops are articulated with laryngeal constriction. The acoustic phonetic features of the stops are consistent with diachronic sources and synchronic distributional patterns.

2. Glottalized Consonants in the World's Languages

Almost all languages have the laryngeal contrast of voiced and voiceless. Other than these, some languages have another type of laryngeal setting, with the larynx involving a tighter constriction of the vocal folds. Consonants with these characteristics are known as glottalized, which can also be called laryngealized.

Maddieson (2005, 2013) classifies glottalized consonants into three different subtypes: ejectives, implosives, and glottalized resonants. An ejective is one of the sounds with a glottalic airstream. It typically involves a complete closure of the vocal folds followed by an upward movement of the larynx. Ejectives are found in Caucasus, Africa, and America. An implosive is another glottalic airstream sound but is made with an ingressive airflow. Some languages of West Africa and Asia contain implosives.

The third group is glottalized resonants. Resonant consonants are usually voiced in the normal manner, but they can be produced with a tighter constriction of the vocal folds which interrupts or modifies the normal voicing. Glottalized resonants are produced neither with a raising movement of the larynx nor with a lowering of the larynx. Glottalized resonants are not so frequent cross-linguistically, most found in North America, such as Nez Perce, Shuswap, Kwak'wala, Nuuchahnulth, Tsimshian, Wappo, Klamath, Kutenai, and Haida (Mithun 1999: 19).

In addition to the three subtypes of glottalized consonants above, there are some consonants with glottal constriction. One of the examples is commonly cited by the Korean voiceless 'fortis' stops. A 'fortis' stop of Korean has more constricted glottis than the other two series of voiceless stop: 'lenis' and aspirated. In Maddieson (2005), this type of glottalized consonants is considered in ejectives. However, the Korean 'fortis' stop is different from an ejective in that it lacks the raising movement of the larynx (Maddieson 2005: 34). If we regard the Korean types of glottalized consonants as the other subtype distinguished from an ejective, i.e. glottalized obstruent, we can classify four types of glottalized consonants as shown in Table 1.

We can figure out that a glottalized consonant is relatively a rare sound cross-linguistically from Maddieson (2013). According to the World Atlas of Language Structures Online, 158

Table 1 Classification of glottalized consonants

Class	Raise the larynx	Lower the larynx	No movement of the larynx
Obstruents	(A) Ejectives Ex. /p’ t’ k’/	(B) Implosives Ex. /ɓ ɗ ʄ/	(C) Glottalized obstruents Ex. /pˀ tˀ kˀ/
Resonants	-	-	(D) Glottalized resonants Ex. /wˀ nˀ rˀ/

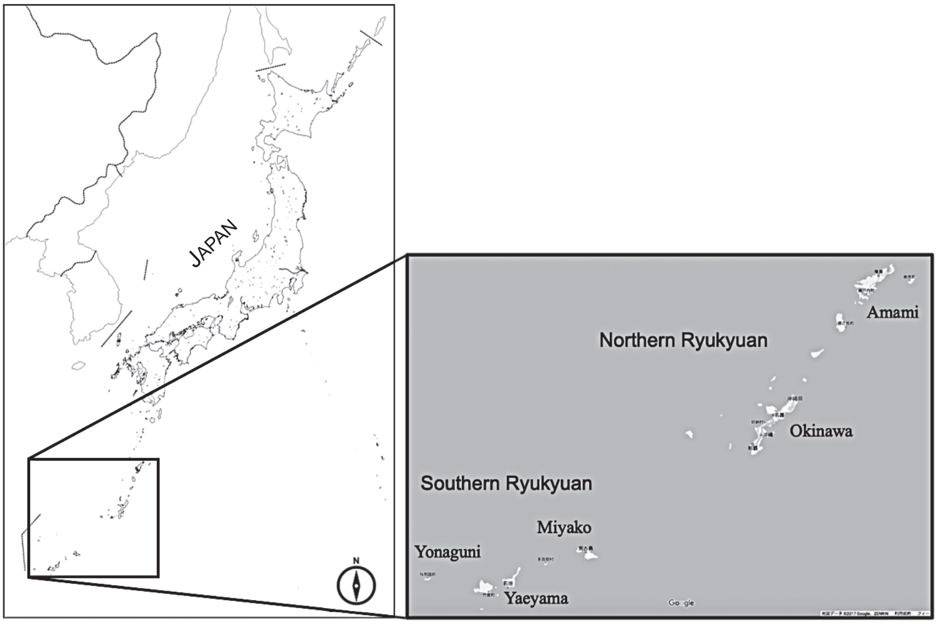
languages (27.8%) of 567 languages have glottalized consonants. Of these, 127 languages (22.3%) have only glottalized stops and 27 (4.8%) have both stops and resonants, while 4 languages (0.7%) have only glottalized resonants.

As Maddieson (2005, 2013) pointed out, glottalized consonants are sporadically distributed in various regions of the world. Ejectives are found densely in the Americas, the east and southern areas of Africa, and the Caucasus. Implosives are especially found in Africa, except for the northernmost area, and in South-East Asia. Glottalized resonants are mostly found in the Americas. According to the WALS, the number of languages with glottalized resonants is 31, out of that, 22 are found in America. Other than the Americas, they are distributed sparsely in South-East Asia, Austronesia, and Africa. Maddieson (2005, 2013) made an important statement on the distribution of glottalized resonants and ejectives. In the Americas, 20 languages out of 57 with ejectives also have glottalized resonants. We might think from the fact that any particular linguistic dependence between the occurrence of ejectives and glottalized resonants. However, after pointing out that only two languages with ejectives have glottalized resonants in the African-area languages, and no language has in the Caucasus, Maddieson (2005: 35) goes on to say: “The areal restriction suggests that the association between glottalized resonants and ejectives might best be viewed as a result of overlapping patterns of spread in a single area.”

Since there is a continuum of laryngeal settings, the phonetic details of glottalization vary from language to language. Ladefoged and Maddieson (1996: 55) proposed the following three aspects to describe stop sounds with creaky voices (i.e. laryngealized or glottalized stops) (1).

(1) Aspects concerning the phonetic details of glottalized stops
  a. The degree of glottal constriction:
    modified voicing (eg. [d]) ~ a simultaneously produced glottal stop (e.g. [d͡ʔ])
  b. The timing of oral and laryngeal movements:
    a single consonant (e.g. [d̰]) ~ a sequence (e.g. [ʔd] or [dʔ])
  c. Oropharyngeal expansion:
    modal voiced stops (e.g. [d]) ~ implosives (e.g. [ɗ])

Ladefoged and Maddieson (1996) also described laryngealized nasals, and commented from the viewpoints (1) above, except for (1c) which does not matter for resonant sounds. The laryngeal constriction gesture may precede (e.g. !Xoo,), follow (e.g. Montana Salish,), or appear centrally at the nasal (e.g. Kwakw’ala,), and the glottis may entirely be closed or



Map 1 Ryukyuan Languages

narrowed producing creaky voicing. Laryngealized laterals are observed in some languages, such as Tiddim Chin, Nez Perce, Chemehuevi, Haida, Sedang, and Klamath (Ladefoged and Maddieson 1996: 200). In Montana Sailish, laryngealized laterals may be produced in a creaky voice, which affects the neighboring vowels.

3. Glottalised Consonants in the Ryukyuan Languages

Ryukyuan is a sister language of Japanese, and Ryukyuan and Japanese together form the Japonic language family (Map 1). Ryukyuan languages are spoken in the Ryukyu archipelago, the southmost islands of Japan. Ryukyuan can be divided into two major groups, Northern Ryukyuan and Southern Ryukyuan. Northern Ryukyuan consists of Amami and Okinawa, and Southern Ryukyuan consists of Miyako, Yaeyama, and Yonaguni.

It has been recognized that glottalized consonants are one of the phonetic characteristics of the Northern Ryukyuan languages. We can recognize four subclasses of glottalized consonants in the Northern Ryukyuan: stops, approximants, nasals, and a liquid. In which classes glottalized oppositions are found varies from dialect to dialect (Uemura 1993). Table 2 summarizes which classes the glottalized consonants are observed in each dialect based on Uemura (1993: 16). For example, in the Yoron dialect of Amami and the Shuri dialect of Okinawa, glottalization is found only in approximants. On the contrary, the Kikai dialect of Amami has glottalized stops but not any glottalized resonants.

Some languages have both glottalized stops and resonants in Northern Ryukyuan. In the Kakeroma dialect of Amami and most of the South Amami-Oshima dialects, glottalization is found in stops and approximants. Naze (Amami), Tokunoshima (Amami),

Table 2 Glottalized consonants in Northern Ryukyuan (based on Uemura 1993: 16). A check in the table indicates that the language has the type of glottalized consonant in question.

Language	Dialect	Stops	Approximants	Nasals	Liquid
Amami	Kikai	✓	-	-	-
	Naze	✓	✓	✓	-
	Kakeroma	✓	✓	-	-
	Tokunoshima	✓	✓	✓	-
	Okinoerabu	✓	✓	✓	-
	Yoron	-	✓	-	-
Okinawa	Ie	✓	✓	✓	✓
	Nakijin	✓	✓	✓	-
	Shuri	-	✓	-	-

Okinoerabu (Amami), and Nakijin (Okinawa) have glottalized consonants in stops, approximants, and nasals, and most of the Amami and the northern Okinawa dialects belong to this type. The only dialect that has a glottalized liquid is the Ie dialect of Okinawa.

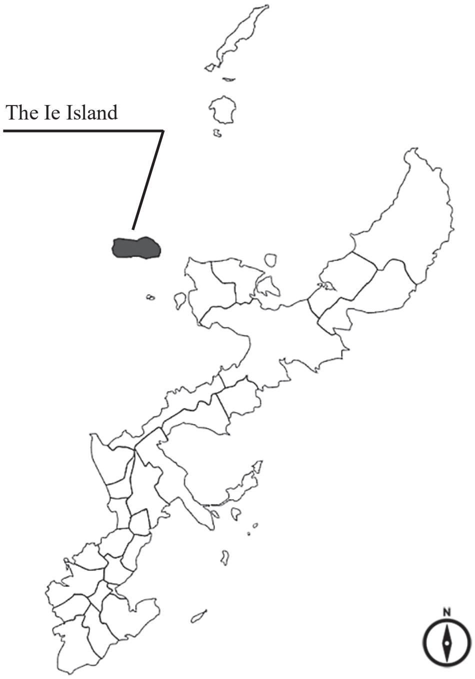
From Table 2, we realize that there is an implicational hierarchy related to the glottalized consonants in Northern Ryukyuan. That is, if the language has glottalized nasals, it also has glottalized approximants and stops, such as Naze, Tokunoshima, Okinoerabu, and Nakijin. If the language has a glottalized liquid, to which only Ie is applicable, the other types of glottalized consonants are also found. On the other hand, there seems to be no implicational relation between glottalized stops and glottalized approximants. All three logically possible types can be seen: the language which has only glottalized stops (e.g. Kikai) or glottalized approximants (e.g. Yoron and Naha), or both (e.g. Kakeroma).

Previous studies, such as Uemura (2000: 15), have described the glottalized consonants of Northern Ryukyuan as "consonants in which the larynx is tense." But what precisely is meant by the word 'tense' is not clear. In other words, important questions about the phonetic details of glottalized consonants, as listed in (1), have not been answered. Whether there are any phonetic differences between the subtypes (stops, approximants, nasals, and a liquid) of glottalized consonants is also unknown. In Section 4, the result of a pilot acoustic study conducted on the glottalized consonants of the Ie dialect will be shown.

4. The Ie Dialect

The Ie dialect belongs to the Okinawan, in the group of Northern Ryukyuan languages, of the Ryukyuan branch of the Japonic language family. The Ie Island, where the Ie dialect is spoken, is a small island located in the northwestern part of the main Okinawan islands (Map 2).

In Section 4.1, I briefly introduce the consonant inventory of the Ie dialect based on Oshio (2009) and my field research. Sections 4.2 and 4.3 describe the acoustic properties of the glottalized stops and resonants respectively. Data were gathered from the field trip



Map 2 The Main Okinawan Islands

on the Ie Island in June 2017 and February 2018. The participants in this survey were two male speakers from the Agarie-Ue region of the Ie village, who were born in 1930 and 1942. The recording sessions took the form of having the participants repeat dialect words, from a vocabulary list, in isolation. Recordings were made using a Marantz PMD661 solid state recorder and an Audio-technica PRO8HE microphone. A sample rate of 44.1 kHz and a bit depth of 16-bit was used, and the recordings were saved as WAVE files. Acoustic analysis was performed using Praat (ver. 6.0.40).

4.1 Consonant Inventory

The consonant inventory of the Ie dialect is shown in Table 3. In Table 3, a glottalized consonant is indicated with [ˀ] on its upper right. Oshio recognized 10 glottalized consonants in Ie, which of five are glottalized stops (plosives and affricates) and the rest are glottalized resonants (nasals, a liquid, and approximants).

All the glottalized stops have two other counterparts of their own, namely voiced and aspirated. The three-way contrastive stops can be distinguished by two phonological features out of three, [voiced], [aspirated], and [glottalized]. In other words, one of those three features might be redundant. In traditional descriptions, including Oshio (2009), it is said that [voiced] and [glottalized] are distinctive features, and [aspirated] is redundant. Considering the phonological behaviors, however, it seems to be valid to interpret [glottalized] as redundant. As I will discuss in Section 5.1, glottalized stops in Ie show more unmarked behavior than aspirated: glottalized stops can be distributed in both word-

Table 3 The consonant inventory of the Ie Dialect (cf. Oshio 2009). Some phonemic symbols have been changed.

	Labial			Coronal			Palatal			Dorsal			
Plosive	pˀ	pʰ	b	tˀ	tʰ	d				kˀ	kʰ	ɡ	
Affricate				cˀ [t͡sˀ]	cʰ [t͡sʰ]	z [d͡z]	cjˀ [t͡ɕˀ]	cjʰ [t͡ɕʰ]	zj [d͡ʑ]				
Fricative					s			sj [ɕ]					h
Nasal	mˀ		m	nˀ		n							
Liquid				lˀ		l							
Approximant	wˀ		w				jˀ		j				

initial and word-medial positions, while aspirated can be distributed only word-initially.

The resonant classes have two distinctions, glottalized and non-glottalized (plain). While the same notion of 'glottalized' is used for stops and resonants in the previous studies, we must discuss carefully whether it is valid that those consonants can be classified as [glottalized] across two classes. It is well known that glottalizing processes are different between the glottalized stops and the glottalized resonants in the Ryukyuan languages (Uemura 1993, 2000, Karimata 2007), thus these differences may reflect in their phonological behaviors. To conclude the issue, as mentioned in Niinaga *et al.* (2011: 295), it is necessary to examine their phonological behaviors. I will take up this point again in Section 5.2.

As mentioned in Section 3, little is known about the phonetic details of the glottalized consonants in the Ryukyuan languages. Oshio (2009: 6), the introductory notes in the dictionary of the Ie dialect, just describes its glottalized consonants as 'the sound articulated with the constriction of the glottis.' In Uemura (2000: 15), overview of the Ryukyuan phonetics, it is only said that the 'glottal constriction' is the phonetic characteristics of the glottalized consonants in the Northern Ryukyuan, and articulatory and acoustic details of the 'glottal constriction' is under described. In the following two sections, I will describe the acoustic phonetic details of the glottalized consonants based on the survey on the Ie dialect. In the beginning, in Section 4.2, I will examine the resonants in the Ie dialect, and then, in Section 4.3, focus my attention on the stops.

4.2 Acoustic Properties of the Glottalized Resonants

For the analysis of the glottalized resonants in the Ie dialect, I referred to the findings of Niinaga *et al.* (2011). Niinaga *et al.* (2011) revealed that the glottalized resonants in the Yuwan dialect of Amami had the following two acoustic properties listed in (2), compared with the non-glottalized resonants.

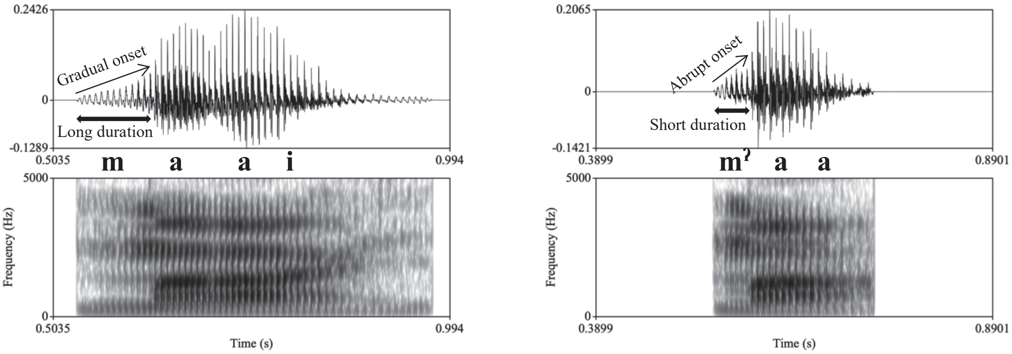


Figure 1 [Left] ***maai*** **'surround'; [right]** ***mˀaa*** **'here'**

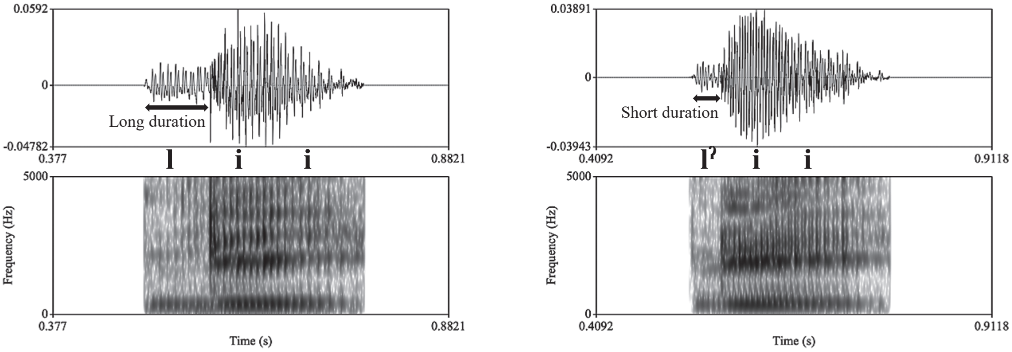


Figure 2 [Left] ***lii*** **'benefit'; [right]** ***lˀii*** **'you (pl.)'**

(2) Acoustic properties of the glottalized resonants (Niinaga *et al.* 2011)
  a. Abrupt onset: The glottalized resonant has a greater amplitude on its onset.
  b. Shorter duration: The duration of the glottalized resonant is shorter.

As we will see below, the same properties of (2) can be found in the Ie dialect. Figure 1 tells us the acoustic differences between the glottalized nasal [mˀ] and the non-glottalized nasal [m].

Comparing the two waveforms, what we initially realize is the difference in duration. That is, in contrast to the non-glottalized nasal, the duration of the glottalized nasal is shorter. In the case of the examples of Figure 1, the duration of a non-glottalized is 74 ms, and that of a glottalized is 33 ms. Also, a difference exists between the beginning of the two sounds. Compared with a non-glottalized, a glottalized nasal begins with a little larger amplitude. On the other hand, the non-glottalized nasal begin with a smaller amplitude and grows gradually. In other words, the glottalized nasal has an abrupt onset, in contrast to the non-glottalized nasal.

Waveforms and spectrograms of the glottalized liquid [lˀ] and the non-glottalized liquid [l] appear in Figure 2. We see from the figure, the distinction on abrupt onset is not clear between a glottalized liquid and a non-glottalized liquid. There is no significant difference in the amplitude of the initial part. In addition, there is no amplification toward

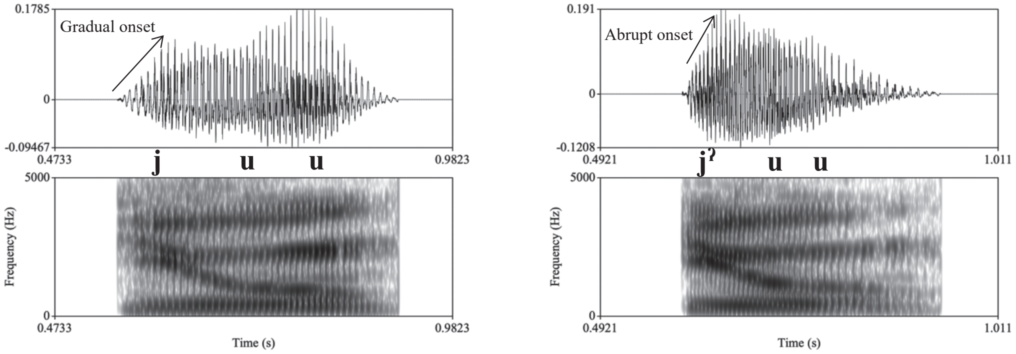


Figure 3 [Left] *juu* 'hot water'; [right] *jˀuu* 'fish'

the vowel in either case. However, by paying attention to the duration, we can see a clear difference between the two. The duration of non-glottalized liquid is relatively long, while that of glottalized one is relatively short.

Figure 3 shows the acoustic properties of a glottalized approximant [jˀ] and a non-glottalized approximant [j]. Approximants are more difficult to measure in duration than nasals and liquids. Still, if we divide an approximant and a following vowel by focusing on the formant transition, it appears that the duration of the steady state part of the formant is comparatively longer for a non-glottalized. When we focus on the onset, on the contrary, we can observe a more vivid difference between the two. The amplitude of the onset of the non-glottalized approximant is smaller, and the amplitude gradually increases toward the vowel. On the other hand, in the glottalized approximant, the amplitude of the onset is relatively large, and the amplitude increases rapidly toward the vowel.

It was found from the result that the glottalized resonants in the Ie dialect have the same acoustic properties as the Yuwan dialect of Amami, reported by Niinaga *et al.* (2011). The acoustic properties of shorter duration and abrupt onset may be said to be general properties of glottalized resonants. Because they have been identified in other languages as well. For example, in Niinaga *et al.* (2011), we reported that similar characteristics were observed in the glottalized resonants not only of the Yuwan dialect of Amami but also of the Khoisan languages, which are of a completely different language family from the Ryukyuan languages.

4.3 Acoustic Properties of the Glottalized Stops

While stops are characterized by their low amplitude, the criteria for observing resonants, as listed in (2), are primarily focused on temporal aspects. This focus on temporal characteristics in the analysis of glottalized resonants means that these criteria are not suitable for analyzing stops. The abrupt onset and shorter duration, key temporal features listed in (2), cannot be directly applied to the analysis of stops. Thus, I will examine the following vowel instead. If the consonant is articulated with a glottal constriction, we can infer that the following vowel will show (a) relatively high pitch and (b) creaky voicing (Ladefoged and Maddieson 1996: 53–57; 74). I will not take up (a) in detail

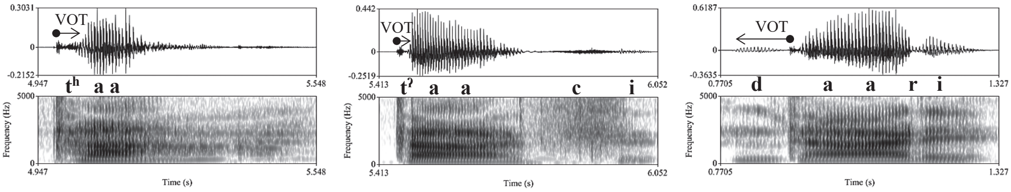


Figure 4 [Left] ***thaa*** **'who'; [center]** ***taatsi*** **'two'; [right]** ***daari*** **'neighborhood'**

this time. The Ie dialect is a tone language, and we need to take into account the tone type to choose test words. It is difficult, however, to pick up meaningful words belonging to the same tone type.

In addition to (b), I will measure (c) VOT (Voice Onset Time), the duration between the release of the stop and the onset of voicing. In Ryukyuan phonetics, a distinctive pattern has been observed regarding glottalization and aspiration. Specifically, glottalized consonants are characterized by their lack of aspiration, contrasting with non-glottalized consonants, which are typically aspirated. This indicates a clear inverse relationship between glottalization and aspiration in the sound system. Thus, it seems quite probable that the glottalized and non-glottalized are acoustically distinguished by VOT.

We will now examine the acoustic properties of glottalized stops more closely. Figure 4 shows the waveforms and spectrograms of three dental stops, aspirated [tʰ], glottalized [tˀ], and voiced [d].

We first paid attention to VOT. Those three kinds of stops distinguished each other by VOT. That is, in the left illustration, an aspirated stop [tʰ], the voice onset delays the release of the stop (VOT = 55 ms). On the contrary, in the right illustration, a voiced stop [d], the voice onset precedes the release of the stop (VOT = −108 ms). And in the central illustration, a glottalized stop [tˀ], vocal cords' vibration starts soon after the stop release (VOT = 36 ms).

Our next focus is on the pulse of the following vowel. Stops with tensed larynx can influence the vowel quality of the following vowel. It is hypothesized that the voice quality of the vowel is similar to a laryngealized or creaky voice after the glottalized stop. If so, the pulse of the vowel following a glottalized stop is expected to be irregular, while it is regular when following a non-glottalized stop. However, there is no significant difference in the pulse between the three contrastive stops. Figure 5 shows an enlarged

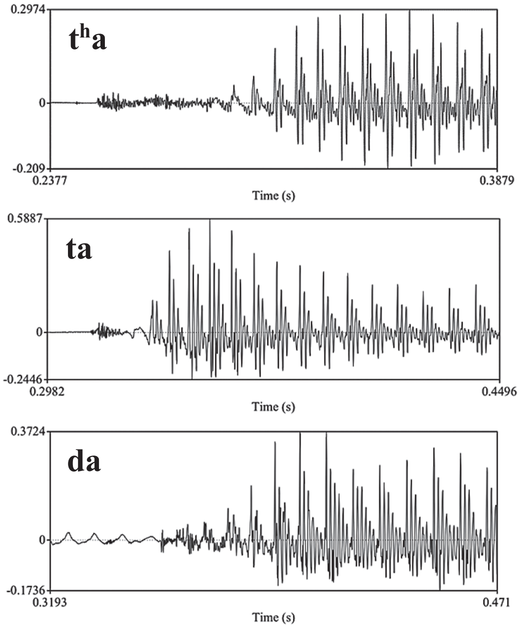


Figure 5 Pulses of vowels after stops

view of the beginning of each of the three stops in Figure 4. We see from the figure that regular pulses were observed no matter which kind of stop was preceded by the vowel.

Laryngealized or creaky voice can be indicated by the negative value of the difference in amplitude between the first and the second harmonics because creaky voicing is produced with a smaller open quotient with the vocal folds remaining closed for a longer time. However, there was no clear difference between the three kinds of stops in this respect either (Figure 6). That is, vowels after glottalized stops do not have negative H1-H2 values, and it is the same as vowels after aspirated or voiced stops.

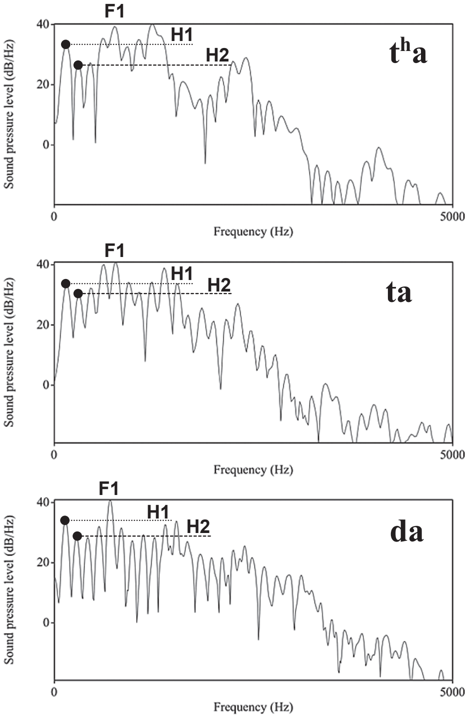


Figure 6 Spectral slices of vowels after stops

4.4 Summary of the Results

In comparison with the non-glottalized nasals, the glottalized nasals have a shorter duration and an abrupt onset. At least one of those characteristics can also be observed with the other resonant consonants, liquids, and approximants. From those acoustic observation, we can infer that the articulation of the glottalized resonants as the resonants with the glottal stop at the beginning like [ʔm̲]. It is consistent with the auditory impression of the glottalized sounds as well.

In light of the observations, there is no acoustic evidence that the voiceless unaspirated stops in Ie are glottalized. When we focus on the voice quality of the following vowel, the distinction between the three was not observed. There was no acoustic trait that the vowels after the glottalized stops were pronounced with creaky voicing. This is not the result expected from traditional descriptions that voiceless unaspirated stops are pronounced with laryngeal constriction. On the other hand, focusing on the VOT, a three-way distinction of stops was observed. This result is consistent with the observation in traditional descriptions that glottalized stops are pronounced as voiceless unaspirated, which have been regarded as a redundant phonetic feature. However, the result of the survey leads us to the conclusion that the stops of the Ie dialect are acoustically distinguished into three series: aspirated, voiceless, and voiced. As we will find in the next section, these acoustic interpretations are consistent with the synchronic and diachronic phonological patterns in the Ie dialect.

5. Discussion

In the field of Ryukyuan phonetics, the notion of a ‘glottalized’ class encompassing both stops and resonants has been a prevailing assumption (Uemura 1993, 2000, Karimata 1999, 2009). However, the acoustic observations presented in this report lead to a reconsideration of these established categorizations in Ryukyuan phonetics. Based on our acoustic findings, it has become apparent that a reevaluation of the conventional assumptions regarding the 'glottalized' class, which includes both stops and resonants, is necessary. As this paper represents an initial report, the discussion will not delve into detailed arguments but will rather focus on identifying and organizing key points of consideration. In this section, I will point out that the 'glottalized' stops are different from the glottalized resonants in the Northern Ryukyuan, particularly in terms of their diachronic origins and synchronic distribution patterns.

5.1 Diachronic Origins of the Glottalized Consonants

It has been recognized that diachronic passes of glottalization differ in the stops from resonants (Uemura 1993: 16–17, Karimata 2009: 323–324). The glottalized resonant derives from a resonant preceded by a word-initial high vowel *i, *u, e.g. *uma > [ʔm]*a* 'horse' (Naze, Amami). Expanding on Karimata (2007: 31–32), the development of the glottalized resonant can be explained as in (3).

(3) Development of the glottalized resonants
  a. *uma > ʔuma: Insertion of a word-initial glottal stop
  b. ʔuma > ʔma: Loss of a high vowel
  c. ʔma > mˀa: Reinterpretation of a cluster of a glottal stop and a resonant consonant

In the Northern Ryukyan language with glottalized resonants, a glottal stop [ʔ] inserts before a word beginning with a vowel, e.g., *ami* [ʔami] 'rain' (Nakizin, Okinawa) (Uemura 1997: 333–334). When the vowel between a glottal stop and a resonant consonant was high, it was deleted (3b). Then glottal-resonant clusters have been reinterpreted as a single consonant, i.e., glottalized resonant (3c).

On the other hand, the 'glottalized' stop is derived from a word-initial stop followed by a high vowel *i, *u, e.g. *kumo > [kˀ]*umu* 'cloud' (Naze, Amami). According to Ohala (2011: 64), however, a high vowel may trigger an aspiration of a preceded voiceless stop, not glottalization nor laryngealization. He explained its mechanism as follows: a close constriction attenuates the rate of airflow exiting the vocal tract after a stop and thus delays the time when Po (oral pressure) is low enough to initiate voicing; Occasionally this longer VOT before close vowels leads to a sound change where aspiration becomes distinctive. For example, in the Bantu language Ikalanga, 'super-close' high vowel *i in the proto-language led to the aspiration of stop: *tima > *ts͡hima* 'well.'

In light of Ohala's explanation, the 'glottalization' of stops in the Ryukyuan languages seemed to be triggered by the opposed condition. On the mechanism of 'glottalization' of stops in the Ryukyuan languages, the hypothesis proposed by Uemura is well known. Uemura (1987, 1989, 1993) assumed that airflow causes changes in articulation and maintained that exceeds airflow can cause articulatory strength, i.e. fortition or '*humbari*' of Uemura's term. Supporting Uemura's view, Karimata (1999, 2009) argues the glottalization in the Ryukyuan languages. However, Uemura's hypothesis remains a matter to be discussed further. Uemura and Karimata have not provided any conclusive evidence that airflow can cause articulatory changes.

(4) shows the processes of the 'glottalization' of the Northern Ryukyuan stops, which is expanded Uemura's explanation.

(4) Development of the 'glottalized' stops (cf. Uemura 1997: 34–35)
  a. *kumu > ʔkumu: Epenthesis of a glottal closure before a stop
  b. ʔkumu > ʔk̰umu: laryngealization of a word-initial stop followed by a high vowel

I must point out the two key problems with the explanation in (4) above. First, Uemura does not detail why and how a glottal stop inserts. As mentioned in Gordon (2016: 162), it is well known that glottal epenthesis can attest before vowels in the initial position of prosodic domains. Thus, (3a) can be interpreted as a natural process but (4a) cannot.

Second, Uemura (1993) has described that a laryngealized articulation is fortified compared with non-laryngealized. However, phonetic details of the fortification in articulation are under description. Results of Section 4.3 do not suggest that stops in the Ie dialect are distinguished by glottalization.

It remains unanswered that the process and mechanism of the 'glottalization' of stops in the Northern Ryukyuan. Yet, at least, we can point out that the historical processes are different between the glottalized resonants and the 'glottalized' stops. Despite that, these two types of consonants have been treated as the same phonological class, i.e. the glottalized consonants, in previous studies. However, as Niinaga *et al.* (2011: 295) stated, it needs to be examined with caution whether the resonants and the stops can be classified into the same class labeled [glottalized]. As detailed in the next section, we can find that distributional patterns are different between the two. To put it briefly, the glottalized resonants behave markedly, but the 'glottalized' stops do unmarked.

5.2 Synchronic Distributional Patterns of the Glottalized Consonants

In previous research, this has been described as if a type of sound called a 'glottalized sound' that exists beyond sonorants and obstruents were identified. However, the acoustic facts provided in this report indicate that a class of 'glottalized sounds' that transverses sonorants and obstruents is not identifiable. To verify this, a detailed examination of phonological behavior, rather than a detailed observation of phonetic facts, is necessary.

Oshio (2009) has pointed out that glottalized consonants of the Ie dialect show an asymmetric behavior on distributions between resonants and obstruents. That is, in the sonorant class, the glottalized sounds show marked behavior. Glottalized resonants only appear on the word-initial position, but non-glottalized sounds attest either at the initial or medial of words.

On the other hand, while the ‘glottalized’ stops and affricates can appear on either the initial or medial position of words, the ‘non-glottalized’ sounds only observe at the beginning of words. According to this fact, in the obstruent class, we can interpret the ‘non-glottalized’, not ‘glottalized’, sounds as marked because they have constraints on distributions. So, in Ie, it is more valid that [aspirated] is a marked feature, and [glottalized] is not.

6. Conclusion

This study conducted a pilot acoustic-phonetic analysis of the glottalized consonants of the Northern Ryukyuan languages, focusing specifically on the Ie dialect. The study revealed that glottalized resonants have relatively shorter durations and abrupt onset compared to their non-glottalized counterparts. On the other hand, no conclusive evidence was found to suggest that glottalization plays a role in the ternary distinction of stops. The ternary distinction in stops could be phonetically interpreted as voiceless unaspirated, voiceless aspirated, and voiced, which aligns with the diachronic and synchronic phonological patterns of glottalized sounds. On the other hand, the processes and mechanisms for deaspiration in front of high vowels (or aspiration in front of non-high vowels) remain unclear, as existing explanations from prior research do not apply. Further investigation is required to address these questions.

References

Gordon, Matthew K. 2016. *Phonological Typology*. Oxford: Oxford University Press.

Karimata, Shigehisa. 1999.「音声から見た琉球諸方言」[The Ryukyuan Languages from Phonetic Aspects].『ことばの科学』[*Science of Language*], 9: 13–85.

Karimata, Shigehisa. 2009.「琉球語音韻変化の研究」[Studies of Phonological Changes in the Ryukyuan Languages].『ことばの科学』[*Sciences of Language*], 12: 275–354.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World’s Languages*, Blackwell.

Maddieson, Ian. 2005. “Glottalized Consonants.” *The World Atlas of Language Structures* (Martin Haspelmath *et al.*, eds.), Oxford: 33–34.

Maddieson, Ian. 2013. “Glottalized Consonants”, *The World Atlas of Language Structures Online* (Matthew S. Dryer and Martin Haspelmath, eds.). Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at https://wals.info/chapter/7, Accessed on 2023-09-05)

Mithun, Marianne. 1999. *The Languages of Native North America*, Cambridge: Cambridge University Press.

Niinaga, Yuto, Hayato Aoi, and Hirosi Nakagawa. 2011「喉頭化共鳴音の音響音声学的特徴：北琉球奄美大島湯湾方言の事例研究」[Acoustic Phonetic Features of Laryngealized Resonants: A Case Study on the Yuwan dialect of Amami, the Northern Ryukyuan Language].『コーパスに基づく言語学教育研究報告』[*Working Papers in Corpus-based Linguistics and Language Education*] 7: 285–299.

Ohala, John J. 2011. “Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance”, *Proceedings of the 17th International Congress of Phonetic Sciences* (Hong Kong, 17–21 August 2011).

Oshio, Mutsuko. 2009.『新版 沖縄伊江島方言辞典』[*Dictionary of the Ie Dialect, Okinawan, the Second Edition*]. Okinawa: Ie Village Board of Education.

Uemura, Yukio. 1987.「音韻変化はどのように引き起こされるか：英語の Great Vowel Shift についての考察」[How Phonological Changes are caused: A Case Study of the Great Vowel Shift in English].

『沖縄言語研究センター資料』[*Reports of Okinawa Center of Language Study*] No.67.

Uemura, Yukio. 1989.「音韻変化はどのように引き起こされるか (2)：琉球方言の場合」[How Phonological Changes are caused (2): A Case of the Ryukyuan Languages].『沖縄言語研究センター資料』[*Reports of Okinawa Center of Language Study*] No.79.

Uemura, Yukio. 1993.「琉球列島諸方言における喉頭化をめぐる子音の音韻的対立の諸相」[Aspects on the Phonological Oppositions on Laryngealization in the Ryukyuan Languages].『「日本語音声」琉球列島班研究成果報告書：琉球列島における音声の収集と研究 I』[*Working Papers by the Group of Ryukyuan Phoneticians: Collection and Analysis on the Phonetic Features of the Ryukyuan Languages I*], 14–26.

Uemura, Yukio. 1997.「琉球列島の言語（総説）」[Languages in the Ryukyu Archipelago (general remarks)].『日本列島の言語』[*Languages in the Japanese Islands*], Sanseedoo, Tokyo: 311–354.

Uemura, Yukio. 2000.「琉球音声学の概説」[Overview of the Ryukyuan Phonetics]. *Journal of Phonetic Society of Japan*, 4(1): 4–18.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Phonetics of Aspirated and Tense Fricatives in Jinghpaw and Burmese[*]

KURABE, Keita
LEE, Seunghun J.

Aspirated fricatives are typologically rare in the world's languages. Building upon characteristics of aspirated fricatives in Korean, this paper examines the phonetics of aspirated fricatives in Jinghpaw and Burmese. Our results show that the duration of aspiration in Jinghpaw and Burmese ranges from about 10 to 20% of the entire fricative duration, which is much shorter than the aspiration duration in Korean (with 50% of the fricative duration being produced with aspiration). These cross-linguistic results demonstrate that aspirated fricatives may not have identical aspiration duration. Our results also support the point mentioned in previous studies that the aspirated-tense fricative contrast in Burmese is disappearing in some speakers. Only half of the Burmese participants in this study maintain the contrast between aspirated and tense fricatives. The other half displays a merger of aspirated fricatives with tense fricatives.

Keywords: aspiration, fricative, Jinghpaw, Burmese, Tibeto-Burman

DOI: https://doi.org/10.15026/0002000627

[*] We thank Shigeto Kamano who assisted us with data collection and processing. This research is written as part of the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023). The research has been funded by a Kakenhi Grant (C) 16K02641. This work was also supported by JSPS KAKENHI Promotion of Joint International Research (Fostering Joint International Research (B)) Grant Number 21KK0005, and the Kakenhi (B) Grant 23H00622. Any errors remain our own.

1. Introduction

Fricatives have the most variety among the manners of articulation. Despite this diversity, the aspirated fricative is typologically uncommon. In Maddieson's 1984 study, only three languages were identified to exhibit /sʰ/, as mentioned on p.229 (see also Craioveanu 2013). Its typological rarity and instability are further discussed by Jacques (2011), who expanded the scope by cataloging around 20 additional languages that distinguish aspiration in their fricatives. This is also illustrated by the limited representation of aspirated fricatives in PHOIBLE 2.0, a cross-linguistic phonological database comprising 3,020 inventories found across 2,186 languages (Moran and McCloy 2019). The number of doculects with aspirated fricatives is quite marginal, as shown in (1).[1] For example, the limited representation of voiceless aspirated alveolar fricatives is in stark contrast to the voiceless alveolar fricative, which is particularly prevalent among fricatives, represented in 2,020 doculects in the database (67%).

(1) Representation of some fricatives in PHOIBLE 2.0

s	h	f	ʃ	x	ɸ	…	sʰ	ɕʰ	xʰ
2020	1703	1329	1104	576	153	…	26	7	5
67%	56%	44%	37%	19%	6%	…	1%	0%	0%

Aspirated segments also show uneven distribution across obstruents; they are particularly prevalent in plosives and affricates, but not so for fricatives. The voiceless aspirated velar plosive stands out, being represented in 605 doculects (20%). This can be contrasted with the marginality of voiceless aspirated fricatives, which are significantly less common. The most frequent aspirated fricatives are the voiceless aspirated alveolar fricatives, yet only 1% of doculects have the sound. Other aspirated fricatives are much rarer (see also Jacques 2011: 1519–20).

(2) Representation of some aspirated segments in PHOIBLE 2.0

kʰ	pʰ	tʰ	tʃʰ	tsʰ	t̪ʰ	…	sʰ	ɕʰ	xʰ
605	592	403	229	195	179	…	26	7	5
20%	20%	13%	8%	6%	6%	…	1%	0%	0%

The inventories of aspirated fricatives, as shown in (3), vary across languages. Burmese possesses a single aspirated fricative, which is in contrast with its unaspirated counterpart (Watkins 2001). Puxi is characterized by a larger inventory of aspirated fricatives (Sun

---

[1] The figures presented in (1) and (2) should be interpreted as approximations since the PHOIBLE 2.0 database includes multiple doculects (individual language documents or sources) for the same language. The figures in (1) and (2) represent the count of doculects and thus do not directly correspond to the actual number of distinct languages. Note also that the PHOIBLE 2.0 database compiles data from various source documents and tertiary databases, leading to varying degrees of phonetic detail.

2010). The Heqing dialect of Bai shows a more robust set of aspirated fricatives, exhibiting contrasts in labiodental, alveolar, alveolo-palatal, and velar fricatives (Xi and Li 1997). Languages with aspirated fricatives typically include the aspirated alveolar fricative. Jacques (2011: 1519) notes that when a language possesses only one aspirated fricative, it is always alveolar, a pattern that seems to be observed in unaspirated fricatives as well.

(3) Inventories of aspirated-unaspirated fricative contrasts

s
sʰ

Burmese

s	ʃ	ç
sʰ	ʃʰ	çʰ

Puxi

f	s	ɕ	x
fʰ	sʰ	ɕʰ	xʰ

Heqing Bai

Nearly all languages and doculects listed in PHOIBLE 2.0 that possess aspirated fricatives also have their unaspirated counterparts, including the tense *s'* in Korean and the ejective *s'* in Emberá-Catío. The phonetic typology from this database suggests that the presence of an aspirated fricative largely implies the presence of an unaspirated counterpart. This is also the case in languages surveyed by Jacques (2011: 1519). Thus, it is fair to say that the marginality of aspirated fricatives equals the marginality in the contrast between aspirated and unaspirated fricatives.

The marginality of the contrast would partly reflect the unmarked laryngeal specification for voiceless fricatives. Kingston (1990) and Stevens (1998) suggest that voiceless fricatives typically involve a larynx status where the glottis is spread. Stevens (1998) emphasizes that this glottal configuration is essential to generate the required pressure for producing such fricatives. Vaux (1998) further explores the laryngeal specifications of fricatives based on data from several languages with extensive laryngeal contrast systems. He shows that the unmarked state for voiceless fricatives is [+spread glottis] whereas that for voiced fricatives is [−spread glottis]. Thus, the marginality in the contrast between aspirated and unaspirated fricatives has a phonetic and phonological underpinning: It is difficult to form a laryngeal contrast in voiceless fricatives in terms of aspiration because they are intrinsically aspirated in many languages (cf. Carveth 2013: 147).

That being said, as noted by Vaux (1998), it is entirely feasible for languages to exhibit contrasts in [spread] values among voiceless fricatives allophonically (e.g., English) or phonemically (e.g., Burmese). Although typologically uncommon, aspirated fricatives are sporadically distributed in languages around the world. They are documented in several Oto-Manguean languages and isolated cases such as Ofo, Barbareño, Emberá-Catío, !Xũ, and Cross River Mbembe (Jacques 2011, Moran and McCloy 2019). A majority of languages featuring aspirated fricatives are found in Asia (Jacques 2011, Carveth 2013). Data from PHOIBLE 2.0 also point to a significant presence of this sound in Asian languages. It appears across multiple language families, such as Korean, Tai-Kadai (e.g., Shan), Hmong-Mien (Hmong Njua), and Sino-Tibetan (e.g., Burmese, Akha, Sgaw Karen, Khams Tibetan, Amdo Tibetan, Cone Tibetan, Baima, Tamang, Shumcho). More detailed data obtained from language descriptions reveal that the voiceless aspirated alveolar

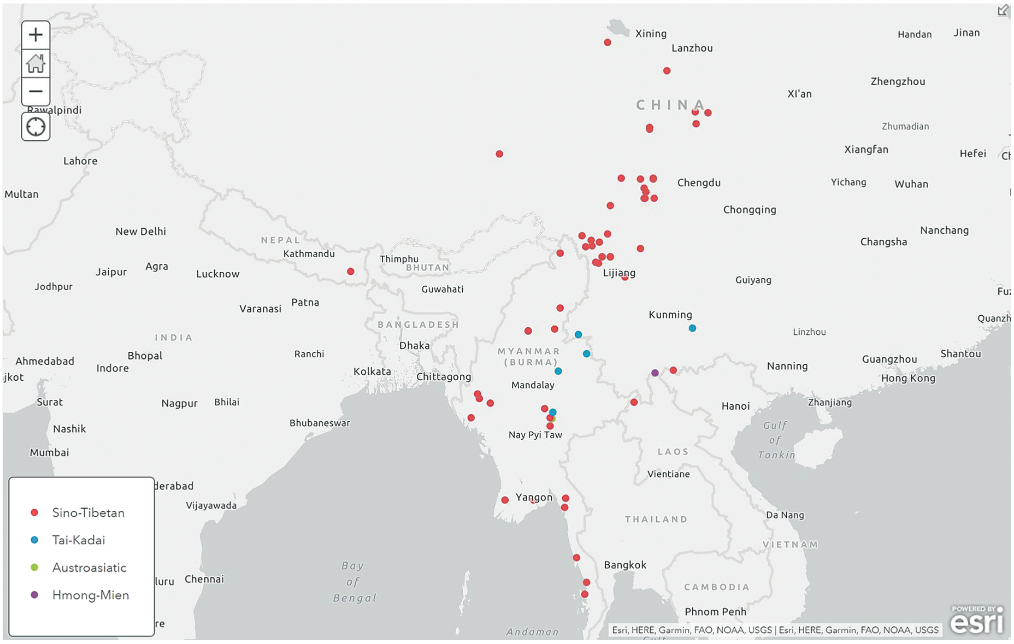


Figure 1 Distribution of the voiceless aspirated alveolar fricatives.

fricative is distributed in at least the areas shown in Figure 1. The concentration of this atypical sound across multiple language families implies its nature as an areal feature.

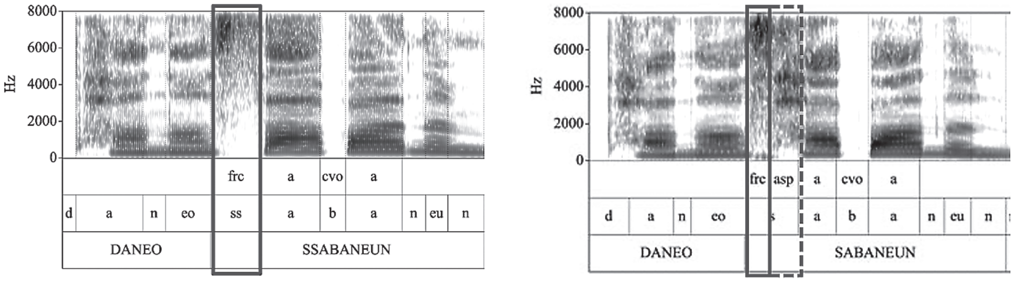
Languages of Southeast Asia and adjacent areas provide good samples of this cross-linguistically atypical phonetic feature. It is crucial to note that many of these languages exhibit phonemic contrasts between aspirated and unaspirated fricatives. However, little attention has been given to the phonetic nature of aspirated fricatives with a few exceptions.

In this paper, we explore the contrast in Jinghpaw and Burmese, the two Tibeto-Burman languages previously reported to exhibit the contrast. It has also been reported that Burmese is in the process of neutralizing this contrast (Okano 2011). Our investigation primarily revolves around three central questions: (a) What are the acoustic correlates of aspirated versus unaspirated fricatives? (b) Does Burmese neutralize the contrast between aspirated and unaspirated fricatives? (c) From a typological perspective, how do aspirated fricatives in Jinghpaw and Burmese compare and contrast with those in Korean and other languages?

2. Background: Aspirated and Tense Fricatives in Korean

This section will first discuss studies of Korean fricatives before reporting on the aspirated fricatives in Jinghpaw and Burmese because Korean has a robust phonetic contrast between the aspirated and tense fricatives.

Korean has two fricatives that are produced in the alveolar region: aspirated [s^h] and tense [s’]. This section provides an overview of the phonetic characteristics of these two



a. tense fricative **b. aspirated fricative**

Figure 2 An illustration of tense and aspirated fricatives in Korean. Solid square encloses frication noise, and dotted square encloses aspiration.

fricatives. In (4), minimal pairs of this contrast are shown.

(4) Korean fricatives contrast
a. aspirated: [sʰal] ‘skin’ [sʰada] ‘to buy’ [sʰi] ‘poem’
b. tense: [s’al] ‘rice’ [s’ada] ‘to wrap’ [s’i] ‘seed’

Cho *et al.* (2002) report that tense fricatives have a longer frication duration than aspirated fricatives, even though the fricative duration (frication + aspiration) is longer in aspirated fricatives. This paper mostly reports frication duration which only refers to the duration of the frication noise in a fricative. This difference is illustrated in Figure 2. The duration of a fricative includes laryngeal features such as aspiration in addition to the frication noise.

Perkins *et al.* (2023a) analyze data from 24 young Seoul Korean speakers who produced disyllabic C1V1CV words where C1 was either an aspirated fricative or a tense fricative before the [a] vowel in the V1 position. A machine learning model, the random forest (RF) model, was used to identify the difference between aspirated and tense fricatives. Data points were annotated for frication noise and aspiration for aspirated fricatives and only frication noise for tense fricatives. The Random Forest model was used to perform the learning algorithm where the relative importance of phonetic variables is examined when learning phonological categories (e.g., aspirated and tense fricatives). This method can produce model accuracy in addition to relative importance values for multiple factors, allowing researchers to identify which measures matter in a given contrast.

For the acoustic analysis, various acoustic measures were extracted. At the vowel onset following the fricatives, they obtained psychoacoustic roughness (Villegas *et al.* 2020) and spectral tilt measures (such as H1*–H2*, H1*–A1*, H1*–A2*, H1*–A3*, H2*–H4*). F0 at the vowel onset following fricatives was normalized within each speaker relative to that speaker’s median f0, which was then converted to cents. They also extracted the duration of frication, aspiration as well as the entire fricative.

During the RF analysis, models were created with all possible permutations of acoustic measures. When a model achieved overall accuracy of 95%, the model was deemed to be successful. A measure or group of measures is sufficient to learn the frication contrast if all models that contained those measures were successful. A measure or group of measures is necessary to learn the contrast if the only models that were successful contained those measures. Importance measures range from 0 to 1, quantifying the relative contribution of each measure to the overall model, and these importance measures allow insights into whether a measure is a likely acoustic correlate when learning a specific contrast.

The acoustic results in Perkins *et al.* (2023b) are as follows. Tense fricatives had longer frication duration than aspirated fricatives, echoing the finding of Cho *et al.* (2002). However, aspirated fricatives included aspiration duration, resulting in a similar total duration with tense fricatives. The RF model results indicated that frication duration is the most likely primary correlate, with total duration, spectral tilt, and psychoacoustic roughness as likely candidates for secondary roles. The only necessary measure for the fricative contrast was frication duration. Sufficient measures for the fricative contrast were (a) frication duration with total duration, which achieved 99.3% accuracy, and (b) frication duration with spectral tilt and roughness, which reached 97.1% accuracy. Frication duration was necessary but not sufficient on its own for the fricative contrast (89.6% accuracy on its own). Thus, either total duration or spectral tilt is required in addition to a secondary acoustic correlate to properly learn the contrast.

In the perception study that explored the perception of the fricative contrast, Ko (2023) utilized a gating method to test perceptual cues for identifying the fricative contrast. The results showed that listeners displayed sensitivity to the onset of the aspiration noise, and the results of the gating study suggest that speakers better identify aspirated fricatives once they hear the presence of aspiration noise. The duration of the frication noise was not explicitly discussed in Ko (2023), but it is possible that the participants in her study may have become sensitive to the shorter frication noise duration in aspirated fricatives as well.

In sum, Korean has a robust contrast between aspirated and tense fricatives based on the frication duration coupled with the fricative duration or spectral tilt (Perkins *et al.* 2023a), or the onset of aspiration (Ko 2023). The frication noise is shorter in aspirated fricatives, and the aspiration takes about 40–50 % of the fricative duration.

3. Fricatives in Jinghpaw and Burmese

Jinghpaw and Burmese are Tibeto-Burman languages mainly spoken in Myanmar and adjacent areas. Within Tibeto-Burman, Jinghpaw belongs to the Sal branch, related to the languages spoken in northern Myanmar and northeastern India, while Burmese belongs to the Lolo-Burmese branch, closely related to the languages of Myanmar and southwestern China. Both languages are syllable-tone languages. Jinghpaw has four distinct tones in its smooth syllables, whereas Burmese features three contrasting tones in its smooth syllables.

Both languages have a three-way laryngeal contrast in fricatives that are orthographically distinct: voiceless aspirated, voiceless tense, and voiced. The contrast between voiced and voiceless fricatives is robust, but the contrast between aspirated and

Table 1 Jinghpaw stimuli.

Stimuli (orthography)	Meaning	IPA	Stimuli (orthography)	Meaning	IPA
sa	go	sʰa	tsa	hundred	sa
si	fruit	sʰì	tsi	drag	sì
su	be awake	sʰù	tsu	be stale	sù
sa	rest	sʰáʔ	tsa	fermented rice	sáʔ
sit	move	sʰìt	tsit	be green	sìt
sut	wealth	sʰùt	tsut	draw a sword	sùt
masa	method	məsʰa	matsa	swear	məsa
kasi	model	kəsʰi	matsi	yeast	məsì
lasu	funeral	ləsʰu	atsu	disembodied spirit	ʔəsù
nsa	not go	nsʰa	ntsan	not be far	nsan
nsi	not bear fruit	ńsʰî	ntsi	not cure	ńsî
nsu	not be awake	ńsʰû	ntsu	not be stale	ńsû

tense fricatives is undergoing changes. In Burmese aspirated fricatives are being merged to tense fricatives, whereas the contrast between aspirated and tense fricatives is relatively salient in Jinghpaw.

3.1 Jinghpaw

3.1.1 Speakers

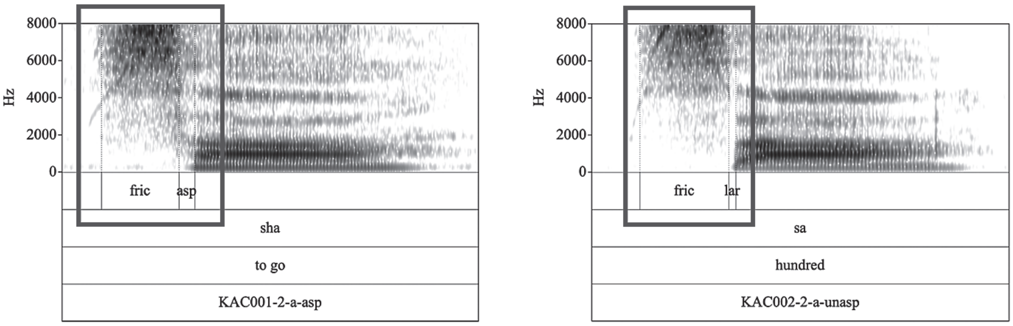
The Jinghpaw language has 630,000 to 940,000 speakers in and around the Kachin State, the northernmost state in Myanmar (Bradley 1996, Lewis *et al.* 2014). The data were collected from two female speakers (aged 27 and 33 years) in an audio recording session. Both speakers are from Kachin State. They were living outside of Myanmar at the time of the recording. Both are fluent in Burmese, as are most Jinghpaw speakers in Myanmar.

3.1.2 Methods

During the recording session, the speakers read a word list consisting of sets of minimal or near-minimal pairs, each differing in the aspect of aspiration. Stimuli items varying in phonological conditions resulted in 24 target items in Table 1: three vowels ([a], [i], [u]), two fricatives (aspirated and tense), two syllable types (with or without a coda consonant), and two positions (word-initial, word-medial). The final number of tokens was 240 (24 items x 5 repetitions x 2 speakers).

The target words were embedded in a frame sentence *ndai ga si* _______ *hpe hti u*, which rendered in English as "Please read this word _______." The experiment was conducted by presenting a list of stimuli in their orthography, as detailed in Table 1. The recording was carried out with a RØDE NTG2 shotgun condenser microphone attached to a ZOOM H5 linear PCM recorder. The stimuli items were recorded at 44,100 Hz in 16-bit in a frame sentence.

Acoustic recordings were processed using Praat (Boersma and Weenink 2023), and a series of Praat scripts were used to manually annotate frication noise and aspiration. Figure



a. aspirated fricative **b. tense fricative**

Figure 3 An illustration of aspirated and tense fricatives in Jinghpaw; frication noise (fric), aspiration (asp), and laryngealized part (lar) are annotated.

3 shows how we annotated the two types of fricatives. In Figure 3a, aspirated fricatives show the frication noise with energy concentration in the region above 4000 Hz followed by the aspiration before the beginning of the vowel. The tense fricative in Figure 3b is characterized by a vertical white stripe between the frication noise and the vowel, which indicates the brief stop of airflow during the production of the glottalized part.

Temporal measures of the frication noise were extracted using a Praat script. For each speaker, the duration was normalized in R (R Core Team 2023) using the scale function that centers and scales numeric values. Fundamental frequency (F0) data were also extracted from 10 equidistant points of a vowel following the fricative.

3.1.3 Results

In our corpus, all 60 word-initial aspirated fricatives are further analyzed. Thirty-four of them showed distinct aspiration noise between the frication noise and the vowel ([a] in open syllables or syllables with a nasal coda, and [a], [i], [u] in syllables with a plosive coda). The aspiration duration had a median of 25% of the fricative duration (ranging from 8.7% to 43%). The aspiration duration in Jinghpaw aspirated fricatives is shorter than that in Korean where aspiration is on average 50% of the fricative duration.

Tense fricatives, as noted above, are often realized with laryngealization between the frication noise and the vowel. This phonetic characteristic is not merely an isolated feature but is closely related to its diachronic source; diachronically, tense fricatives have evolved from an affricate [ts], a feature still evident in the orthography and predominantly maintained by elderly speakers. Meanwhile, younger speakers, including the two participants in this study, exhibit a notable trend towards deaffrication.2) The laryngealization in tense fricatives aligns with phenomena observed in voiceless unaspirated plosives and affricates in Jinghpaw, where laryngealization is a common production trait

---

2) As a reviewer suggested, a synchronic relation between deaffrication and laryngealization could explain the sound change: place features of the original affricates are dropped, with laryngealization acting as a remnant, something like debuccalization.

Table 2 Speaker information.

ID	Gender	Age	Location
BRM528	Female	36	Yangon
BRM529	Female	34	Kachin State
BRM530	Female	29	Yangon
BRM531	Male	34	Mandalay
BRM532	Female	26	Mandalay
BRM533	Female	45	Yangon

(Dai and Xu 1992, Kurabe 2016). Furthermore, a similar sound change is also found in the history of Shan and Burmese, languages with which Jinghpaw has been in contact, as well as other neighboring languages such as the Qiandong dialects of Hmongic (Jacques 2011, Carveth 2013).

3.2 Burmese

Burmese is the official language in Myanmar, spoken by around 30 million native (L1) speakers and up to 10 million second language (L2) speakers (Watkins 2001). The aspirated-tense fricative contrast is reported to have been disappearing in some speakers (Okano 2011, Jenny and San San Hnin Tun 2016: 16). Preliminary observation of a few recordings shows no clear phonetic distinction between the two fricatives.

3.2.1 Speakers

The data were collected from six Burmese speakers (five females and one male). At the time of the recording, all speakers were living outside of Myanmar. Except for one who comes from the Kachin State, all other speakers come from Burmese-speaking cities (Yangon and Mandalay). Details of our participants are shown in Table 2. The age refers to the speakers' age at the time of the data collection.

3.2.2 Methods

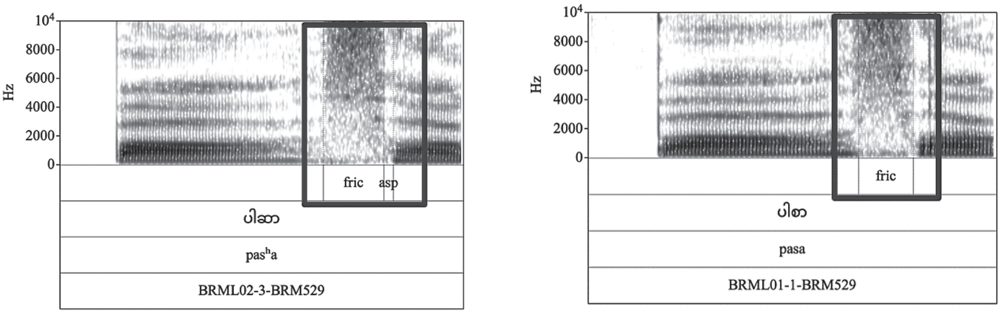
We created a stimuli list with a total of 27 target items (nonce words) involving three types of fricatives (tense, aspirated, and voiced) in the word-medial position, as shown in Table 3.[3] These items were constructed by varying three different tones (low, high, and creaky) with the vowel [a], in the context of three word-initial onsets ([p, t, k]). Our Burmese corpus consists of 648 tokens which were produced by repeating 27 items four times each by 6 participants.

---

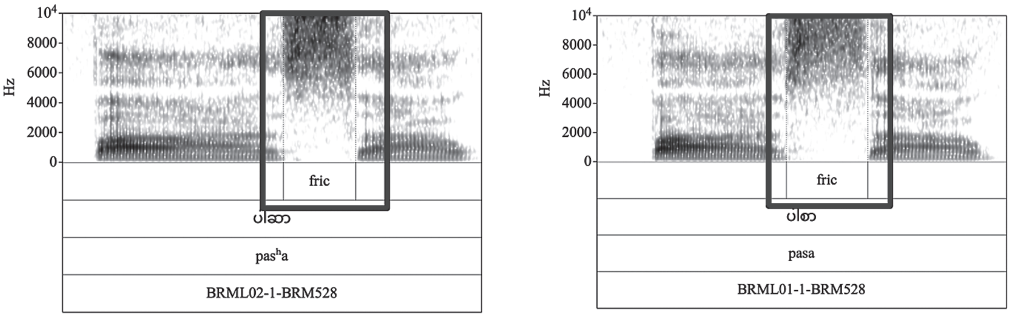
[3] We agree with a reviewer who pointed out that phonological contrasts across languages tend to be preserved word-initially rather than word-medially. Aspirated fricatives in the word-medial position were recorded because we tested whether the contrast of aspirated fricatives is salient in a prosodically weak position (cf. about the contrast being neutralized, see Okano 2011, Jenny and San San Hnin Tun 2016: 16). We found that half of the speakers retained aspiration, suggesting that aspirated fricatives were salient for these individuals. In a follow-up study, we will examine the aspirated fricatives in the word-initial position.

Table 3 Burmese target stimuli.

Burmese	IPA	Burmese	IPA	Burmese	IPA	fricative	tone
ပါစာ	pasa	တာစာ	tasa	ကာစာ	kasa	*tense*	*low*
ပါဆာ	pasʰa	တာဆာ	tasʰa	ကာဆာ	kasʰa	*aspirated*	
ပါဇာ	paza	တာဇာ	taza	ကာဇာ	kaza	*voiced*	
ပါးစား	pásá	တားစား	tásá	ကားစား	kásá	*tense*	*high*
ပါးဆား	pásʰá	တားဆား	tásʰá	ကားဆား	kásʰá	*aspirated*	
ပါးဇား	pázá	တားဇား	tázá	ကားဇား	kázá	*voiced*	
ပစ	pa̰sa̰	တစ	ta̰sa̰	ကစ	ka̰sa̰	*tense*	*creaky*
ပဆ	pa̰sʰa̰	တဆ	ta̰sʰa̰	ကဆ	ka̰sʰa̰	*aspirated*	
ပဇ	pa̰za̰	တဇ	ta̰za̰	ကဇ	ka̰za̰	*voiced*	
labials		*coronals*		*dorsals*			



a. aspirated fricative (ဆာ) **b. tense fricative (စာ)**



c. aspirated fricative (ဆာ) **d. tense fricative (စာ)**

Figure 4 An Illustration of Aspirated and Tense Fricatives in Burmese. Speaker BRM529 in (a) and (b) show the contrast, but BRM528 in (c) and (d) does not.

The target words were embedded in a frame sentence ဒီစာလုံး_______ကိုဖတ်ပါ။ /di salóuɴ _______ go phaʔ pa/ "Please read this character _______." The experiment was conducted by advancing PowerPoint slides by the experimenter in a soundproof booth. Recording was done on a TASCAM-MKIII with a headworn-microphone Shure WD-30.

Tokens were processed and annotated with a series of Praat scripts that were used to process Jinghpaw data. Following the same criteria, fricatives were annotated. In aspirated fricatives, the frication noise is followed by aspiration, and in tense fricatives, the frication

Table 4 The frequency of the realization of aspirated and tense fricatives in Burmese.

Orthography	realized with aspiration	realized as tense
aspirated ဆာ	92	123
tense စာ	8	208

Table 5 Realization of aspirated fricatives by speakers.

Group	ID	gender	aspiration	tense
Merger Group	BRM528	Female	2	34
	BRM531	Male	2	34
	BRM530	Female	1	35
Faithful Group	BRM529	Female	33	3
	BRM532	Female	28	8
	BRM533	Female	26	9

noise is followed by a laryngeal component.

Temporal measures of the frication noise and fundamental frequency (F0) of the vowel following fricatives were extracted in the same way as in Jinghpaw (cf. Section 3.1.2).

3.2.3 Results

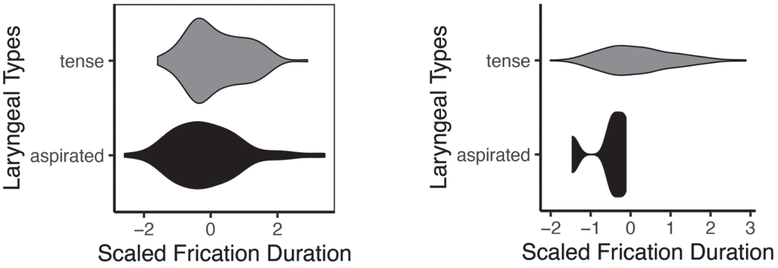
Data in our Burmese corpus show that most tense fricatives are produced as expected: the frication noise is followed by a laryngealized part. Aspirated fricatives (n = 215) are produced either with aspiration or with laryngealization, suggesting that aspirated fricatives are produced as tense fricatives in the latter case.

In Table 5, we present the count data of the realization of aspirated fricatives by speakers. A group of three speakers produced nearly all of the aspirated fricatives as tense, whereas another group of three speakers produced most of the aspirated fricatives as aspirated. The merger group produced most fricatives as tense with some degree of laryngealization, but no aspiration. The faithful group produced aspirated fricatives with aspiration. The rest of this section presents separate analyses of these two groups.[4)]

The faithful group of Burmese produced the most aspirated fricatives with aspiration. When the normalized frication duration was compared between aspirated and tense fricatives, no difference was observed ($t(206.19) = -0.59$, p. 0.55), see Figure 5a. This comparable frication duration is what makes Burmese different from Korean where frication duration is shorter in aspirated fricatives than in tense fricatives. We also examined the aspiration duration of aspirated fricatives that were produced from the tense fricatives. The aspiration duration of these fricatives was not different from the aspiration duration of the aspirated fricatives in the faithful group ($t(7.6) = -0.78$, $p = 0.45$). The Burmese merger

---

[4)] Our analysis has not revealed any clear sociolinguistic patterns that differentiate between the merger and faithful groups. Sociolinguistic variables such as age, gender, and geographical origin do not appear to significantly influence the observed pattern in our data.



a. Burmese faithful group **b. Burmese merger group (only 5 tokens in the aspirated category)**

Figure 5 Normalized frication duration in Burmese

group in Figure 5b shows that the frication duration of tense fricatives (recall nearly all voiceless fricatives were produced as tense) is long. The frication data of aspirated fricatives with aspiration is not informative due to few data points (n = 5).

The faithful group displays about 20% (range of 5% to 44%) of the aspiration ratio, which was obtained by dividing the aspiration duration by the duration of the whole fricative (frication and aspiration). The merger group produced 5 tokens with aspiration even though they were exposed to 108 tokens with the orthography of aspirated fricatives. These tokens have shorter aspirations with a median of 17% (with a range of 12 to 22%).

3.3 Summary

This section has demonstrated that Jinghpaw speakers, though only two participants, retain the contrast between aspiration and tense. Burmese speakers, on the other hand, show patterns that divide them into two groups: one with aspiration in aspirated fricatives, and the other with the majority of tokens produced as tense fricatives.

4. Discussion

4.1 Other Acoustic Correlates

Korean has a salient phonetic difference between the aspirated fricative and the tense fricative, and the frication noise is significantly shorter in aspirated fricatives. It would be in our interest to compare duration data in Jinghpaw and Burmese.

In Jinghpaw, 34 out of 60 tokens are realized with distinct aspiration, and on average 25% of the fricative duration was aspirated. The relatively small amount of aspiration has less effect on the frication duration (unlike Korean). A two-tailed t-test shows that the frication duration of aspirated fricatives is not significantly different from that of tense fricatives (t(105.72) = 1.28, p = 0.2). Out of the two Burmese groups, aspirated fricatives in the faithful group have on average 20% of the fricative as aspiration. When the frication duration in this group was compared, the duration was not significantly different: t(206.19) = −0.59, p. 0.55.

We also examined data of F0 of the following vowel in the two types of fricatives in the faithful group because aspirated fricatives may have higher F0 than tense ones. The rationale was to examine whether acoustic correlates distinguishing the two fricatives shifted to an adjacent segment. The result was not significant in Jinghpaw (t(85.41) = −0.18, p = 0.85). In Burmese, the difference was significant (t(203.5) = 2.17, $p < 0.05$) in the expected direction. The F0 in the interval of the vowel following aspirated fricatives (mean = 213.05 Hz) was higher than the vowel following tense fricatives (mean = 202.44 Hz).

4.2 Neutralization of Laryngeal Features in Fricatives

As demonstrated in this paper, the contrast between aspirated and tense fricatives is becoming neutralized among certain Burmese speakers. Our study's participants also acknowledge that the contrast is fading among younger generations, a phenomenon similarly noted in previous studies such as Jenny and San San Hnin Tun (2016: 16), which notes the neutralization of laryngeal features. Diachronically, the contrast has its source in the contrast in palatal affricate (i.e., between /tɕ/ and /tɕ^{h}/) during the Pagan period, as indicated in the orthography. This distinction evolved into /ts/ and /ts^{h}/ before 1780, and further into /s/ and /s^{h}/ after the 1850s (Bradley 2011). The contrast was documented as the contrast in aspiration in early records (Armstrong and Pe Maung Tin 1925: 17). Despite its clear representation in orthography, numerous spelling errors are found in internet blogs, attributed to the merger (Okano 2011). As a result of the merger, many minimal pairs such as given in (5) have become homophones.

(5) Burmese minimal pairs

/sá/ 'eat'	/sì/ 'set up'	/sṵ/ 'gather'	/sé/ 'sticky'	/sò/ 'wet'	/seiʔ/ 'mind'
/s^{h}á/ 'salt'	/s^{h}ì/ 'oil'	/s^{h}ṵ/ 'award'	/s^{h}é/ 'medicine'	/s^{h}ò/ 'speak'	/s^{h}eiʔ/ 'goat'

Since both aspirated and tense fricatives feature prominently in a wide range of lexical items, from basic vocabulary terms to more complex ones, the impact of the merger is believed to be significant. What occurs as a result of this merger is an area for future exploration. It remains to be seen whether non-phonetic contexts contribute to the contrast. Furthermore, to investigate whether the phonetic aspect has indeed become obsolete, follow-up studies, such as phonetic studies with minimal pairs incorporating a range of vowel combinations and a perception study, may prove to be beneficial.

5. Conclusion

This paper explored the phonetics of aspirated fricatives in Jinghpaw and Burmese. It showed that the duration of aspiration constitutes approximately 25% of the total fricative duration in Jinghpaw, which is much shorter compared to the aspiration duration observed in Korean aspirated fricatives. In the case of Burmese, despite the orthographic presence of aspirated fricatives, there is a noticeable trend of the phonetic distinction fading away in some speakers. Cross-linguistic comparisons have revealed that languages may not have identical aspiration duration in aspirated fricatives. For instance, Jinghpaw and the faithful

Burmese group demonstrate shorter aspiration in aspirated fricatives, whereas Korean shows a longer aspiration duration in them. Frication duration with a high frequency fricative noise is a salient acoustic correlate when distinguishing aspirated fricatives from tense ones in Korean, where tense and aspirated fricatives are distinguished primarily via frication duration, with laryngeal constriction playing a secondary role (Perkins *et al.* 2023b). However, differences in the frication duration were not significant in Jinghpaw and the faithful Burmese group.

References

Armstrong, Lilias E. and Pe Maung Tin. 1925. *A Burmese Phonetic Reader*. London: University of London Press.

Bradley, David. 1996. "Kachin." *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas*, vol. 2.1 (Stephen A. Wurm, Peter Mühlhäusler, and Darrell T. Tryon, eds.), 749–751, Berlin: Mouton de Gruyter.

Bradley, David. 2011. Changes in Burmese Phonology and Orthography. Keynote talk at the 21st Annual Meeting of the Southeast Asian Linguistics Society. Kasetsart University, 12 May 2011.

Boersma, Paul and David Weenink. 2023. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.3.17. http://www.praat.org/ [accessed September 2023].

Carveth, Vincent. 2013. "Pathways of Development for Qiándōng Hmongic Aspirated Fricatives." *Journal of the Southeast Asian Linguistics Society*, 6: 146–157.

Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged. 2002. "Acoustic and Aerodynamic Correlates of Korean Stops and Fricatives." *Journal of Phonetics*, 30(2): 193–228.

Craioveanu, Radu. 2013. "The Rise and Fall of Aspirated Fricatives." *Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association* (Shan Luo, ed.), Toronto: Canadian Linguistic Association.

Dai, Qingxia and Xijian Xu. 1992. *Jingpoyu yufa* [A Jingpo grammar]. Beijing: Minzu University of China.

Jacques, Guillaume. 2011. "A Panchronic Study of Aspirated Fricatives, with New Evidence from Pumi." *Lingua*, 121: 1518–1538.

Jenny, Mathias and San San Hnin Tun. 2016. *Burmese: A Comprehensive Grammar*. London and New York: Routledge.

Kingston, John. 1990. "Articulatory Binding." *Papers in Laboratory Phonology I* (John Kingston and Mary E. Beckman, eds.), 406–434, Cambridge: Cambridge University Press.

Ko, Joohee. 2023. Korean Sibilant Fricative Identification and Vowel Anticipation: A Gating Experiment. Paper presented at the 9th International Conference on Phonology and Morphology. Seoul National University, 30 June 2023.

Kurabe, Keita. 2016. A Grammar of Jinghpaw, from Northern Burma. Ph.D. dissertation, Kyoto University.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.) 2014. *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International.

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

Moran, Steven and Daniel McCloy (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. http://phoible.org [accessed September 2023].

Okano, Kenji. 2011. Is the Aspirated Fricative "s(h)-" Really a Phoneme in Colloquial Burmese? Paper presented at the 17th Himalayan Languages Symposium. Kobe City University of Foreign Studies, 6–9 September 2011.

Perkins, Jeremy, Dahm Lee, and Seunghun J. Lee. 2023a. "A Production Study of Korean Consonants." *Proceedings of the 20th International Congress of Phonetic Sciences* (Radek Skarnitzl & Jan Volín, eds.), 778–782, Guarant International.

Perkins, Jeremy, Yu Yan, Dahm Lee, and Seunghun J. Lee. 2023b. "Using Machine Learning to Model the Three-way Laryngeal Contrast in Korean.” *Proceedings of the 20th International Congress of Phonetic Sciences* (Radek Skarnitzl & Jan Volín, eds.), 783–787, Prague: Guarant International.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

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

Sun, Jackson T.-S. 2000. “Stem Alternations in Puxi Verb Inflection: Toward Validating the rGyalrongic Subgroup in Qiangic.” *Language and Linguistics*, 1(2): 211–232.

Vaux, Bert. 1998. “The Laryngeal Specifications of Fricatives.” *Linguistic Inquiry*, 29(3): 497–511.

Villegas, Julián, Konstantin Markov, Jeremy Perkins, and Seunghun J. Lee. 2020. “Prediction of Creaky Speech by Recurrent Neural Networks Using Psychoacoustic Roughness.” *IEEE J. of Selected Topics in Signal Processing*, 14(2): 355–366.

Watkins, Justin W. 2001. “Illustrations of the IPA: Burmese.” *Journal of the International Phonetic Association*, 31(2): 291–295.

Xi, Xingcan and Shaoni Li. 1997. “Heqing baiyu de songqi cayin [The Aspirated Fricatives in the Heqing Dialect of Bai].” *Zhongyang minzu daxue xuebao*, 111(2): 102–106.

Journal of Asian and African Studies, Supplement, No. 4, 2024

A Typological Overview of Lateral Fricatives in Southern Bantu Languages*

Shinagawa, Daisuke
Lee, Seunghun J.

This paper aims to provide a typological overview on the group-internal variation of lateral fricatives in selected southern Bantu languages. Based on phonetic observations about attested realisations in sample languages and their distributional patterns, we propose several hypothetical principles that explain observable variations, which include overall preference of voiceless lateral fricatives over voiced counterparts, implicational hierarchy that defines possible configuration of phonemic lateral fricatives in attested consonant inventories, and a phonotactic restriction that blocks a voiceless lateral fricatives to occur in a post-nasal position. We further discuss that through these principles the attested variation can be recognised as a continuum reflecting degrees of restrictiveness on lateral fricatives. We conclude the paper by addressing possible directions of further investigation especially on the emergence of lateral fricatives as a yet unsolved question.

Keywords: lateral fricatives, Bantu languages, phonetic typology, S Zone languages

DOI: https://doi.org/10.15026/0002000628

---

* This research is written as part of the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023). The research has been funded by a Kakenhi Grant (C) 20K00578. This work was also supported by JSPS KAKENHI Promotion of Joint International Research (Fostering Joint International Research (B)) Grant Number 21KK0005, and the Kakenhi (B) Grant 23H00622. Any errors remain our own.

1. Introduction

It is widely recognised that the Bantu language family, consisting of about 500 languages spread in vast areas of Sub-Saharan Africa (cf. Hammerström 2019), generally has a relatively small consonant inventory with a limited set of articulatory features (cf. Maddieson & Sands 2019). The simplicity of the consonant inventory reconstructed for Proto Bantu, which consists of two series of obstruents contrasting voiceless /p, t, c, k/ vs. voiced /b, d, j, ɡ/, latter of which could have been a series of continuants, with the nasal series /m, n, ɲ, ŋ/ (cf. Meeussen 1967: 83, Hyman 2019: 128), is still a common basis of a variety of phonemic systems found in many present-day Bantu languages. Most of the contemporary languages also have developed a series of fricatives through the process known as Bantu Spirantisation (Shadeberg 1994/95), where stop consonants are fricativised when followed by the first-grade high vowels, traditionally called 'super-high vowels' (cf. Maddieson & Sands 2019: 89). The final outcomes of the process, which are /f, v/ triggered by the highest back vowel and /s, z/ by the highest front vowel, are the most typical, and in many cases only distinctive fricatives in the phonemic inventory (Maddieson & Sands 2019: 90, see also 2.1 below).

However, this may not be the case in southern Bantu languages.[1] Especially those classified into zone S in the referential classification by Guthrie (1967–71), which is a southern-most group in Eastern Bantu languages, are said to have developed a complex system of consonants probably due to contact with surrounding non-Bantu languages including the languages spoken in Kalahari Basin area (traditionally called Khoesan lanaguages, mentioned as KBA languages hereafter), or to internal innovation (cf. Maddieson & Sands 2019: 90, Sands & Gunnink 2019, Gunnink, Chousou-Polydouri and Bostoen 2023, among others). At least three types of relatively complicated fricative sounds can be identified as those exclusively observed in zone S languages, which are i) 'whistled' fricatives[2] (cf. Lee-Kim *et al.* 2014), ii) a hetero-organic sequence of fricatives[3] (Ladefoged & Maddieson 1996: 330–331), and iii) lateral fricatives, the last of which is the focus of this article. In the following, we will discuss phonetic details of the lateral fricatives observed in the selected zone S languages, namely Northern Sotho [S32], Southern Ndebele [S407], Xhosa [S41], Zulu [S42], Swati [S43], and Tsonga [S53] (see Figure 1 for their geographical locations), based on the first-hand data collected in a series of field recording sessions with

---

[1] Not only zone S languages but south-western Savannah languages including languages in zones K and L also tend to have developed a relatively complex system especially in terms of fricatives. This topic, however, is beyond the scope of this paper.

[2] According to Lee-Kim *et al.* (2014), 'whistled' (or 'whistling') fricatives are observed in a zone S language Tsonga [S53], where the sounds are produced through a vertical narrowing of the lips, rather than a simple labialization with protrusion of the lips, with a retroflex lingual gesture.

[3] Hetero-organic fricatives are the segments which have been described as double-articulated fricatives in the literature, e.g., Lombard (1985) for Northern Sotho [S32]: [f͡s] as in *βofsa* 'youth', [f͡ʃ] as in *leffeːra* 'coward', and [β͡ʒ] as in *βʒalwa* 'beer'. However, Ladefoged & Maddieson (1996: 330–331) recognise the sound as a phonetic sequence rather than a single sound with two simultaneous fricative articulations.

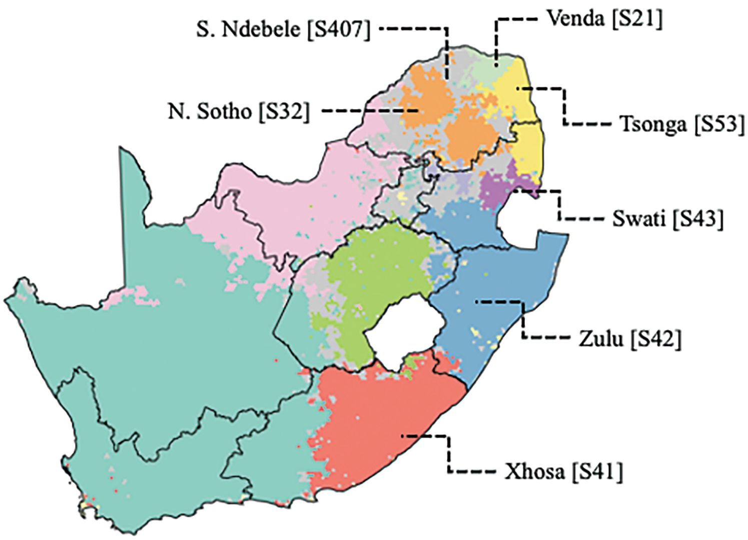


Figure 1 Map of Bantu languages in South Africa[4]

relatively young speakers of each target language. Based on the empirical data, we will provide a tentative overview of the typological variation of the lateral fricatives attested in the sample languages.

The remaining part of this paper is organised as follows. In Section 2, we will introduce information on fricatives in Bantu from an intra-genetic and areal typological point of view. Section 3 provides basic observations of the attested phonetic realisations of lateral fricatives in the sample languages and summarises the distribution of lateral fricatives in terms of different phonetic environments. Based on the observation, in Section 4 we will argue on what kind of typological generalisation can be made from the attested variation of lateral fricatives throughout zone S languages. Section 5 concludes the article with a summary of the discussions and further questions to be investigated in the future research.

2. Fricatives in Bantu

2.1 Cross-Bantu Overview of the Fricative Series

The general tendency is that Bantu languages have relatively a small set of fricative consonants as Maddieson & Sands (2019: 90) point out: “[m]ost of the languages have relatively limited sets of fricatives of the cross-linguistically common types”. Table. 1 shows a list of fricative consonants in the phonemic inventory of the selected sample languages consisting of at least two languages from all Guthrie zones.[5]

---

[4] Source: https://www.reddit.com/media?url=https%3A%2F%2Fi.redd.it%2F1uhm4svry5661.png

[5] The primary source of the phonemic inventory of each sample language is collected from grammatical sketches included in Nurse & Philippson (2003) and Van de Velde *et al.* (2019). The secondary source, which is to fill the gap of geographical zones that are not covered by the primary sources, is supplied by the PHOIBLE database (Moran and McCloy 2019) and other additional sources shown in the table.

Table 1 List of fricative phonemes in sample languages[6]

language name	bilab	lab-den	den	alv	pal-alv/pal	lat	vel/uvu	glt	source
Basaá [A43]				s				h	Hyman (2003: 259)
Nen [A44]		f		s			x	h	Mous (2003: 284)
Kpāʔ [A53]		f, v		s, z					Guarisma (2003: 308)
Makaa [A83]		f, v		s, z	ʃ, ʒ			h	Heath (2003: 336)
Kwakum [A91]		f, (v)			ʃ, ʃʱ				Njantcho & van de Velde (2019: 384)
Tsogo [B31]	β	f		s			ɣ		van der Veen (2003: 378–379)
Nsong [B85d]		f, v		s, z	ʃ, ʒ			h	Koni Muluwa & Bostoen (2019: 416)
Babole [C101]				s				h	Leitch (2003: 394)
Pagibete [C401]		f, v		s, z					Reeder (2019: 452)
Lega [D25]		(f, v)		s, z	(ʃ)				Botne (2003: 425–426)
Zimba [D26]		f		s					Kutsch Lojenga (2019: 476)
Bila [D32]	ɸ			s				h	Kutsch Lojenga (2003: 456)
Chuka [E541]	(β)		ð				(ɣ)		Kanana (2011)
Rombo [E623]		f, (v)		s, (z)	ʃ			h	DS's Fieldnote
Hangaza [JD65]	β	f, v		s, z	ʃ, ʒ			h	Rubagumya (2006) in UPSID[7]
Rwanda [JD61]	β	f, v		s, z, ʂ, ʐ	ç			h	Walker *et al.* (2008)
Jita [JE25]		f		s	(ʃ)				Kagaya (2005)
Mara [JE40]		(f)		s	(ʃ)			h	Aunio *et al.* (2019: 507)
Langi [F33]		f, v		s, z	ʃ			h	Dunham (2005) in UPSID
Mbugwe [F34]		f		s	ʃ				Wilhelmsen (2019: 535)
Kami [G36]		f, v		s, z				h	Petzell & Aunio (2019: 565)
Ngazidja [G44a]	β	f, v	(θ, ð)	s, z	ʃ, (ʒ)		(x, ɣ)	h	Patin *et al.* (2019: 593)
Vwanji [G66]		f		s					Eaton (2019: 618)
Kituba [H10a]		f, v		s, z				h	Chanard (2006) in UPSID
Hungana [H42]		f		s, z				h	Takizala (1974) in UPSID
Fipa [M30]		f, v		s, z	ʃ, ʒ				Riedel & Bickmore (2013) in UPSID
Bemba [M42]	β	f		s	ʃ				Hamann & Kula (2015)

Chimpoto [N14]		v		s, z	ʃ, ʃ^{w}		ɣ	h, h^{w}, h^{y}	Botne (2019: 698)
Manda [N11]		f, v		s	ʃ		ɣ	h	Bernander (2017)
Matuumbi [P13]				(s)					Odden (2003: 532)
Yao [P21]		(f)		s					Odden (2003: 532)
Makonde [P23]				s	(ʃ)			h	Odden (2003: 532)
Makhuwa [P31]		v		s	ʃ			h	Kisseberth (2003: 549)
Cuwabo [P34]		f, v	ð	s, z	(ʃ)				Guérois (2019: 734)
Samba [L12a]		f, v		s, (z)	ʃ			h	van Acker & Bostoen (2020)
Lunda [L52]		f, v		s, z	ʃ, ʒ			h	Kawasha (2006)
Luvale [K14]		f, v		s, z	ʃ, ʒ			h	Sommer (2003: 568)
Luyana [K31]	β	f		s, z	ʃ				Sommer (2003: 568)
Kwangari [K33]	β	f, v		s	ʃ			h	Sommer (2003: 568)
Gciriku [K38b]	β	f, v			ʃ		ɣ	h	Sommer (2003: 568)
Totela [K41]	β	f		s, z	(ʃ)			(h), h^{w}	Crane (2019: 650)
Mbukushu [K43]		f, v	θ, ð		ʃ		ɣ	h	Sommer (2003: 568)
Umbundu [R11]		f, v		s				h	Sommer (2003: 568)
Ndonga [R22]		f, v	θ, ð	s, z	ʃ, (ʒ)		x, ɣ	h	Sommer (2003: 568)
Herero [R31]		v	θ, ð					h	Elderkin (2003: 582)
Yeyi [R41]		f, v		s, z	ʃ, ʒ			h	Sommer (2003: 568)
Venda [S21]	ɸ, β	f, v		s, z, ʂ, ʐ	ʃ, ʒ		x, (ɣ)	ɦ	Nemakhavhani (2002)
Tswana [S31]		(f)		s	ʃ		χ	h	Bennett *et al.* (2016)
Xhosa [S41]		f, v		s, z	ʃ	ɬ, ɮ̤	x, ɣ̤	h, ɦ̤	Gowlett (2003: 615)
Tsonga[8] [S53]	ɸ, β	v, v̤		s, z	ʃ, ʒ	ɬ	x	ɦ	Gowlett (2003: 615), Baumbach (1987: 3–16)
Copi [S61]		f		s, s͡w	(ʃ)	(ɬ)		ɦ̤	Gowlett (2003: 615)

---

6) The abbreviations in Table 1 are as follows: bilab for bilabials, lab-den for labio-dentals, den for dentals, alv for alveolars, pal-alv for palatal-alveolars, pal for palatals, lat for laterals, vel for velars, uvu for uvulars, and glt for glottals.

7) UPSID data is taken from https://phoible.org/contributors/UPSID.

8) /v̤/ and /ɬ/ are missing in Baumbach (1987).

Table 2 Geographical distribution of fricative phonemes: comparison between south (zones K, R, and S) and non-south

	bilab	lab-den	den	alv	pal-alv/ pal	lat	vel/uvu	glt
lgs attested (total)	**13**	45	**6**	46	34	**3**	**13**	35
percentage (total)	0.25	0.88	0.12	0.90	0.67	0.06	0.25	0.69
lgs attested (south: K, R, S)[9]	6	15	3	12	13	3	7	14
percentage (south)	**0.40**	1.00	**0.20**	0.80	0.87	**0.20**	**0.47**	0.93
percentage (south/total)	0.46	0.33	0.50	0.26	0.38	**1.00**	0.54	0.40
lgs attested (non-south)	7	30	3	34	21	0	6	21
percentage (non-south)	**0.19**	0.83	**0.08**	0.94	0.58	**0.00**	**0.17**	0.58
percentage (non-south/total)	0.54	0.67	0.50	0.74	0.62	0.00	0.46	0.60

As summarised in Table 2, the most typical fricative types found across Bantu languages are alveolars /s, z/ (90%, 46 out of the 51 sample languages) and labiodentals /f, v/ (88%), both of which are cross-linguistically common. Palato-alveolars /ʃ, ʒ/ and the glottal /h/ are the second common group, which are geographically distributed in wide areas. What is striking is the remaining types of fricatives. Bilabials /ɸ, β/ and velars + uvulars /x, ɣ; χ, ʁ/ are less common across Bantu languages, i.e., both appear in 25% of sample languages, but also the distribution in non-south languages are further limited (19% for bilabials and 17% for velars + uvulars). Dentals /θ, ð/ show the similar distribution but they must have been introduced through lexical borrowing from languages in recent contact at least in some cases e.g., Ngazidja [G44a] from Arabic (Patin *et al.* 2019: 592). Most importantly, lateral fricatives /ɬ, ɮ/, the rarest throughout the Bantu area, are exclusively attested in zone S.

2.2 South as a Phonological Area

Lateral fricatives being attested exclusively in zone S has been argued in a wider context of inter-genetic areal typology. Discussing segmental as well as suprasegmental features that characterise South as one of the phonological areas in Africa, among which are North, East, Rift, Center and Sudanic, Clements and Rialland (2008: 82)[10] propose that ‘lateral obstruents’ as one of such features shared among languages spoken in the Kalahari Basin area (KBA languages). It should also be pointed out that interestingly KBA languages have not developed a rich system of fricatives and most of them do not have lateral fricatives in their phonemic inventory (cf. Nakagawa 2014). However, it is still

---

[9] This classification is simply based on geographical adjacency and is not based on any historical consideration about genetic affinity, for which there seems no controversial classification has been established so far (cf. Naumann & Bibiko 2016). For example, Nurse & Philippson (2003b: 171) tentatively propose “Zone S: S20–30–40–?S50–?S60, plus P30” as ‘Southeast’ according to Janson’s (1991/2) proposal.

[10] Clements & Rialland (2008: 82) “A third zone, the South, is sharply delineated by the remaining features [...]: ejective and aspirated stops, clicks, and slack voiced stops. To these features we could add their characteristic series of lateral affricates and fricatives. All these features are widely shared by Khoisan and Bantu languages in the region.”

plausible that manipulation of the lateral as a phonological areal feature, which is directly associated with articulatory control of both or either side(s) of the tongue body, can be shared by languages across phylogenetic boundaries since such areal features may not be necessarily fully reflected on every language within the area, especially on 'newcomers' which are less influenced by such contact effect[11] (cf. Güldemann & Fehn 2017, Güldemann 2011. as cited in Naumann & Bibiko 2016).

Based on their own survey on Bantu and non-Bantu KBA language of the area, Naumann & Bibiko (2016) argue that lateral fricatives themselves may not be included in a set of areal features of South as a meta-phylogenetic linguistic area but are southeastern 'Bantu' areal feature. Besides the issue whether lateral fricatives, or more broadly lateral obstruents, should be part of areal features or not, what is significant in the scope of this paper is that there is internal variation in terms of distribution of different types of lateral fricatives in different phonetic environment, which in turn may shed light on typological generalisation of diversification and diffusion as well as diachronic processes of innovation.

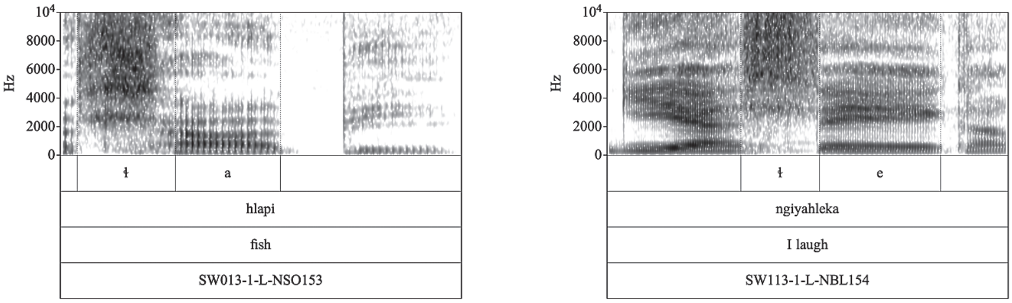
3. Data from Southern Bantu Languages

The data in this section comes from a project about Bantu languages in Southern Africa, which collects phonetic data from multiple speakers to investigate segmental and suprasegmental patterns of the six sample languages, namely Venda [S21; VEN], Northern Sotho [S32; NSO], Southern Ndebele [S407; NBL], Xhosa [S41; XHO], Zulu [S42; ZUL], Swati [S43; SSW], and Tsonga [S53; TSO]. We conducted recording sessions in local environments where each language is spoken in its speech communities in 2022 and 2023. Data are collected through an organised questionnaire consisting of sample sentences composed of the basic vocabulary (mainly from Swadesh's 200-word list), which were embedded in frame sentences varying in tone yielding various combinations of segmental and suprasegmental features, such as syllable structures, phonotactic variation as well as tonal properties of lexical items. The use of Swadesh list was decided to collect a set of samples, which enables a comparative study of the seven target languages. As for the lateral fricative sounds, which are the main focus of this article, only the Venda language does not have a lateral fricative in its phoneme inventory. To allow readers to compare phonetic realizations of the fricatives in the same position, this section shares representative examples from the six languages with lateral fricatives[12]. A note about the entire corpus is in order. The stimuli were constructed from a list of 176 items that include target words (nouns, verbs, adjectives from the Swadesh list). These targets were embedded in low and

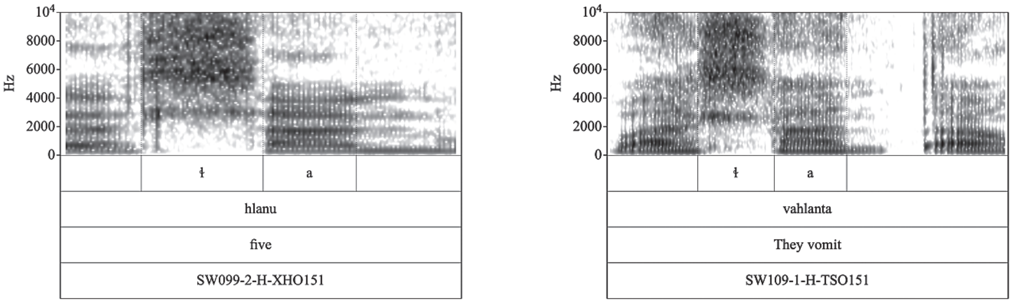
---

[11] Naumann & Bibiko (2016) "... [S]ubstrate interference contributed repeatedly to creating linguistic similarities [in Nguni, Tswana, and Afrikaans] with Kalahari Basin languages (or at least maintaining existing ones) but has not been strong enough to make the newcomers "full" members of the area." Güldemann & Fehn (in prep.: 18; cf. also Güldemann 2010: 572f.)

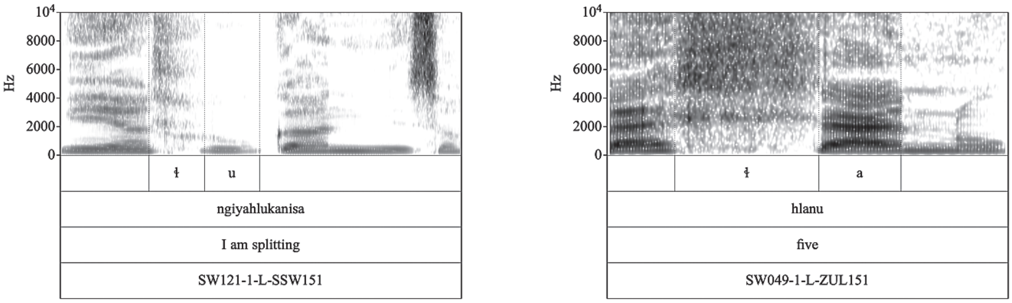
[12] Plots that appear in this section are taken from a larger set of recorded data of 8 South African Bantu languages. We show representative realisations of words that contain target lateral fricatives. Detailed phonetic studies of these lateral fricatives await future studies.



Northern Sotho **Southern Ndebele**



Xhosa **Tsonga**



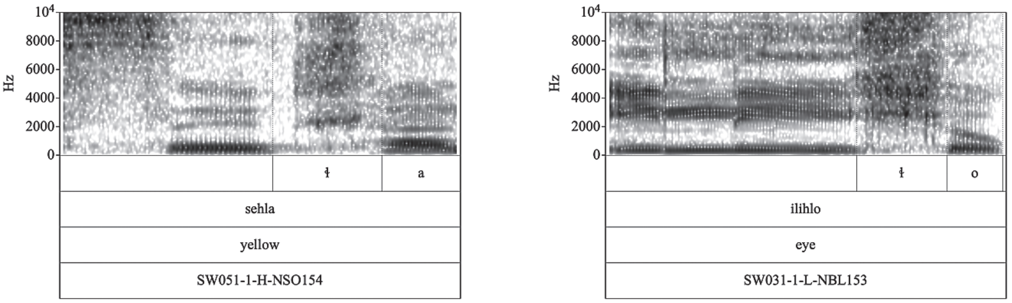
Swati **Zulu**

Figure 2 Voiceless lateral fricatives in the stem-initial position in six languages

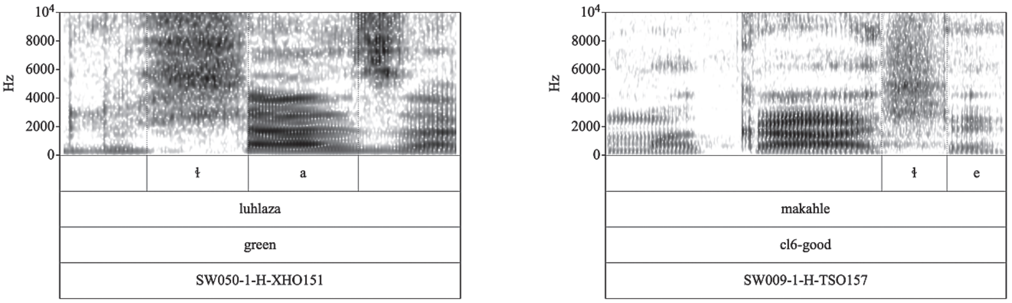
high tone contexts and were produced twice, yielding 704 tokens per participant (176 items * 2 tonal contexts * 2 repetitions). The stimuli were recorded from 8 Venda and 8 Northern Sotho speakers (5682 tokens per language), as well as 12 Southern Ndebele and 12 Tsonga speakers (8448 per language). The recordings were made by one Xhosa speaker and one Zulu speaker. Altogether, 29668 tokens with or without lateral fricatives consist of our corpus.

3.1 Voiceless Lateral Fricative [ɬ]

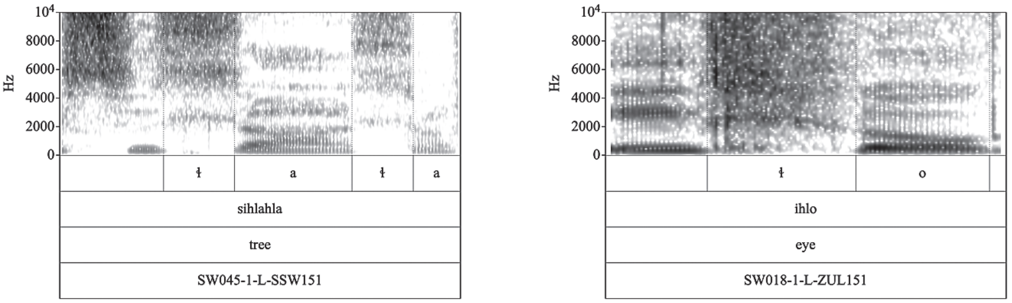
The voiceless lateral fricative [ɬ] in Northern Sotho, Southern Ndebele, Xhosa, Zulu, Swati, and Tsonga shows distributional differences. While all six languages allow voiceless lateral fricatives in the stem-initial position and the word-medial position, only Zulu, Swati



Northern Sotho **Southern Ndebele**



Xhosa **Tsonga**



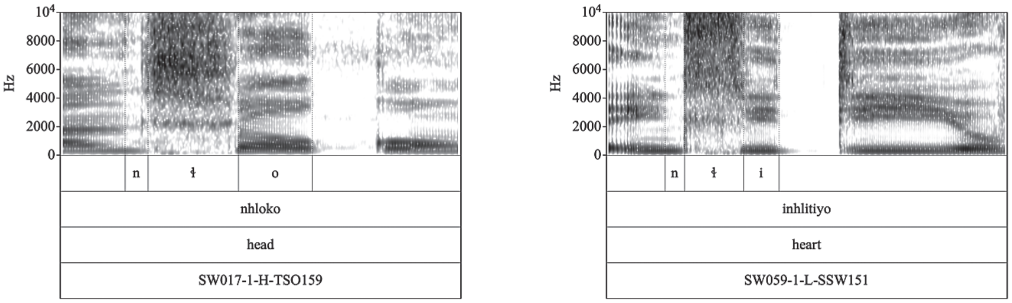
Swati **Zulu**

Figure 3 Voiceless lateral fricatives in the word-medial position in six languages

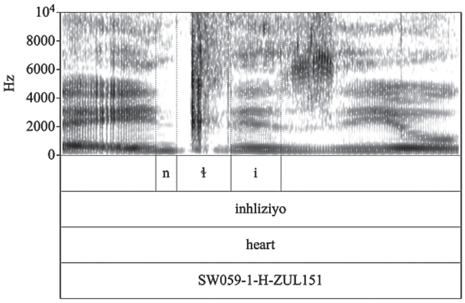
and Tsonga have [ɬ] after a nasal consonant.

3.1.1 Stem-initial

Examples from Northern Sotho, Southern Ndebele, Xhosa, Zulu, Swati, and Tsonga are shown in figure 2. In each plot, the first tier is annotated for a lateral fricative and the following vowel, which is taken from a target embedded in a frame sentence. The voiceless fricative is spelt as *hl* in the orthographic description given in the second tier. The second to the last tier is a token ID, and the bottom tier is the ID number for each token when it is stored in an archive.



a. Tsonga **b. Swati**



c. Zulu

Figure 4 Voiceless lateral fricatives in the post-nasal position in three languages

3.1.2 Word-medial Position

Voiceless lateral fricatives in Northern Sotho, Southern Ndebele, Xhosa, Zulu, Swati, and Tsonga also appear in word-medial position as shown in figure 3. When compared with the acoustic signals in figure 2, the lateral fricatives in the word-medial position show similarities with those in the stem-initial position.

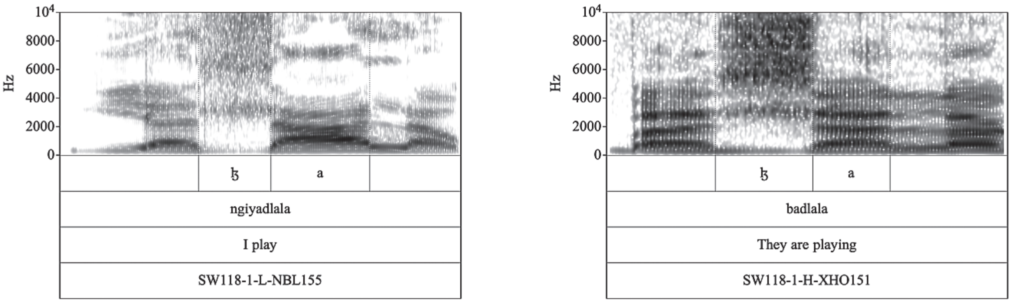
3.1.3 Post-nasal Voiceless Lateral Fricatives

Only three languages (Zulu, Swati, Tsonga) exhibit lateral fricatives in the post-nasal position. As studies on post-nasal plosives have shown, voiceless plosives are less preferred after a nasal (cf. Pater 1999). This tendency to avoid voiceless obstruents after a nasal extends to lateral fricatives.

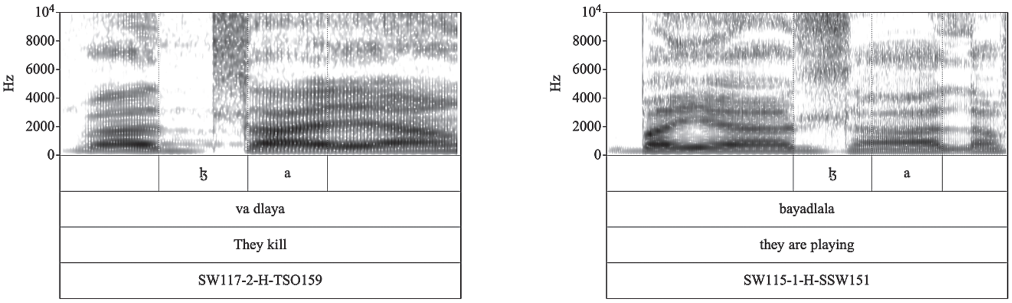
3.2 Voiced Lateral Fricative [ɮ]

3.2.1 Stem-initial

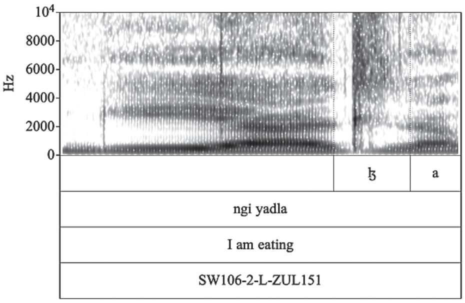
Except for Northern Sotho, the voiced lateral fricative is observed in Southern Ndebele, Xhosa, Zulu, and Swati, and Tsonga, as shown in figure 5 (the sound is spelled as *dl* in the orthographic description). The realization of voiced lateral fricative shows voicing throughout the entire fricative (as in Xhosa and S. Ndebele), or partially (Tsonga, Swati).



Southern Ndebele **Xhosa**



Tsonga **Swati**



Zulu

Figure 5 Voiced lateral fricatives in the stem-initial position in five languages

3.2.2 Word-medial Position

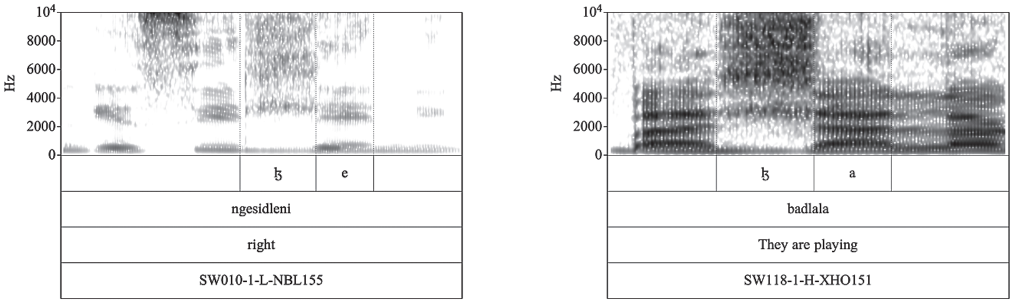
Four languages, namely Southern Ndebele, Xhosa, Zulu, and Swati have voiced lateral fricatives in the word-medial position. In our data set, Northern Sotho and Tsonga13) do not have such examples.

3.2.3 Post-nasal Position

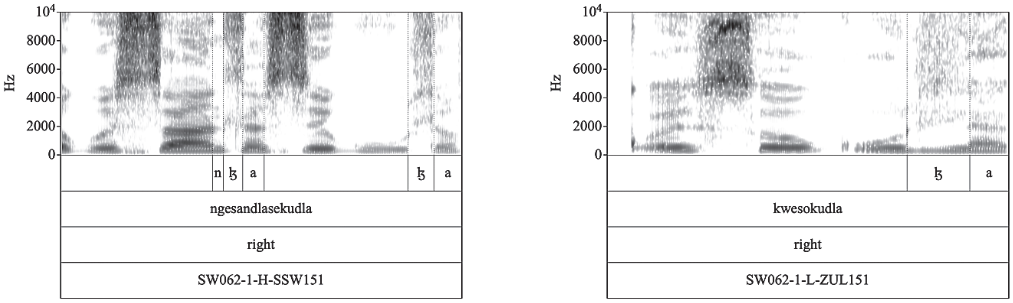
Five languages in figure 7, namely Southern Ndebele, Xhosa, Zulu, Swati, and Tsonga

---

13) Our data has an accidental gap because we only took words that appear in the Swadesh list of the languages. Examining Cuenod (1967) reveals that Tsonga has a voiced lateral fricative between vowels as in *-bádlama* ‘to lie in one’s stomach’, *swidlikí* ‘residue of fat rendering’.



Southern Ndebele **Xhosa**



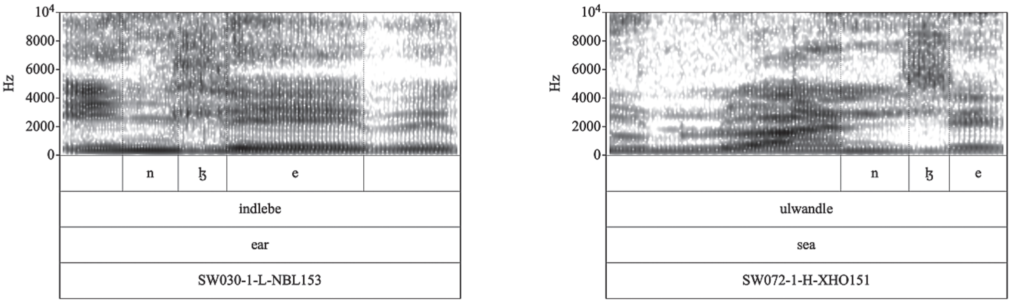
Swati **Zulu**

Figure 6 Voiced lateral fricatives in the word-medial position in four languages

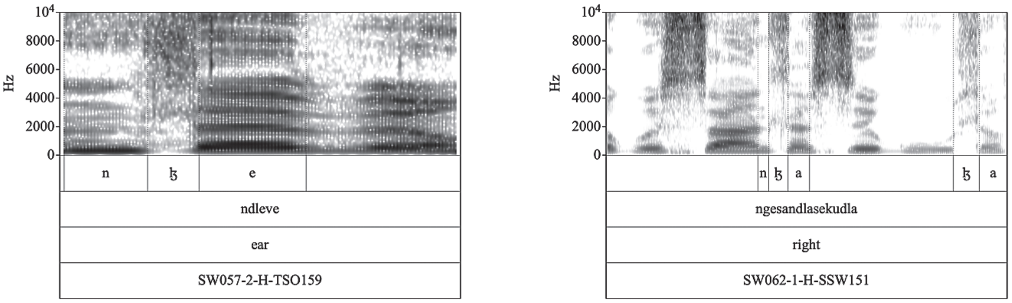
have examples with voiced lateral fricatives after a nasal consonant. Compared to the nasal consonant before a voiceless lateral fricative in figure 4, the duration of the nasal consonant is longer[14].

---

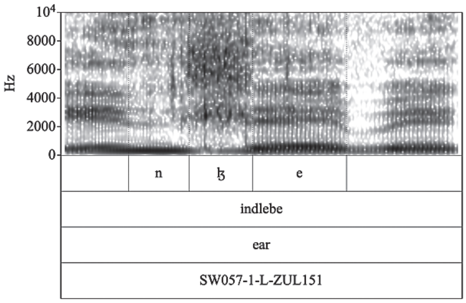
[14] The average nasal duration before a voiced lateral fricative of the five languages was 71 ms, whereas the average duration of the nasal before a voiceless lateral fricative was 49 ms. Detailed phonetic studies on the duration of the nasal consonant awaits future research.



Southern Ndebele **Xhosa**



Tsonga **Swati**



Zulu

Figure 7 Voiced lateral fricatives in the post-nasal position in five languages

3.3 Summary

The distribution of lateral fricatives in each language is shown in table 3. For both segments (voiceless or voiced), three environments are examined. For example, in our Zulu data, voiceless lateral fricative [ɬ] appears 22% of the time in the stem-initial position, 33% in word-medial position, and 11% in post-nasal position, but voiced lateral fricative appears more after a nasal (17%) than stem-initial or word-medial positions. In Northern Sotho, Southern Ndebele and Xhosa, voiceless lateral fricative does not appear in post-nasal position. The distribution of the voiceless lateral fricative [ɬ] seems to be roughly equal between stem-initial and word-medial position in Northern Sotho, which does not have the voiced lateral fricative. When both [ɬ] and [ɮ] are possible in a language, voiceless lateral

Table 3 Distribution of the voice and voiceless lateral fricatives in different Phonological environments

	[ɬ]			[ɮ]		
	#R[	V_V	N_	#R[	V_V	N_
Venda [S21]	0.00	0.00	0.00	0.00	0.00	0.00
N. Sotho [S32]	0.53	0.47	0.00	0.00	0.00	0.00
S. Ndebele [S407]	0.46	0.18	0.00	0.04	0.14	0.18
Xhosa [S41]	0.20	0.40	0.00	0.07	0.07	0.27
Zulu [S42]	0.22	0.33	0.11	0.11	0.06	0.17
Swati [S43]	0.28	0.17	0.17	0.11	0.06	0.22
Tsonga [S53]	0.42	0.26	0.16	0.05	N/A	0.11

Table 4 Implicational hierarchy about the distribution of voiced and voiceless lateral fricatives[15]

no lateral fricatives[16]	> /ɬ/[17]	> /ɬ, ɮ/
Venda [S21] (Tswana [S31])	N. Sotho [S32] (Copi [S61])	S. Ndebele [S407], Xhosa [S41], Zulu [S42], Swati [S43], Tsonga [S53]

fricative is more frequent (about 60–70% of the time) as in Southern Ndebele, Xhosa, and Tsonga. Nguni languages such as Zulu, Swati, Xhosa and Southern Ndebele show one fifth of the voiced lateral fricative appear in the word-medial position or in post-nasal position. In Tsonga, [ɬ] is found in more than 80%, with less frequency of the voiced lateral fricative.

4. Tentative Typological Generalisation

Based on the observation of six Bantu languages from zone S presented in Section 3, some salient typological tendencies about the distribution patterns of lateral fricatives can be generalised as in (1).

(1) Possible generalisation of the typological tendencies of lateral fricatives in southern Bantu

a. A language that has a voiced lateral fricatives in its phonemic inventory, it must also have a voiceless lateral fricatives and not vice versa.

b. In all languages that allow both lateral fricatives, a voiceless lateral fricatives occurs more frequently than a voiced counterpart.

c. In some languages, there is a positional restriction that a voiceless lateral fricatives does not occur in a post-nasal position.

d. In all languages that allow both lateral fricatives, voiced lateral fricatives show preference to occur in post-nasal positions.

---

15) We agree with the reviewer pointed out that this implicational relationship may only hold for Bantu languages that have lateral fricatives. We do not mean to implicate all Bantu languages in this hierarchy.

16) The absence of phonemic lateral fricatives in Venda and Tswana can also be confirmed by Nemakhavhani (2002) and Bennett *et al.* (2016), respectively.

17) The absence of phonemic voiced lateral fricatives in Copi can also be confirmed by Gowlett (2003: 615).

Table 5 Positional restrictions on a voiceless lateral fricatives in post-nasal positions

/N-ɬ/ disallowed			/N-ɬ/ allowed		
N. Sotho [S32]	S. Ndebele [S407]	Xhosa [S41]	Zulu [S42]	Swati [S43]	Tsonga [S53]

In this section, we will further discuss the natures of each generalised phenomenon and possible typological implications suggested by them.

4.1 Implicational Relation between Voiceless and Voiced Lateral Fricatives

The first generalization (1)a about the distribution of voiced and voiceless lateral fricatives within a single system, i.e., there is no such a language that has only voiced lateral fricatives and lacks voiceless lateral fricatives, can be formalised as an implicational statement in (2).

(2) If a language has a phonemic voiced lateral fricative, it also has a voiceless counterpart.

Table 4 shows actual distribution based on our observation with additional information from Naumann & Bibiko (2016).

This generalisation can be seen as a typical realisation of a wider cross-linguistic implicational tendency about voice contrast of obstruents, i.e., in a phonemic system where obstruents do not have voice contrast, it is voiceless that is expected to be present in such a system (cf. Maddieson 2013).

What is significant in the context of intra-genetic variation is that the cutting line between the presence and absence of the lesser counterpart /ɮ/ is drawn within a single group of the referential classification. It should be noted, however, that the referential classification, Guthrie's coding system is not meant to be a classification directly reflecting genealogical branching (Schadeberg 2003, Nurse & Philippson 2003b, Philippson & Grollemund 2019, among others). Nevertheless, it is still significant that the systematic variation can be observed within the same unit, S30 in this case, which is geographically distributed between southern S40 where majority of languages have both lateral fricatives and the remaining northern groups including S20 where lateral fricatives are not entirely attested.

4.2 Frequency of Voiceless Lateral Fricatives over Voiced Counterpart

The generalisation in (1)b is a qualitative tendency that in languages with both voiced and voiceless lateral fricatives, a voiceless counterpart tends to appear more frequently.

(3) In a language where both voiceless and voiced lateral fricatives are present, voiceless lateral fricatives tend to appear more frequently than the voiced counterpart.

In all languages with both lateral fricatives, more than 60% of occurrences in our sample tokens are voiceless, i.e., S. Ndebele 64%, Xhosa 60%, Tsonga 84%, Swati 61%, Zulu 67%. Whereas the sample tokens are quite limited in number, this tendency of relative frequency can be seen as a natural reflection of the cross-linguistic tendency of higher occurrence rate of a voiceless lateral fricative than a voiced one as stated in (2).

4.3 Occurrence Frequency in Terms of Post-nasal Position

As discussed in 3.1.3, a positional restriction that blocks a voiceless lateral fricative to occur in post-nasal positions is attested in some languages (1)c. This restriction can be paraphrased as a statement of implicational tendency as in (4).

(4) If a language allows a voiceless lateral fricative to occur in post-nasal positions, it also occurs in stem-initial and inter-vocalic positions.

This restriction seems to work not only in N. Sotho [S32], which only has a voiceless lateral fricatives and lacks a voiced lateral fricatives in its inventory, but also in languages with both lateral fricatives such as S. Ndebele [S407] and Xhosa [S41]. On the other hand, voiceless lateral fricatives do occur in post-nasal positions in Zulu [S42], Swati [S43], and Tsonga [S53], all of which have both types of lateral fricatives in its phonemic inventory.

Just as in the case of availability of voiced lateral fricatives addressed in 4.1, whether the restriction rule is in effect or not may vary even within a single unit of classification. Moreover, it should be worth noting that the unit in question, Nguni (S40), is clearly a group of a genetic unity rather than of mere geographical vicinity (cf. Nurse and Philippson 2003b: 169–170). In this respect, it is necessary to investigate possible factors that cause the internal variation, which, in turn, would provide us significant insights to understand the historical process of development or loss of lateral fricatives.

4.4 Relative Preference of the Post-nasal Position

On the other hand, the post-nasal is the preferred position for a voiced lateral fricative. As shown in Table 3, in all five languages in our sample that have a phonemic voiced lateral fricatives, the number of its occurrence in the post-nasal position exceeds the total numbers of occurrence in all other phonotactic environments, i.e., in word initial and word-medial positions.

(5) If voiceless fricatives appear in the post-nasal positions, voiced lateral fricatives occur in the same environment as well.

What may be directly suggested by this tendency is that lateral fricatives may also be the target of voicing effect triggered by the preceding homorganic nasal, i.e., post-nasal voicing as a typical segmental process of Bantu phonology. The typological tendency in (5) accounts for the pattern in Southern Ndebele and Xhosa since voiceless lateral fricatives do not occur after the nasal even though the voiceless version can occur elsewhere. In Zulu, Swati and Tsonga, however, the statement in (5) seems to be strong because voiceless lateral fricatives do occur after nasals. Nasals can be followed by a voiceless sound as it was shown during a debate about clusters when a nasal is followed by an obstruent.

Pater (1999) proposes the constraint against a sequence of nasal and a voiceless obstruent (*NC̥), which is based on observations where Australian languages satisfy the constraint by deleting the nasal, by denasalizing the nasal or by voicing the post-nasal obstruent. Responding to this proposal, Hyman (2001) raises doubts and suggests that the *NC̥ constraint is not phonetically grounded because it is possible to have voiceless obstruents after a nasal such as in Sotho-Tswana. Follow-up studies (Solé *et al.* 2010 on Shekgalagari, and Gouskova *et al.* 2011 on Tswana) argue that the post-nasal devoicing can be explained by phonetically grounded constraints. If lateral fricatives behave akin to obstruents, the distribution of voiceless and voiced lateral fricatives in Table 3 is compatible with (5); the presence of voiceless lateral fricative after a nasal implies that voiced lateral fricatives are also allowed.

Phonemic stability of the voiced lateral fricatives may be measured by the dependency ratio of its occurrence in post-nasal positions, i.e., in the languages where the occurrence of voiced lateral fricatives depends relatively highly on post-nasal environments, e.g., Xhosa (27%) and Swati (22%), voiced lateral fricatives may be less distinctive than other phonemes. This kind of phonemic stability is also a crucial part of the phonemic status of lateral fricatives and thus would be a significant topic of further investigation to better understand the synchronic diversity as well as to shed light on the diachronic process of emergence, development, and loss of lateral fricatives in those languages.

5. Conclusion

We have discussed basic phonetic and phonotactic features of lateral fricatives as one of the relatively complex types of fricative sounds only attested in specific zone S languages across Bantu, aiming to present its typological overview. The general observation is that the lateral fricatives in selected Bantu S languages follow cross-linguistically typical distribution patterns of fricatives, or obstruents in general, e.g., there holds an implicational hierarchy that explains non-existence of a language with voiced lateral fricatives and lacking voiceless lateral fricatives, which is motivated by the general principle of the voicing contrast of obstruents, i.e., the preference of the voiceless over voiced when lacking the contrast. We also argued that the realisation of lateral fricatives can be significantly affected by positional restrictions associated with a post-nasal position where voiceless lateral fricatives can be blocked while a voiced counterpart prefers to occur.

Whereas such cross-linguistically common tendencies hold in general, we have also pointed out internal variation in terms of attested types of lateral fricatives and their phonotactic distribution. First, there is a typological continuum between languages lacking lateral fricatives at all, which include Venda [S21] and Tswana [S31] spoken in the north, and those with both a voiceless and voiced lateral fricatives, which are entire S40 languages spoken in further south. The languages having a voiceless lateral fricatives only are found sporadically in S30 (and S60 according to Naumann & Bibiko (2016)). Second, there is further variation in terms of a positional restriction by which a voiceless lateral fricatives is blocked in post-nasal positions. This restriction applies not only to N. Sotho [S32] as a language that has only a voiceless lateral fricatives, but also to those spoken in the northern area of S40, namely S. Ndebele [S407] and Xhosa [S41]. These distributional patterns can be summarised in Table 6a and 6b.

Table 6a Relative restrictedness of lateral fricatives in sample languages

N. Sotho [S32]	S. Ndebele [S407]	Xhosa [S41]	Zulu [S42]	Swati [S43]	Tsonga [S53]
highly restricted	> restricted		> no restriction		
/ɬ/	> /ɬ, ɮ/				
/N-ɬ/ disallowed			> /N-ɬ/ allowed		

Table 6b Relative restrictedness of lateral fricatives in sample languages (another visualisation)

	/ɬ/	/ɮ/	/N-ɬ/	
Venda [S21]	*	*	*	no lateral fricatives
N. Sotho [S32]	✓	*	*	highly restricted
S. Ndebele [S407]	✓	✓	*	restricted
Xhosa [S41]	✓	✓	*	
Zulu [S42]	✓	✓	✓	no restriction
Swati [S43]	✓	✓	✓	
Tsonga [S53]	✓	✓	✓	

As shown in the table, this summarised distribution pattern may suggest a cline of the typological restriction on the phonological status of the lateral fricatives across languages. At one extreme is the most liberal languages that allow both voiceless and voiced lateral fricatives without positional restrictions, while the other extreme is the language lacking lateral fricatives at all. In the middle are two categories. One is the languages where both types of lateral fricatives are attested but the positional restriction should be applied to voiceless lateral fricatives. The other is a more restricted type in that only a voiceless lateral fricative is available in the phonemic inventory.

To conclude the paper, it would be worth addressing remaining questions and directions of the further investigation. The first and foremost is to collect an extensive range of data from as many dialects/varieties as possible, in order not only to draw a more fine-grained picture of internal variation, but also to find out more precise typological parameters that better explain the attested variation, which in turn shed light a diachronic process of emergence, development, and (possible) loss of lateral fricatives.

Another crucial point is exactly about the emergence of the lateral fricatives. Maddieson & Sands (2019: 90)[18] address that the development of relatively complex fricatives including lateral fricatives can be accounted for either by innovation or by contact with surrounding languages including phylogenetically distinct KBA languages.

---

[18] "Most of the languages have relatively limited sets of fricatives of the cross-linguistically common types, although lateral fricatives (and affricates) have developed in or been borrowed into a number of the southern languages..."

Concerning the development process, Blench (2006: 3) explicitly claims that lateral fricatives must have developed as local innovation. Major argumentations are based in the fact that i) at least several realizations of lateral fricatives can be identified as regular correspondence with the specific PB phoneme */c/, suggesting that lateral fricatives emerged through a local process of internal sound change. Another piece of evidence is that interestingly lateral fricatives are missing in the huge inventory of the consonantal systems of KBA languages, i.e., there is no possibility of direct borrowing. On the other hand, as briefly mentioned in 2.2, given that the areal features may not have to be fully inherited, there remains a possibility that only relevant features, in this case those related to lateral articulation, are 'borrowed' and, in turn, triggered the emergence of the lateral fricatives by combining with other pre-existing features in the original consonantal system of the languages in question. To decipher a more detailed picture of the historical process of emergence of lateral fricatives, more extensive data are necessary to carry out an extensive investigation from historical-comparative as well as micro-typological perspectives.

References

Aunio, Lotta, Holly Robinson, Tim Roth, Oliver Stegen, and John B. Walker. 2019. "The Mara Languages JE40". (Van de Velde *et al.*, eds.), 501–532.

Baumbach, Ernst. J. M. 1987. *Analytical Tsonga Grammar*. Pretoria: University of South Africa.

Bennett, William G., Maxine Diemer, Justine Kerford, Tracy Probert, and Tsholofelo Wesi. 2016. "Setswana (South African)." *Journal of the International Phonetic Association*, 46(2): 235–246. doi: 10.1017/S0025100316000050.

Bernander, Rasmus. 2017. Grammar and Grammaticalization in Manda: An Analysis of the Wider TAM Domain in a Tanzanian Bantu Language. A doctoral dissertation presented to the University of Gothenburg.

Blench, Roger. 2006. *Lateral Fricatives in Africa and Worldwide*. m.s. available at https://rogerblench.info/Language/World/Lateral%20Fricatives.pdf

Bostoen, Koen. 2008. "Bantu Spirantization: Morphologization, Lexicalization and Historical Classification." *Diachronica*, 25(3): 299–356.

Botne, Robert. 2003. "Lega (Beya dialect) (D25)." (Nurse and Philippson, eds.), 422–449.

Botne, Robert. 2019. "Chimpoto N14." (Van de Velde, eds.), 692–732.

Chanard, Christian. 2006. Systèmes *Alphabétiques Des Langues Africaines*. Online: urlhttp://sumale.vjf.cnrs.fr/phono/ in UPSID

Clements, George N. and Annie Rialland. 2008. "Africa as a Phonological Area." *A Linguistic Geography of Africa* (Heine, Bernd and Derek Nurse, eds.), 36–85, Cambridge: Cambridge University Press.

Crane, Thera. 2019. "Totela K41. Vwanji G66." (In Van de Velde, eds.), 615–644.

Elderkin, Edward D. 2003. "Herero (R31)." (Nurse and Philippson, eds.), 581–608.

Gouskova, Maria, Elizabeth Zsiga, and One Tlale Boyer. 2011. "Grounded Constraints and the Consonants of Setswana." *Lingua*, 121(15): 2120–2152. DOI: https://doi.org/10.1016/j.lingua.2011.09.003

Gowlett, Derek. 2003. "Zone S." (Nurse and Philippson, eds.), 609–638.

Guarisma, Gladys. 2003. "Kpāʔ (A53)." (Nurse and Philippson, eds.), 307–334.

Guérois, Rozenn. 2019. "Chuwabo P34." (Van de Velde *et al.*, eds.), 733–775.

Gunnink, Hilde, Natalia Chousou-Polydouri, and Koen Bostoen. 2023. "Divergence and Contact in Southern Bantu Language and Population History: A New Phylogeny in Cross-Disciplinary Perspective." *Language Dynamics and Change*, 13(1): 74–131. https://doi.org/10.1163/22105832-bja10022

Guthrie, Malcolm. 1967–1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. 4 vols. Farnborough: Gregg International.

Güldemann, Tom. 2011. "Sprachraum and Geography: Linguistic Macro-areas in Africa." *Language*

Mapping: An International Handbook of Linguistic Variation Volume 2 (Lameli, Alfred, Roland Kehrein and Stefan Rabanus, eds.), 561–593, Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110219166.1.561

Güldemann, Tom and Anne-Maria Fehn. 2017. "The Kalahari Basin Area as a 'Sprachbund' before the Bantu Expansion." *The Cambridge Handbook of Areal Linguistics* (Hickey, Raymond, ed.), 500–526, Cambridge: Cambridge University Press.

Hamann, Silke and Nancy C. "Kula. 2015." Bemba. Journal of the International Phonetic Association, 45: 61–69.

Hammarström, Harald. 2019. "An Inventory of Bantu Languages." (Van de Velde *et al.*, eds.), 17–78.

Heath, Teresa. 2003. "Makaa (A83)." (Nurse and Philippson, eds.), 335–348.

Hyman, Larry M. 2001. "The Limits of Phonetic Determinism in Phonology: *NC Revisited." *The Role of Speech Perception in Phonology* (Elizabeth Hume & Keith Johnson, eds.), 141–185, Leiden: Brill.

Hyman, Larry M. 2003. "Basaá (A43)." (Nurse and Philippson, eds.), 257–282.

Hyman, Larry M. 2019. "Segmental Phonology." (Van de Velde *et al.*, eds.), 128–149.

Janson, Tore. 1991/2. *Southern Bantu and Makua. SUGIA*, 12/13: 63–106

Kagaya, Ryohei. 2005. *A Jita Vocabulary*. Fuchu, Tokyo: ILCAA.

Kanana, Fridah Erastus. 2011. "Meru Dialects: The Linguistic Evidence." *Nordic Journal of African Studies*, 20(4): 300–327.

Kawasha, Boniface Kaumba. 2003. Lunda Grammar: A Morphosyntactic and Semantic Analysis. A doctoral dissertation presented to the University of Oregon.

Kisseberth, Charles. 2003. "Makhuwa (P30)." (Nurse and Philippson, eds.), 546–565.

Koni Muluwa, Joseph and Koen Bostoen. 2019. "Nsong (B85d)." (Van de Velde *et al.*, eds.), 414–448.

Kutsch Lojenga, Constance. 2003. "Bila (D32)." (Nurse and Philippson, eds.), 450–474.

Kutsch Lojenga, Constance. 2019. "Zimba D26." (Van de Velde *et al.*, eds.), 472–500.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford and Cambridge: Blackwell.

Lee-Kim, Sang-Im, Shigeto Kawahara, and Seunghun J. Lee. 2014. "The 'Whistled' Fricative in Xitsonga: Its Articulation and Acoustics." *Phonetica*, (71)1: 50–81. https://doi.org/10.1159/000362672

Leitch, Myles. 2003. "Babole (C101)." (Nurse and Philippson, eds.), 392–421.

Lombard, Daan. 1985. *An Introduction to the Grammar of Northern Sotho*. Pretoria: Van Schaik.

Maddieson, Ian. 2013. "Voicing in Plosives and Fricatives." (Dryer, Matthew S. and Marten Haspelmath, eds.), WALS Online (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/4, Accessed on 2023-09-23)

Maddieson, Ian and Bonny Sands. 2019. "The Sound of the Bantu Languages." (Van de Velde *et al.*, eds.), 79–127.

Maho, Jouni. 2001. "The Bantu Area: (Towards Clearing up) a Mess." *Africa & Asia*, 1: 40–49. Dept of Oriental and African Languages, Göteborg University.

Meeussen, Achille Emile. 1967. "Bantu Grammatical Reconstructions." *Africana Linguistica*, 3: 81–121.

Moran, Steven and Daniel McCloy (eds.) 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. Available online at http://phoible.org, Accessed on 2023-09-23. doi: 10.5281/zenodo.2593234

Mous, Maarten. 2003. "Nen (A44)." (Nurse and Philippson, eds.), 283–306.

Nakagawa, Hirosi. 2014. "Khoisan Comparative Phonology Questionnaire: Provisional Version." *Area and Cultural Studies, Tokyo University of Foreign Studies*, 88: 145–158.

Naumann, Christfried and Hans-Jörg Bibiko. 2016. "Southern Africa as a Phonological Area." Paper presented at "Speaking (of) Khoisan", a *Symposium Reviewing Southern African Prehistory*, Leipzig, 14–16 May 2015.

Nemakhavhani, Daniel Phuluwani. 2002. *Palatalization and Labialization in Tshivenda: A Linear and Non-linear Phonological Analysis*. M.A. thesis submitted to Stellenbosch University.

Njantcho Kouagang, Elisabeth and Mark van de Velde. 2019. "Kwakum A91." (Van de Velde *et al.*, eds.), 383–413.

Nurse, Derek and Gérard Philippson (eds.) 2003a. *The Bantu Languages*. London: Routledge.

Nurse, Derek and Gérard Philippson. 2003b. "Towards a Historical Classification of the Bantu languages." (Nurse and Philippson, eds.), 164–181.

Odden, David. 2003. "Rufiji-Ruvuma (N10, P10-20)." (Nurse and Philippson, eds.), 529–545.

Patin, Cédric, Kassim Mohamed-Soyir, and Charles Kisseberth. 2019. "Ngazidja G44a." (Van de Velde *et al.*, eds.), 591–614.

Pater, Joe. 1999. "Austronesian Nasal Substitution and Other NC Effect." *The Prosody-Morphology Interface* (René Kager, Harry van der Hulst, and Wim Zonneveld, eds.), 310–343, Cambridge: Cambridge University Press.

Petzell, Malin and Lotta Aunio. 2019. "Kami G36." (Van de Velde *et al.*, eds.), 563–590.

Philippson, Gérard and Marie-Laure Montlahuc. 2003. "Kilimanjaro Bantu (E60 and E74)." (Nurse and Philippson, eds.), 475–500.

Philippson, Gerárd and Rebecca Grollemund. 2019. "Classifying Bantu Languages." (Van de Velde *et al.*, eds.), 335–354.

Reeder, JeDene. 2019. "Pagibete C401." (Van de Velde *et al.*, eds.), 449–471.

Sands, Bonny and Hilde Gunnink. 2019. "Clicks on the Fringes of the Kalahari Basin Area." *Theory and Description in African Linguistics: Selected Papers from the 47th Annual Conference on African Linguistics* (Clem, Emily, Peter Jenks and Hannah Sande, eds.), 703–724, Berlin: Language Science Press.

Schadeberg, Thilo C. 1994–1995. "Spirantization and the 7-to-5 Vowel Merger in Bantu." *Sound Change* (Dominicy, Mark, and Didier Demolin, eds.), 73–84, Amsterdam: John Benjamins.

Solé, Maria-Josep, Larry M. Hyman, and Kemmonye C. Monaka. 2010. "More on Post-nasal Devoicing: The Case of Shekgalagari." *Journal of Phonetics*, 38(4): 604–615. DOI: https://doi.org/10.1016/j.wocn.2010.09.002

Sommer, Gabriele. 2003. "Western Savannah (K, R)." (Nurse and Philippson, eds.), 566–580.

Van Acker, Sifra and Koen Bostoen. 2022. "Inheritance and Contact in the Genesis of Gisamba (Bantu, L12a, DRC): A Diachronic Phonological Approach." *Linguistique et langues africaines* 6: 73–129. DOI: https://doi.org/10.4000/lla.299

Van de Velde, Mark, Koen Bostoen, Derek Nurse, and Gérard Philippson (eds.) 2019. *The Bantu Languages* Second edition. London: Routledge.

Van der Veen, Lolke J. 2003. "The B30 Language Group." (Nurse and Philippson, eds.), 371–391.

Walker, Rachel, Dani Byrd, and Fidèle Mpiranya. 2008. "An Articulatory View of Kinyarwanda Coronal Harmony." *Phonology*, 25(3): 499–535. https://doi.org/10.1017/S0952675708001619

Wilhelmsen, Vera. 2019. "Mbugwe F34." (Van de Velde, eds.), 533–562.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Acoustic Characteristics of Voiced and Voiceless Nasals in Mizo*

SARMAH, Priyankoo
LALHMINGHLUI, Wendy

Tibeto-Burman languages are characterized by voicing distinction in the nasal phonemes and in recent years, several languages of the family have reported the phenomenon. Nevertheless, in-depth studies on the voicelessness in nasal consonants are still scarce. Mizo (ISO: LUS) is a Tibeto-Burman language primarily spoken in the state of Mizoram of North-East India which also demonstrates the existence of voicing distinction phonemically. The Mizo nasals, both voiced and voiceless, can occur in three places of articulation: bilabial, alveolar and velar. In the current study, we analyze speech data collected from 10 Mizo speakers producing the voiced and voiceless nasals in isolation and in CV syllables containing all the five Mizo vowels. The first four nasal formants (N1-N4), their bandwidths (BW1-BW4), spectral Center of Gravity (CoG) and Standard Deviation (SD) are used to characterize the Mizo voiced and voiceless nasals. Apart from that, the effect of the voiced and the voiceless nasals is also investigated by means of nasality estimation in the vowels immediately following the voiced and voiceless nasals. Nasality measures, A1P0, A1P1 and A3P0 are performed on the vowels following the voiced and voiceless nasals. The results showed that nasal formants N1-N4, BW1 are reliable measures for distinguishing voiced and voiceless nasals. In terms of nasality, the vowels following the nasals are not significantly influenced by voicing differences. Rather than voicing, in case of Mizo, A1P0, A1P1 and A3P0 are indicators of the place of the articulation of the nasals.

Keywords: Mizo, Lushai, Speech sounds, Nasals, Voiceless nasals

DOI: https://doi.org/10.15026/0002000629

* The authors would like to thank the anonymous reviewers for their suggestions. The authors would also like to acknowledge the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021–2023), which facilitated several discussions that helped in the preparation of this manuscript.

Table 1 List of words elicited from Mizo speakers contrasting in voicing in nasals.

Sl. No.	Voiced		Voiceless		Tone
	IPA	*Meaning*	*IPA*	*Meaning*	
1.	maŋ	*burnt wood*	m̥aŋ	*to use*	High
2.	meʔ	*shave*	m̥eʔ	*dish*	Low
3.	muʔ	*roasted*	m̥uʔ	*to show*	Low
4.	nan	*to come upon*	n̥an	*over*	Falling
5.	nu	*mother*	n̥u	*a print*	Falling
6.	nam	*push*	n̥am	*clan, tribe, nation*	Rising
7.	nɔʔ	*to rub, wipe*	n̥ɔʔ	*to fill*	Low
8.	nur	*sad expression*	n̥ur	*plenty of*	High
9.	ŋal	*shin*	ŋ̊al	*naughty*	High
10.	ŋaʔ	*to have much*	ŋ̊aʔ	*to wait*	Low

1. Introduction

Mizo is a Tibeto-Burman language of the Kuki-Chin subfamily, primarily spoken in the state of Mizoram in North-East India. Apart from India, it is also spoken by a sizable population in Myanmar and a small population in the Chittagong Hill Tracts region of Bangladesh. In India, according to the population census of 2011, the number of Mizo speakers in India is 825,900 (Registrar General & Census Commissioner 2011: 10).

Mizo is a tone language, and a substantial number of discussions are dedicated towards the tonal phenomena in Mizo (Sarmah and Wiltshire 2010). Apart from the suprasegmental features, the segmental features of Mizo, specifically the nasals, are worth exploring. Mizo distinguishes phonemically depending on the voicing feature of the nasals, as illustrated in Table 1. Bhaskararao and Ladefoged (1991) investigate the voiceless nasals in Mizo and conclude that phonetically they are similar to the ones produced in Burmese where the voiceless counterpart of the voiced nasal is preceded by a preaspiration. This prompts Bhaskararao and Ladefoged (1991) to argue that the location of aspiration in voiceless nasals may be different in languages. Bhaskararao and Ladefoged (1991) proposed two types of voiceless nasals in languages, namely the Burmese type (preaspirated voiceless nasal) and the Angami type (aspirated voiceless nasal). They have confirmed with acoustic data that in Khonoma Angami voiceless nasals, aspiration follows the voiceless nasals, unlike in Mizo or Burmese. This also prompted several researchers to argue that the nasal distinction in voicing is not between voiced and voiceless, per se, but between (pre) aspirated and unaspirated nasals (Breit 2013).

Voiceless nasals are typologically rare, and their distribution is limited to only a few regions in the world. The Tibeto-Burman languages are well known for their inventory of voiceless nasals. Five different subgroups of the Tibeto-Burman language family—Lolo-Burmese (Burmese, Achang, Nuosu Yi), Qiangic (Xumi, Pumi, Zhaba), Bodish (Chepang, Dhimal, Drenjongke, and Kham Tibetic), Nungish (Anong), and Kuki-chin (Angami, Chokri, Khezha, Lushai, Lai, Laizo, Kom Rem, etc.) are reported to have voicing contrasts in nasals (Matisoff 2003; Guillemot and Lee, this volume; Edmondson *et al.* 2017). In addition to the Tibeto-Burman languages, Welsh and Ikema Ryukyuan have also reported nasal voicing contrasts (Ford 2016; Hammond 2019).

The Burmese voiceless nasals are relatively well studied, and they are characterized as made of a voiceless nasal friction and a voiced nasal region (Dantsuji 1984; Dantsuji 1986). Bhaskararao and Ladefoged (1991) also report that the Burmese voiceless nasals are initially voiceless but voiced towards their termination. As stated previously, Burmese nasals differ from the Angami or Xumi voiceless nasals that have entirely voiceless nasal portions followed by aspiration (Terhiija and Sarmah 2020; Chirkova *et al.* 2019).

Burmese is one of the few languages where acoustic properties of the voiceless nasals have been studied in-depth (Dantsuji 1984). In his acoustic study of Burmese nasals Dantsuji reports several prominent peaks of energy in the spectrum of the voiceless nasal friction portion. He also reports that the voiced portions of the voiceless nasals are shorter than the voiced nasal stops. In a follow up study, Dantsuji (1986) analyzes the acoustic properties of the nasal murmur spectra (nasal part of the voiceless nasal), namely, the first, second, third nasal formants, known as N1, N2 and N3, respectively and their bandwidths (B1, B2 and B3). He used these features to find cues for place of articulation for voiceless nasals. He reported that cues for places of articulation could be signalled by the nasal formants and their bandwidths in the nasal portion of the voiceless nasals. These features were later used by Tabain *et al.* in their study of the voiced nasal consonants in three central Australian languages where they reported that nasal formants, their bandwidths along with the Center of Gravity (CoG) and Standard Deviations (SD) calculated from the nasal spectrum provided reliable characterization of nasals with different places of articulation (Tabain *et al.* 2016). In one such study on Xumi and Burmese the differences in the rate of voicing for voiced and voiceless nasals were also investigated and they were found to be significantly different (Chirkova *et al.* 2019). In two recent studies on the properties of voiceless nasals, Nasometer II was used to show the fine-grained acoustic characteristics of the nasals in Angami and Mizo (Terhiija and Sarmah 2020; Lalhminghlui and Sarmah 2021).

Following the previous studies such as Dantsuji (1984, 1986) and Tabain *et al.* (2016), in the current study we investigate the first four nasal formants (N1 - N4) and their bandwidths (BW1 - BW4) and the Standard Deviation (SD) and the Center of Gravity (CoG) of the nasal spectrum of the voiced and voiceless nasals in Mizo. It is to be noted that the temporal characteristics of Mizo voiced and voiceless nasals are discussed in a previous work (Lalhminghlui and Sarmah 2021) and hence, in the current work, the focus is laid only on the spectral properties of the voicing contrast in Mizo nasals.

In the current work, we also investigate the effect of voiced and voiceless nasals on following vowels in terms of vowel nasality. Vowel nasality occurs when the velum is lowered during vowel production, which opens the velopharyngeal port and allows air to flow through the nose and mouth (Chen 1997; Styler 2017). While, nasalization in vowels maybe phonemic, as in French or Hindi, vowels may also be nasalized due to coarticulatory effects of the preceding nasal (Malécot 1960; Ohala & Ohala 1975; Maeda 1993; Delvaux *et al.* 2002). Hence, we assume that vowels in Mizo followed by nasals will be nasalized, and nasalization will be different for voiced and voiceless nasals. As we expect the effect of the nasals on the following vowels, we also investigate the vowel nasality using A1-P0, A1-P0 compensated, A1-P1, A3-P0 (Styler 2013).

In case of the nasality measures, A1 is the amplitude of the first formant (F1), and P0 is the amplitude of the nasal peak that occurs around 250 Hz. Hence, as nasality increases, P0 rises and A1 lowers, resulting in lower A1-P0 values (Styler and Scarborough 2014). In case of the A1-P1 measure, P1 is the second nasal pole that appears between 790 to 1100 Hz (Chen 1995). The third measure of spectral tilt, A3-P0 compares the amplitude of the third nasal formant with P0. Previous studies have shown that these three measures have language-specific and speaker-specific properties (Styler 2017). The measures associated with the nasal consonants, namely, nasal formants (N1-N4) and their corresponding bandwidths (BW1-BW4) had strong correlation with the place of articulation of the nasals (Tabain *et al.* 2016). Similarly, the spectral features, Center of Gravity (CoG) and standard deviation of the nasal spectra showed place of articulation effects.

2. Methodology

The current study is an experimental acoustic phonetic study where we have collected a corpus of speech data with instances of Mizo voiced and voiceless nasals produced by 10 Mizo speakers. The Mizo speakers were instructed to produce the nasals by reading from a printed text material in a controlled environment. The details of the procedures followed in this study are provided in the subsections to follow.

2.1 Materials

The text material to collect nasal consonant production contained 20 words of Mizo with lexical meanings. As seen in Table 1, each word had a minimally contrasting counterpart in terms of voicing of the nasal. As Mizo is a tone language (see Sarmah *et al.* 2015) it was made sure that each pair of syllables contrasting in voicing of the nasals had the same lexical tone assigned. Each of the words was pronounced three times by each speaker, which should result in (20 words x 3 repetitions x 10 speakers) 600 total iterations. However, as some of the speakers repeated a couple of more repetitions, the final count of iterations to be analyzed stood at 604.

2.2 Speakers

Ten Mizo speakers (5 male and 5 female) participated in the production experiment. The speakers were provided the wordlist with their meanings in English, so that there is no ambiguity about the tonal affiliation of the syllables. All ten Mizo speakers were native speakers of the language, and the average age was 26 (age range: 16–40) years. Out of the ten participants, five hold a bachelor's degree in various disciplines including arts, science and engineering; three obtained master's degrees in business administration and arts; and two were in higher secondary schools. All of them could understand English and none of them reported any problem with speech production.

2.3 Recording, Annotation, and Analyses

The speech data was recorded in a noiseless environment using a Tascam DR 100 MK II solid state recorder. Data was acquired using a cardioid close talk microphone (Shure SM10A). The recorded data was transferred from the SD card in the recorder to a laptop computer for annotation. Praat 6.1.12 was used to read and annotate the speech data (Boersma and Weenink 2023). Speech data was annotated at the phoneme level and the annotations were saved in the TextGrid format for further automatic analyses. Later a Praat script was run to extract the first four nasal formants (N1 - N4) at every 10% of the total duration of the nasal consonant. For each formant, the nasal bandwidth and Standard Deviation (SD) was also extracted. For estimating nasality using A1P0, A1P1 and A3P0 measures, Styler (2017)'s *nasality automeasure* script was used with minor modifications. All the extracted values were saved to a spreadsheet for visualization and for statistical measurements. The raw values of the acoustic measures were subjected to Linear Mixed Effects analysis using the *lme4* package on the R statistical software (Bates *et al.* 2015; R core team 2022). Further each model was subjected to an analysis of deviance test to see the size of the various effects using the *car* and *emmeans* packages (Fox and Weisberg 2019; Lenth 2022).

3. Results

3.1 Acoustics of Nasal Consonants

In the following sections we first provide the results of the acoustic analyses on the nasal consonants using the methodology described in section 2.3. Following that we provide the results of the acoustic analyses on the nasality of the vowels following voiced and voiceless nasals. Apart from the visualizations, the following sections are complemented by statistical analyses of the acoustic parameters used to assess nasals and nasality.

3.1.1 Nasal Formants and Bandwidths

The plot in Figure 1 shows the formant trajectories of the nasal formant measures across the duration of the nasal consonants. The plots are made with nasal formants calculated at every 10% of the total duration of the nasals. Visually, the voice and voiceless distinction is evident only in N1 and N3. A series of LME models were performed with the nasal formant values at the midpoint as dependent variables and nasal phonemes (6 nasals) as fixed effects and speaker and word type as random effects. The models were subjected to an analysis of deviance (ANOVA) using type II Wald chi-square tests that showed that phonemes had a significant effect on the nasal formant values. The results of the ANOVA tests are provided in Table 2. Additionally, we also observed that the random effect, speaker,

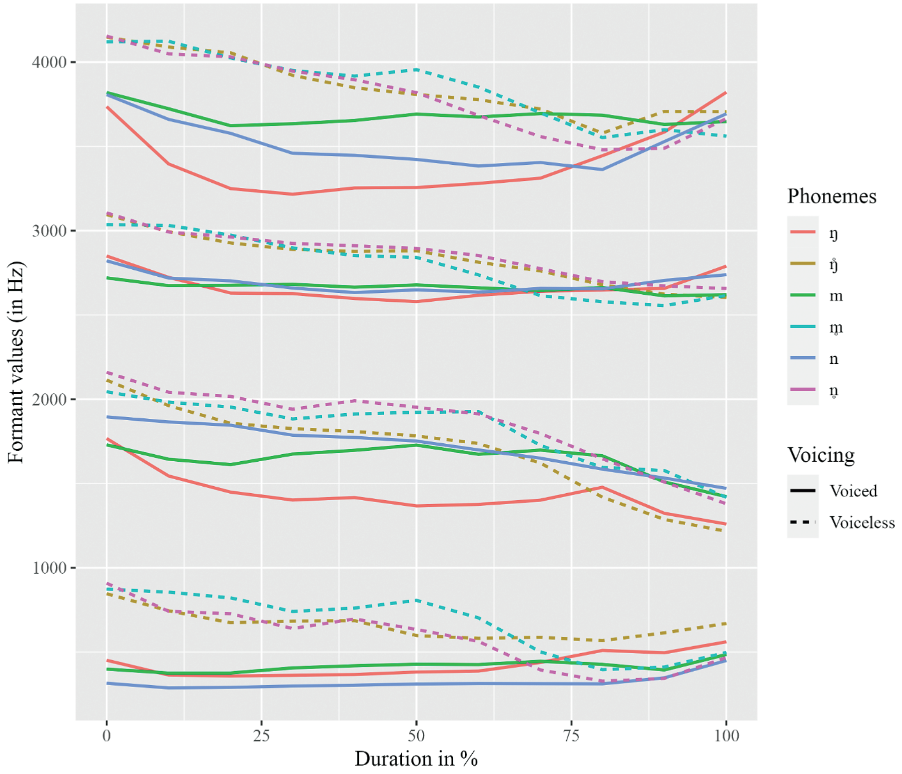


Figure 1 The first four nasal formants across the duration of the nasal consonants.

Table 2 Chi-square and p values of ANOVA conducted on LME models for four nasal formants.

Formants	χ^2	df	p-value
N1	131.2	5	<0.001
N2	104.8	5	<0.001
N3	76.2	5	<0.001
N4	137.0	5	<0.001

Table 3 Estimates and p values of pairwise comparisons of nasal formant values at midpoint.

	N1		N2		N3		N4	
Contrasts	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
m-m̥	−353.86	<0.001	−264.13	<0.001	−210.77	<0.001	−271.30	<0.01
n-n̥	−394.88	<0.0001	−275.44	<0.0001	−302.77	<0.0001	−404.60	<0.0001
ŋ-ŋ̊	−306.07	<0.05	−478.97	<0.0001	−355.75	<0.0001	−599.30	<0.0001

had a significant effect on the nasal formants, indicating that nasal formants may have speaker-specific effects that may need to be normalized as in the case of vowel formants. To see the significance of difference between the voiced and the voiceless nasals, we subjected the LME models to a post-hoc Bonferroni analysis and the results are provided in Table 3. The results show that all the four nasal formants have significant effect of the nasal types.

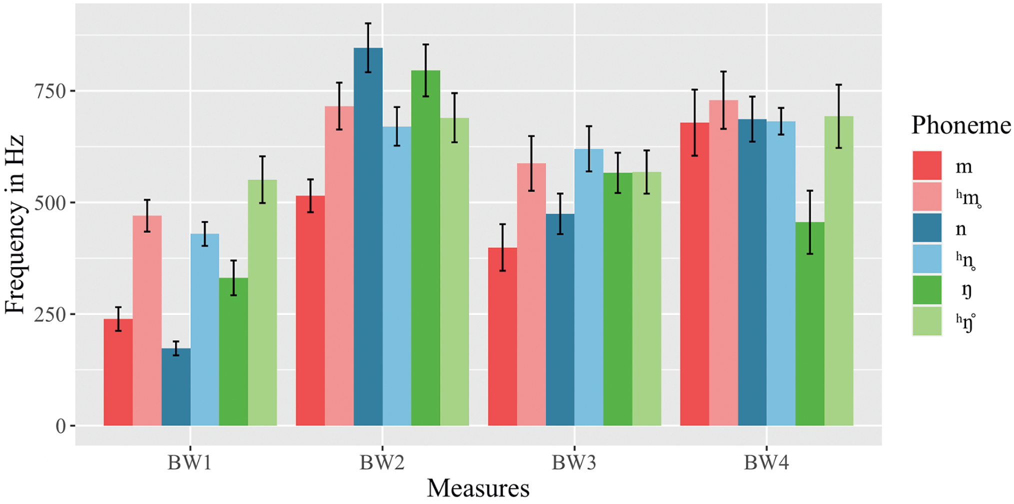


Figure 2 Four nasal formants bandwidth averaged across the duration of the nasal consonants.

Table 4 Chi-square and p values of ANOVA conducted on LME models for nasal bandwidths.

Bandwidths	χ^2	df	p-value
BW1	90.7	5	<0.001
BW2	20.8	5	<0.001
BW3	10.0	5	0.07
BW4	11.0	5	<0.05

Table 5 Estimates and p values of pairwise comparisons of nasal formant bandwidths.

Contrasts	BW1		BW2		BW3		BW4	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
m-m̥	−232.00	<0.0001	−201.50	0.13	−187.70	0.21	−50.36	1.00
n-n̥	−256.60	<0.0001	175.40	0.05	−144.77	0.22	6.54	1.00
ŋ-ŋ̊	−220.00	<0.001	105.90	1.00	−1.82	1.00	−252.98	0.19

Additionally, it is also confirmed that for all places of articulation, the distinction between voiced and voiceless nasals is captured well by the nasal formants.

In case of nasal formant bandwidths, Figure 2 shows that there is a noticeable difference in the average first formant bandwidth (BW1) of the nasal formants. The first formant bandwidth of the voiceless nasals is consistently larger than the voiced nasals. However, such consistent differences are not noticed in case of the higher bandwidths (BW2 to BW4). To confirm the assumptions, a series of LME models were built with the nasal bandwidth as dependent variables and nasal phonemes as fixed effects, speaker and word type were introduced as random effects. As seen in Table 4, the LME models for BW1, BW2 and BW4 showed statistical significance.

Table 5 showed the results of a Bonferroni post-hoc test where contrasts between voiced and voiceless nasals were considered. As the table shows, only BW1 showed significance

Table 6 Chi-square and p values of ANOVA conducted on LME models for six Mizo nasals.

Dependent variable	χ^2	df	p-value
SD	17.45	5	<0.01
CoG	76.74	5	<0.001

differences in terms of voicing in nasals. As mentioned earlier, the voiceless nasals had consistently higher bandwidths than their voiced counterparts. Higher bandwidth indicates that the signal is dampened, and it is less sinusoidal in nature. Hence, it is not surprising that the voiceless nasals have larger bandwidths as they primarily consist of nasal and oral friction and are devoid of voiced nasal-specific periodic waveforms (Lalhminghlui and Sarmah 2021).

3.1.2 Nasal Spectrum SD and Center of Gravity (CoG)

Spectral based measures, Standard Deviation (SD) and Center of Gravity (CoG) represent the distribution of energy in the nasal spectrum. Spectral CoG shows the center of mass of the speech spectrum. In case of aspirated and unaspirated consonants, it was shown that the CoG values of aspirated consonants differ significantly from their unaspirated counterparts (Rabha *et al.* 2019). Occurrence of fricative-like characteristics usually lead to a difference in CoG values. Tabain *et al.* 2016 showed that Center of Gravity (CoG) and standard deviation of the nasal spectra showed place of articulation (PoA) effects in nasal consonants. However, in case of the Mizo nasals, as noticed in Figure A-1 and Figure A-2 in Appendix A, there is no discernable difference in nasal CoG or SD in terms of voicing of the nasals. LME models constructed with CoG and SD as dependent variables and nasal types (six nasals) as fixed effects, showed significance ($p<0.01$ and $p<0.001$, respectively) as in Table 6. However, post-hoc Bonferroni tests showed no significant difference in terms of voicing among the homorganic nasals as seen in Table A-1, Appendix A. Hence, it can be concluded that while CoG and SD of the nasal spectrum may have POA-specific values, voicing contrast does not lead to any difference in terms of CoG and SD. It may be noted that CoG and SD are considered reliable measures for POA, specifically for consonants involving friction, such as fricatives (Zygis *et al.* 2015, McCarthy 2019).

3.2 Acoustics of Vowel Nasality

In this section, we provide the results of spectral tilt measures, comparing the amplitudes of harmonics with nasal poles. Primarily, A1P0, A1P1 and A3P0 measures were used to determine the amount of nasalization on the vowels following the Mizo nasals. The extracted A1P0, A1P0 (compensated)[1], A1P1, A1P1 (compensated) and A3P0 on the vowels following the nasals are plotted in Figure 3 to Figure 5. As seen in the figures, the measures seem not to vary according to the voicing of the nasals but according to the place of articulation of the nasals.

---

[1] The compensated measures were proposed by Chen (1997) for adjusting A1-P0 and A1-P1 based on formants and bandwidth.

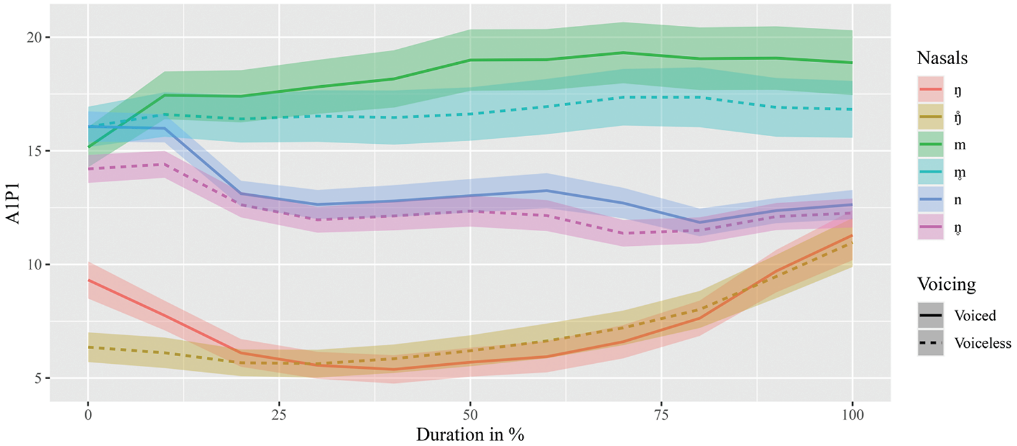


Figure 3 Nasality measured with A1P1 in the vowels following nasals.

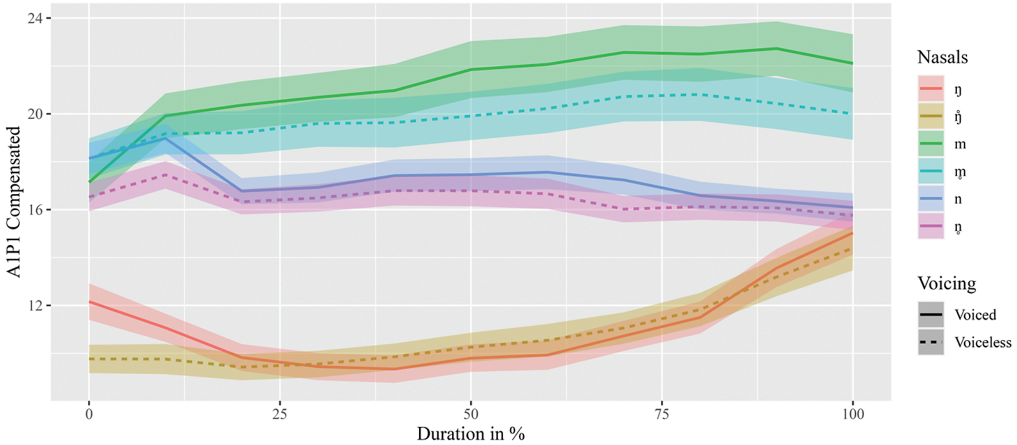


Figure 4 Nasality measured with A1P1 (compensated) in the vowels following nasals.

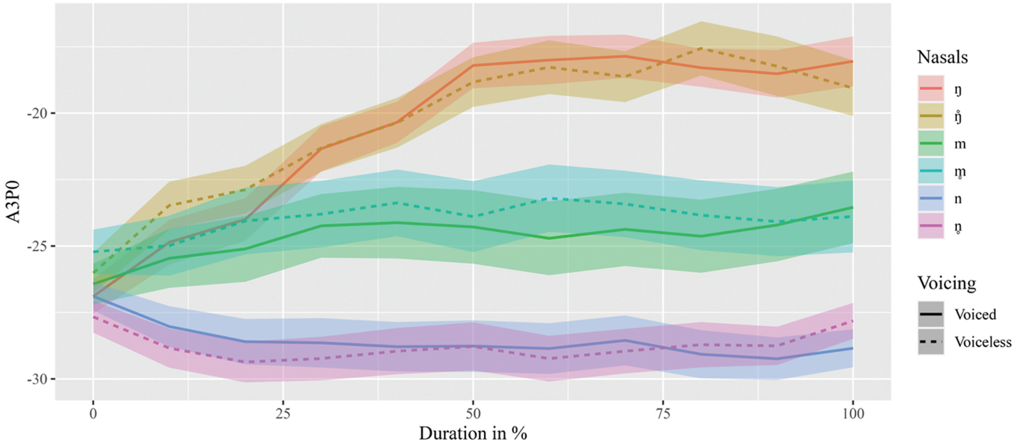


Figure 5 Nasality measured with A3P0 in the vowels following nasals.

Table 7 Chi-square and p-values from Type II Wald χ^2 tests conducted on LME models.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
AIP0	09.7	17.8*	13.6*	4.2	0.5	4.5	8.3	12.2*	19.0*	28.2*	27.1*
AIP1	94.6*	198.1*	199.1*	172.0*	158.8*	146.5*	146.0*	152.9*	124.8*	84.6*	41.7*
A3P0	10.1	59.3*	86.8*	118.3*	139.1*	156.8*	153.0*	149.6*	131.6*	111.0*	90.8*

As seen in Figure A-3 and Figure A-4 in Appendix A, A1P0, A1P0 (compensated) values across the total duration of the vowels do not seem to be different until about the 80% of the total duration. In case of A1P1 and A1P1 (compensated) as seen in Figure 3 and Figure 4, we can see the A1P1 values are good discriminators of place of articulation but not so for voicing. Finally, for A3P0 values too, there is better discrimination for place of articulation than for voicing (Figure 5).

To confirm the assumptions made from the Figure 3 to Figure 5, we created separate LME models with A1P0 (compensated), A1P1 (compensated) and A3P0 as the dependent variables and six nasal consonants (/m/, /m̥/, /n/, /n̥/, /ŋ/, /ŋ̊/) as fixed effects. Speaker and tones were added as random effects. For each nasality measure, there were eleven dependent variables, as nasality values were extracted from every 10% of the total duration of the vowels following the nasals. Each model was subjected to Type II Wald χ^2 tests, so that effect of the fixed effect in terms of *p*-values can be obtained. Table 7 shows the χ^2 values and statistical significances indicated by asterisk. While A1P1 and A3P0 showed significance almost through the whole duration of the vowels, post-hoc tests showed that the nasal pairs did not have any significant difference in terms of voicing. The significant χ^2 values were due to the significant interaction of place of articulation with the nasality measures. To substantiate this, a series of LME tests were conducted with the nasality measures (A1P0, A1P0 compensated, A1P1 compensated and A3P0) in 11 equidistant points in the vowels as dependent variables. Speaker and tones were kept as random effects. All the 55 models were subjected to ANOVA tests to determine the statistical significance of the fixed effects and their interactions. The results of the ANOVA tests are reported in Table A-2 in Appendix A, with significant effects indicated with an asterisk.

The results in Table A-2 for the ANOVA tests on the LME models clearly show that A1P0 and A1P0 compensated do not interact systematically with any of the fixed effects. On the other hand, A1P1, A1P1 compensated and A3P0 significantly interact with the POA of the nasals without any interaction with the voicing of the nasals. Hence, in case of Mizo nasals, A1P1 and A3P0 can be considered robust correlates of place of articulation.

4. Discussion and Conclusion

This study has attempted to characterize the voiced and voiceless nasals in Mizo by using spectral features collected from the nasal consonant and following vowels. In this study, spectral characteristics of the nasal consonants and of the vowels following them were investigated. It was expected that due to coarticulatory effects, the vowels following nasals will contain characteristics which may also give information about the voicing and place of articulation of the preceding nasals. In case of the measurements conducted on the nasal consonants, we show that both nasal formants and their bandwidths provide information about the voicing of the nasals. While statistical tests have confirmed a significant effect of voicing on the nasal formants, visually, the first and the third nasal formants (N1 and N3) seem to be distinct in terms of voicing of the nasals. The first formant bandwidth (BW1) showed systematic differences between voiced and voiceless nasals. The higher formant bandwidths did not show any systematic difference in terms of voicing of the nasals. On the other hand, CoG and SD of the nasal spectrum did not provide any clues about the voicing of the nasal consonants.

In terms of the nasal formants, one interesting observation is regarding the merger of the voiced and voiceless specific formants from about 75% of the total duration of the nasals, until their end. This is possibly because voiceless nasals tend to be voiced towards the end of its production as shown in Bhaskararao and Ladefoged (1991) and Lalhminghlui and Sarmah (2021). Lalhminghlui and Sarmah (2021) have shown in a nasometric study that the nasalance values at the end of the voiceless nasal is very similar to that of a voiced nasal. This is probably the reason why from about 75% of the total duration of the nasal formant, voiced and voiceless nasal formants start merging.

The current study also investigated the amount of nasality in the vowels following the voiced and voiceless nasals in Mizo. It was shown in Terhiija and Sarmah (this volume) that in terms of A1P0, A1P1 and A3P0, the nasality in the vowels following voiced and voiceless nasals is different. However, there is a fundamental difference between the voiceless nasals in Angami and in Mizo. While Mizo has preaspiration in the voiceless nasals, the Angami voiceless nasals are characterized by post-aspiration. Hence, the differences arising in the following vowel due to post-aspiration are acoustically significant in case of Angami. At the same time, it is not the case that preaspiration is the only feature that characterizes Mizo voiceless nasals. As shown earlier (Lalhminghlui and Sarmah 2021), Mizo nasals have substantial voiceless nasal murmur, visible in nasal waveforms. On the other hand, as mentioned earlier, towards the end of the Mizo voiceless nasals, they tend to be more akin the voiced nasals. Therefore, the vowels following the Mizo nasals may not have any voicing specific differences permeate into them. Considering that, it is not surprising that in the current study we did not find any difference in terms of voicing in the nasality measures extracted from the vowels following Mizo nasals. It also makes us realize that the difference between the languages with preaspirated and the languages with postaspirated voiceless nasals is very fundamental and robust. In that sense, the measures for voiceless nasals may need to be specific and in some cases language specific.

References

Bates, Douglas, Maechler Martin, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1): 1–48.

Bhaskararao, Peri and Peter Ladefoged. 1991. "Two Types of Voiceless Nasals." *Journal of the International Phonetic Association*, 21(2): 80–88.

Boersma, Paul and David Weenink. 2023. "Praat: Doing Phonetics by Computer [Computer program]." Version 6.3.14, retrieved 4 August 2023 from http://www.praat.org/

Breit, Florian. 2013. "Voice–Nasality Interaction and Headedness in Voiceless Nasals." *UCL Working Papers in Linguistics*, 25: 1–21, University College London.

Chen, Marilyn Y. 1995. "Acoustic Parameters of Nasalized Vowels in Hearing Impaired and Normal-hearing Speakers." *The Journal of the Acoustical Society of America*, 98: 2443–2453.

Chen, Marilyn Y. 1997. "Acoustic Correlates of English and French Nasalized Vowels." *The Journal of the Acoustical Society of America*, 102(4): 2360–2370.

Chirkova, Katie, Patricia Basset, and Angélique Amelot. 2019. "Voiceless Nasal Sounds in Three Tibeto-Burman Languages." *Journal of the International Phonetic Association*, 49(1): 1–32.

Dantsuji, Masatake. 1984. "A Study on Voiceless Nasals in Burmese." 音声科学研究 *(Studia Phonologica)*, 18: 1–14.

Dantsuji, Masatake. 1986. "Some Acoustic Observations on the Distinction of Place of Articulation for Voiceless Nasals in Burmese." 音声科学研究 (*Studia Phonologica*), 20: 1–11.

Delvaux, Véronique, Didier Demolin, Bernard Harmegnies, and Alain Soquet. 2008. "The Aerodynamics of Nasalization in French." *Journal of Phonetics*, 36(4): 578–606.

Edmondson, Jerold A., John H. Esling, and Lama Ziwo. 2017. "Nuosu Yi." *Journal of the International Phonetic Association*, 47(1): 87–97.

Ford, Catherine. 2016. *The Status of Voiceless Nasals in Ikema Ryukyuan*. BA Thesis, Department of Linguistics and Department of East Asian Studies, University of Alberta, Edmonton, Alberta.

Fox, John and Sanford Weisberg. 2019. An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion

Guillemot, Celeste and Seunghun J. Lee. (in press). "Phonetic variation of voiceless nasals in Drenjongke." *Journal of Asian and African Studies, Supplement.*

Hammond, Michael. 2019. "Voiceless Nasals in Welsh." *Journal of Celtic Linguistics*, 20(1): 31–60.

Lalhminghlui, Wendy and Sarmah, Priyankoo. 2021. "Characterizing Voiced and Voiceless Nasals in Mizo." *Proceedings of INTERSPEECH 2021*: 426–430.

Lenth, Russell. 2022. *emmeans*: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.1-1, retrieved on August 16, 2023 from: https://CRAN.R-project.org/package=emmeans.

Maeda, Shinji. 1993. "Acoustics of Vowel Nasalization and Articulatory Shifts in French Nasal Vowels." In *Nasals, Nasalization, and the Velum*, 147–167, Academic Press.

Malécot, André. 1960. "Vowel Nasality as a Distinctive Feature in American English." *Language*, 36(2): 222–229.

Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. University of California Press.

McCarthy, Daniel Timothy Pio Denis. 2019. *The Acoustics of Place of Articulation in English Plosives*. Doctoral dissertation, Newcastle University.

Ohala, John J. and Manjari Ohala. 1975. "Nasals and Nasalization in Hindi." *Nasalfest: Papers from a Symposium and Nasals and Nasalization* (Charles A. Ferguson, Larry M. Hyman and John J. Ohala, eds.), 317–332, Stanford: Language Universals Project, Dept. of Linguistics, Stanford University.

Rabha, Saswati, Priyankoo Sarmah, and S. R. Mahadeva Prasanna. 2019. "Acoustic Correlates of Aspiration in Fricatives and Nasals." In *TENCON 2019–2019 IEEE Region 10 Conference*, 986–990, IEEE.

R Core Team. 2022. "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing*, Vienna, Austria. URL: https://www.R-project.org/.

Registrar General and Census Commissioner India, Office of the. 2011. "Statement-1, Part-B, Languages not specified in the eighth schedule (Non-Scheduled Languages)." *The Population Census of India 2011.*

Sarmah, Priyankoo and Caroline R. Wiltshire. 2010. "A Preliminary Acoustic Study of Mizo Vowels and Tones." *Journal of Acoustical Society of India*, 37(3): 121–129.

Sarmah, Priyankoo, Leena Dihingia, and Wendy Lalhminghlui. 2015. "Contextual Variation of Tones in Mizo." *Proceedings of Interspeech 2015*, 983–986.

Styler, Will. 2017. "On the Acoustical Features of Vowel Nasality in English and French." *Journal of the*

Acoustical Society of America, 142(4): 2469–2482.

Styler, Will. 2013. "Using Praat for Linguistic Research." *University of Colorado at Boulder Phonetics Lab*.

Styler, Will and Rebecca Scarborough. 2014. "Surveying the Nasal Peak: A1 and P0 in Nasal and Nasalized Vowels." *The Journal of the Acoustical Society of America*, 136(4_Supplement): 2083–2083.

Tabain, Marija, Andrew Butcher, Gavan Breen, and Richard Beare. 2016. "An Acoustic Study of Nasal Consonants in Three Central Australian Languages." *The Journal of the Acoustical Society of America*, 139(2): 890–903.

Terhiija, Viyazonuo and Priyankoo Sarmah. 2020. "Aspiration in Voiceless Nasals in Angami." In *Proceedings of Meetings on Acoustics*, 42(1). AIP Publishing.

Żygis, Marzena, Daniel Pape, Luis MT Jesus, and Marek Jaskula. 2015. "The Effects of Intonation on Acoustic Properties of Fricatives." In *Proceedings of the 18th International Congress of Phonetic Sciences*, 1-4.

Appendix A

Table A-1: Estimates and p values of pairwise comparisons of nasal formant SD and CoG

	SD		CoG	
Contrasts	Estimate	p-value	Estimate	p-value
m-m̥	7.37	1.00	−3.31	1.00
n-n̥	−5.05	1.00	192.30	0.24
ŋ-ŋ̊	−74.28	1.00	211.47	1.00

Table A-2: Chi-square and p-values from Type II Wald χ2 tests conducted on LME models.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A1P0											
POA	n.s.	*	*	n.s.	n.s.	n.s.	*	n.s.	*	*	*
Voicing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
POA x Voicing	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
AIP0 (compensated)											
POA	n.s.	*	*	n.s.	n.s.	n.s.	*	n.s.	*	*	*
Voicing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
POA x Voicing	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
AIP1											
POA	*	*	*	*	*	*	*	*	*	*	*
Voicing	*	*	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
POA x Voicing	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.
AIP1 (compensated)											
POA	*	*	*	*	*	*	*	*	*	*	*
Voicing	n.s.	*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
POA x Voicing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
A3P0											
POA	*	*	*	*	*	*	*	*	*	*	*
Voicing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
POA x Voicing	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

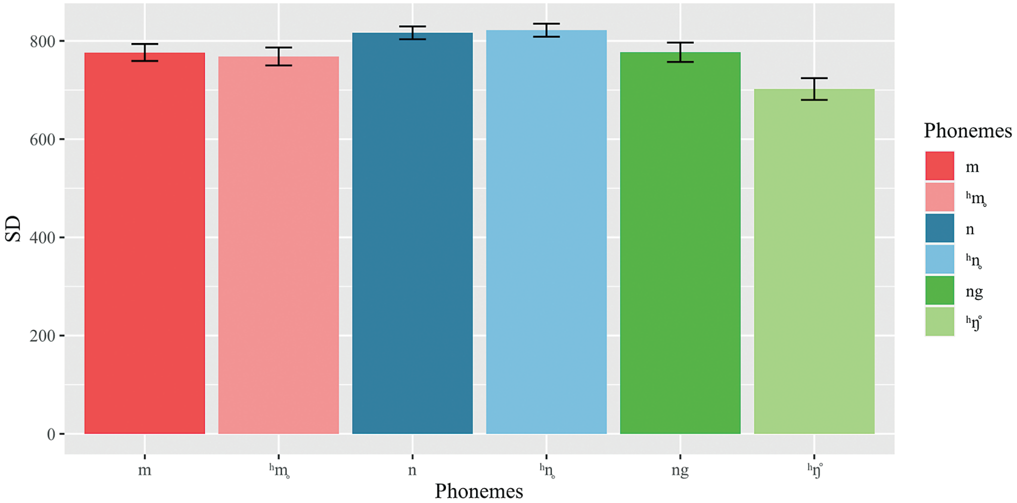


Figure A-1. Standard Deviation (SD) of the nasal spectrum at nasal midpoint

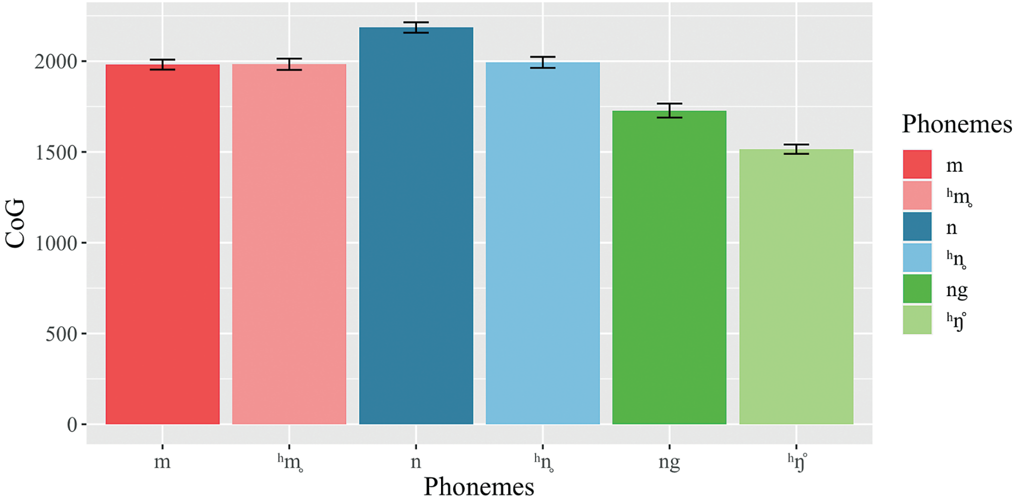


Figure A-2. Center of Gravity (CoG) of the nasal spectrum at nasal midpoint

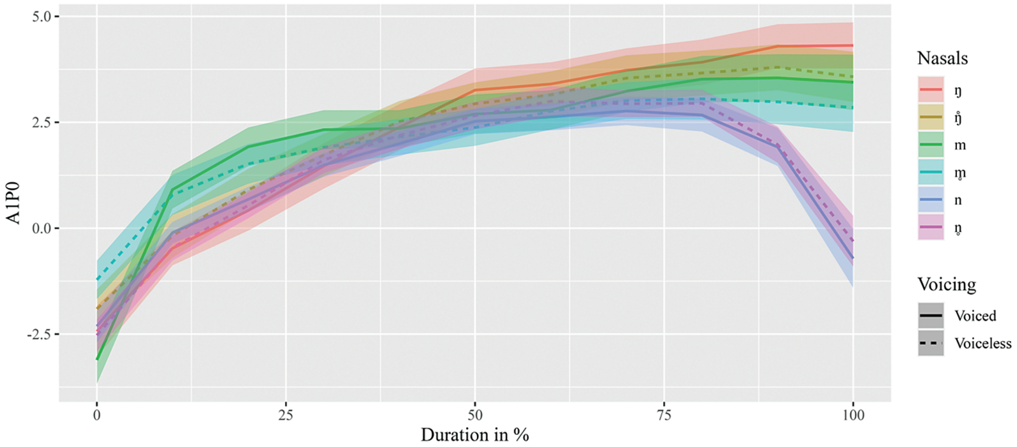


Figure A-3. Nasality measured with A1P0 in the vowels following nasals.

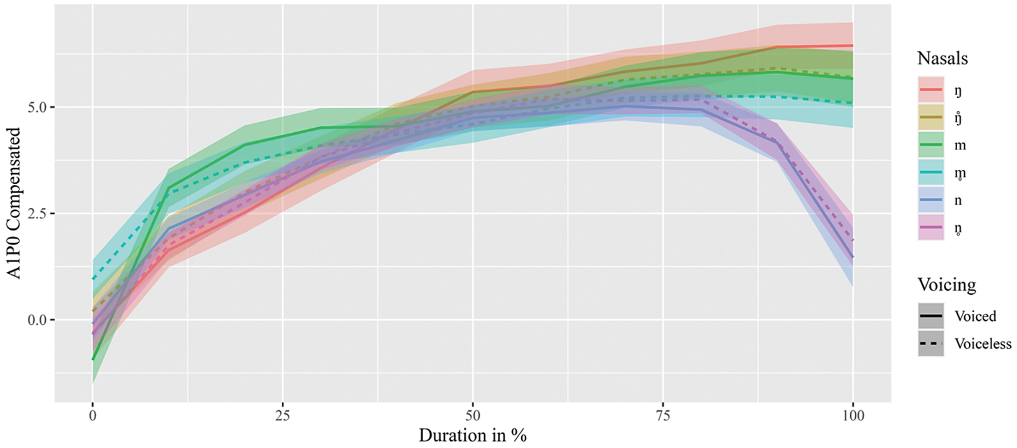


Figure A-4. Nasality measured with A1P0 (compensated) in the vowels following nasals.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Phonetic Variation of Voiceless Nasals in Drenjongke (Bhutia)*

GUILLEMOT, Céleste
LEE, Seunghun J.

Voiceless nasals in Drenjongke (Bhutia), a Tibeto-Burman language, are innovative segments that display variable realizations, not found in neighboring languages. In this paper, we present novel acoustic data that allows to identify four distinct phonetic realization patterns for voiceless nasals. Building upon the gestural model (Browman and Goldstein 1986, 1992), we analyze these variations by considering differences in relative gesture timing. Furthermore, we propose that temporal restrictions on the timing of nasal and laryngeal gestures such that nasals do not immediately follow laryngeals. This gestural restriction is supported by cross-linguistic data which show that the sequence of nasality and laryngeal gesture should be in that order, but not vice versa. The innovative circumstances of voiceless nasals in Drenjongke provided a testing ground to understand how phonetic variations reveal the nature of phonological processes underlying a phonological target (i.e. voiceless nasals).

Keywords: voiceless nasals, Drenjongke (Bhutia), phonetic variation, gestural model, gesture timing

DOI: https://doi.org/10.15026/0002000630

* We thank speakers of Drenjongke (Bhutia) and coordinators of the project, Kunzang Namgyal and Jigmee Wangchuk Bhutia. This research is written as part of the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021–2023). The research has been funded by the Strategic Japanese-Swiss Science and Technology Programme of JSPS and SNSF. We also thank George van Driem and Priyankoo Sarmah for discussing aspects of this paper. All errors are of our own.

1. Introduction

Voiceless nasals have an 'unusual' nature, in which nasal sounds with inherent voicing are produced as voiceless sounds. Acoustic studies such as Bhaskararao and Ladefoged 1991 present a phonetic typology of voiceless nasals demonstrating two different phonetic realizations for voiceless nasals, the first one consists in a voiceless portion followed by a voiced one, and the second is realized without voicing, which they call respectively Type 1 and Type 2. Subsequent studies have advanced our understanding of the phonetic nature of voiceless nasals; they are nasal sounds with a voiceless part. Based on novel acoustic data on Drenjongke, an understudied Tibeto-Burman language, this paper proposes to deepen our understanding of Drenjongke voiceless nasals: Our data suggests that variable realizations of Drenjongke voiceless nasals are constrained by the coordination of nasal gesture and glottal gesture, in which the glottal gesture cannot precede the nasal gesture. This is further supported by cross-linguistic patterns on the relationship between a nasal and a glottal fricative.

Voiceless nasals (also called 'breathy' or 'aspirated' nasals) are relatively rare segments in the world's languages. Only about 4% of the 451 languages listed in the UPSID Database (UCLA Phonological Segment Inventory Database, Maddieson 1984) comprise such segments in their phonological inventory (Chirkova 2019). Languages in which voiceless nasals are contrastive are much less common (Ohala and Ohala, 1993), yet contrastive voiceless nasals are found in a number of language families such as Tibeto-Burman, Hmong-Mien, Tai-Kadai or Mon-Khmer (Matisoff 2003, Chirkova *et al.* 2013, 2019): Burmese (Ladefoged 1971, Dantsuji 1984, 1986), Mizo (Bhaskararao and Ladefoged 1991), Angami (Bhaskararao and Ladefoged 1991, Blankenship *et al.* 1993), Xumi (Chirkova *et al.* 2019), Sre (Manley 1972), Achang (Dai 1985), Sui (Wei & Edmondson 2008: 586), Chadong (Li 2008: 598), Prinmi (Ding 2014), Hakha Lai (Peterson 2003), Anong (Sun and Liu 2009) and Niuwozi (Ding 2003). Bantu languages show voiceless nasals when a nasal prefix precedes a root-initial voiceless stop (Maddieson & Sands 2019): Sukuma (Maddieson 1991), Pokomo (E.71) and Bondei (G.24) (Huffman and Hinnebusch 1998), Ikalanga (Mathangwane 1998), as well as Nyarwanda (Demolin and Delvaux 2001). Voiceless nasals are also reported in the Ikema dialect spoken in Miyako, Okinawa in Japan (Hayashi 2013).

The acoustic investigation of Burmese voiceless nasals by Dantsuji 1984 led to the conclusion that in Burmese, voiceless nasals are composed of a voiceless nasal friction of weak intensity at the beginning, and a voiced nasal portion with higher intensity at the end. The voiced portion of the voiceless nasal is much shorter than that of regular voiced nasals, such that the durations of the voiceless and the voiced portions are negatively correlated. Based on this finding, Dantsuji argues for only one nasal phoneme, which comprises both a voiced and a voiceless part.

Ohala and Ohala 1993 show that voiceless nasals are only 'half-voiceless'; they have a voiced nasal portion in the end. According to them, this suggests that voiceless nasals lose their place feature during consonant constriction. As such, they claim that the place

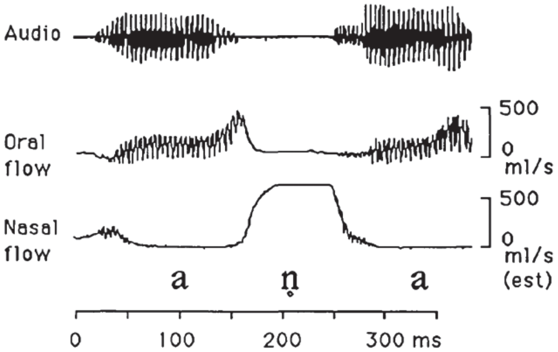


Figure 1 An illustration of Type 1 voiceless nasal, from Bhaskararao and Ladefoged 1991

of articulation of voiceless nasals such as [n̥, m̥, ŋ̊] can be differentiated only based on the transitions in the surrounding vowels. A study by Ladefoged and Maddieson 1996 suggests that in Burmese, since the glottis is mostly spread during the voiceless nasal but not in the voiced nasal, the primary cue for the nasal voicing contrast is whether the glottis is spread or not.

What these studies point out is that the aerodynamics of voiceless nasals is typically binary: one part is produced with nasal airflow only and the other part exhibits both nasal and oral airflow. Bhaskararao and Ladefoged 1991 and Blankenship *et al.* 1993 identify two distinct types of voiceless nasals. In Figure 1, reproduced from Bhaskararao and Ladefoged 1991, the voiceless nasal begins with nasal airflow with no oral airflow. Just before the production of the vowel, the nasal airflow abruptly decreases, and the oral airflow increases.

This Type 1 voiceless nasal in Figure 1 is the kind of voiceless nasal that is typically observed in Burmese (spoken in Myanmar) or Mizo (spoken in Mizoram, India). It involves a two-step process. First, oral airflow is blocked by the tongue tip against the alveolar region, the velum is lowered to allow a peak in nasal airflow and the glottis is spread. Second, the vocal folds gradually become closer to generate glottal vibration while the velum becomes raised and the tongue tip lowered, thus decreasing nasal airflow while increasing oral airflow. This second part in the articulation of voiceless nasals, whether it be lowering the tongue tip for alveolars, or opening the mouth for labials, or lowering the velum for velars, enables formant transition information to cue the place distinctions in voiceless nasals (cf. Ohala and Ohala 1993).

Unlike Type 1 voiceless nasals that are produced with a two-step process, the second type of voiceless nasals in Figure 2 is produced in a single step. Voiceless nasals in Angami (spoken in Nagaland, India) show no voicing throughout their production. The acoustic signal shows glottal vibration due to the spread glottis, but place distinctions between voiceless nasals can still be made based on the varying degree of nasal airflow (Bhaskararao & Ladefoged 1991). The first portion of the articulation sees nasal air flow increase to reach a peak, which then progressively decreases during the second part, while a sudden increase in oral air flow indicates the release of the closure.

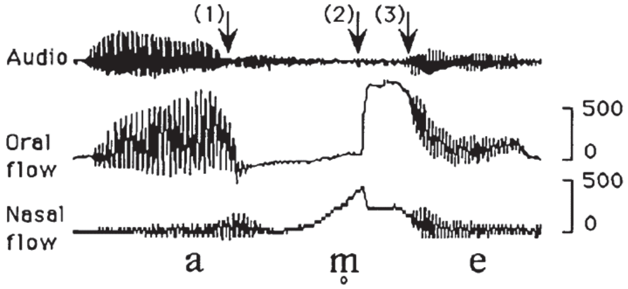


Figure 2 An illustration of Type 2 voiceless nasal, from Bhaskararao and Ladefoged 1991

Table 1 Comparison between Proto Tibeto-Burman form and Burmese written form. Superscripts mark tonal categories. The entry numbers of PTB Forms are based on STEDT1).

gloss	Proto Tibeto Burman (PTB)	written Burmese	
‘to give’	*g/s-naŋ (#5820)	*hnang*^{2}	(high tone)
‘to borrow’	*r/s-ŋ(y)a (#2541)	*hnga*^{2}	(high tone)
‘nose’	*s-na ~ *s-naːr (#803)	*hna*	(low tone)

The emergence of voiceless nasals is often attributed to a diachronic sound change where a consonant + nasal (Cn) cluster turns into a voiceless nasal (Ohala and Ohala 1993). In Burmese, voiceless nasals probably originate from the historic /s/+nasal cluster: [n̥a] ‘nose’ in Burmese corresponds to *sna* in orthographic Tibetan. Sturtevant 1940 reconstructs *sn and *sm for the voiceless nasals [n̥] and [m̥] in Primitive Greek, as does Thurneysen 1946 when reconstructing proto forms for voiceless nasals. Ohala and Ohala 1993 find in these diachronic explanations that a sonorant that loses its voicing becomes similar to an obstruent, e.g., a fricative. In two-step articulation cases, they argue that when a [-sonorant][+sonorant] sequence, where the [-sonorant] part is a fricative, is substituted for a similar [-sonorant][+sonorant] sequence, then the [-sonorant] part is a voiceless nasal.

Voiceless nasals may emerge from consonant clusters even if the first segment is not the sibilant /s/. Nishida 1970, 1975 hypothesizes that Burmese voiceless nasals derive from *Cn clusters based on comparisons with early Tibetan forms. Data in Table 1 demonstrates that some voiceless nasals in Burmese stem from consonant clusters such as **sn-*, **gn-* or **rn-*.

Examples from Icelandic in Jessen and Péturrsson 1998 also suggest that voiceless nasals can arise from non-sibilant sounds. The Icelandic voiceless nasals in (1) show that a historic **kn* cluster corresponds to a voiceless nasal.

(1) Voiceless nasals in Icelandic and cognates in German

a. Icelandic	[n̥iːvʏr]	‘knife’	[n̥jeː]	‘knee’
b. German	[knif]	‘knife (low German)’	[kni]	‘knee’

---

1) The Sino-Tibetan Etymological Dictionary and Thesaurus (https://stedt.berkeley.edu/)

Although a variety of studies investigate the acoustic properties of these sounds, the issue of the phonological interpretation of voiceless nasal is still, as far as we are aware, left unsolved. Indeed, the two-step articulation of these consonants raises the issue of whether they should be treated as a single segment with a single feature or a cluster of two distinct entities with two separate phonological features (cf. Ohala and Ohala 1993: 233). The widespread view treats voiceless nasals as a single phoneme (Cornym 1944, Sprigg 1965, Okell 1969, Nishida 1972, Dantsuji 1984 among others); voiceless nasals are voiceless segments with a low-level phonetic rule inserting voicing at the end (Bhaskararao and Ladefoged 1991). An opposing view is found in McDavid 1945 who proposes that voiceless nasals should be regarded as consonant clusters, since diachronic changes indicate that voiceless nasals have emerged from consonant clusters, but McDavid's proposal results in confounding diachronic evidence with synchronic phonological patterns.

The present study has several goals. The first is a report on fieldwork data of production of voiceless nasals in Drenjongke, an understudied Tibeto-Burman language (Namgyal *et al.* 2020). The acoustic characteristics of voiceless nasal consonants show inter- and intra-speaker variation regarding the realization of glottal frication and nasality. Building upon the Gestural Model (Browman and Goldstein 1986), we claim that this phonetic variation can be accounted for by a difference in the relative timing of the different gestures involved in the articulation of voiceless nasals (i.e., velic and glottal aperture). The second goal is proposing a restriction concerning the gestural timing between the velic aperture and the glottal aperture. We argue that at least in Drenjongke, the glottal aperture cannot precede the velic aperture in the production of voiceless nasals. We support our claim about this restriction by examining phonotactics of cross-linguistic data from English, French, and Korean among others.

The remainder of this paper is organized as follows. After reviewing the phonological inventory of Drenjongke with a focus on previous studies on voiceless nasals, section 2 shows an analysis of novel Drenjongke acoustic data. Section 3 is an attempt to account for the phonetic variation observed in the realization of Drenjongke voiceless nasals by proposing an articulatory phonology analysis within the gestural model framework (Browman and Goldstein, 1986). The discussion in section 4 addresses the issue of directionality biases in fricative-sonorant clusters and also examines cross-linguistic phonotactic patterns.

2. Voiceless Nasals in Drenjongke

2.1 Drenjongke Phonology

Drenjongke (also known as "Bhutia", "Hloke" or "Sikkimese"), is a Tibeto-Burman language spoken by about 80,000 speakers in the state of Sikkim, in the north of India. Impressionistic descriptions of the phonological inventory of the language have been provided in van Driem 2016 and Yliniemi 2005, 2019. Acoustic and articulatory properties have been investigated in detail in a series of papers which main findings are summarized in Namgyal *et al.* 2020.

	bilabial		alveolar		retroflex		alveolo-palatal		palatal		velar		glottal	
voiceless/voiced plosive	p	b	t	d	ʈ	ɖ					k	ɡ	ʔ	
aspirated/devoiced plosive	pʰ	b̥	tʰ	d̥	ʈʰ	ɖ̥					kʰ	ɡ̊		
voiceless/voiced affricate			ʦ	ʣ			ʨ	ʥ						
aspirated/devoiced affricate			ʦʰ				ʨʰ	dʑ̥						
voiceless/voiced nasal	m̥	m	n̥	n					ɲ̊	ɲ	ŋ̊	ŋ		
voiceless/voiced tap or trill			ɾ̥	ɾ										
voiceless/voiced fricative			s	z			ɕ	ʑ						
devoiced sibilant fricative				z̥				ʑ̥						
voiced approximant		w								j				
voiceless/voiced lateral			ɬ	l										

Figure 3 Consonant inventory of Drenjongke, reproduced from Namgyal *et al.* 2020

As presented in Figure 3, a chart of Drenjongke consonants reproduced from Namgyal *et al.* 2020, plosives have four places of articulation (bilabial, alveolar, retroflex and velar). Together with affricates they exhibit a four-way laryngeal contrast, which was investigated in Lee *et al.* 2019a, 2019b. Namely, they can be voiceless, aspirated, voiced and voiced aspirates (also called devoiced[2]). For fricatives on the other hand, only three laryngeal categories can be observed (voiceless, voiced, voiced aspirates), as presented in Guillemot *et al.* 2019a. The articulatory properties of retroflex consonants have been investigated with ultrasound data in Lee *et al.* 2019c and Guillemot *et al.* 2020.

Drenjongke has 5 short and 8 long vowels, which are contrastive, and differ both in terms of duration and quality (van Driem 2016). A remarkable property in Drenjongke long vowels is that their phonetic realization is subject to variations (e.g., alternation between long vowel and vowel with a coda consonant inserted), a phenomenon that has been specifically investigated in Lee *et al.* 2019d and Guillemot *et al.* 2019b.

Lastly, tone is contrastive in syllables with a vowel only or with a nasal onset; Drenjongke has a two-tone system, high and low. Syllables can also be high or low register. In the transcriptions in this paper, high tone is indicated by an acute accent, and low by a grave accent on the vowel. Tone patterns with laryngeal categories are post-lexical; a syllable with an onset that is a voiceless or an aspirated consonant is followed by a high tone, while when

---

2) Although the literature usually tends to use the term “devoiced” in order to reflect the historical origins of the laryngeal category in the South Bodish language, based on the results of acoustic analyses we choose to use the term “voiced aspirates”, following the descriptions in languages with a four-way laryngeal contrast.

the onset is a voiced and voiced aspirate consonant it bears a low tone. Issues pertaining to prosody and tone in Drenjongke are discussed in Lee *et al.* 2018, 2019e and 2020.

Although a preliminary description of the phonological inventory of Drenjongke was proposed in Yliniemi 2019, to the best of our knowledge, there is no study accounting for the phonetic characteristics of Drenjongke voiceless nasals. The present study aims to fill this gap by providing and analyzing acoustic data for Drenjongke voiceless nasals.

2.2 Voiceless Nasals in Drenjongke

Previous impressionistic descriptions of the phonological inventory of Drenjongke report that the language has eight nasal phonemes, which contrast in terms of voicing (Yliniemi 2019, van Driem p.c.), such as [+voice] /m, n, ɲ, ŋ/ and [-voice] /m̥, n̥, ŋ̊, ɲ̊/. While voiced nasals can be found in word-initial, word-medial and in coda position[3], voiceless nasals appear only in word initial position. Yliniemi 2019 points out that aspiration is reduced in word-medial position, causing all the "breathy" phoneme series (i.e., voiceless liquids and voiceless nasals) to not occur at all. The distribution of /ŋ̊/ and /ɲ̊/ seems to be complementary as the former occurs only preceding non-front vowels and the latter precedes front ones only[4].

The voicing contrast in nasals is illustrated in the minimal pairs in (2) from Yliniemi 2019. Voiced nasals can bear either a high or low tone (as exemplified in 2a). However, voiceless nasals, like other voiceless phonemes in Drenjongke, exclusively belong to the high register.

(2) Voiced and voiceless nasals in Drenjongke

a. **labials**	/m/ vs. /m̊/	/mà/ 'mother'	/má/ 'wound	/m̥a/ 'down, lower'
b. **alveolars**	/n/ vs. /n̊/	/nà:/ 'here'	/ná/ 'ear'	
c. **palatals**	/ɲ/ vs. /ɲ̊/	/ɲìm/ 'sun, day'	/ɲ̥im/ 'sister-in-law'	
d. **velars**	/ŋ/ vs. /ŋ̊/	/ŋàk/ [ŋàʔ]'speech'	/ŋ̥aʔ/ 'invocation'	/ŋ̥a/ 'nose'

Yliniemi 2019 describes Drenjongke voiceless nasals as complex segments, in which the voiced part is preceded by the voiceless part, resulting in phonetic realization of [m̊m], [n̊n], [ɲ̊ɲ], and [ŋ̊ŋ]. This impressionistic description corresponds to what Bhaskararao and Ladefoged 1991 identify as a Type 1 voiceless nasals found in Burmese or Mizo.

Other sonorants such as the rhotic and the lateral in Drenjongke also exhibit a similar contrast in word-initial position: [là] 'pass' vs. [ɬá] 'deity' or [ràm] 'be broken' vs. [ɾ̥ám] 'break (trans.)' (Yliniemi 2019, Namgyal *et al.* 2020). These sonorants are described as having realizations akin to the nasals where the voiced component is preceded by a voiceless part. Both voiceless rhotic and lateral have cognate sounds in other Tibeto-

---

3) Note that the palatal nasal /ɲ/ is the only one that does not occur in the coda position (Yliniemi 2019). A reviewer pointed out that this distributional restriction of /ɲ/ is common across languages in Southeast Asia, except for the Austroasiatic languages.

4) The distribution of these two voiceless nasals may suggest that they are variants of the same phoneme, but further exploration is required to establish the relationship between the two sounds.

Table 2 Experiment stimuli with voiceless nasals[5]

Drenjongke	Gloss	POA
m̥e	‘lower’	labial
n̥abe	‘pillow’	alveolar
n̥o	‘snot’ [Lachen dialect]	alveolar
ɲ̊ap[6]	‘to claim, seize’	palatal
ɲ̊a	‘borrowed’	palatal
ɲ̊eː	‘trap’	palatal
ɲ̊ik	‘to squeeze’	palatal
ɲ̊ima	‘impure’	palatal
ŋ̊aːle	‘early’	velar

Burman languages in the areas such as Tibetan and Dzongkha. When comparing Drenjongke with these other languages, voiceless nasals may be a novel group of sounds unique to Drenjongke. Acoustic results reported in this paper address the characteristics of these newly emerged sounds.

2.3 Acoustic Analysis of Drenjongke Voiceless Nasals

In this subsection, results based on analyses of acoustic data on voiceless nasals recorded from twelve Drenjongke native speakers are reported. Drenjongke speakers show that they aim for a phonological target for voiceless nasals, but that phonetic realizations of voiceless nasals are variable. We were able to identify and categorize at least four different realizations: (I) voiceless nasal, (II) nasal, (III) aspiration and (IV) inversion. After describing the data collection process, each pattern is introduced in detail.

2.3.1 Data Collection

Data collection sessions for the analysis of Drenjongke voiceless nasals are based on recordings collected in Sikkim, India in 2018 and 2019. Twelve speakers were recruited by the local coordinators; all of them were teaching Drenjongke at primary or secondary schools. Recordings were made using a TASCAM Linear PCM Recorder (DR-100MK III) and Shure WH30-XLR head-worn microphone with a 44.1 kHz sampling frequency. Participants were asked to take part in a reading task. Stimuli were presented on powerpoint slides in English to the participants, who had to translate them to the corresponding Drenjongke word in a frame sentence (/ŋa X lap to ĩ/, ‘*I say X.*’)[7]. A set of randomized wordlists was recorded five times. All stimuli are listed in Table 2 below.

---

[5] A reviewer pointed out that the dataset has a bias toward palatal sounds. The frequency of voiceless nasals is not high, and our consultants came up with examples that were biased toward palatals. The recordings of this data are available online by accessing the project archive website (PhoPhoNO Digital Archive 2020).

[6] /ɲ̊ap/ and /ɲ̊a/do not pattern with the description in Yliniemi 2019, as he reports that /ɲ̊/ should be preceding front vowels only. However, this data was provided to one of the project collaborators by native speakers.

[7] Recordings took place after a training session during which participants checked the meaning of the English words with the experimenter.

Table 3 Intra-speaker and inter-speaker variations for voiceless nasals

Speaker	Gender	Pattern I	Pattern II	Pattern III	Pattern IV	Total
SIP048	F	0	9	1	21	31
SIP050	F	0	16	1	21	34
SIP051	F	25	3	2	12	42
SIP052	M	6	24	0	5	35
SIP053	M	14	13	0	0	27
SIP054	M	2	5	9	27	43
SIP055	M	9	13	17	4	43
SIP057	M	22	17	1	2	42
SIP058	M	12	13	0	9	34
SIP071	F	0	15	4	3	22
SIP072	F	20	13	1	0	34
Total		110	141	36	100	395

The data obtained from the recordings was processed using a series of Praat scripts (Boersma and Weenink 2020). Boundaries for preceding and following vowels of target sounds were annotated manually. The preceding vowel in the frame sentence was fixed as [a].

Errors found in the data (e.g., the speaker did not produce the intended word, recording noise) were excluded manually. Note that the experimental methodology (i.e., words presented in English only) is inherently responsible for the accuracy errors, as not all speakers were familiar with the stimuli words. One speaker was excluded of the dataset due to his low accuracy. On the other hand, variability being often observed in the phonetic realization of Drenjongke codas (e.g., alternation between glottal stop and long vowel; Yliniemi 2019, Lee *et al.* 2019d, words which diverging realization could be attributed to phonetic variation were included in the dataset for analysis.

2.3.2 Results

The analysis of the data obtained from our recordings suggests that there is no unique phonetic realization of the voiceless nasal in Drenjongke but at least four different patterns: (i) a voiceless nasal corresponding to Bhaskararao and Ladefoged 1991's Type 1 that articulates the nasality at the end of glottal frication, (ii) a nasal without voicelessness, (iii) glottal frication but no nasality on the consonant portion, and (iv) an inversion in the consonantal portion: a voiced nasal consonant followed by glottal frication. In addition, we observed intra-speaker and inter-speaker variations (see Table 3). That is, a single speaker can produce several different phonetic realizations for the same phoneme, sometimes across repetitions of the same word. None of the items examined were realized with a single pattern type only. While the phonetic realization expected was pattern I, among the four patterns observed, it was pattern II (i.e., the voiced nasal) which had the highest frequency. Inversely, pattern III (i.e., aspiration only) was rare. Moreover, there seems to be a tendency among speakers to prefer some pattern over the others, with inter-speaker differences (e.g., pattern I and II for SIP053 and 57, and pattern III and IV for SIP054).

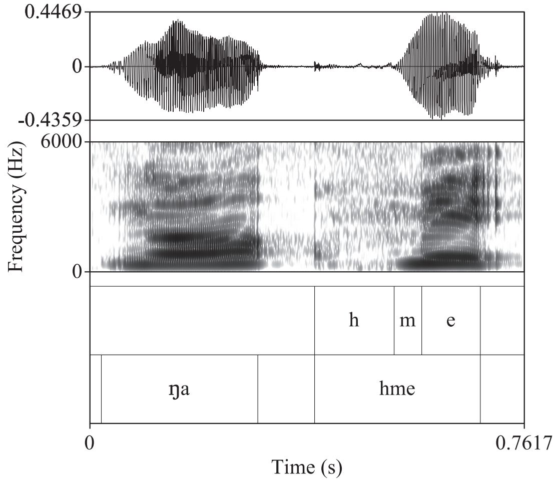


Figure 4 Spectrogram of a voiceless nasal pattern (pattern I) that is produced from a voiceless nasal stimulus /m̥e/ in a frame sentence by a female speaker (SIP072)

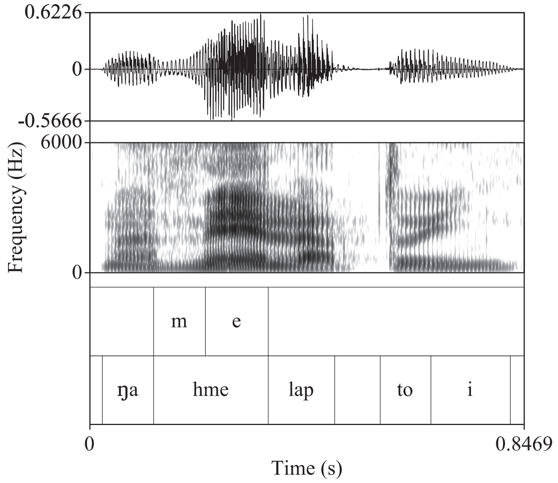


Figure 5 Spectrogram of a voicing pattern (pattern II) that is produced from a voiceless nasal stimulus /m̥e/ in a frame sentence by a male speaker (SIP058)

The first pattern corresponds to what Bhaskararao and Ladefoged 1991 describe as a Type 1 voiceless nasal. This pattern is represented in Figure 4 produced by a female participant. After the vowel in the frame sentence, no voicing is observed on the first part of the articulation ([h]), and the spectrogram shows glottal frication, corresponding to the opening of the glottis described by Bhaskararao and Ladefoged. Although the present data does not include nasal airflow measurements, the spectrogram shows aspiration coming from the nasal cavity during articulation. In the second step of the articulation (represented as [m]), we observe a voicing bar at the foot of the spectrogram with a weaker formant structure that contrasts with the following vowel [e].

Figure 5 illustrates the second pattern. The target sound is a voiceless nasal, but the

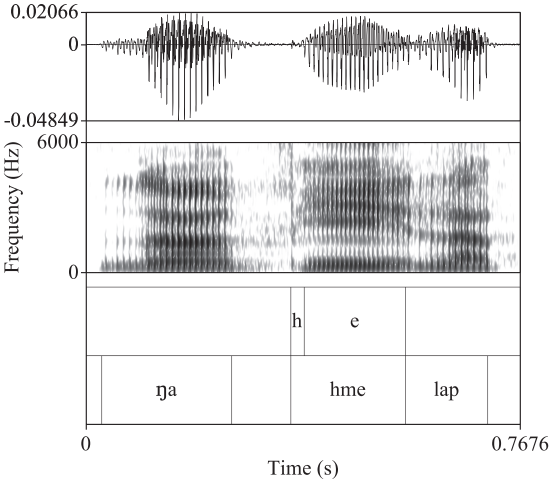


Figure 6 Spectrogram of a voiceless pattern (pattern III) that is produced from a voiceless nasal stimulus /m̥e/ in a frame sentence by a male speaker (SIP055)

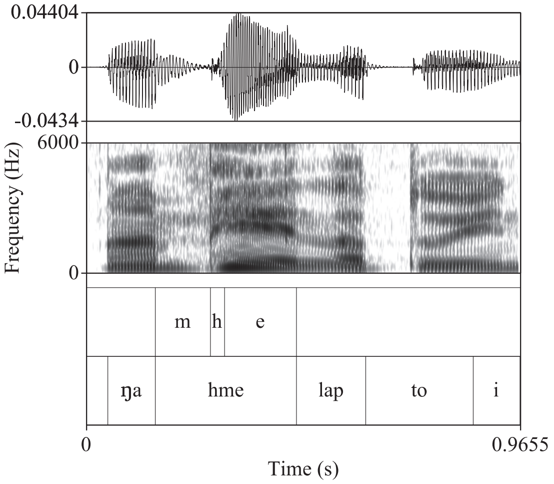


Figure 7 Spectrogram of an inversion pattern (pattern IV) that is produced from a voiceless nasal stimulus /m̥e/ in a frame sentence by a male speaker (SIP054)

spectrogram in Figure 5 produced by a male speaker suggests that no glottal gesture precedes the onset of voicing of the nasal consonant ([m]); that is, there is no aspiration and only the nasal portion remains in the acoustic signal making it similar to voiced nasals.

In the third pattern, nasality is absent in the consonant target and only the aspiration is realized at the target syllable onset. In Figure 6, a voiceless nasal is produced by a male speaker; we can observe a stronger frication in the consonant portion ([h]) with an absence of voicing component, differing from the first two patterns.

The last pattern is illustrated in the spectrogram in Figure 7 pronounced by a male speaker. The spectrogram shows a voicing bar as well as weaker formant structure right after the preceding vowel, suggesting the nasal sound is produced ([m]) before glottal aperture. The onset of the target word, annotated as [h], only shows strong glottal frication and no voicing: a realization found also in the third pattern.

This pattern is in contrast with the first pattern, as the glottal gesture and the nasal gesture are reversed. While the first pattern has voicelessness followed by voicing, in the fourth pattern, the voiced part of the nasal is realized first, followed by the glottal gesture. It is also in contrast with previous descriptions of the voiceless nasals in Drenjongke ([m̥m], in Yliniemi 2019, for example), and the phonetic definitions (for Type 1 and Type 2) proposed by Bhaskararao and Ladefoged 1991, which suggest that voicelessness should appear first in voiceless nasals.

2.4 Summary

What the acoustic data presented in section 2.3 suggests is that although there is a single phonological target (i.e. the voiceless nasal consonant), its phonetic realization is not uniform. These results raise a question about phonological targets concerning voiceless nasals. Are voiceless nasals a single phoneme or a consonant cluster (i.e., /h/+/N/)? Our findings can be interpreted that Drenjongke speakers possess a coalesced double target for the phonological representation of voiceless nasals. If Drenjongke speakers have a single target for voiceless nasals, we would expect uniform phonetic realization with no inversion-type pattern, since inversion implies the presence of a double target. We also expect that intra-speaker variation would be minimal, although inter-speaker variation may occur. The variability observed in our data can be accounted for only if we consider the hypothesis that there is a double target, that is, a voiceless glottal frication (i.e. /h/) and a voiced nasal, and that speakers variably realize one, the other, or both of the targets.

Here we take a short excursion to voiced and voiceless rhotics and laterals in Drenjongke. The voicing contrast in rhotics and laterals is more stable; as far as we know, no variations are observed in their phonetic realizations. Voiceless rhotics and laterals are also found in related languages such as Tibetan and Dzongkha (van Driem 1992), suggesting that these sonorants have corresponding reflexes in other Tibeto-Burman languages. A survey of forty Tibeto-Burman languages shows that languages can have voiceless rhotics and laterals even though they do not have voiceless nasals. Voiceless nasals appear to be innovative segments in Drenjongke, as they cannot be found in adjacent languages such as Dzongkha or Tibetan. For example, the Drenjongke word for ‘pillow’ is [ŋ̊abø] with a voiceless nasal, but ‘pillow’ in Dzongkha is [haŋbo]. We claim that the non-uniformity of the phonetic implementation of voiceless nasals reflects their relative novelty. It may be the case that we are witnessing a sound change in progress, that is, the process of shifting from a double target (patterns II, III and IV) to a coalesced single target (I) voiceless nasal, with analogous properties to similar segments in other languages.

3. Phonetic Variation in the Realization of Voiceless Nasals

The phonetic implementation of Drenjongke voiceless nasals shows non-uniformity in our acoustic data. This raises the questions of why we observe such variability in the realization of these segments, and how we can account for this phenomenon. The inter-item and inter-speaker variation in our results suggest that voiceless nasals are in the process of being lexicalized; otherwise we would expect uniformed phonetic realizations of them. In this section, we attempt to account for these variable phonetic patterns in a gestural model framework (Browman and Goldstein 1986, 1992) analysis. What we suggest is that the variability in the phonetic realization of voiceless nasals can be attributed to a difference in the timing of the gestures involved in their articulation.

3.1 Voiceless Nasals in the Gestural Model

Further insights about voiceless nasals that display four variable realizations can be gained from analysis based on articulatory phonology. Building on the featural analysis in section 3.1 as a representation of voiceless nasals, the articulatory phonology analysis we propose demonstrates how two types of gestures, velic aperture and glottal aperture, coordinate in generating voiceless nasals.

The gestural model in articulatory phonology proposed by Browman and Goldstein 1986, 1992 suggests that phonological processes originate from the change in the timing between articulations and their magnitude. They introduce a way to deal with various phonological phenomena that remained unexplained until now by emphasizing the link between the phonological and physical structures of speech in order to account for the organization of speech in 'both space and time'. In (relative) opposition to more traditional approaches to phonological representation and autosegmental phonology (Goldsmith 1976, Clements 1980), in which phonological representations are made based on one (or several) linear sequences of non-overlapping segments in terms of tiers, the gestural framework does not make use of phonological features. Instead, it proposes gestures, which the authors define as 'events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators', in order to represent utterances phonologically. If phonological patterns are different realizations of gestures that depend on a change in the relative timing of each gesture, voiceless nasals are good candidates for examining the theory, because two idiosyncratic gestures are at work in the production of voiceless nasals.

In autosegmental theory, a spreading or narrowing of the glottis corresponds to the presence or absence of the privative [spread glottis] feature; in the gestural model, the glottal movement is a function of the (gradient) timing and magnitude of the glottal gesture. In Browman and Goldstein 1986, 1992's analysis, allophonic variation results from the overlap of invariant gestural units. In the case of the allophonic variation between aspirated and unaspirated stops in English, the gestural theory analyzes that the variation is not due to the opening or the closure of the glottis itself, but to the timing and magnitude of the glottal opening. Likewise, clusters with /s/ and a following stop cannot be aspirated because English has a constraint that restricts the glottal opening gesture in the word-initial position to once only.

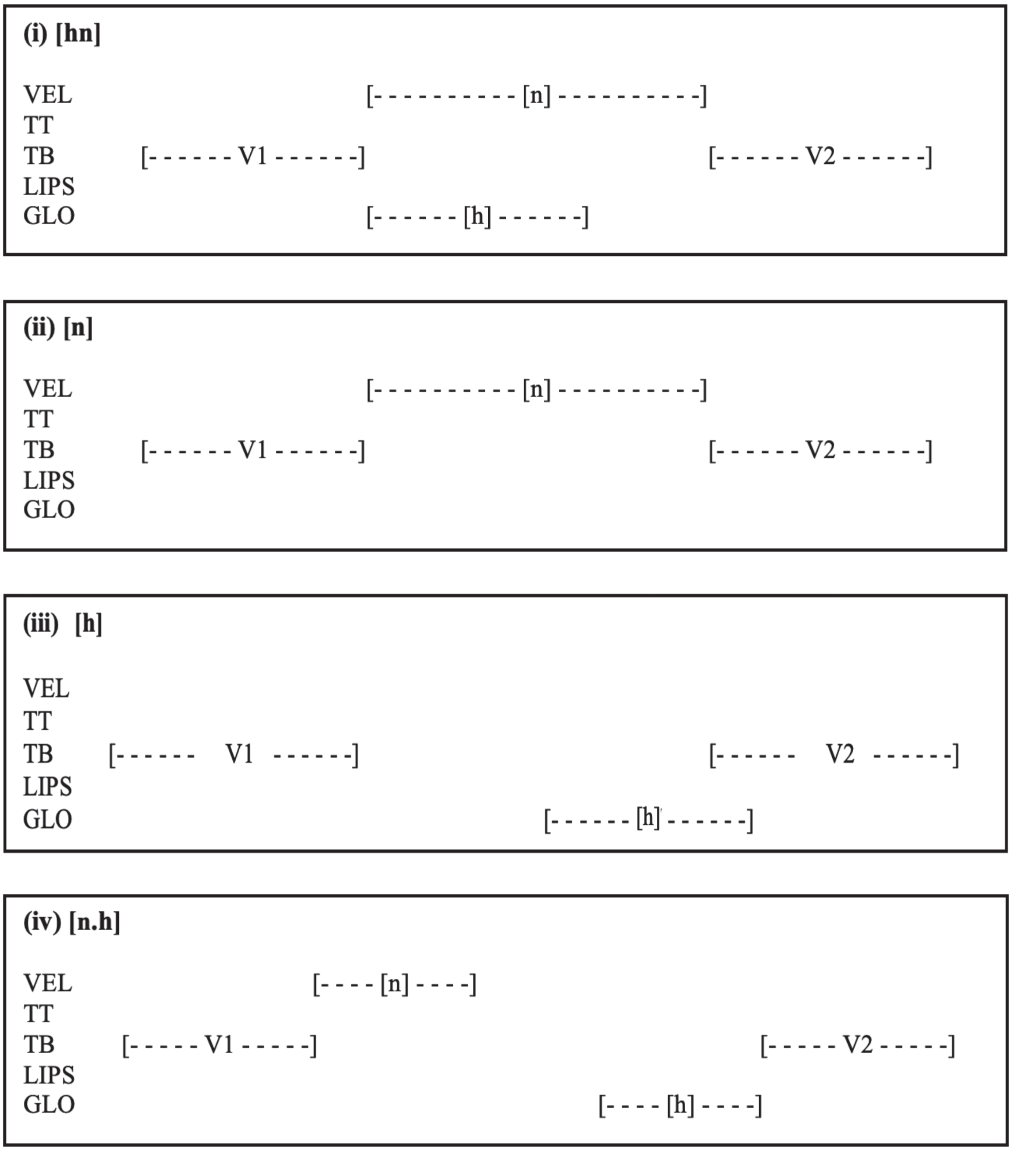


Figure 8 The four voiceless nasal patterns in the gestural model framework. Adapted from representations in Browman and Goldstein 1992 and Steriade 1990
(VEL: velic aperture, TT: tongue tip, TB: tongue body, LIPS: lips, GLO: glottal aperture)

The four patterns of variation in voiceless nasals are represented in Figure 8 based on a standard gestural model framework. The velic aperture (VEL) produces a nasal sound as it allows airflow into the nasal cavity. The tongue tip (TT) gesture and the tongue body (TB) gesture are responsible for vowel production and place of articulation of consonants. The lips gesture (LIPS) is active when labial sounds are produced, and the glottalic aperture (GLO) gesture controls the opening of the glottis.

Pattern I represents voiceless nasals of Type 1 in Bhaskararao and Ladefoged 1991, which correspond to a simultaneous timing of the velic aperture and glottal aperture gesture at the beginning. Towards the end of the consonant, the glottalic aperture gesture is narrowed for the production of the following vowel, while the velic aperture gesture remains open until the vowel production begins. In pattern II, the glottal aperture gesture is not present while the glottis is narrowly open to produce the vibration typically associated with nasal sounds. In pattern III, the velic aperture gesture is deleted, and the glottal aperture gesture only is articulated as the word onset. Pattern IV shows an inversion pattern, as the velic aperture gesture precedes the glottal aperture after completion of the tongue body gesture for the preceding vowel.

The variation in the realization of voiceless nasals shows an asymmetry between the timing of the velic aperture gesture and the glottal aperture gesture. While the velic aperture gesture may co-occur with the glottal gesture, it never is initiated after the onset of the glottal aperture gesture. We propose that the variation observed in Drenjongke voiceless nasals follows a constraint, in which the velic aperture gesture should not follow the glottal aperture gesture: the tautosyllabic [hn] sequence is banned. If Drenjongke allowed a violation of this constraint, we would expect to see variation patterns where a preceding vowel becomes aspirated or where a following vowel becomes nasalized. The absence of these two patterns validates our proposed constraint in Drenjongke. In non-simultaneous variations, such as pattern IV, the glottal aperture gesture is articulated after a prosodic boundary, further corroborating that Drenjongke speakers do not favor a contiguous production of the velic aperture and the glottal aperture gestures.

In Dzongkha, a Tibeto-burman language spoken in Bhutan and closely related to Drenjongke, the word for ‘pillow’ is [haŋbo]. The cognate in Drenjongke is /hŋabø/ [ŋ̊abø] ‘pillow’. The first syllable with the /h/+/V/+/ŋ/ sequence in Dzongkha is realized with a voiceless nasal in Drenjongke. If Dzongkha also follows the constraint where the velic aperture gesture should not follow the glottal aperture gesture, the syllable [haŋ] avoids the environment by the presence of a vowel between the two gestures.

4. Articulatory Constraints in Drenjongke and Other Languages

The variation patterns found in the phonetic realization of Drenjongke voiceless nasals are proposed to be restricted by articulatory constraints of the velic aperture gesture and the glottal aperture gesture. Namely, these constraints can be interpreted as a directionality in the order of the two gestures involved in the phonetic realization of voiceless nasals. This idea of directional restrictions of articulatory gestures was raised in Browman and Goldstein 1986, 1992. They argue that restricted sequences in English are the results of constraints that apply on the gestures themselves. Their account of allophonic variation of English aspiration is an example of a constraint on glottal aperture gesture: only one glottal gesture is allowed in the word-initial position.

For voiceless nasals, if the two gestures were allowed to be combined in any order, we would expect to find phonetic realizations that reflect such a free order in languages: both velic → glottal gesture (i.e. /nh/ sequence) and glottal → velic (i.e. /hn/sequence) would occur. On the other hand, if an asymmetry in the realization of these two features is a norm, we would expect a bias against the glottal → velic sequence, as we found in Drenjongke voiceless nasals. This gestural bias would result in variations that rearrange the timing of the gesture so that the velic aperture gesture precedes the glottalic aperture gesture, as we observe in the present case.

Cross-linguistic patterns suggest that this potential asymmetry between the velic and glottal aperture gestures is not limited to Drenjongke; when a nasal and a glottal fricative appear sequentially languages prefer the nasal-fricative sequence to the fricative-nasal sequence.

Korean offers an interesting example of the bias against /hn/ sequences. An /h/ occurring between sonorants undergoes optional deletion, while an /h/ before or after a plosive becomes instead an aspirated segment. Underlying /h/ before or after a stop consonant merges into an aspirated consonant: (a) after a stop, /pap + hana/ [pap^{h}ana] ‘a bowl of rice’, /kuk + hana/ [kuk^{h}ana] ‘a bowl of soup’, and (b) before a stop /noh + ta/ [not^{h}a] ‘to put down’, /noh + ko/ [nok^{h}o] ‘to put down and’. When an /h/ appears before a sonorant, the /h/ is deleted: /noh + ɨnik’a/ [noɨnik’a] ‘because of putting down’. When /h/ appears after a nasal, either /h/ is deleted or a boundary is inserted before the glottal fricative: /pam + hako/ [pamaɡo] ~ [pam#haɡo] ‘with a chestnut’ (Kim-Renaud 1975, Kang 2003, Kim 2005, Cha *et al.* 2005, Park 2015). The /h/ deletion environment can be comparable to environments that trigger voiceless nasals. Korean phonotactics does not allow voiceless nasals on the surface. As such, /h/ after a nasal can only be realized after a short pause, otherwise the /h/ is deleted.

In English, tautosyllabic sequences with /h/ preceding a nasal sound are absent from the phonotactics.[8] Sequences with a nasal preceding an /h/ are found in words such as *un-healthy* or *un-happy*, but the sequence is separated by a morpheme boundary. Fricatives such as /s/ can precede nasals word-internally in words such as *small* and *snail*.

As for French, the /h/ sound is not part of the French phoneme inventory, and it does not have surface realization. Even so, restrictions concerning nasal and fricative sequences are observed. Fricatives in French may appear after a nasal under two conditions: (a) nasal-fricative sequences such as /ns/, /ms/, /nf/, /mf/ are only allowed when a word boundary is present: e.g. *bonne soirée* [bɔnswaʁe] ‘good evening’, or (b) fricatives may follow a nasalized vowel: e.g. *bonsoir* [bɔ̃swaːʁ] ‘good evening’. The fricative-nasal sequences are more restricted because /sn/, /sm/ clusters are banned in the native French lexicon, and found only in loanwords such as *snowboard* [snobɔʁd], *smiley* [smajle] or *schnaps* [ʃnaps] (Dell 1995).

Going back to Drenjongke, even though there is variability in the phonetic realizations, we do not see any merger pattern (which would correspond to a Type II in Bhaskararao and Ladefoged 1991’s description) but inversion (i.e., pattern IV) or deletion of one of the gestures (i.e., pattern II and III) only, which shows that Drenjongke speakers have a clear

---

[8] Greenlee 1973, Hooper 1977, Smith 1973 and Ohala and Ohala 1993 report that children acquiring English have a tendency to make mistakes in the production of /sn/ and /sm/ clusters and pronounce them as voiceless nasals (e.g. ‘Smith’ as [m̥it] or ‘sneeze’ as [n̥id]).

control of the laryngeal gesture. They are treating it like a sequence of gestures as if they were moving from a two-steps progress (i.e., /h/+/n/), and therefore voiceless nasals are realized as a Type I where glottis control is still there. The analysis of variability as different realizations of glottal timing is consistent with that of Kingston 1990 who proposes perceptual reasons for articulatory alignment.

A possible answer to the question of why the phonetic realizations observed in our data match Bhaskararao and Ladefoged 1991's Type I (and not Type II) can be found based on the proposal made by Silverman 1996. In his analysis of voiceless nasals, he claims that Type I (that he calls "pre-voicelessness") is optimal articulatorily, because it is more economic based on recoverability. He proposes three constraints to account for the variations in voiceless nasals: economize, recover and overlap. Based on these constraints, a Type II involving a merger, that is, a breathy nasal, is articulatorily costlier, which explains why Type I is preferred. This provides further support for a view where voiceless nasals in Drenjongke are an on-going process: Voiceless nasals can arise as a Type I but not a Type II due to articulatory costliness. This is also why the variable patterns in the current data do not include a breathy nasal "merger" type. If we consider a typology of diachronic change based on Silverman 1996 we postulate that a Type I voiceless nasal would always arise first (or as a first stage). Type II on the other hand might arise later, or emerge from a different process.

Lastly, while we propose that differences in glottal gesture timing are responsible for the variations observed), an alternative account to this kind of variability is offered in Howe and Pulleybank 2001. In their view, it is not glottal timing but syllable structure that plays a crucial role. However, although our data does not dispute the argument presented in Howe and Pulleybank 2001; it does also not provide further evidence for it.

5. Conclusion

This paper has presented new data on voiceless nasals from Drenjongke. Based on the analysis of the acoustic data, we identified four different patterns of phonetic realization which are subject to inter- and intra-speaker variation. While the first pattern corresponds to the expected realization of a voiceless nasal segment based on the literature, the three others suggest a more complex status of these consonants in the phonological inventory. Pattern II and III are characterized by the realization of nasality only for the former and aspiration only for the latter. In the fourth pattern, although the phonetic implementation of the voiceless nasal segment includes both aspiration and nasalization, these appear in reverse order compared to what is expected, that is, nasality precedes aspiration. These four patterns were used variably by speakers in their production of voiceless nasals, and although some preferences for specific patterns could be observed among speakers, no speaker used a single pattern exclusively. In addition, no specific pattern could be associated with a specific item, as all four patterns could be observed for each stimulus. In the second part of this paper, we attempted to account for phonetic variation by an analysis of the patterns of realization following the framework of gestural phonology.

Drenjongke has innovative voiceless nasals that are not found in related languages in the region. Our data also suggests that voiceless nasals are not lexicalized yet in Drenjongke as they have various intra- and inter-speaker realizations in terms of phonetic implementation of the phonological target. Examination of the four distinct variations within the gestural phonology framework (Browman and Goldstein 1986, 1992) reveals that the glottal gesture can occur before (or simultaneously with) the velic gesture in Drenjongke, but once the velic gesture ends, it is not possible to have a glottal gesture; this restriction limits the variation that we observe in our data, and, which is also supported by cross-linguistic examination of various nasal and glottal fricative sequences in a variety of languages.

As a final comment, we acknowledge that Drenjongke voiceless nasals might possibly be characterized as 'aspirated' nasals, rather than 'voiceless' nasals. Detailed phonetic underpinning of voiceless nasals must be accompanied with articulatory data measures nasal airflow or vocal cords vibration. Such data would reveal acoustic parameters that can be used for predicting the presence of voiceless nasals. As this phonetic work is beyond the scope of the current paper, we defer these issues until future work.

References

Ali, Latif H. 1979. "Observations on English Nasal Fricatives." *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, 119–122. New York: Acoustical Society of America.

Ali, Latif H. 1982. "Some Observations on the Nasal-fricative Sequences in English." *Journal of Phonetics*, 10: 315–323.

Ali, Latif H., Ray G. Daniloff, and Robert Hammarberg. 1979. "Intrusive stops in nasal-fricative clusters: An aero- dynamic and acoustic investigation." *Phonetica*, 36: 85–97.

Archangeli, Diana. 1984. Underspecification in Yawelmani Phonology and Morphology. MIT dissertation.

Avery, Peter and William J. Idsardi. 2001. "Laryngeal Dimensions, Completion and Enhancement." *Distinctive Feature Theory* (T. Alan Hall, ed.), 41–70, Berlin: Walter De Gruyter Inc.

Berns, Janine. 2016. "The phonological Representation of Affricates." *Language and Linguistics Compass*, 10(3): 142–156.

Bhaskararao, Peri and Peter Ladefoged. 1991. "Two Types of Voiceless Nasals." *Journal of the International Phonetic Association*, 21(2): 80–88.

Blankenship, Barbara, Peter Ladefoged, Peri Bhaskararao, and Nichumeno Chase. 1993. "Phonetic Structures of Khonama Angami." *Fieldwork studies of targeted languages* (UCLA Working Papers in Phonetics), 84: 127–141.

Boersma, Paul and David Weenink. 2020. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.1.09, retrieved 26 January 2020 from from http://www.praat.org/.

Browman, Catherine and Louis Goldstein. 1986. "Towards an Articulatory Phonology." *Phonology*, 3: 219–252.

Browman, Catherine and Louis Goldstein. 1992. "Articulatory Phonology: An Overview." *Haskins Laboratories Status Report and Speech Research*, SR111/112: 23–42.

Cha, Jaeeun, Myungsook Jung, and Jiyoung Shin. 2003. "A Phonetic and Phonological Study on /h/-realization between Sonorant Sounds." *Korean Journal of Linguistics*, 28(4): 765–783.

Chirkova, Katia and Zev Handel. 2013. "Diachronic Developments of Voiceless Nasals: The Case of Ersu, Lizu, and Related Language." Talk presented at the *46th* International Conference on Sino-Tibetan Languages and Linguistics, Dartmouth.

Chirkova, Katia. 2016. "Lizu (Ersu)." *The Sino-Tibetan Languages*, Second Edition (G. Thurgood and R. LaPolla, eds.), 823–839, New York: Routledge.

Chirkova, Katia, Patricia Bassot, and Angelique Amelot. 2019. "Voiceless Nasal Sounds in Three

Tibeto-Burman Languages." *Journal of the International Phonetic Association*, 49: 1–32.

Clements, George. N. 1985. "The Geometry of Phonological Features." *Phonology Yearbook*, 2: 225–252.

Cornym, William. 1994. "Outline of Burmese Grammar." *Language*, 20(4): Supplement.

Dai, Qinxia. 1985. "Achangyude qingbiyin [Voiceless Nasals in the Achang Language]." *Minzu Yuwen*: 11–15.

Dantsuji, Masatake. 1984. "A Study on Voiceless Nasals in Burmese." *Studia Phonologica*, 18: 1–14.

Dantsuji, Masatake. 1986. "Some Acoustic Observations on the Distinction of Place of Articulation for Voiceless Nasals in Burmese." *Studia Phonologica*, 20: 1–11.

Davenport, Michael and S. J. Hannahs. 1998. *Introducing Phonetics and Phonology*. London, UK: Arnold.

Dell, François. 1995. "Consonant Clusters and Phonological Syllables in French." *Lingua*, 95: 5–26.

Demolin, Didier and Véronique Delvaux. 2001. "Whispery Voiced Nasal Stops in Rwanda." Paper presented at the 32nd Annual Conference on African Linguistics. University of California, Berkeley.

Ding, Picus Sizhi. 2003. "Prinmi: A sketch of Niuwozi." *The Sino-Tibetan Languages* (Thurgood, G. and LaPolla, R.J., eds.), 588–601, London and New York: Routledge.

Ding, Picus Sizhi. 2014. *A Grammar of Prinmi: Based on the Central Dialect of Northwest Yunnan, China*. (Brill's Tibetan Studies Library: Languages of the Greater Himalayan Region.) xxi. 383. Leiden: Brill.

Esposito, Christina, Sameer U. Khan, and Alex Hurst. 2005. "Breathy Nasals and /Nh/ Clusters in Bengali, Hindi, and Marathi." *UCLA Working Papers in Phonetics*, 104: 82–106.

Fourakis, Marios and Port, Robert. 1986. "Stop Epenthesis in English." *Journal of Phonetics*, 14: 197–221.

Goldsmith, John A. 1976. Autosegmental Phonology. MIT dissertation.

Greenlee. Mel. 1973. "Some Observations on English Initial Consonant Clusters in a Child Two to Three Years Old." *Papers and Reports on Child Language Development*, 6: 97–106.

Guillemot, Céleste, Seunghun J. Lee, and Fuminobu Nishida. 2019a. "An Acoustic and Articulatory Study of Drenjongke Fricatives." *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (Calhoun, S., Escudero, P., Tabain, M., and Warren, P., eds.), 1179–1183, Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Guillemot, Céleste, Tomoko Monou, Shigeto Kawahara, and Seunghun J. Lee. 2019b. "Prosody Comes First? Phonetic Realization of Long Vowels in Drenjongke." Talk presented at the 6th NINJAL International Conference on Phonetics and Phonology. NINJAL. December 13–15, 2019.

Guillemot, Céleste, Seunghun J. Lee, and Jeremy Perkins. 2020. "The Retroflex Tongue Position in Drenjongke (Bhutia)." Talk presented at the 2020 Summer Southeast Asia Conference Program. Hankuk University of Foreign Studies, Seoul, Korea, June 19, 2020.

Hall, Nancy Elizabeth. 2003. Gestures and Segments: Vowel Intrusion as Overlap. University of Massachusets dissertation.

Halle, Morris. 1986. *The Immanent Structure of Speech Sounds*. MIT, ms.

Halle, Morris and Kenneth N. Stevens. 1971. "A Note on Laryngeal Features." *MIT Research Laboratory of Electronics Quarterly Progress Report*, 101: 198–213.

Harms, Robert T. 1973. *Some Nonrules of English*. Bloomington: Indiana University Linguistics Club.

Hayashi, Yuka. 2013. *Minami Ryukyuu Miyako Ikema hoogen no bunpoo* [Grammar of the Ikema variety of Miyako, a southern ryukyuan language). Kyoto University dissertation.

Howe, Darin and Douglas Pulleyblank. 2001. "Patterns and Timing of Glottalisation." *Phonology*, 18(1): 45–80.

Huffman, Marie K. and Thomas J. Hinnebusch. 1998. "The Phonetic Nature of the "voiceless" Nasals in Pokomo: Implications for Sound Change." *Journal of African Languages and Linguistics*, 19: 1–19.

Iverson, Gregory K. and Joseph C. Salmons. 2003. "Laryngeal Enhancement in Early Germanic." *Phonology*, 20: 43–74.

Jessen, Michael and Magnús Petursson. 1998. "Voiceless Nasal Phonemes in Icelandic." *Journal of the International Phonetic Association*, 28(1): 43–53.

Kang, Ji-Soo. 2003. The Study on Variation and Phonological Phenomena of the Laryngeal /h/. Chungnam National Univeristy, MA thesis.

Keating, Patricia. 1988. "Underspecification in Phonetics." *Phonology*, 5(2): 275–292.

Kim, Chin-Wu. 1970. "A Theory of Aspiration." *Phonetica*, 21: 107–116.

Kingston, John. 1990. "Articulatory Binding." *Papers in Laboratory Phonology I* (Kingston, J. and

Beckman, M.E., eds.), 406–434, Cambridge: Cambridge University Press.

Kim, Ok-Young. 2005. “A Study of /h/ Deletion in Korean.” *Korean Journal of Linguistics*, 30(4): 631–650.

Kim, Hyunsoon. 1997. The Phonological Representation of Affricates: Evidence from Korean and Other Languages. Cornell University dissertation.

Kim-Renaud, Young-Key. 1975. “On h-deletion in Korean.” *Journal of Korean Linguistics*, 3: 45–64.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World’s Languages*. Oxford and Malden, MA: Blackwell Publishers.

Lee, Seunghun. J., Hyun kyung Hwang, Tomoko Monou, and Shigeto Kawahara. 2018. “The Phonetic Realisation of Tonal Contrast in Dränjongke.” *TAL2018*, *Sixth International Symposium on Tonal Aspects of Languages*: 217–221.

Lee, Seunghun J., Shigeto Kawahara, Céleste Guillemot, and Tomoko Monou. 2019a. “Acoustics of the Four-way Laryngeal Contrast in Drenjongke (Bhutia): Observations and Implications.” *Journal of the Phonetic Society of Japan*, 23(1): 65–75.

Lee, Seunghun J., Shigeto Kawahara, Céleste Guillemot, and Tomoko Monou. 2019b. “The Acoustic Correlates of the Four-way Laryngeal Contrast in Drenjongke Stops.” *Proceedings of the 19th International Congress of Phonetic Sciences*, *Melbourne, Australia 2019* (Calhoun, S., Escudero, P., Tabain, M., and Warren, P., eds.), 1445–1449, Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Lee, Seunghun J., Jeremy Perkins, and Céleste Guillemot. 2019c. “A Preliminary Ultrasound Study of the Retroflex Contrast in Drenjongke.” Talk presented at the Seoul International Conference on Speech Sciences 2019. Seoul National University. November 15–16, 2019.

Lee, Seunghun J., Céleste Guillemot, Honoka Asai, Audrey Lai, and Kotone Sato. 2019d. “A Preliminary Study of the Vowel Length Contrast in Drenjongke.” *Proceedings of the Thirty-third General Meeting of the Phonetic Society of Japan* (PSJ 33). Seisen University. September 28, 2018.

Languages, 217–221.

Lee, Seunghun J. Lee, Céleste Guillemot, Kunzang Namgyal, and Jigmee Wangchuk Bhutia. 2019e. “Prosodic Cues of Polar Questions in Drenjongke (Bhutia).” Talk presented at the 6th NINJAL International Conference on Phonetics and Phonology. NINJAL. December 13–15, 2019.

Lee, Seunghun J., Céleste Guillemot, Le Xuan Chan, and Mana Ashida. 2020. “Intonation of Questions in Drenjongke (Bhutia).” Talk presented at Prosody and Grammar Festa 4. February 15–16, 2020, Kobe University.

Li, Jinfang. 2008. “Chadong, a Newly-discovered Kam-Sui Language in Guangxi, China.” *The Tai-Kadai Languages* (Anthony V. N. Diller, Jerold A. Edmondson, and Yongxian Luo, eds.), 596–620, London & New York: Routledge.

Lombardi, Linda. 1990. “The Nonlinear Organization of the Affricate.” *Natural Language & Linguistic Theory*, 8(3): 375–425.

Lombardi, Linda. 1991. *Laryngeal Features and Laryngeal Neutralization*. Amherst, MA: GLSA Publications.

Maddieson, Ian. 1991. “Articulatory Phonology and Sukuma “Aspirated Nasals”.” *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: Special Session on African Language Structures*: 145–154.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge studies in speech science and communication. Cambridge: Cambridge University Press.

Maddieson, Ian. 1984. “The Effects on F0 of a Voicing Distinction in Sonorants and their Implications for a Theory of Tonogenesis.” *Journal of Phonetics*, 12(1): 9–15.

Maddieson, Ian and Bonny Sands. 2019. “The Sounds of the Bantu Languages.” *The Bantu Languages* (Van de Velde, M., Bostoen, K., Nurse, D. and Philippson, G., eds.), London: Routledge.

Manley, Timothy. M. 1972. “Outline of Sre Structure.” *Oceanic Linguistics Special Publications*, 12: 1–239.

Mathangwane, Joyce. 1998. “Aspirates: Their Development and Depression in Ikalanga.” *Journal of African Languages and Linguistics*, 19: 113–136.

Matisoff, James, A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press.

McDavid, Raven. I. 1945. “Burmese Phonemics.” *Studies in Linguistics*, 3(1): 6–18.

Michailovsky, Boyd and Martine Mazaudon. 1988. “Lost Syllables and Tone Contour in Dzongkha (Bhutan).” Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg. *Pacific Linguistics*, 104: 115–136.

Namgyal, Kunzang, Seunghun J. Lee, George Van Driem, and Shigeto Kawahara. 2020. "A Phonetic Analysis of Drenjongke: A First Critical Assessment." *Indian Linguistics*, 81(1-2): 1–14.

Nishida, Tatsuo. 1970. *Seibankan-Yakugo no Kenkyuu* [A Study of the Tibetan-Chinese Vocabulary: Mien-Tien-Kuan I-Yu). Kyoto: Shokado.

Nishida, Tatsuo. 1972. *Mentenkan-Yakugo no Kenkyu* [A Study of the Burmese-Chinese Vocabulary: Mien-Tien-Kuan I-Yu). Kyoto: Shokado.

Nishida, Tatsuo. 1975. "On the Development of Tones in Tibetan." *Acta Asiatica* 29: 43–55.

Ohala, John J. 1975. "Phonetic Explanations for Nasal Sound Patterns." *Nasalfest: Papers from a Symposium on Nasals and Nasalization*: 289–316. Stanford, CA: Language Universals Project, Department of Linguistics, Stanford University.

Ohala, John, J. and Manjari Ohala. 1993. "The Phonetics of Nasal Phonology: Theorems and Data." *Nasals, Nasalization and the Velum* (Huffman, M.K, and Krakow, R.A, eds.), 225–249. San Diego, CA: Academic Press.

Okell, John. 1969. *A Reference Grammar of Colloquial Burmese*. Oxford University Press.

Park, Seon. 2015. Phonological variation of Korean /h/-deletion. Seoul National University MA thesis.

Peterson, David A. 2003. "Hakha Lai." *The Sino-Tibetan Languages* (Thurgood, G. and LaPolla, R. J., eds.), 409–426, London: Routledge.

Sagey, Elisabeth. 1986. The Representation of Features and Relations in Non-linear Phonology. Cambridge, MA: MIT dissertation.

Schein, Barry and Donca Steriade. 1986. "On Geminates." *Linguistic Inquiry*, 17: 691–744.

Silverman, Daniel. 1996. "Voiceless Nasals in Auditory Phonology." *Proceedings of the Twenty-second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory*: 363–374.

Silverman, Daniel. 1997. *Phasing and Recoverability*. Series: Outstanding Dissertations in Linguistics. New York, NY, USA: Garland.

Sprigg, Richard Keith 1965. "Prosodic Analysis and Burmese Syllable Initial Features." *Anthropological Linguistics*, 7(6): 59–81.

Steriade, Donca. 1990. "Gestures and Autosegments." *Papers in Laboratory Phonology* I (John Kingston and Mary Beckman, eds.), 382–397, Cambridge: Cambridge University Press.

Steriade, Donca. 1995. "Underspecification and Markedness." *The Handbook of Phonological Theory* (John Goldsmith, ed.), 114–174, Oxford: Basil Blackwell.

Sturtevant, E. 1940. *The Pronunciation of Greek and Latin*. Philadelphia: Linguistic Society of America.

Sun, Hongkai and Guangkun Liu. 2009. *A Grammar of Anong: Language Death under Intense Contact*. Brill's Tibetan Studies Library xiv. Leiden and Boston: Brill.

Thurneysen, Rudolf. 1946. *A Grammar of Old Irish*. Dublin: Institute for Advanced Studies.

Trask, Robert L. 1996. *A Dictionary of Phonetic and Phonology*. London and New York: Routledge.

van Driem, George 1992. *The Grammar of Dzongkha*. Dzongkha Development Dommission. Royal Government of Bhutan.

van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language*. Leiden: Brill.

van Driem, George. 2016. *The Phonology of Dränjoke*. ms.

van Driem, George, Seunghun J. Lee, Shigeto Kawahara, and Karma Tshering. 2017. *Dränjongke*. ms.

Vaux, Bert 1998. "The Laryngeal Specifications of Fricatives." *Linguistic Inquiry*, 29(3): 497–511.

Wei, James and Jerold A. Edmondson. 2008. "Sui." *The Tai-Kadai Languages* (Diller, A.V.N., Edmondson, J.A. and Luo, Y., eds.), 585–595. London and New York: Routledge.

Wheatley, Julian K. 2003. "Burmese - Background, Phonology, Word Classes, Syntax (constructions and particles)." *The Sino-Tibetan languages* (Thurgood, G. and LaPolla, R. J., eds.), London: Routledge.

Yliniemi, Juha. 2005. Preliminary Phonological Analysis of Denjongka of Sikkim. University of Helsinki MA thesis.

Yliniemi, Juha. 2019. *A Descriptive Grammar of Denjongke (Sikkimese Bhutia)*. University of Helsinki (with Sikkims University) dissertation.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Acoustics of Voiceless Nasals in Angami*

TERHIIJA, Viyazonuo
SARMAH, Priyankoo

Angami (exonym), also known as Tenyidie (endonym), is one of the Tibeto-Burman languages which makes voicing contrast in nasals. Voiceless nasals are relatively rare and less studied cross-linguistically. This study attempts to study the acoustic properties of voiced and voiceless nasals in Angami, produced in three places of articulation: bilabial, alveolar, and palatal. The study also investigates the nasalization of vowels following voiced and voiceless nasals. Acoustic properties such as the first four nasal formants (N1-N4), their bandwidths (BW1-BW4), Center of Gravity (CoG), and Standard Deviation (SD) of the nasal spectrum were used to characterize the Angami voiced and voiceless nasals. Nasality was measured in the vowels following the nasals, using A1-P0, A1-P1, and A3-P0 features. Results showed that all four nasal formants demonstrated significant voicing-specific effects. Vowel nasality measures also showed significant differences depending on the voicing of the preceding nasal.

1. Introduction

Voiceless nasals are less studied (cross-linguistically), and their distribution is localized—for instance, they are more prevalent in Tibeto-Burman languages. The focus language of the current study, Angami, also known as Tenyidie, is a Tibeto-Burman language that phonemically contrasts voiced and voiceless nasals in three places of articulation. In the

---

Keywords: voiceless nasals, Angami, Tibeto-Burman, Nasalization

DOI: https://doi.org/10.15026/0002000631

* The authors would like to acknowledge the ILCAA joint research project *"Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023)*, which facilitated several discussions that helped in the preparation of this paper. The authors would also like to thank Asano Zhünyü, teacher at Government Higher Secondary School Sechü (Zubza, Nagaland) for the valuable input in cross-checking the data list with the first author prior to speech data collection.

current study, we examine the acoustic properties of the voiceless nasals and compare them with their voiced counterparts. Secondly, we also investigate the nasalization-related acoustic properties of the vowels following the Angami nasals.

Angami is spoken in the Nagaland state of India and belongs to the Tibeto-Burman language family. Angami is also known as Tenyidie (ISO 639–3: njm), spoken by about 152,796 people (Burling 2003; Marrison 1967; Office of the Registrar General & Census Commissioner of India 2011). Angami is a level-tone language that has six vowels and thirty-nine consonants. Angami is a fairly well-described language with general description of grammar (Burling 1960; Giridhar 1980; Ravindran 1974; Kuolie 2006; Blankenship *et al.* 1993) and a few acoustic studies have been reported as well (Bhaskararao and Ladefoged 1991; Lalhminghlui *et al.* 2019; Terhiija *et al.* 2018; Terhiija and Sarmah 2020). There are seven nasals in Angami, namely, m, m̥ʰ, n, n̥ʰ, ɲ, ɲ̊ʰ and ŋ. As seen here, among the nasals in Angami, the velar nasal does not have a voiceless counterpart. For the current study, we took only the nasals with both voiced and voiceless counterparts.

The Tibeto-Burman languages have a prevalence of voicing contrasts in nasals, as in Angami. Five different subgroups of the Tibeto-Burman language family — Himalayish, Qiangic, Lolo-Burmese, Nungish, and Kamarupan are reported to have voicing contrasts in nasals (Matisoff 2003). In addition to the Tibeto-Burman languages, Welsh and Ikema Ryukyuan have also reported voicing contrasts in nasals (Ford 2016; Hammond 2019). Among the Tibeto-Burman languages, Burmese is relatively well-studied in terms of nasal voicing contrasts. Several instrumental studies in Burmese confirmed that a voiceless nasal consists of a voiceless nasal friction and a voiced nasal region (Dantsuji 1984; 1986). The Burmese voiceless nasals are initially voiceless but voiced towards the termination of the nasals (Bhaskararao and Ladefoged 1991). Burmese nasals differ from Angami nasals, as the latter has continuous nasal airflow with no voicing throughout the production of the voiceless nasal. Hence, the languages with voiceless nasals show two patterns, one with pre-aspiration followed by the voiceless nasals terminating in a voiced nasal. Languages such as Mizo and Burmese belong to this type. On the other hand, languages such as Angami and Xumi consist of voiceless nasals characterized by entirely voiceless nasal portions followed by aspiration (Bhaskararao and Ladefoged 1991; Chirkova *et al.* 2019).

Burmese is one of the few languages where the acoustic properties of the voiceless nasals have been studied (Dantsuji 1984). In his study of Burmese nasals, Dantsuji has analyzed the spectral and temporal characteristics of the voiced and voiceless nasals. He reports several prominent energy peaks in the spectrum of the voiceless nasal friction portion. The intensity of the voiceless portion of the Burmese voiceless nasal is higher than the intensity of the voiced portion of the voiceless nasal. It is also noticed that the voiced portions of the voiceless nasals are shorter than the voiced nasal stops (Dantsuji 1984). In a subsequent study on the Burmese voiceless nasals by Dantsuji, acoustic properties of the nasal murmur spectra (nasal part of the voiceless nasal) such as the first nasal formants (N1), second nasal formants (N2), third nasal formants (N3) along bandwidths (B1, B2, B3) were used to find cues for place of articulation for voiceless nasals (Dantsuji 1986). It is reported

Table 1 Example of nasal minimal pairs in Angami

Voiced		Voiceless		
IPA	Meaning	IPA	Meaning	Tone[1]
ma	price	m̥ʰa	something	T4
na	budding	n̥ʰa	plant	T1
ɲa	crazy	ɲ̊ʰa	messy	T1

that cues of places of articulation could be distinguished by the nasal formants and their bandwidths using the nasal portion of the voiceless nasals (Dantsuji 1986).

Besides the nasal formants and nasal formant bandwidths, previous works have reported measurements of the voicing rate for voiceless nasals. In one such study, Xumi and Burmese showed significant differences in the rate of voicing for voiced and voiceless nasals (Chirkova *et al.* 2019). In the case of Mizo and Angami, instrumental studies using the Nasometer II were also conducted, which showed fined grained characteristics of Mizo and Angami voiceless nasals (Lalhminghlui and Sarmah 2021; Terhiija and Sarmah 2020). Nevertheless, nasal formants and their bandwidths have been shown to be reliable measures in characterizing nasals. For instance, in an acoustic study of the voiced nasal consonants in three central Australian languages, it was reported that nasal formants, their bandwidths along with the Center of Gravity (CoG), and Standard Deviations (SD) calculated from the nasal spectrum provided reliable characterization of nasals with different places of articulation (PoA) (Tabain *et al.* 2016).

This study also looks into the effect of voiced and voiceless nasals on following vowels in terms of vowel nasality. Vowel nasality occurs when the velum is lowered during vowel production, which opens the velopharyngeal port and allows air to flow through the nose and mouth (Styler 2015; Chen 1997). While nasalization in vowels may be phonemic, as in French or Hindi, vowels may also be nasalized due to coarticulatory effects of the preceding nasal (Maeda 1993; Delvaux *et al.* 2002; Ohala and Ohala 1975; Malécot 1960). Hence, we assume that vowels in Angami that are preceded by nasals will be nasalized, and the nasalization will differ for voiced and voiceless nasals.

Considering the discussion in this section, we present the results of acoustic studies conducted on Angami voiced and voiceless nasals. An example of nasal minimal pairs in Angami is shown in Table 1. In the current study, we consider acoustic properties directly extracted from the nasal consonants, such as four nasal formants (N1, N2, N3, N4) and their bandwidths, CoG, and SD of the nasal spectrum. As we expect the effect of the nasals on the following vowels, we also investigate the vowel nasality using A1-P0, compensated A1-P0, A1-P1, and A3-P0 (Styler 2013).

A detailed discussion of these features is provided in Section 2.4. The rest of the paper is organized as follows: Section 2 describes the methodology, which includes a description

---

[1] Here, T1 represents the high tone, T5 represents the low tone, whereas T2, T3 & T4 are the intermediate tones. The corresponding representation of tones in Chao's tone numerals are 55 (T1), 44 (T2), 33 (T3), 22 (T4) and 11 (T5).

Table 2 Example of a nasal target sound /ne/ produced in the three environment

Context	IPA	Meaning
Sentence	*puo a ne ʃə*	s/he push me
Carrier	*a ne puba*	I said push
Isolation	*ne*	push

of the speech corpus, and the acoustic and statistical analyses. Section 3 discusses the study's results, and finally, the conclusion is in Section 4.

2. Methodology

The participants, materials, the recording procedures, and the acoustic and statistical analyses are all addressed in this section.

2.1 Speakers

13 native Angami speakers (six males and seven females) participated in the study. The participants are all residents of Kohima village, except for one female speaker. However, all the participants spoke the Kohima variety which is considered the standard variety of the Angami language (Ezung 2018). The average speaker's age is 40 (Standard Deviation = 15.5) at the recording time. In addition to Angami, all participants are fluent in English (the state's official language) and Nagamese (*lingua franca*). All the participants had at least a bachelor's degree, and some were working professionals.

2.2 Materials

The material was designed in a way to capture the voiced and voiceless nasals (/m, m̥ʰ, n, n̥ʰ, ɲ & ɲ̊ʰ/) in the context of six vowels (/i, e, a, o, u & ə/) whenever possible. The combination of nasals and vowels were meaningful words in the language. The nasals occur in three places of articulation, namely, bilabial, alveolar, and palatal. The target nasal sounds were monosyllabic (CV) in structure. There are two sets of data: the first set produced by six speakers includes 74 unique nasal tokens produced in three environments, namely, sentence, isolation, and carrier phrase, as seen in Table 2.

The second data set is a subset of the first, where 28 nasal tokens were produced only in isolation and repeated three times by seven speakers. The number of tokens elicited in the first and second data sets is 1332 (n = 6 speakers) and 588 (n = 7 speakers), respectively. Due to mispronunciation, 43 tokens were removed. A total of 1,877 tokens are considered for analysis, combining both data sets. The distribution of the six nasals with all possible vowel combinations in the data set is shown in Table 3.

2.3 Recording of the Speech Data

The recordings were conducted in a noise-controlled environment at the speaker' home. Each person took about an hour for the recording session. The data were collected

Table 3 Distribution of Nasals and Vowels in the data

	a	e	i	ə	o	u	Total
/m/	76	76	73	0	76	38	339
/m̥ʰ/	87	39	39	0	21	124	310
/n/	72	76	75	0	75	94	392
/n̥ʰ/	89	89	21	0	87	21	307
/ɲ/	74	16	75	73	65	40	343
/ɲ̊ʰ/	36	0	19	111	0	20	186
Grand Total							1877

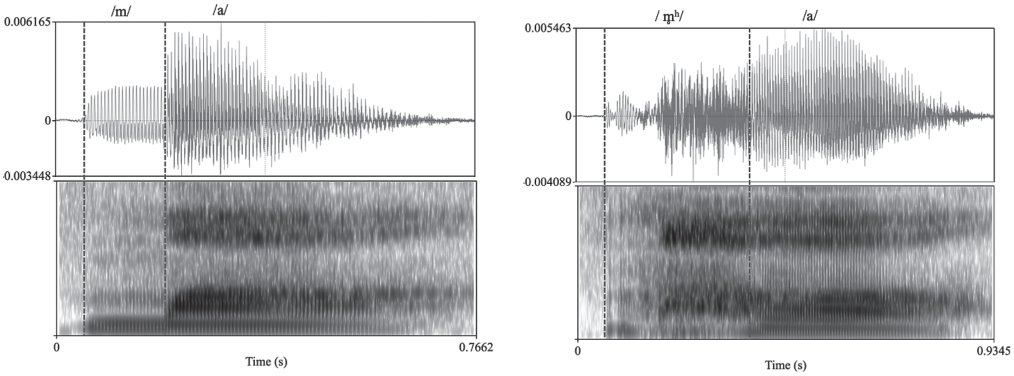


Figure 1 An example of the nasal and vowel boundary for the words /ma/ 'price' and /m̥ʰa/ 'things' as produced by a female speaker.

at two different times, first in September 2018 and later in October 2019. The recording was collected using a Shure unidirectional head-worn microphone (model: Shure SM10-CN) connected to a Tascam linear PCM recorder (model: TASCAM DR-100MKII) through an XLR jack. The sampling frequency at the recording was 44.1 kHz, 24 bits in WAV format.

2.4 Acoustic and Statistical Analysis

The speech recordings were transferred to a computer for analysis. The sound files were segmented and annotated by the first author, a native speaker of Angami, by means of listening and visual examination of the waveforms and spectrograms using Praat 6.0.43 (Boersma and Weenink 2018). An example of the nasal and vowel boundary for the words /ma/ and /m̥ʰa/ produced by a female speaker is shown in Figure 1. A Praat script was used to extract the first four nasal formants, namely, N1, N2, N3, and N4 in hertz (HZ), along with their bandwidths (BW1, BW2, BW3, BW4) were calculated at the temporal midpoint of the nasals. Further, Center of Gravity (CoG) and Standard Deviation (SD) were also extracted. To visualize the nasal formant contours, nasal formants were extracted at every 10% of the total duration.

To study vowel nasality, the nasality auto-measure Praat script developed by Styler was used (Styler 2013). The spectral measurements and vowel duration were taken at 11 points (0–100%) with 10% intervals. The four spectral measurements taken as adapted for the current study are A1-P0, compensated A1-P0, A1-P1, and A3-P0. A1-P0 is the amplitude of the first formant (A1) minus the amplitude of the first nasal peak (P0). Usually, P0 occurs between the first two harmonics (H1, H2). A1-P1 is calculated by subtracting the amplitude of the second nasal peak that occurs in the 850–1050 Hz range from A1. A3-P0 is the difference in amplitude between the third formant and P0. A3-P0 captures the spectral tilt that may help characterize nasality. As seen in the study of French (nasalized vowels) and English (phonetic nasality), the values in features such as A1-P0, A1-P1, and A3-P0 decreased with increased nasality (Chen 1997; Styler 2017). A detailed discussion on the effectiveness of these features in measuring vowel nasality can be found in Styler (2013, 2017) and Chen (1997).

R's statistical packages were used for exploratory statistics. Linear Mixed Effects (LME) models were carried out using the *lme4* package on R, and descriptive statistics like mean, standard deviation, and standard errors were calculated using R's core functions (Bates *et al.* 2015; R Core Team 2022). Using the Anova function in the *car* package, Type II Wald chi-square tests were performed on the linear mixed models to determine the significance of the fixed effects (Fox and Weisberg 2018). The plots for visualizing nasal and nasality data were produced on R with the *ggplot2* package (Wickham 2016).

3. Results

The acoustic properties of voiced and voiceless nasals in Angami and the nasalization-related acoustic properties will be discussed in the following section.

3.1 Acoustic Properties of Voiced and Voiceless Nasals

The nasal formant contours have been plotted for visualization in Figure 2. As mentioned in section 2.4, the nasal contours (N1-N4) were calculated at every 10% of the total duration. Figure 2(a) clearly distinguishes between voiced and voiceless nasals for the first nasal formant (N1). The voiced nasals have low N1 values (< 400 Hz), while the voiceless have high N1 (> 600 Hz). We observed that in the voiceless nasal, the N1 value is highest for bilabial, followed by alveolar and palatal (/m̥ʰ/ > /n̥ʰ/ > /ɲ̥ʰ/). Whereas, amongst voiced nasal, bilabial nasal has higher N1 values than alveolar and palatal (/m/ > /n/ & /ɲ/). In Figure 2(b), similar to N1, the N2 values of voiced are lower than voiceless nasals, but the differences between the two are relatively less. Considering the mid 50% of the total duration, the N2 values for voiceless nasals are about 1800–2400, while the voiced is about 1600–1800. The N2 values for voiceless nasals, the palatal is highest, followed by alveolar and bilabial (/ɲ̥ʰ/ > /n̥ʰ/ > /m̥ʰ/). The voiced bilabial /m/ is the lowest, followed by palatal nasal /ɲ/ and alveolar nasal /n/. Figure 2(c) shows that for the N3 values, there is a clear distinction between voiceless nasals and voiced nasals. The range of voiceless nasal is 2850 Hz and above, while voiced nasal is 2750 Hz and below. The PoA for the voiceless nasals for N3 is not obvious, as all the nasals have merged or overlapped. However, for voiced nasals, we see similar patterns with N2.

There is a marginal difference between voiceless and voiced nasals, as seen in Figure

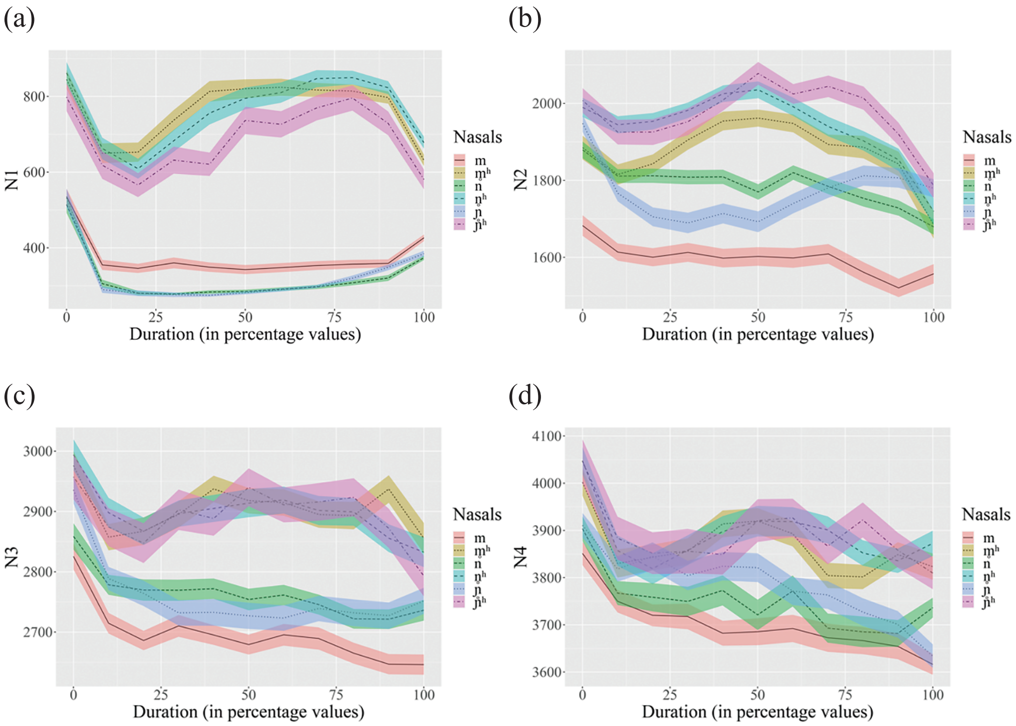


Figure 2 Formant contours (N1-N4) in Hertz for the Angami nasal consonants

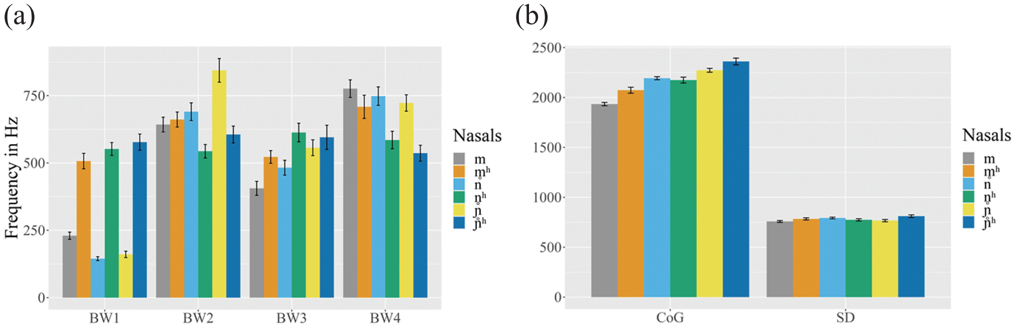


Figure 3 Bar plots of (a) the mean of nasal formants bandwidths (BW1-BW4), (b) SD and CoG calculated from the nasal spectrum.

2(d). The N4 values for voiced nasal ranges from 3680 Hz and below, while voiceless nasal ranges from > 3850 and above. There is a clear PoA pattern in voiced nasals (/m/ > /n/ & /ɲ/). However, in voiceless nasals, there is an overlapping.

Furthermore, the mean and standard error of the frequencies of nasal formant bandwidths, along with CoG and SD, is displayed in Figure 3. Figure 3(a) shows that the frequency values of BW1 and BW3 are lower than BW2 and BW4. The distinction between voiced nasal and voiceless nasal is seen in the values of BW1 and BW3. Figure 3(b), the CoG and SD mean values are displayed, where it is observed that the CoG values of voiced nasals are relatively lesser than voiceless nasals. CoG depends on PoA (bilabial > alveolar > palatal) in both the contrasting nasals. As for SD, there is no significant difference in the nasals.

Table 4 Results of Analysis of Deviance (ANOVA) tests on the LME models for estimating the effect of nasals consonants and estimates of post-hoc Bonferroni tests for pair-wise contrasts.

Feature	χ^2	Contrasts	Estimate	SE	df	*t*-ratio	*p-value*
N1	Nasals: 1119.6***	m - m̥ʰ	−519.4	26.3	1899	−19.7	< 0.0001
	Context: 13.5*	n - n̥ʰ	−543.5	26.1	1899	−20.8	< 0.0001
	Nasals x Context: 81.5***	ɲ - ɲ̊ʰ	−463.3	30.3	1899	−15.3	< 0.0001
N2	Nasals: 379.1***	m - m̥ʰ	−426.0	32.8	1899	−13.0	< 0.0001
	Context: 5.2	n - n̥ʰ	−299.5	32.6	1899	−9.1	< 0.0001
	Nasals x Context: 74.1***	ɲ - ɲ̊ʰ	−438.4	37.8	1899	−11.5	< 0.0001
N3	Nasals: 204.8***	m - m̥ʰ	−276.0	25.3	1899	−11.0	< 0.0001
	Context: 8.0	n - n̥ʰ	−175.0	25.2	1899	−7.0	< 0.0001
	Nasals x Context: 61.1***	ɲ - ɲ̊ʰ	−225.4	29.2	1899	−7.7	< 0.0001
N4	Nasals: 73.1***	m - m̥ʰ	−249.0	40.1	1871	−6.1	< 0.0001
	Context: 3.2	n - n̥ʰ	−174.4	40.2	1871	−4.4	< 0.001
	Nasals x Context: 30.4	ɲ - ɲ̊ʰ	−132.5	47.0	1871	−3.0	0.0708
BW1	Nasals: 605.1***	m - m̥ʰ	−262.0	27.1	1899	−9.6	< 0.0001
	Context: 7.3	n - n̥ʰ	−417.6	27.0	1899	−15.5	< 0.0001
	Nasals x Context: 46.9**	ɲ - ɲ̊ʰ	−424.3	31.2	1899	−13.6	< 0.0001
BW2	Nasals: 49.5***	m - m̥ʰ	−40.2	48.5	1899	−0.8	1.0000
	Context: 2.5	n - n̥ʰ	128.3	48.2	1899	2.7	0.1185
	Nasals x Context: 23.8	ɲ - ɲ̊ʰ	255.0	56.0	1899	4.5	< 0.0001
BW3	Nasals: 32.1***	m - m̥ʰ	−123.6	44.2	1899	−2.8	0.0779
	Context: 1.3	n - n̥ʰ	−157.6	44.0	1899	−3.6	< 0.05
	Nasals x Context: 41.4*	ɲ - ɲ̊ʰ	−98.5	51.0	1899	−1.9	0.8052
BW4	Nasals: 31.3***	m - m̥ʰ	91.4	51.7	1870	1.8	1.0000
	Context: 2.0	n - n̥ʰ	202.3	51.7	1871	3.9	< 0.05
	Nasals x Context: 26.1	ɲ - ɲ̊ʰ	200.1	60.3	1871	3.3	0.0139
CoG	Nasals: 237.5***	m - m̥ʰ	−197.0	31.5	1899	−6.2	< 0.0001
	Context: 2.2	n - n̥ʰ	−53.9	31.4	1899	−1.7	1.0000
	Nasals x Context: 96.7***	ɲ - ɲ̊ʰ	−102.3	36.4	1899	−2.8	0.0752
SD	Nasals: 16.5**	m - m̥ʰ	−28.6	14.4	1899	−1.9	0.6997
	Context: 4.3	n - n̥ʰ	24.0	14.3	1899	1.7	1.0000
	Nasals x Context: 24.6	ɲ - ɲ̊ʰ	−45.0	16.6	1899	−2.7	0.1032

In order to confirm the statistical significance of the nasal formant values and bandwidths among the six nasals, we constructed LME models. The dependent variables were the nasal features which included the four nasal formant values at midpoint and their bandwidths. Additionally, the average Center of Gravity (CoG) and Standard Deviation (SD) of the nasal spectrum were considered as dependent variables. Hence, nasals, context, and their interaction were the fixed effects, while the speaker was the random effect, as shown in (1).

$$\textit{Nasal features} \sim \textit{Nasals} + \textit{context} + \textit{Nasals} \times \textit{context} + \textit{(1|speaker)} \tag{1}$$

The summary of the LME model is presented in Table 4. The constructed models were subjected to Type II Wald Chi-square tests for analysis of deviance on LME models separately, where it showed that the effect of nasal consonants on nasal formants was significant. It is also observed that the interaction of nasals and context also has a significant effect. The $\chi 2$ value, degrees of freedom (df), and significance for each feature are shown in Table 4 (2nd column). Hence, looking into the pairwise comparisons of nasal features by voicing contrasts of the nasals is pertinent. The LME models were further subjected to post-hoc Bonferroni analysis to see the significant differences between voiced and voiceless nasals. The result of the test is reported in Table 4. The estimate, standard error (SE), degrees of freedom (df), t-ratio, and significance of the voiced and voiceless nasals under each model are reported. As seen in the table, the voicing contrasts in the three places of articulation, that is, bilabial, alveolar, and palatal, significantly affect nasal features such as N1, N2, N3, and BW1. As for N4, the effect is observed in bilabial and alveolar voicing contrasts. It is also noticed that voicing contrast does not have any significant effect on Standard Deviation (SD) and Center of Gravity (CoG).

3.2 Vowel Nasality

Figure 4 provides a visual representation of the nasal effects on the vowel in each context. Figures 4(a) and 4(b), the nasals have been plotted on A1-P0 and compensated A1-P0 values. An upward trajectory of the voiceless nasals is seen from the initial to 25% of the nasal formant contour. There is a distinction between voiced and voiceless nasals, as observed. In carrier and sentence context, the values correlate with PoA in voiceless nasals (m̥ʰ > n̥ʰ> ɲ̊ʰ). A significant effect on A1-P1 is seen in Figure 4(c), where the voiced nasal tends to differentiate from the voiceless nasal across the whole nasal contour. Voiceless alveolar nasal /n̥ʰ/ have the lowest A1-P1, followed by bilabial and palatal. Further, as seen in Figure 4(d), the A3-P0 values overlap and do not show any clear pattern. To see the effect of voicing contrasts on nasality, we conducted exploratory modeling where the average four nasality features, A1-P0, compensated A1-P0, A1-P1, and A3-P0, were the dependent variables, Nasals, context, and their interactions were the fixed effects, and the speaker was the random effects as shown in (2).

$$\textit{Nasality features} \sim \textit{Nasals} + \textit{context} + \textit{Nasals*context} + \textit{(1|speaker)} \tag{2}$$

Type II Wald Chi-square tests for analysis of deviance test conducted on the four LME models separately showed that voicing contrasts have a significant effect on all four nasality features, as seen in Table 5. There is no interaction of nasal types and contexts indicating no change in nasality depending on context. The LME models were further subjected to Bonferroni post-hoc test to see the differences between voiced and voiceless nasals in three places of articulation (PoA). Table 5 shows the estimates, standard errors (SE), degrees of freedom (df), t-ratios, and p-values of homorganic voiced and voiceless nasals. The pairwise comparison of the voiced and voiceless nasals showed that nasality features such as A1-P0, compensated A1-P0, and A1-P1 have a significant effect on voicing. A3-P0 does not show any systematic difference between voiced and voiceless nasals and hence, it can be concluded that A1-P0, compensated A1-P0, and A1-P1 are the salient features that manifest the difference of voicing in nasals in the case of Angami.

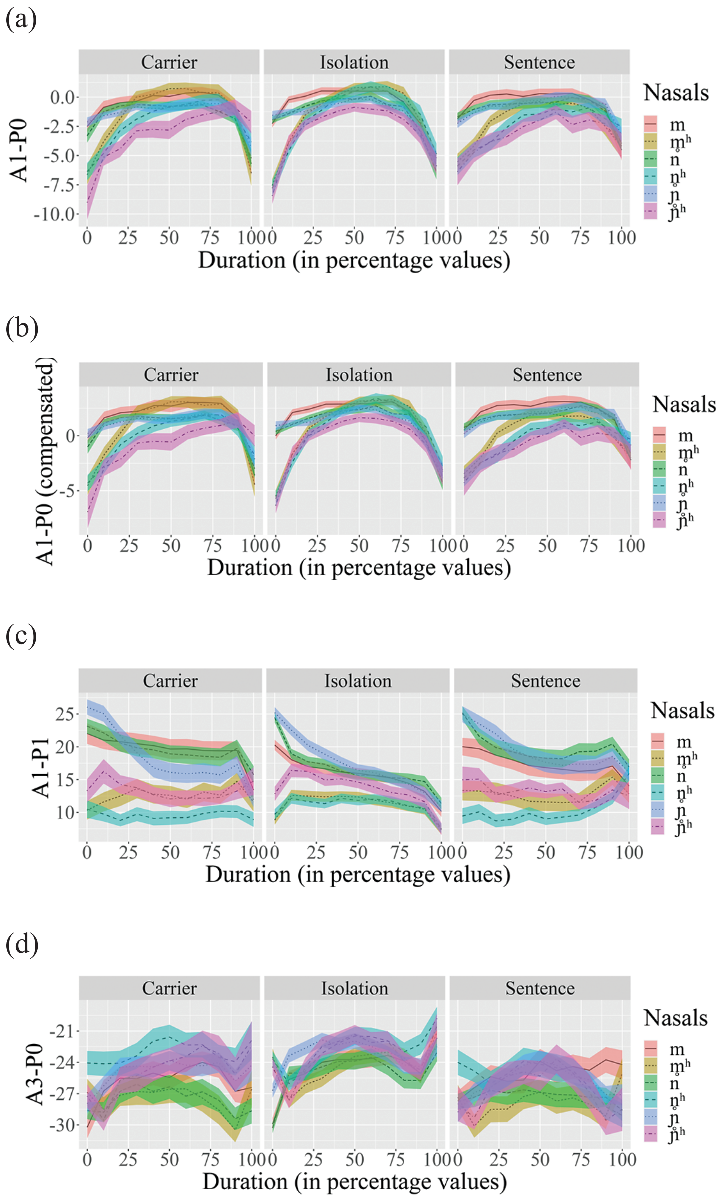


Figure 4 Nasality measures for the vowels following Angami nasal.

While the effect of voicing on nasals is manifested on the following vowels in terms of vowel nasality, it is not clear to what extent into the following vowel is the effect significant. To investigate that, 11 separate LME models were constructed with nasality measures (A1-P0, compensated A1-P0, and A1-P1) at every 10% of the total duration of the vowel following nasals. Only A1-P0 compensated A1-P0, and A1-P1 were considered as they showed a systematic effect of the preceding nasal voicing as discussed previously. All 11 models were subjected to post-hoc Bonferroni tests to obtain pair-wise comparisons. The results of the comparisons are provided in Table 6. As seen in the table, the difference

Table 5 Analysis of Deviance (ANOVA) tests on the LME models with average nasality values as dependent variables and pair-wise contrasts estimated from post-hoc Bonferroni tests.

Model	Contrasts	Estimate	SE	df	*t*-ratio	*p-value*
A1-P0	m - m̥ʰ	0.9	0.3	1511	3.5	< 0.01
$\chi^2(5) = 68.1$*	n - n̥ʰ	0.8	0.3	1511	3.0	< 0.05
	ɲ - ɲ̊ʰ	1.9	0.3	1511	6.0	< 0.0001
com. A1-P0	m - m̥ʰ	1.7	0.3	1511	4.2	< 0.001
$\chi^2(5) = 84.4$*	n - n̥ʰ	1.1	0.3	1511	4.0	< 0.001
	ɲ - ɲ̊ʰ	2.1	0.3	1512	6.5	< 0.0001
A1-P1	m - m̥ʰ	5.8	0.9	1513	6.7	<0.0001
$\chi^2(5) = 193.9$*	n - n̥ʰ	8.9	0.8	1513	10.6	<0.0001
	ɲ - ɲ̊ʰ	5.0	1.0	1515	5.0	<0.0001
A3-P0	m - m̥ʰ	1.6	0.7	1511	2.2	0.3871
$\chi^2(5) = 30.2$*	n - n̥ʰ	−3.1	0.6	1511	−4.6	< 0.001
	ɲ - ɲ̊ʰ	0.1	0.8	1512	0.2	1.000

$p < 0.001$ is indicated by an asterisk

Table 6 Significance of nasality values from Bonferroni post-hoc pairwise comparisons conducted on LME models for 11 equidistant points of vowels following nasals.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A1-P0											
m - m̥ʰ	****	****	****								
n - n̥ʰ	****	****	****	***							
ɲ - ɲ̊ʰ	****	****	****	****	***	**					
com.A1-P0											
m - m̥ʰ	****	****	****								
n - n̥ʰ	****	****	****	****							
ɲ - ɲ̊ʰ	****	****	****	****	****	****	*	*			
A1-P1											
m - m̥ʰ	****	****	****	****	****	****	****	****	****	*	
n - n̥ʰ	****	****	****	****	****	****	****	****	****	****	**
ɲ - ɲ̊ʰ	****	****	****	****	**	*		**	*	*	

**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

between voiced and voiceless nasals in terms of nasality is noticed in varying degrees for different nasality correlates. As far as A1-P0 is concerned, the effect of nasal voicing is significant up to 50% of the total duration of the following vowel. The effect of nasal voicing permeates more into the following vowel if the preceding nasal is palatal. On the other hand, in the case of bilabial nasals, the effect is the least. In the case of A1-P1, the effect is seen in the entire vowel duration.

3.3 Duration of Voiced and Voiceless Nasals

Furthermore, the six nasals' average duration across the contexts was examined, showing that voiceless nasals are longer than their voiced counterparts as shown in Table 7. To confirm the statistical significance of the nasal duration, we constructed an LME model

Table 7 The mean duration and Standard Deviation (SD) in milliseconds (ms) of the six nasals in Angami by contexts.

Nasals	Sentence	Carrier	Isolation
m	149.7 (76)	166.2 (45)	156.2 (59)
m̥ʰ	198.6 (72)	223.6 (73)	190.0 (72)
n	153.0 (57)	198.2 (64)	149.3 (56)
n̥ʰ	212.5 (90)	219.3 (81)	190.1 (67)
ɲ	187.5 (74)	221.8 (78)	169.6 (54)
ɲ̊ʰ	251.4 (105)	234.0 (79)	207.2 (75)

Table 8 Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of voicing, PoA, and contexts on nasal duration.

Fixed effects	χ^2	*df*	*p*-value
Voicing	267.0	1	< 0.001
PoA	86.8	2	< 0.001
Context	136.7	2	< 0.001
Voicing x PoA	1.0	2	n.s
Voicing x Context	13.0	2	< 0.01
PoA x Context	19.1	4	< 0.001
Voicing x PoA x Context	13.7	4	< 0.01

with nasal duration as the dependent variable. Assuming the contextual effects on nasal duration, voicing contrasts, place of articulation (PoA), context, and their interactions were considered fixed effects, whereas, the speaker was considered a random effect as shown in (3).

$$Duration \sim voicing + PoA + contexts + voicing \times PoA \times context + (1/speaker) \tag{3}$$

Results of the Type II Wald chi-square tests for analysis of deviance conducted on the LME model show that voicing contrast, PoA, contexts, and their interactions have a significant effect on the nasal duration as seen in Table 8. Hence, it can be concluded that the voicing of the nasals has significant duration differences, however, the durational differences are also affected by the place of articulation of the nasals and the contexts in which they are produced. The interaction plot shown in Figure 5 confirms that in all cases, the voiceless nasals are longer than their voiced counterparts. However, in the carrier frame context, the difference between the duration of voiced and voiceless nasals with palatal and alveolar PoA seems to be less. Hence, to explore these interactions in detail, the LME model was subjected to a pairwise comparison using a Bonferroni post-hoc test. The results of the post-hoc test as shown in Table 9 confirm that the bilabial nasals maintain statistically significant durational differences across all contexts. However, the alveolar and palatal nasals do not show any significant durational difference in the carrier frame context. In all other contexts, the durational difference between voiced and voiceless nasal

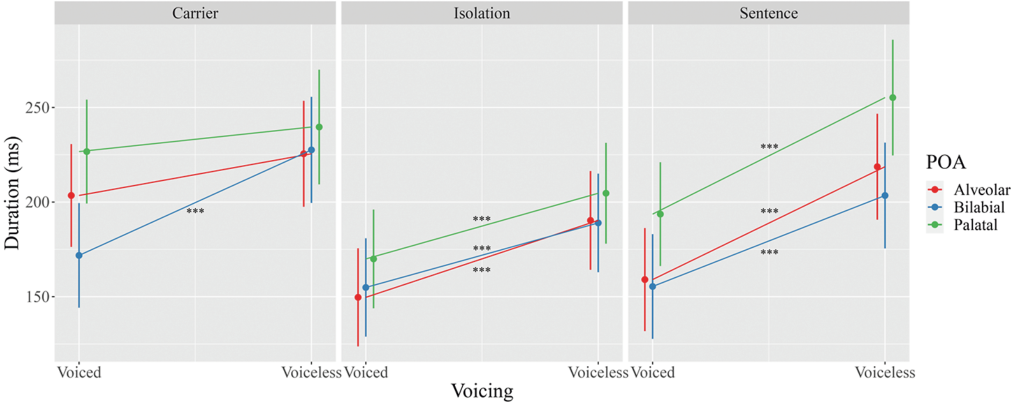


Figure 5 Variable interaction plot generated from an LME model with the duration of nasal as the dependent variable and voicing types, contexts, PoA, and their interactions as fixed effects. The statistically significant interactions are indicated with asterisks.

Table 9 Estimates of the duration of nasal pairs obtained from a Bonferroni post-hoc pairwise comparison conducted on an LME model, with significant estimates (*p* < 0.0001) marked with an asterisk.

Contrast	Sentence	Carrier	Isolation
m - m̥ʰ	−48.1*	−55.8*	−34.1*
n - n̥ʰ	−59.6*	−22.0	−40.7*
ɲ - ɲ̊ʰ	−61.6*	−13.0	−34.7*

is systematic and statistically significant. These results also concur with the findings in Burmese (Dantsuji1984; Chirkova *et al.* 2019), Icelandic (Jessen and Pétursson 1998), and Mizo (Lalhminghlui *et al.* 2019) where it was shown that voiceless nasals are longer than their voiced counterpart.

4. Conclusions

In this paper, we look into Angami's acoustic features of voiced and voiceless nasals. Regarding the nasal consonants, we observed that the voiced and the voiceless nasals have significantly different patterns. As shown in the nasal formant tracks (Figure 2), the voiced nasals consistently have lower formant values than their voiceless counterparts. The difference between voiced and voiceless nasals is visually prominent in the case of N1 and N3. Another cursory observation here is that the nasal formants have PoA-specific effects in both voiced and voiceless nasal contexts. The PoA effect of the voiceless nasal is pronounced in N1 and N2. However, the PoA effect is asymmetric in N1 and N2. Regarding N1, the voiceless bilabial nasal has the highest, and the palatal nasal has the lowest values.

On the other hand, in terms of N2, the reverse is observed. In the case of voiced nasals, a similar pattern is observed. While the voiced bilabial nasal has higher N1 than the voiced alveolar and palatal nasals, in the case of N2, the reverse is observed. In terms of statistical models, all four nasal formants showed significant P-values when all six types of nasals were used as fixed factors. Post-hoc analyses of the LME models confirmed that there was a significant difference between the voiced and the voiceless nasals in each pair of nasals (bilabial, alveolar, palatal). Similar post-hoc analyses on CoG and SD failed to yield any statistically significant for voicing in nasals.

As voiceless nasals are relatively rare in the world's languages, there are no studies that have tried exploring the physiology in the production of such phonemes. The closest comparison that can be drawn is from the observations on voiceless or whispered vowels. In the case of whispered vowels, they are observed to be produced with partially open glottis. Whispered vowels are characterized by higher formant frequencies than their voiced counterparts (Peterson 1961; Smith 1973; Lehiste 1964). In the study of English vowels, Lehiste (1964, 1970) noted that F1 was approximately 200–250 Hz higher, whereas F2 and F3 were 100–150 Hz higher, in whispered vowels than in their voiced counterparts. Lehiste (1964, 1970) postulated that the differences in formant frequencies are due to the degree of glottal openness. During voiced vowel production, the glottis is closed while, during whispered vowel production, the glottis is open resulting in formants that are higher in frequency. This formant raising is due to the fact that the resonant frequencies of an acoustic tube open at both ends are higher than that of a tube closed at one end. Considering this, we argue that voiceless nasals also function in a similar manner. The open glottis in the production of voiceless nasals contributes to the raising of the formant frequencies.

It has been noted that open glottis have wider formant bandwidths than closed glottis (Smith 1973; Fujimura and Lindquist 1971). In the case of whispered vowels, the coupling of the trachea to the supraglottic spaces resulting in the damping of formant frequencies leads to the higher first formant bandwidth (BW1) values (Peterson 1961; Fujimura and Lindquist 1971). Hence, BW1 is larger in voiceless vowels than in voiced vowels. Additionally, it has also been shown that the value of formant frequency is positively correlated with their bandwidths, beyond a certain threshold (Hawks & Miller 1995).

As for vowel nasality, average A1-P0, compensated A1-P0, and A1-P1 showed effects of nasal voicing. A detailed investigation showed that the effect of the preceding nasal voicing contrast is manifested in at least 30% of the following vowels. However, in the case of A1-P0, the effect seems to be dependent on the place of articulation of the nasals. We postulate that the varying changes in nasality by PoA are due to the size of oral occlusion in the production of the nasals. As the oral cavity is larger in the case of the bilabial nasals, more energy is lost in the oral cavity resulting in less nasal energy. This results in shorter nasality in vowels followed by bilabial nasals. On the other hand, the palatal nasals, which have the shortest oral cavity result in the least energy lost in the oral cavity resulting in higher nasal energy leading to higher nasality. Hence, oral cavity size in the production of voiceless nasals is inversely proportionate to the vowel nasality.

Nasality is higher after the voiceless nasals than after the voiced nasals. As studies on voiceless nasals are extremely rare, not many studies have looked into the relationship between the voicing of nasals and nasality in the following vowels. In the current study, it was noticed that nasality in vowels following voiceless nasals is higher than in voiced nasals. It is to be noted that the Angami voiceless nasals are produced with aspiration towards the end of the voiceless nasal. This breathy aspiration increases the airflow mimicking the production of a nasal. It has been seen that the increase associated with a sound segment mimics the production of a contextual vowel with nasalization. Ohala (1993) noticed that when breathiness is associated with a sound segment, the contextual vowel is nasalized as seen in the Hindi word /ãnk^h/ which is derived from the Sanskrit word /akṣii/. Intuitively, we hear more nasality in vowels following voiceless nasals in Angami than after voiced nasals. We postulate that the higher nasalization in the vowels following Angami voiceless nasals may be due to the aspiration followed by the voiceless nasals which may be phonologized by the Angami speakers. However, at this point in time, we would like to conduct more experimental studies to come to a conclusion regarding the pathway that prompted higher nasality in vowels following voiceless nasals in Angami.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1): 1–48.

Bhaskararao, Peri and Peter Ladefoged. 1991. “Two Types of Voiceless Nasals.” *Journal of the International Phonetic Association*, 21(2): 80–88.

Blankenship, Barbara, Peter Ladefoged, Peri Bhaskararao, and Nichumeno Chase. 1993. “Phonetic Structures of Khonoma Angami.” *Linguistics of the Tibeto-Burman Area*, 16(2). 69–88.

Boersma, Paul and David Weenink. 2018. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.0. 37.

Burling, Robbins. 1960. “Angami Naga Phonemics and Word List.” *Indian Linguistics*, 21: 51–60

Burling, Robbins. 2003. “The Tibeto-Burman Languages of Northeastern India.” *The Sino-Tibetan Languages* (Graham Thurgood and Randy J. LaPolla, eds.), 169–191, London, Routledge.

Chen, Marilyn Y. 1997. “Acoustic Correlates of English and French Nasalized Vowels.” *The Journal of the Acoustical Society of America*, 102(4): 2360–2370.

Chirkova, Katia, Patricia Basset, and Angélique Amelot. 2019. “Voiceless Nasal Sounds in Three Tibeto-Burman Languages.” *Journal of the International Phonetic Association*, 49(1): 1–32.

Dantsuji, Masatake. 1984. “A Study on Voiceless Nasals in Burmese.” 音声科学研究, 18: 1–14.

Dantsuji, Masatake. 1986. “Some Acoustic Observations on the Distinction of Place of Articulation for Voiceless Nasals in Burmese.” 音声科学研究, 20: 1–11.

Delvaux, Véronique, Thierry Metens, and Alain Soquet. 2002. “French Nasal Vowels: Acoustic and Articulatory Properties.” *Seventh International Conference on Spoken Language Processing* (John H. L. Hansen and Bryan Pellom, eds.), 53–56, Causal Productions.

Ezung, Mimi Kevichüsa. 2018. *Relative Clause Formation in Tenyidie*. Dimapur, India: Heritage Publishing House. 1–5.

Ford, Catherine. 2016. The Status of Voiceless Nasals in Ikema Ryukyuan. Bachelor’s thesis, University of Alberta, Edmonton, Canada. 1–20.

Fox, John and Sanford Weisberg. 2018. *An R Companion to Applied Regression (3rd edn.)*. Sage Publications. 149–188.

Fujimura, Osamu and Jan Lindqvist. 1971. “Sweep-tone Measurements of Vocal-tract Characteristics.” *The Journal of the Acoustical Society of America*, 49(2B): 541–558.

Giridhar, Puttushetra Puttuswamy. 1980. *Angami Grammar*. Mysore: Central Institute of Indian Languages Mysore. 1–107.

Hammond, Michael. 2019. “Voiceless Nasals in Welsh.” *Journal of Celtic Linguistics*, 20(1): 31–60.

Hawks, John W. and James D. Miller. 1995. “A Formant Bandwidth Estimation Procedure for Vowel Synthesis.” *The Journal of the Acoustical Society of America*, 97(2): 1343–1344.

Jessen, Michael and Magnús Pétursson. 1998. “Voiceless Nasal Phonemes in Icelandic.” *Journal of the International Phonetic Association*, 28(1-2): 43–53.

Kuolie, Duovituo. 2006. *Structural Description of Tenyidie: A Tibeto-Burman Language of Nagaland*. Kohima: Ura Academy, Publication Division. 1–210.

Lehiste, Ilse. 1964. “Acoustical Characteristics of Selected English Consonants.” *International Journal of American Linguistics*, 30(3): 1–197.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, Massachusetts, and London, England: The M.I.T Press.

Lalhminghlui, Wendy and Priyankoo Sarmah. 2021. “Characterizing Voiced and Voiceless Nasals in Mizo.” *In Proceedings of INTERSPEECH 2021*: 426–430.

Lalhminghlui, Wendy, Viyazonuo Terhiija, and Priyankoo Sarmah. 2019. “Vowel-tone Interaction in Two Tibeto-Burman Languages.” *In Proceedings of INTERSPEECH 2019*: 3970–3974.

Maeda, Shinji. 1993. “Acoustics of Vowel Nasalization and Articulatory Shifts in French Nasal Vowels.” *Nasals, Nasalization, and the Velum (Phonetics and Phonology, Vol. 5)* (Marie K. Huffman and Rena A. Krakow, eds.), 147–167, San Diego: Academic Press.

Malécot, André. 1960. “Vowel Nasality as a Distinctive Feature in American English.” *Language*, 36(2): 222–229.

Marrison, Geoffrey Edward. 1967. The Classification of the Naga Languages of North-east India. PhD thesis, School of Oriental and African Studies, University of London.

Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press. 36–40.

Ravindran, N. 1974. *Angami Phonetic Reader*. Mysore: Central Institute of Indian Languages.

Office of the Registrar General & Census Commissioner of India. 2011. *Statement-1: Abstract of Speakers’ Strength of Languages and Mother Tongues*.

Ohala, John J. 1993. “Sound Change as Nature’s Speech Perception Experiment.” *Speech Communication*, 13(1): 155–161.

Ohala, Manjari, and Manjari Ohala. 1975. “Nasals and Nasalization in Hindi.” *Nasalfest: Papers from a Symposium and Nasals and Nasalization* (Charles A. Ferguson, Larry M. Hyman and John J. Ohala, eds.), 317–332, Stanford: Language Universals Project, Dept. of Linguistics, Stanford University.

Peterson, Gordon E. 1961. “Parameters of Vowel Quality.” *Journal of Speech and Hearing Research*, 4(1): 10–29.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*, Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Smith, Wendell Franklin. 1973. A Formant Study of Whispered Vowels. Ph.D. thesis, The University of Oklahoma.

Styler, Will. 2013. *Using praat for Linguistic Research*. University of Colorado at Boulder Phonetics Lab.

Styler, Will. 2015. On the Acoustical and Perceptual Features of Vowel Nasality. Ph.D. thesis, University of Colorado at Boulder.

Styler, Will. 2017. “On the Acoustical Features of Vowel Nasality in English and French.” *The Journal of the Acoustical Society of America*, 142(4): 2469–2482.

Tabain, Marija, Andrew Butcher, Gavan Breen, and Richard Beare. 2016. “An Acoustic Study of Nasal Consonants in Three Central Australian Languages.” *The Journal of the Acoustical Society of America*, 139(2): 890–903.

Terhiija, Viyazonuo, and Priyankoo Sarmah. 2020. “Aspiration in Voiceless Nasals in Angami.” *Proceedings of Meetings on Acoustics*, 42(1): 1–13.

Terhiija, Viyazonuo, Priyankoo Sarmah, and Samudra Vijaya. 2018. “Acoustic Analysis of Vowels in Two Southern Angami Dialects.” *Oriental COCOSDA 2018*, Miyazaki, Japan.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*, Springer New York.

Journal of Asian and African Studies, Supplement, No. 4, 2024

The Interdental Approximant in Kagayanen*

YAMAMOTO, Kyosuke

This paper describes the acoustics of the interdental approximant in Kagayanen, a typologically rare speech sound in which the tongue protrudes from the mouth. This speech sound has been attested in several Austronesian languages in the Philippines and Australian languages and stated to have a lateral perceptual quality. Based on the data from naturally occurring speech, this study shows that the interdental approximant has lower F2 and smaller F2-F1 values compared to the alveolar lateral approximant. Furthermore, I discuss historical-comparative data suggesting that the interdental approximant was previously an allophone of /l/ and underwent a phonemic split from it. I also investigate the behavior of the sound within the current phonological system of Kagayanen.

Keywords: the interdental approximant, acoustic characteristics, segmental phonology, Kagayanen, Austronesian

DOI: https://doi.org/10.15026/0002000632

* I am grateful to Pastor Jehu Cayaon and other Kagayanen speakers, above all, those who provided the naturally occurring speech on which this study is based. Preliminary results from this study were first presented at an online meeting of the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023) in 2022. Many thanks for the discussion and comments received there. My thanks also go to Izumi Ochiai, Motoyasu Nojima, Sho Yamaoka, and Syuntaro Tida, as well as an anonymous reviewer, for commenting on earlier versions of the paper. This work was supported by JSPS KAKENHI Grant Numbers 19H01264 and 20K13042.

1. Introduction

Several Austronesian and Western Australian languages have a cross-linguistically rare speech sound produced with tongue protrusion from the mouth. Based on their elicited data from Philippine languages, Olsen *et al.* 2010 make a precise characterization of the articulation of the sound and argue that it is best described as an interdental approximant. Although brief descriptions of this speech sound are found in a number of descriptive work (e.g., Gieser 1958, McFarland 1974, Wiens 1976, Harmon 1977, Wordick 1982, Dench 1995, Blevins 2001), more work is necessary to document the sound. The purpose of this paper is to provide a more detailed description of the acoustics of the interdental approximant in Kagayanen ([kagajanɨn]; ISO 639-3: cgc). Kagayanen is an Austronesian language of the Philippines spoken on Cagayancillo in the Sulu Sea and coastal communities distributed over the Palawan Island.

The sound that can be described as an interdental approximant is attested at least in four varieties of Kalinga (Butbut Kalinga [kyb], Limos Kalinga [kmk], Lower Tanudan Kalinga [kml], and Lubuagan Kalinga [knb]), three Central Philippine languages (Kalagan [kqe], Mandaya [mry], and Southern Catanduanes Bikol [bln]), one Manobo language (Kagayanen [cgc]), and six Australian languages (Bunuba [bck], Kurrama [vkn], Martuthunira (extinct) [vma], Nhanda [nha], Unggumi [unp], and Yindjibarndi [yij]) (Olson *et al.* 2010 and references cited therein). In the literature of Australian languages, the sound is usually labeled as a 'lamino-dental glide' and transcribed with < yh > (Wordick 1982: 12, Blevins 2001: 10).[1] In the Philippines, on the other hand, the sound is described with different labels, such as a 'central resonant oral' (Gieser 1958: 17), a 'palatal lateral' (Wiens 1976: 41), and an 'L-colored glide' (Harmon 1977: 17).

Following Olson *et al.* 2010 and Mielke *et al.* 2011, I transcribe interdental approximants by using < ð̞ >, constituted by the Latin small letter 'eth' < ð > to represent an interdental articulation and the lowering sign diacritic indicating that the sound is lowered relative to the articulation expressed by the base character.

This paper describes the acoustics of interdental approximants that appear in naturally occurring speech produced by Kagayanen speakers. Furthermore, this study investigates the behavior of the sound within the phonological system of Kagayanen, observing its phonotactics and its involvement in a phonological process.

2. Kagayanen

2.1 The language and its speakers

Kagayanen is classified as a member of the Manobo subgroup within the Austronesian

---

[1] The reason why labeling is made this way and consistent among researchers may be attributed to the convention in Australian linguistics, where the place of articulation for consonants is described by the combination of an active and passive articulator (e.g., Evans 1995, Dixon 2010, Aikhenvald 2015).

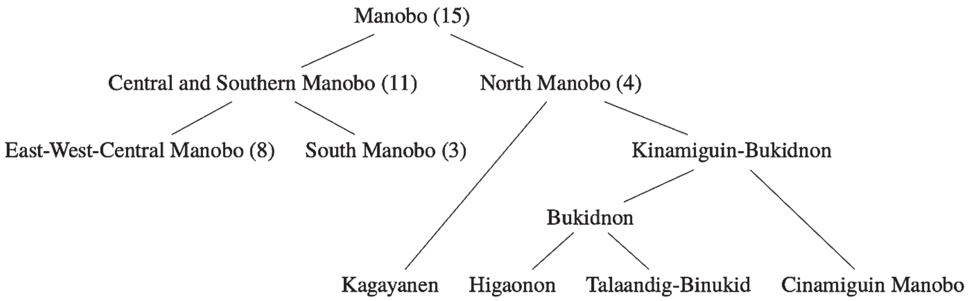


Figure 1 Subgrouping of Manobo languages based on Hammarström *et al.* (2022). The numbers in brackets indicate the number of languages that belong to each subgroup.

language family (Harmon 1977, Hammarström *et al.* 2022). The subgrouping of Manobo languages based on Hammarström *et al.* 2022 is given in Figure 1. Kagayanen is the only language of this subgroup that is not spoken on Mindanao or its neighboring islands. The language serves as the primary language of the people who live on Cagayancillo in the Sulu Sea, and it is also spoken in coastal communities of Palawan due to migration from Cagayancillo. There are about 30,000 speakers of the language (Eberhard *et al.* 2023). The speakers are usually multilingual, often conversant in Tagalog, English, and/or Hiligaynon. Additionally, Kagayanen speakers in Palawan may also speak Cuyonon, which had been the lingua franca of the region until recently.

2.2 The interdental approximant

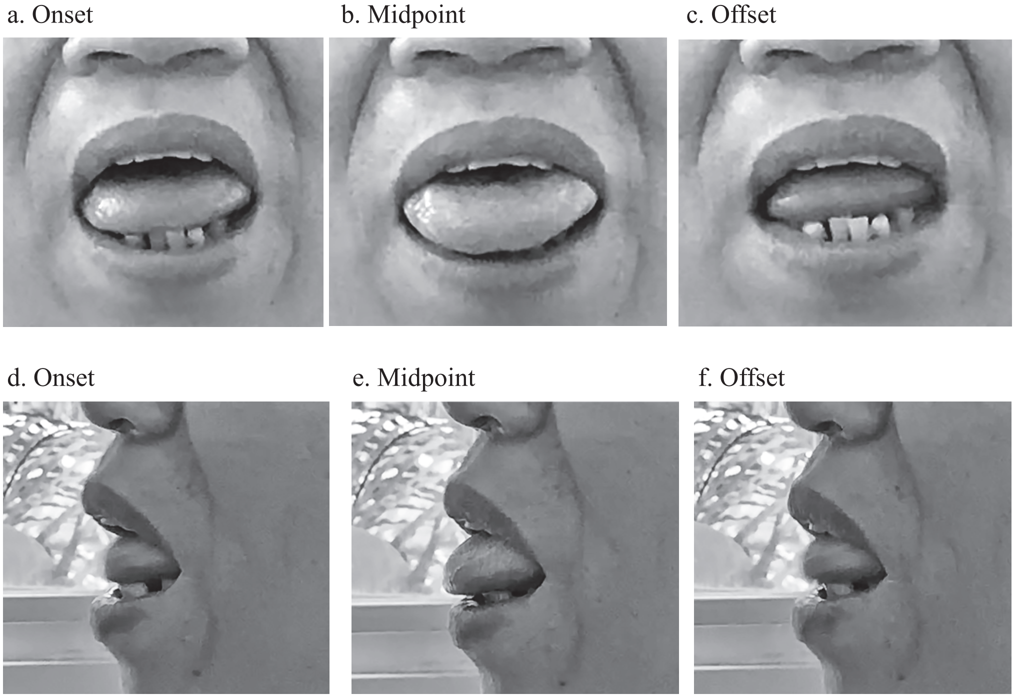
Olson *et al.* 2010 provide a description of phonetic characteristics of interdental approximants in terms of articulatory parameters (p. 202). Their observation is summarized as follows.

(1) a. It is always voiced.
 b. It is oral and does not involve velic lowering.
 c. The lips are neither rounded nor spread.
 d. It is produced with an egressive pulmonic air.
 e. The manner of articulation is an approximant. The tongue blade approaches but does not touch the upper teeth.
 d. The place of articulation is dental or interdental. It depends on the degree of tongue protrusion.[2]

This description is compatible with my observation regarding the articulation of the Kagayanen [ð̞]. The frames extracted from a video recording in Figure 2 show steps in the articulation of an interdental approximant. The video recording was made while the

---

[2] Mielke *et al.* (2011) report that the degree of tongue protrusion is affected by vowel context. More protrusion is observed when the interdental approximant appears with the low vowel [a] than with either of the two high vowels [u] and [ɨ].



a. Onset b. Midpoint c. Offset

d. Onset e. Midpoint f. Offset

Figure 2 Articulation of the interdental approximant in the word [dað̞an] 'street'.

word [dað̞an] 'street' was being produced. Frames (a) and (d) show that the onset of the articulation where the tongue blade moves forward, while frames (b) and (e) shows the maximum advance position of the tongue during the articulation of interdental approximants. Frames (c) and (f) show the tongue returning to the position for the next vowel [a]. Another observation that can be made from the figure is that the tongue blade only approaches the upper teeth, never touching them.

As already mentioned above, Harmon 1977: 17 uses the label 'L-colored lateral' for the interdental approximant in Kagayanen, suggesting that the sound has a lateral perceptual quality and [ð̞] is perceptually similar to the Kagayanen alveolar lateral approximant [l]. Perhaps for this reason, previous studies of the language usually transcribe both /l/ and /ð̞/ with < l > (e.g., Huggins 1989, Pebley 1999a, 1999b, MacGregor and Pebley 1999). The intuition that the Kagayanen [ð̞] sound involves laterality is subsequently supported by an ultrasound study (Mielke *et al.* 2011). Based on ultrasound data from one Kagayanen speaker, the study reveals that [ð̞] shows the same pattern with [l], in that during their articulation, one or both sides of the tongue are lower than during the articulation of coronal stops and fricatives, indicating that they are articulatorily lateral (2011: 409). Building on these observations, I compare the acoustic data of [ð̞] with that of [l] in the next section.

Olson *et al.* 2010 also provide a brief description of acoustic aspects of the interdental approximant. Comparing a waveform and wide-band spectrogram of [sala] 'living room' and [pað̞ad] 'palm (of hand)', they report that F1 of [ð̞] is slightly higher than F1 of [l], while major differences are not found in their F2.

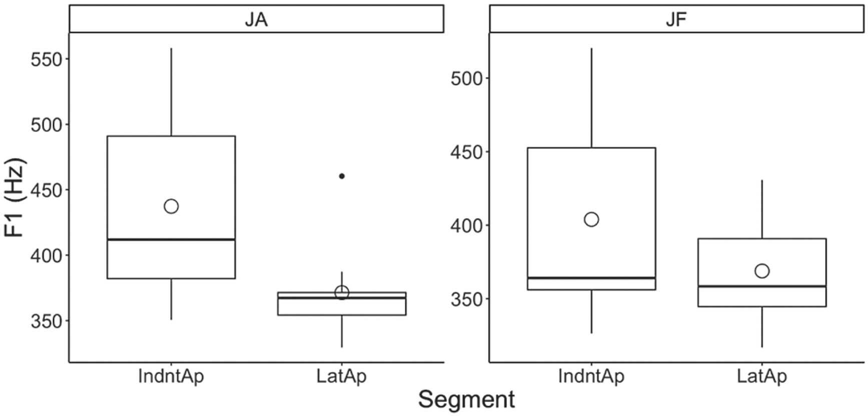


Figure 3 A boxplot representation of the F1 values (midpoint) for [ð̞] (IndntAp) and [l] (LatAp). The white circles represent the means in each condition. The black dot indicates the outlier.

3. Acoustics of the interdental approximant

3.1 Data

The data reported in this section is based on the fieldwork in Roxas, the Province of Palawan, Philippines, which was conducted in August of 2023. The data are naturally occurring speech produced by two male native speakers of Kagayanen. At the time of the recording, the first speaker (referred to as JA hereafter) was 62 years old, while the second speaker (referred to as JF hereafter) was 70 years old. Both were raised in Roxas and spoke Tagalog in addition to Kagayanen. Within each recording session, each speaker was asked to talk freely in Kagayanen about his daily life, work, childhood, and the like. In the narrative by JA, [l] and [ð̞] occurred 10 and 8 times, respectively, while in the narrative by JF, [l] and [ð̞] occurred 10 and 5 times, respectively. The list of the observed words including [l] or [ð̞] for each speaker is given in Appendix A.

All recordings were made using a Zoom H5 digital recorder and an AudioTechnica AT831b lavalier microphone, with a 44100 Hz sampling rate. Acoustic analysis of the data was performed using Praat (Boersma & Weenink 2022). The following section documents the first three formants extracted at the temporal midpoint of each [l] and [ð̞].

3.2 Formants

Figure 3 is a boxplot which shows the F1 values for the interdental approximant (IndntAp) and the alveolar lateral approximant (LatAp) with respect to each speaker. I observe that interdental approximants produced by JA have slightly higher F1 than alveolar lateral approximants. As for JF, in contrast, the F1 values for each speech sound show no significant differences, although F1 of [ð̞] extends to higher frequencies than F1 of [l]. Given this variability, it seems that F1 alone is not a reliable acoustic cue for the interdental approximant, unlike the description in Olson *et al.* 2010: 204–205.

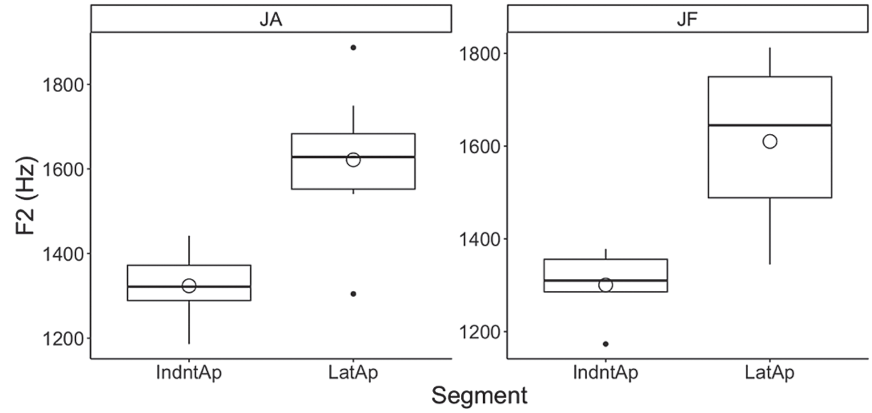


Figure 4 A boxplot representation of the F2 values (midpoint) for [ð̞] (IndntAp) and [l] (LatAp). The white circles represent the means in each condition. The black dots indicate the outliers.

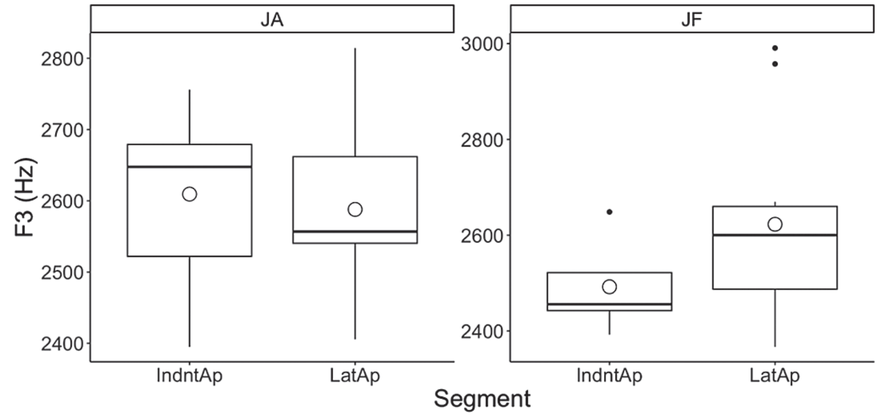


Figure 5 A boxplot representation of the F3 values (midpoint) for [ð̞] (IndntAp) and [l] (LatAp). The white circles represent the means in each condition. The black dots indicate the outliers.

Figure 4 shows the F2 values for the interdental approximant and alveolar lateral approximant. I observe that the interdental approximant entirely has lower F2 compared to the alveolar lateral approximant, and the difference in their F2 values is consistent between the two speakers.

Figure 5 is a boxplot which shows the F3 values for [ð̞] and [l]. For each speaker, the two sounds show no significant differences for their F3 value.

The data presented in this section suggests that the interdental approximant has a compact spectral peak (high F1 and low F2). This is confirmed by Figure 6, which indicates that F2-F1 values are smaller for the interdental approximant than for the alveolar lateral approximant.

3.3 Spectral characteristics

Figures 7 and 8 show the overlaid spectra of the interdental approximant and alveolar lateral approximant with respect to each Kagayanen speaker. These were made in the following procedure. First, I extracted a spectral slice at 5 ms intervals within the central

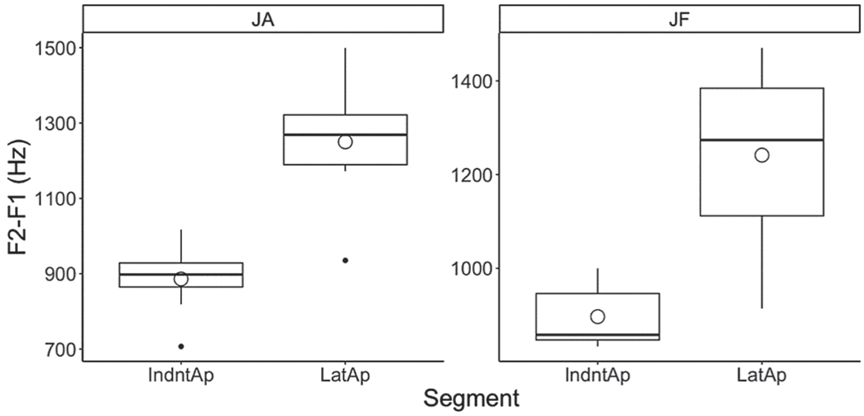


Figure 6 A boxplot representation of the F2-F1 values for [ð̞] (IndntAp) and [l] (LatAp). The white circles represent the means in each condition. The black dots indicate the outliers.

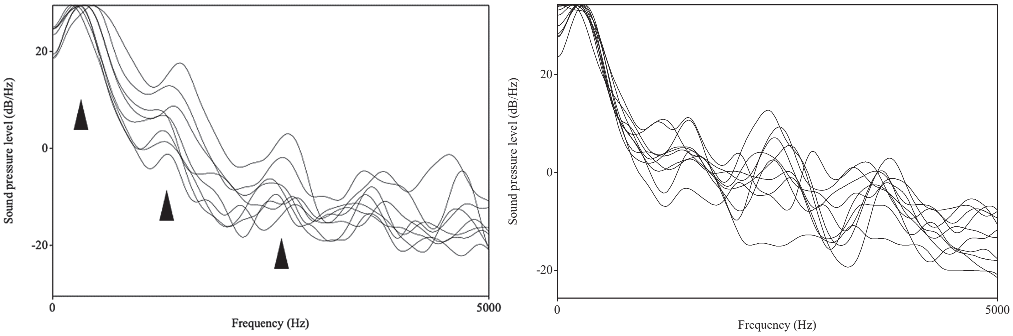


Figure 7 Cepstral spectra of [ð̞] (left) and [l] produced by JA

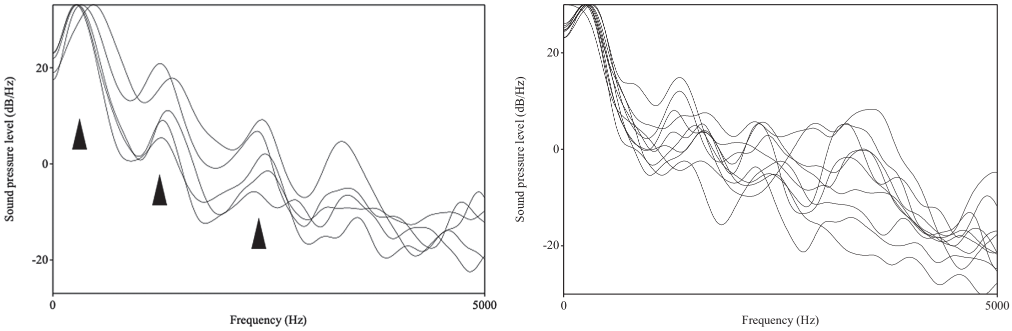


Figure 8 Cepstral spectra of [ð̞] (left) and [l] produced by JF

region for each [ð̞] and [l].[3] Second, I smoothed those spectral slices using Cepstral Smoothing with a 500 Hz bandwidth. Then, the resulting spectra obtained from the same speaker were overlaid.

Although no clear differences can be stated between the two sounds, the interdental approximant generally has multiple spectral peaks in approximately the same position within the low frequency band, as indicated by the black triangles.

---

[3] If the duration was less than 5 ms, the spectral slice was taken at the entire interval of the segment.

Table 1 Contrast with other coronal consonants

Segments	Example	
/ð̞/ vs. /l/	buð̞a 'to bubble'	bula 'to lie'
	kað̞a 'to know (person)'	alaɡa 'care'
/ð̞/ vs. /r/	uð̞a 'not'	sura 'cheek'
/ð̞/ vs. /n/	dað̞an 'street'	danɨn 'they'
/ð̞/ vs. /d/	ð̞að̞a 'to weave'	dað̞a 'to send'
/ð̞/ vs. /t/	aɡað̞ 'to cry'	kaɡat 'to bite'

Table 2 Kagayanen phoneme inventory

p b		t d		k ɡ	ʔ			
		s			h	i	ɨ	u
m		n		ŋ			a	
		r						
		l						
w	ð̞		j					

4. Phonological Characteristics

4.1 Phonemic split

In Kagayanen, the interdental approximant serves as a phoneme and contrasts with other coronal sounds, as shown in Table 1. The phoneme inventory of the languages is given in Table 2 (Harmon 1977: Ch. 2, Olson *et al.* 2010).[4] It should be noted that the coronal plosives /t/ and /d/ are articulated with the tongue against the upper teeth rather than the alveolar ridge.

While the interdental approximant and alveolar lateral approximant are distinctive sounds in the current Kagayanen segmental phonology, historical-comparative evidence suggests that this is the result of a split of [ð̞] from /l/ in the language. The forms shown in Table 3 are cognates found in Kagayanen and Binukid (ISO 639-3: BKD), and the forms of proto-Malayo-Polynesian (PMP) reconstructed by Blust (n.d.) for each cognate set. Kagayanen and Binukid belong to the North Manobo subgroup of the Malayo-Polynesian family (see also Figure 1 in Section 2). The Binukid data comes from *Binukid Dictionary* compiled by Otanes & Wrigglesworth 1992. Table 3 shows that **l* in PMP generally turned into [ð̞] in Kagayanen when it occurs in word-final and intervocalic position (not adjacent to /i/).[5] It is also shown that **l* is realized as [l] in Kagayanen when it follows or precedes /i/, it occurs word-initially, or it is geminated.

---

[4] To be more precise, Harmon's (1977) phoneme inventory differs from Olson *et al.*'s (2010) in that she does not recognize the interdental approximant as a distinct phoneme but considers it to be an allophone of /l/. She argues that [l] and [ð̞] show a complementary distribution in her data.

[5] There are regular exceptions for these generalizations. For example, loan words retain [l] in the positions where [ð̞] is expected. One finds [alas] 'at' (loan from Spanish) rather than *[að̞as].

Table 3 North Manobo cognate sets related to the Kagayanen /ð̞/ and /l/

Gloss	Kagayanen	Binukid	Proto-Malayo-Polynesian
‘palm’	pað̞ad	palad	*palaj
‘widow’	bað̞u	balu	*balu
‘head’	uð̞u	ulu	*qulu
‘deep’	dað̞ɨm	ma-daləm	*daləm
‘medicine’	buð̞uŋ	buluŋ	*buluŋ
‘deaf’	bɨŋŋɨð̞	bəŋəl	*bəŋəl
‘ginger’	lujʔa	lujʔa	*laqia
‘neck’	liʔɨɡ	liʔəɡ	*liqeʀ
‘three’	tallu	tatulu	*tatəlu
‘morning’	sɨllɨm	sələm	*sələm
‘lightning’	kilat	kilat	*kilat
‘return home’	uliʔ	uliʔ	*uliq

4.2 Phonotactics

As discussed in Section 4.1, the interdental approximant is historically related to the alveolar lateral approximant in Kagayanen. The data allows us to address the question of whether or not the interdental approximant forms a natural class with the alveolar lateral approximant in the current Kagayanen sound system.

The phonotactics of Kagayanen indicates that the interdental approximant forms a natural class with the glides /w/ and /j/, rather than with /l/. The possible surface syllable types for Kagayanen are C(C)V(C) in word-initial position and (C)(C)V(C) in word-medal and word-final position. Furthermore, only [ð̞], [w], and [j] can occur as the second consonant of the CC onset cluster, as can be seen in (1) and (2). It should be noted that while all consonants except for [ʔ], [j], and [w] can appear in the first C position of Cw and Cj, only labial or dental/alveolar consonants can form complex onsets together with [ð̞]. This seems to suggest that there is a constraint governing the possible Cð̞ clusters, that is, the interdental approximant cannot cooccur with a sound that has a different tongue gesture specification (e.g., a sound articulated with the dorsum).

(1) Simplex onset
CV ʔa ‘1SG.ABS’ CVC din ‘3SG.ERG’
CV.VC wa.iɡ ‘water’ CV.CVC ju.pan ‘bird’
CVC.CVC wal.lɨŋ ‘face’

(2) Complex onset
CCV.CV bwa.ja ‘crocodile’ CCV.CVC pwi.kan ‘turtle’
CCV dja ‘DIS.ADV’ CV.CCV ki.kju ‘2PL.OBL’
CCV.CV mð̞a.ʔu ‘thirsty’ CCV.CVC dð̞a.ɡan ‘run’
CCV.CVC bð̞a.ŋaw ‘rainbow’

4.3 Phonological process

Further evidence for the groping of the interdental approximant with the glides as a natural class is the presence of a phonological process which treats them in the same way. Kagayanen has the verb prefix *m-*, which forms syntactically intransitive verbs, such as motion verbs (e.g., ‘go’ and ‘enter’), activity verbs (e.g., ‘dance’ and ‘eat’), and bodily process verbs (e.g., ‘sweat’ and ‘breathe’) (Pebley 1998).[6] When *m-* is attached, the root’s initial consonant is reduced, as in (3), but it is retained when the root begins with [j] or [ð̞], as in (4). Although examples in which *m-* is attached to roots that begin with [w] have not been attested, this phonological process corroborates the view that the interdental approximant forms a natural class with the glides but not with the lateral approximant /l/.

(3)	a. maʔan	b. mɨllɨd	c. mitɨm
	m-kaʔan	m-sɨllɨd	m-litɨm
	IRR.AV-eat	IRR.AV-enter	IRR.AV-hungry
	‘to eat’	‘to enter’	‘to become hungry’

(4)	a. mjapun	b. mð̞að̞a	c. mð̞aʔu
	m-japun	m-ð̞að̞a	m-ð̞aʔu
	IRR.AV-eat.supper	IRR.AV-weave	IRR.AV-thirsty
	‘to eat supper’	‘to weave’	‘to become thirsty’

5. Conclusion

I have described the acoustics of the interdental approximant in Kagayanen, based on the data from naturally occurring speech, and have discussed the phonological characteristics of this speech sound. The data has shown that the interdental approximant exhibits lower F2 values and has more compact spectral peaks when compared to the alveolar lateral approximant. In terms of phonology, historical-comparative evidence suggests that the interdental approximant has resulted from a phonemic split from /l/, but I have argued that it should be considered to form a natural class with approximants rather than /l/ in the current phonological system of Kagayanen.

Abbreviations

ABS = absolutive, ADV = adverbial, AV = actor voice, DIS = distal, ERG = ergative, IRR = irrealis, OBL = oblique, PL = plural, SG = singular, 1 = speaker, 2 = addressee, 3 = non-speech-act participant

---

[6] A syntactically intransitive verb is a verb that takes only one core argument. This verb class includes verbs that semantically require two arguments such as an actor and a patient argument.

Appendix A

List of the observed words including [j] or [ð̞] for each speaker.

Speaker	Word	Gloss	Token
JA	alas	‘at’ (used to indicate time, loan from Spanish)	1
	ɡalimpiu	‘to clean’	1
	lunis	‘Monday’ (loan from Spanish)	1
	malik	‘to go back’	3
	miliŋ	‘to go’	1
	mirkulis	‘Wednesday’ (loan from Spanish)	1
	muliʔ	‘to go home’	2
	bað̞aj	‘house’	4
	kuð̞aŋ	‘short’	1
	uð̞a	Negative polarity marker	3
JF	diliʔ	Negative polarity marker	1
	ɡalaktəd	‘to move’	1
	ɡaliŋ	‘to leave’	1
	ɡaliŋkud	‘to serve’	2
	ilintan	‘lord’	1
	iskwila	‘school’	1
	katuliku	‘catholic’ (loan from Spanish’	1
	laiŋ	‘bad’	1
	lunis	‘Monday’ (loan from Spanish)	1
	bað̞aj	‘house’	1
	uð̞a	Negative polarity marker	4

References

Aikhenvald, Alexandra Y. 2015. *The Art of Grammar: A Practical Guide*. Oxford: Oxford University Press.

Blevins, Juliette. 2001. *Nhanda: An Aboriginal Language of Western Australia* (Oceanic Linguistics Special Publication 30). Honolulu, HI: University of Hawaii Press.

Blust, Robert A. N.d. *Austronesian Comparative Dictionary*. (Available online at https://www.trussel2.com/ACD/, Accessed on 2023-09-10.)

Boersma, Paul and David Weenink. 2022. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.2.10.

Dench, Alan. 1995. *Martuthunira: A Language of the Pilbara Region of Western Australia* (Pacific Linguistics C-125). Canberra: Australian National University.

Dixon, Robert M. W. 2011. *The Languages of Australia*. Cambridge and New York: Cambridge University Press.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World. Twenty-sixth edition*. Dallas, Texas: SIL International. (Available online at http://www.ethnologue.com, Accessed on 2023-09-11.)

Evans, Nicholas D. 1995. *A Grammar of Kayardild: With Historical-Comparative Notes on Tangkic*. Berlin, New York: Mouton de Gruyter.

Gieser, C. Richard. 1958. “The Phonemes of Kalinga.” *Studies in Philippine Linguistics by Members of the Summer Institute of Linguistics* (Oceania Linguistic Monographs 3) (Capell, Arthur & Wurm, Stephen A., eds.), 10–23, Sydney: University of Sydney.

Huggins, Jacqueline, Louise A. MacGregor, Scott W. MacGregor & Carol Jean Pebley. 1989. *Mga pangadlaw-adlaw na ambalen* [Everyday conversation]. Manila: Summer Institute of Linguistics.

Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2022. *Glottolog 4.7. Leipzig: Max Planck Institute for Evolutionary Anthropology.* https://doi.org/10.5281/zenodo.7398962 (Available online at http://glottolog.org, Accessed on 2023-09-01.)

Harmon, Carol W. 1977. Kagayanen and the Manobo Subgroup of Philippine Languages. Ph.D. dissertation, University of Hawaii at Manoa.

MacGregor, Louise A. and Carol Jean Pebley. 1999. “Two Kagayanen Texts.” *Studies in Philippine Languages and Cultures*,10: 91–115.

McFarland, Curtis Daniel. 1974. The Dialects of the Bikol Area. Ph.D. dissertation, Yale University.

Mielke, Jeff, Kenneth Olson, Adam Baker & Diana Archangeli. 2011. “Articulation of the Kagayanen Interdental Approximant: An Ultrasound Study.” *Journal of Phonetics*, 39(3): 403–412.

Olson, Kenneth, Jeff Mielke, Josephine Sanicas-Daguman, Carol Pebley, and Hugh Paterson. 2010. “The Phonetic Status of the (inter)dental Approximant.” *Journal of the International Phonetic Association*, 40(2): 199–215.

Otanes, Fe T. and Hazel Wrigglesworth. 1992. *Binukid Dictionary*. Manila: Linguistic Society of the Philippines.

Pebley, Carol. 1998. “Semantic Intransitivity, Causation and the Affix ‘m-’ in Kagayanen.” *Localist Case Grammar and Philippine Verbs* (Sherri Brainard, ed.), 44–53, Manila: Linguistic Society of the Philippines.

Pebley, Carol Jean. 1999a. “A Sketch of Kagayanen Clause Structures.” *Studies in Philippine Languages and Cultures*, 10(2): 1–20.

Pebley, Carol Jean. 1999b. “Kagayanen Enclitic Demonstratives ‘i’, ‘an’, and ’ya’.” *Studies in Philippine Languages and Cultures*, 10(2): 50–90.

Wiens, Hartmut. 1976. “Phonological Features of Limos Kalinga, with Comments on Affected Speech.” *Philippine Journal of Linguistics*, 7: 38–47.

Wordick, Frank J. F. 1982. *The Yindjibarndi Language* (Pacific Linguistics C-71). Canberra: Australian National University.

Journal of Asian and African Studies, Supplement, No. 4, 2024

A Preliminary Study of the Prosody of Zanzibar Swahili *tu**

Abe, Yuko
Lee, Seunghun J.
Kamano, Shigeto
Miyazaki, Kumiko

The exclusive particle *tu* ‘only’ in contemporary Zanzibar Swahili shows the distribution and prosodic patterns different from Ashton (1944) and Wilson (1985). This paper reports prosodic patterns of this exclusive particle *tu* in final and post-verbal positions produced in statements and questions and shows that *tu* is predominantly produced with the most prominent pitch of a sentence, contrary to Ashton (1944) who describes a Swahili variety with lowering of pitch on *tu*. In the absence of *tu*, statements and questions differ from one another in that questions end with a final rise. Distributionally, *tu* can appear in the sentence-final position or the post-verbal position, that is, the position immediately after the verb. While *tu* is realized with the highest pitch, the pitch excursion patterns surrounding *tu* in the rest of the sentences are the same.

Keywords: sentence prosody, Zanzibar Swahili, exclusive particle *tu*, statement, yes-no question

DOI: https://doi.org/10.15026/0002000633

* This research is written as part of the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021–2023). The research has been funded by a Kakenhi Grant (C) 20K00578. This work was also supported by JSPS KAKENHI Promotion of Joint International Research (Fostering Joint International Research (B)) Grant Number 21KK0005, Kakenhi (B) Grant 23H00622, as well as an undergraduate research grant from Japan ICU Foundation to Shigeto Kamano. The main idea of this paper is based on papers by Kamano (2023) and Lee *et al.* (2023) and revised by supplemented data. Any errors remain our own.

1. Introduction

Swahili is widely used in East Africa and parts of Central Africa as a lingua franca with a wide variety (Shinagawa and Nassenstein 2020: 4). Since 19th century, the Zanzibar variety with its economic power carried its dialects out of Zanzibar Island, to adjacent islands, along the adjacent coast, and as far as Congo (Nurse, Hinnebusch, and Philipson 1993: 31), then became the basis for setting Standard Swahili during the colonial times.

Compared to most other Bantu languages, Swahili does not have tone but has a penultimate stress on a word or on a phonological word (Downing 2004: 121). As for Swahili yes-no question intonations, a sentence-final HL tone melody is reported as a question marker, which is cross-linguistically common, especially in non-tone languages such as Swahili (Clements and Rialland 2008: 75–76). In this type of language, the intonation of a question takes the form of a sentence-final rise. Ashton (1944: 23a–23b) and Polomé (1967: 55), however, explain more detailed intonation patterns in Swahili statements and questions, in that a statement as in (1a) has the pitch fall from the penultimate syllable, while a question as in (1b), has the penultimate syllable with a stressed mid (or slightly rising) [–], followed by a long stressed final syllable which generally has a high falling tone [\].

(1) Intonation patterns of a statement and a question (Ashton 1944: 23b)
a. Statement
Imekwisha. 'They are finished'

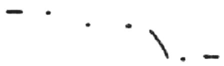
b. Question
Imekwisha? 'Are they finished?'

The exclusive particle *tu* in Standard Swahili is considered to be one of the few stems that originate from adverbial Bantu roots (Ashton 1944: 158). The reconstructed form of *tu* is **túpú* 'only, in vain' in Proto-Bantu (Meeussen 1967: 115). In Ashton (1944), the distribution of *tu* is described to be restrictive as *tu* appears only in the sentence-final position as in (2). Wilson (1985: 72) further notes that "*tu* must always come at the end of its phrase".

(2) Exclusive particle *tu* in a conversation (Ashton 1944: 158)
A: *Wataka viti vingapi?* 'How many chairs do you want?'
B: *Nataka viwili <u>tu</u>.* 'I want <u>only</u> two (chairs).'

Ashton (1944: 75) describes the tone of the exclusive particle *tu* as low when it is not emphasized, in a sentence such as (3), *tu* is realized with a regular falling intonation.

(3) Low realization of *tu* in the falling intonation pattern (Ashton 1944: 75)
Watu wamelala <u>tu</u>. 'The people are <u>just</u> sleeping.'



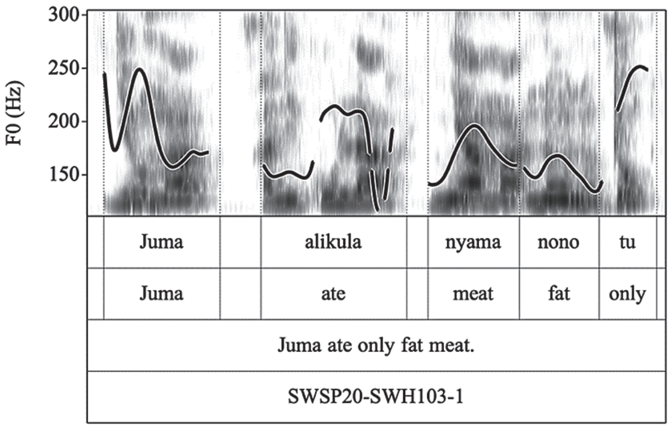


Figure 1 Prominent high pitch of *tu* in Zanzibar Swahili (our own recordings)

Contrary to Ashton's descriptions of the low tone associated with the particle *tu*, instructions in Swahili classes tend to ask learners to raise *tu* extremely prominent as shown in Figure 1, where the particle is pronounced with an extremely high pitch[1]. Note that the pitch of *tu* can be even higher than the preceding elements in the sentence, suggesting that descriptions about the prosodic status of *tu* need revisiting.

This paper focuses on three aspects of the Zanzibar variety of Swahili. First, we explain the statement intonation in Zanzibar Swahili. Second, we report the prosodic pattern of yes-no questions and the prosody with an exclusive particle *tu* 'only, just'. Our data show a wider distribution of *tu* in a sentence; a departure from previous studies. *Tu* may appear in two positions in a typical SVO sentence; in the final position of a sentence (SVO *tu*), or in the post-verbal position before the object of the sentence (SV *tu* O). As far as we know, the position of *tu* belonging to the latter type has not been reported nor explained in previous studies. The third aspect of *tu* focuses on how intonation patterns vary depending on the position of *tu* in a sentence: post-verbal or sentence-final position.

The rest of the paper is organized as follows. Data collection methods and stimuli, as well as analytical methods, are described in Section 2. Results of various data points are explained in Section 3, followed by discussions in Section 4.

2. Methods

2.1 Stimuli

Three conditions were included in the stimuli: sentences without *tu* (No-*tu* condition), sentences with the post-verbal *tu* (PV-*tu* condition), and sentences with a final *tu* (Final-*tu* condition). In the No-*tu* condition, a sentence has a subject, a verb and a noun with a

---

[1] Contrary to (3) in which *tu* modifies the preceding verb *wamelala* 'they are sleeping', while *tu* in the sentence Figure 1 modifies the object *nyama nono* 'fat meat' (but not anything else). That suggests two different interpretations of *tu* an exclusive focus marker as in (3) and a contrastive focus marker as in Figure 1.

modifier as in (4a). Sentences in the PV-*tu* condition have the exclusive particle between the verb and the object as in (4b), and sentences with the Final-*tu* condition have the particle at the end of the sentence. The prosodic structure of the stimuli was matched as all of them were composed of four phonological words[2)] before adding the particle *tu*.

(4) Example question sentences derived from *Juma alikula nyama nono*. 'Juma ate fat meat.'

a. No-*tu* sentence
Juma a-li-kula nyama nono. / ?
Juma 3SG-PST-eat meat fat
'Juma ate fat meat. / Did Juma eat fat meat?'

b. PV-*tu* sentence
Juma alikula ***tu*** *nyama nono. / ?*
'Juma just ate fat meat. / Did Juma just eat fat meat?'

c. Final-*tu* sentence
Juma alikula nyama nono ***tu****. / ?*
'Juma ate only fat meat. / Did Juma eat only fat meat?'

Each stimulus contains five types of object DP consisting of a noun and an adjectival modifier, which differ from one another in the vowel quality of the penultimate syllable in the sentence-final word. The main verbs preceding each object DP are selected for semantic saliency. In (5), the rest of the four stimuli object DPs are shown. Each base sentence is modified to create sentence types in three conditions: No-*tu*, PV-*tu*, and Final-*tu*. Each condition has five different items yielding 15 sentences that were read twice in a statement and a question. The resulting tokens were 60 per participant.

(5) Stimuli sentences

a. *Juma a-li-wa-fuata watoto wanane.*
Juma 3SG-PST-3PL-follow children eight 'Juma followed eight children.'
b. *Juma a-na-soma vitabu vinene.*
Juma 3SG-PRS-read book thick 'Juma reads thick books.'
c. *Juma a-na-penda shepu ya kifuniko.*
Jume 3SG-PRS-like shape of lid 'Juma likes the shape of the lid.'
d. *Juma a-li-tafuta kiwango cha nuru.*
Juma 3SG-PST-search level of light 'Juma checked the level of light.'

---

2) The stimuli sentences (4) and (5) are composed of five words, but *ya* of *ya kifuniko* 'of the lid' in (5) and *cha* of *cha nuru* 'of light' in (5) are not stressed and do not affect the prosodic patterns under investigation. We can regard the phrases *ya kifuniko* 'of lid' or *cha nuru* 'of light' as one phonological unit.

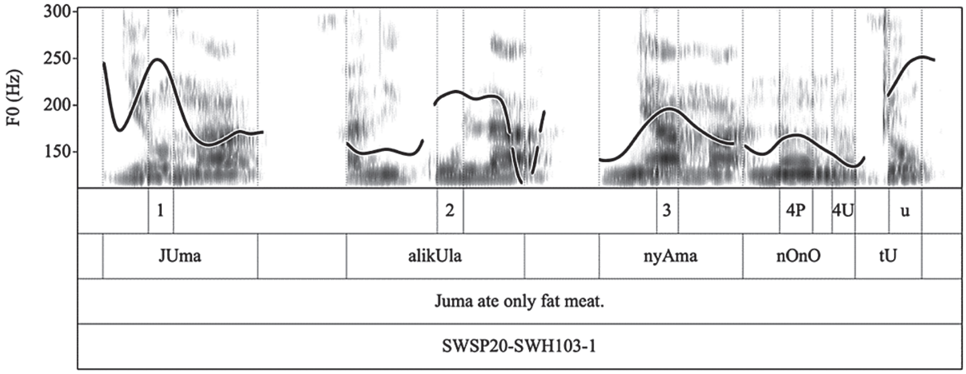


Figure 2 Example of the annotations of a sentence indicating stressed vowels (= Figure 1)

2.2 Participants and Recordings

Six native speakers of Zanzibar Swahili (three males and three females, aged 27 to 49) were asked to read the stimuli sentences. The stimuli sentences were presented one by one by the 3rd author. All recordings were conducted in a quiet room September 2022 using Audacity on a MacBook computer set at 44.1 kHz sampling rate and 16-bit depth.

2.3 Annotation

All the recordings are processed and annotated by using Praat (Boersma and Weenink 2023). As in Figure 2, we annotate the penultimate vowel of each word where Swahili usually has a high pitch: for example, “1” is the penultimate of the first word, and “2” indicates the penultimate vowel of the second word. The final word is further segmented with the ultimate vowel (4U) and the penultimate vowel (4P). Also, the high vowel of the exclusive particle *tu* in the post-verbal position or sentence-final position is marked with “u”.

Using the Praat algorithm, we extracted the maximum F0 from each annotated interval and the duration of the annotated points in the final word. These acoustic data are further processed by R (R Core Team 2023).

3. Results

3.1 Intonation of Statements in Zanzibar Swahili

Ashton (1944: 11, 16), while describing the sentence intonation, states that “the trend of the sentence is downwards”, and “words at the beginning or middle of a sentence have level tone for both penultimate and final syllables”3), but the last two syllables in the last word are of “Stressed + Low unstressed [ˋ.]” pattern as in (6). Polomé (1967: 52) describes the final two syllables in the final position as “high + low” or “high-falling + low”.

---

3) Underlines are added by the authors.

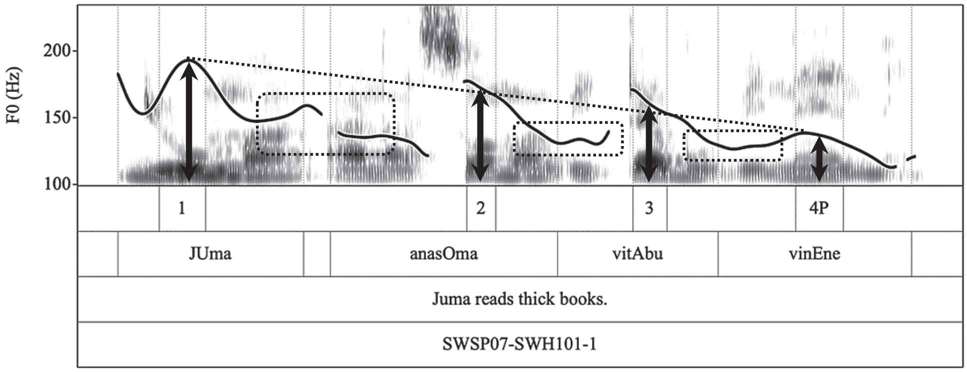


Figure 3 The downstep of a sentence (SWH101) (c.f., Kamano 2023)

(6) Ashton's description of statement intonation (Ashton 1944: 16)

Kisu kimeanguka. 'A knife has fallen.'

In our data, we also observe that this "trend of the sentence is downwards", i.e. the downstep/downdrift[4] in a sentence. The falling from the penultimate syllable to the final syllable is observed in all four locations of the word prominence. The pitch then plateaus into the following word until its penultimate syllable, as shown in Figure 3. The plateauing is marked by dotted boxes, and the downstep of the pitch of each penultimate syllable is marked with a solid line connecting the prominent penultimate (adopted from Kamano 2023: 98).

3.2 Intonation of Yes-No Questions of Zanzibar Swahili

This section comes from Lee *et al.* (2023: 1490–1492). Most tokens produced by four speakers out of six (SWH102, 103, 104, 106) have a higher pitch of the final word in questions, although the two remaining speakers did not display any pitch differences in both statements and questions.

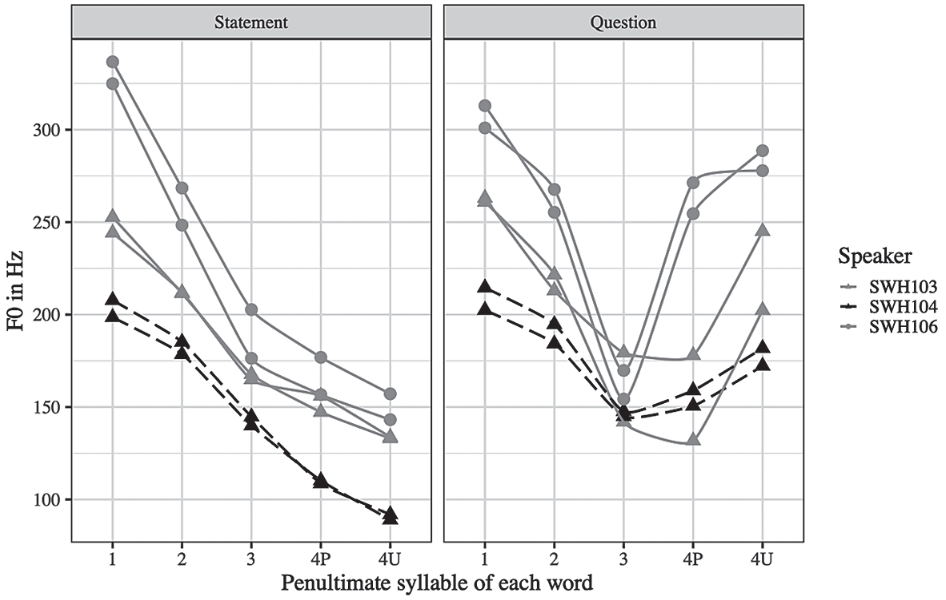
Previous studies (Ashton 1944: 23, Polomé 1967: 55) illustrate that in the yes-no questions, the penultimate syllable has a stressed mid (or slightly rising) [–], followed by a long stressed final syllable which generally has a high falling tone [╲] in the sentence-final word as was illustrated in (7).

---

[4] One thing to note is that downstep here is defined as "the lowering of a high tone in certain specifiable circumstances" (Connell 2001). It would be different from downstep in Bantu tonal languages because it is said to be influenced by a low tone or floating low tone (Maddieson and Sands 2019: 115). Since Swahili is a stress language, the pitch-lowering pattern in Swahili might be just a phonetic effect of natural declination.

(7) Intonation patterns of a question (Ashton 1944: 23b) = (1)
Imekwisha? ‘Are they finished?’

The last two syllables of a sentence in questions are expected to have a “mid + high-falling” tone [- ˋ], while in statements they have “stressed (high or high-falling) + low unstressed [ˋ.]”. In Figure 4, we examine the pattern of the penultimate vowel of the fourth word (4P) and the ultimate vowel of the fourth word (4U) in statements of the sentence *Juma alikula nyama nono*. ‘Juma ate fat meat’5). The result of statement prosody matches the expected “stressed (high or high-falling) + low unstressed [ˋ.]” pattern described in previous studies. In the question prosody, however, we observe that the pitch pattern of the final word shows variability. Although we expect the prosody to exhibit a high falling tone on the final word, individual tokens also show a higher pitch on the ultimate vowel (4U) than on the penultimate vowel (4P) in Figure 4. In particular, when asking the question *Juma alikula nyama nono?* ‘Did Juma eat fat meat?’, three speakers (SWH103, 104, 106) have a higher pitch on the final vowel (4U) than on the penultimate vowel of that word (4P).



* 4P = penultimate syllable of the fourth word, 4U = ultimate syllable of the fourth word

Figure 4 F0 contrast of 4P and 4U in statements and questions for three speakers

In (8), two speakers (SWH103 and SWH102) are zoomed in to compare the two types produced by the participants. Both speakers demonstrate the expected downstep

---

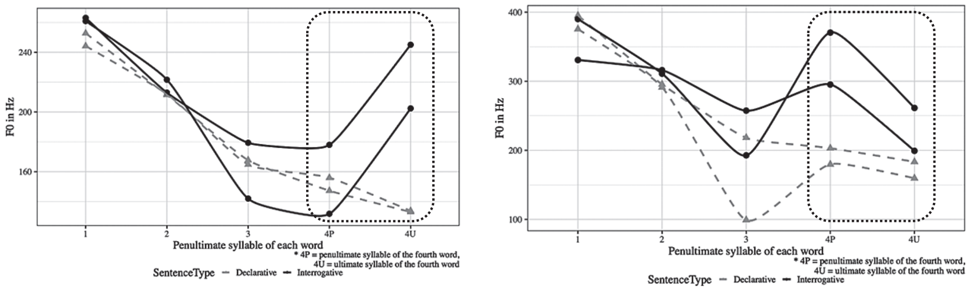
5) Target vowels (the penultimate vowel of each word and the ultimate vowel of the final word) are underlined.

prosody in statements, whereas, in questions, the pitch excursion of 4P and 4U displays discrepancies. In the (8a) type, 4U has a higher pitch than 4P, but the pattern is reversed in the (8b) type. At least in this preliminary dataset, the pattern in (8a) is dominant compared to the (8b), whereas previous studies report “mid + high-falling” [⁻ ˋ] prosody in questions, i.e. Swahili uses a sentence-final high-falling as a question marker (Clements and Rialland 2008: 75–76).

(8) Pitch excursion in statements and questions (Lee *et al.* 2023: 1490–1491)

a. Higher pitch on the ultimate syllable
Juma alikula nyama nono?
‘Did Juma eat fat meat?’ (SWH103)

b. Higher pitch on the penultimate syllable
Juma alikula nyama nono?
‘Did Juma eat fat meat?’ (SWH102)



3.3 Intonations of Statements and Yes-No Questions with an Exclusive Particle *tu*

The exclusive particle *tu* ‘only, just’ appears in two possible positions in a typical SVO sentence. One type is called the ‘PV-*tu* sentence’ in the S + V + *tu* + O structure, where the *tu* occurs in the post-verbal position, that is, the position immediately after the verb. In the second type, the ‘Final-*tu* sentence’ in the S + V + O + *tu* structure, the particle appears at the end of a sentence. The particle *tu* bears an inherent high pitch, therefore, in this subsection, we report how the high pitch of *tu* is realized in the prosody of the PV-*tu* and Final-*tu* sentences both in statements and questions.

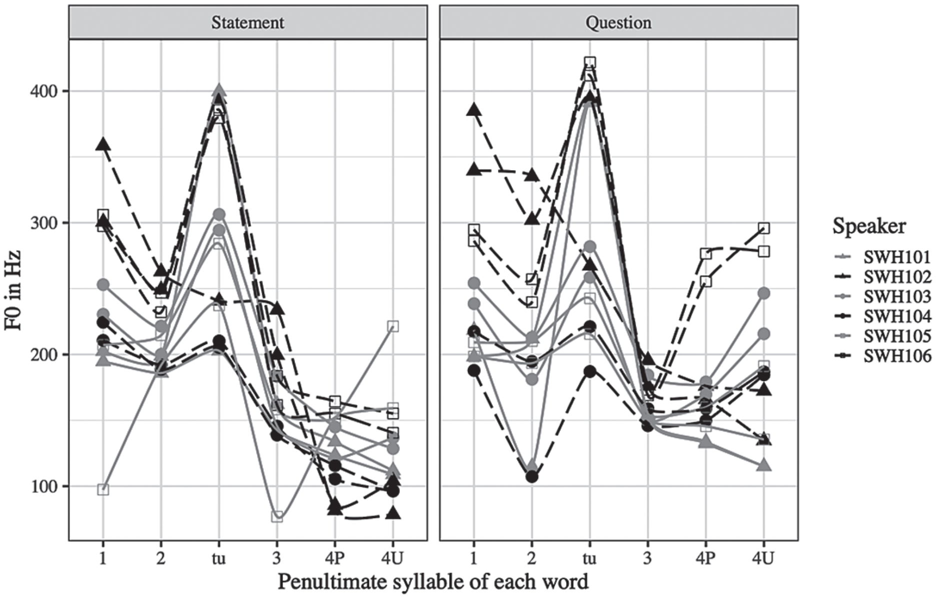
3.3.1 PV-*tu* Sentences in Statements and Questions

Recall that all the stimuli sentences have a subject and a verb followed by a noun with a modifier. The particle *tu* in PV-*tu* sentences between the verb *alikula* ‘ate’ and the noun *nyama* ‘meat’ in (9); *Juma alikula tu nyama nono.* ‘Juma just ate fat meat.’ for the statement and *Juma alikula tu nyama nono?* ‘Did Juma just eat fat meat?’ for the question. The most prominent syllable in each utterance is the particle *tu* regardless of the sentence type.

(9) Penultimate syllables in the PV-*tu* sentences in statements and questions

a. *Juma alikula tu nyama nono.* b. *Juma alikula tu nyama nono?*

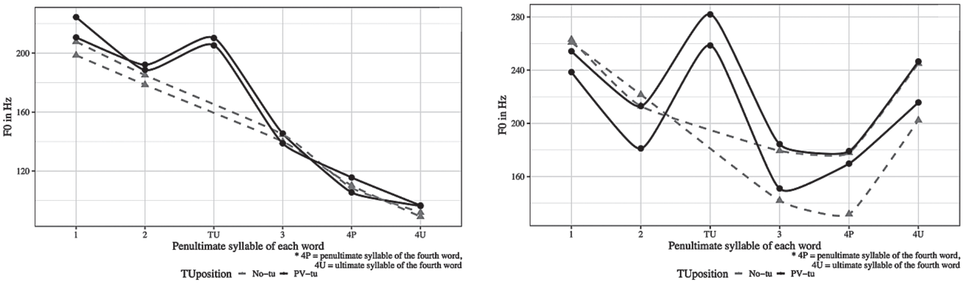
'Juma just ate fat meat.' 'Did Juma just eat fat meat?'



* 4P = penultimate syllable of the fourth word, 4U = ultimate syllable of the fourth word

Except for the pitch boost shown in *tu* in the prosody in (9), we observe speakers show similar pitch excursion in No-*tu* sentences and PV-*tu* sentences, regardless of the sentence types (also see, Kamano 2023: 97, Lee *et al.* 2023: 1491). In examples (10a) and (10b), the pitch excursion of sentences with *tu* (PV-*tu*) or without *tu* (No-*tu*) is plotted together from speakers SWH104 and 103. The high pitch of the particle *tu* is observed regardless of the sentence type. This pattern shows that the presence of *tu* prominence does not affect the intonation of the whole sentence.

(10) Individual pitch transition for No-*tu* and PV-*tu* sentences (Kamano 2023: 97, Lee *et al.* 2023: 1491)



a. Statements (c.f., Kamano 2023)
Ju̲ma aliku̲la (tu̲) nya̲ma no̲no̲
'Juma just ate fat meat.' (SWH104)

b. Questions (c.f., Lee *et al.* 2023)
Ju̲ma aliku̲la (tu̲) nya̲ma no̲no̲?
'Did Juma just eat fat meat?' (SWH103)

3.3.2 Final-*tu* in Statements and Questions

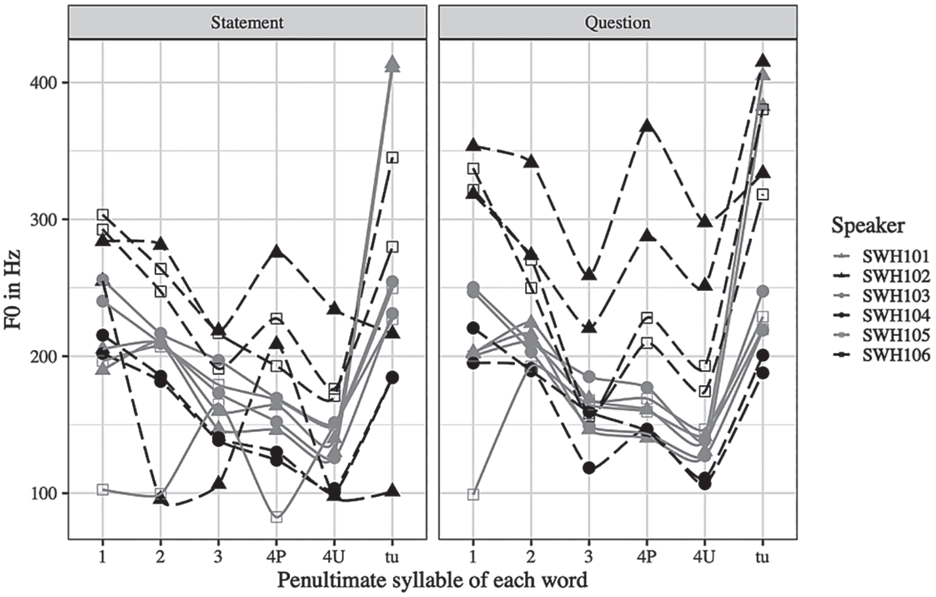
Final-*tu* sentences also display a prominent high pitch on the particle *tu*. In both the statement *Juma alikula nyama nono tu*. 'Juma ate only fat meat.' and the question *Juma alikula nyama nono tu?* 'Did Juma eat only fat meat?', the highest pitch is on the particle. The high pitch on *tu* creates a conflict with the sentence-level prosody because Swahili questions are suggested to be marked by a sentence-final high-falling (Clements and Rialland 2008: 75–76). When in conflict with the sentence-level pitch rise, the pitch boost of sentence-final *tu* would create ambiguity when Swahili speakers produce or perceive the two sentence types (statements and questions).

In (11), questions have overall higher F0 values by the speaker SWH102, but the pitch excursion and pitch height are comparable between statements and questions in the rest of the speakers.

(11) Penultimate syllables in the Final-*tu* sentences in statements and questions

a. *Juma alikula nyama nono tu* ‘Juma ate fat meat only.’

b. *Juma alikula nyama nono tu?* ‘Did Juma eat fat meat only?’



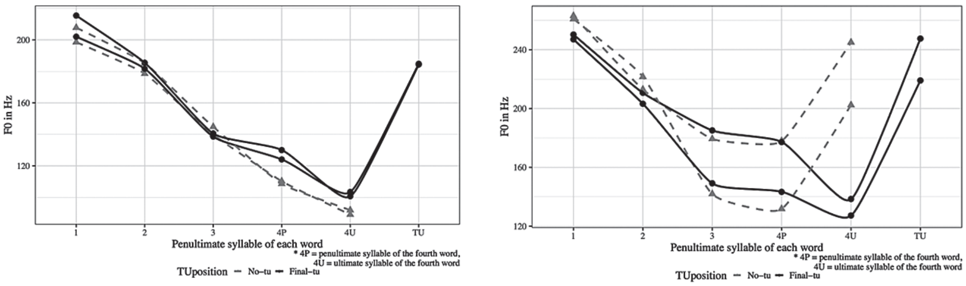
* 4P = penultimate syllable of the fourth word, 4U = ultimate syllable of the fourth word

In (12), plots of pitch excursion of sentences with and without a final *tu* produced by two speakers (SWH103 and SWH104) are shown. The pitch excursion before the *tu* is neutralized and the high-falling pitch suggested in Clements and Rialland (2008: 75–76) is not present in the questions (Kamano 2023: 97, Lee *et al.* 2023: 1491). The sentence prosody behaves independent from the pitch boost in *tu* as it was the case in the PV-*tu* sentences.

(12) Individual pitch transition for No-*tu* and Final-*tu* sentences

a. Statements (c.f., Kamano 2023)
Juma alikula nyama nono (tu).
‘Juma ate (only) fat meat.’ (SWH104)

b. Questions
Juma alikula nyama nono (tu)?
‘Did Juma eat (only) fat meat?’ (SWH103)



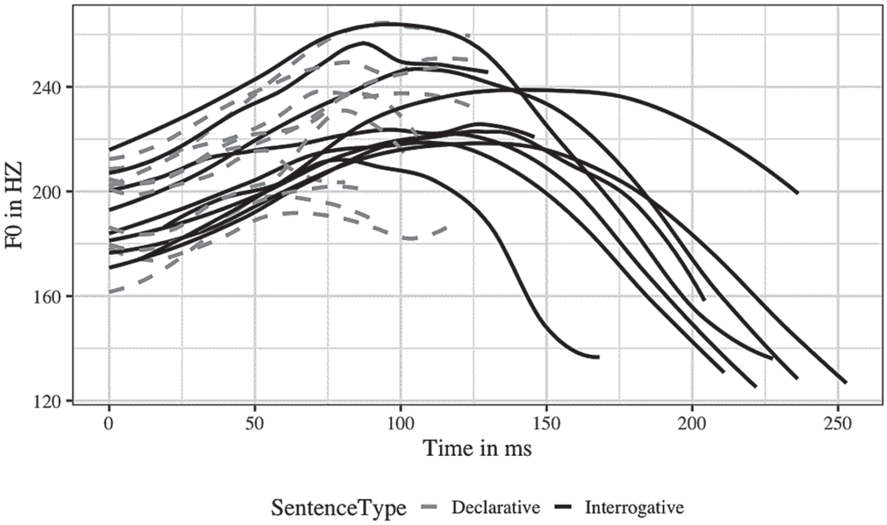


Figure 5 Pitch excursion and duration of the vocalic part of *tu* (not normalized) from SWH103

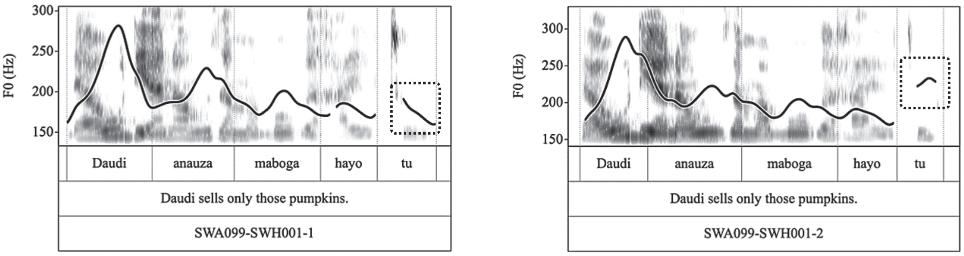
In addition to the F0 values, we also examined the duration of *tu* in the Final-*tu* sentences. The maximum and minimum F0 of each vowel interval was extracted alongside the F0 of 10 equidistant points throughout the interval of the vocalic part of *tu*. Some speakers, but not all, utilize the duration of the final *tu* as an alternative prosodic cue for questions. The pitch excursion shown in Figure 5 comes from all stimuli by SWH103. The vowel duration of the final *tu* is more than twice as long as the duration of *tu* in statements (Lee *et al.* 2023: 1491). Three out of six speakers produce a distinct pitch fall in the final *tu* when it appears in questions, as shown in Figure 5 (Lee *et al.* 2023: 1492).

4. Discussion

4.1 Two Prosodic Realizations of *tu*; Low or High Pitch

In our data of five stimuli, *tu* appears with a high pitch, while previous studies described the prosody of *tu* being low as in (3) *Watu wamelala tu*. ‘The people are just sleeping.’ (Ashton 1944: 75). During our data collection of other various sentences involving *tu*, we observe some *tu*s without a high pitch even in the sentence of the same speaker of Zanzibar Swahili as in (13).

(13) Two types of pitch realization of *tu*



a. *tu* with a low pitch

b. *tu* with a high pitch

The prosody of the example (13a) is consistent with Ashton's example (Ashton 1944: 75), which were observed in non-trivial number of examples, even though the dominant pattern in our corpus shows that *tu* has a high pitch as in (13b).

The presence of these two patterns can be understood in the context of the relationship between sentence structures and intonation in Swahili (Maw 1969, Maw and Kelly 1975), in which two dialogue texts (each with a length of 7.5 minutes) were analyzed. The texts were not controlled to study the particle *tu*, and it does not offer any phonetic data, but they include examples with *tu* and some comments that suggest what the prosody of *tu* may be in natural dialogues. Maw and Kelly (1975: 12) note that some adjuncts[6] in Swahili often appear after the predicate (called 'the Predicator' in their study), and the adjuncts in this position are realized in neutral intonation without prominence ('the tonic'). Their observation suggests that the particle *tu* is not necessarily realized with prominence, especially when they are immediately following a predicate; this distribution of particle *tu* is closely related to the prosodic realization of the particle.

In Maw (1969) and Maw and Kelly (1975), all 12 instances of *tu* occur in the sentence-final position (or 'tone group' in their study), and not in the post-verbal position (PV-*tu*). Nine out of these 12 *tu*s follow a predicate. The prosodic realization of *tu* may reflect some type of information structure, and the difference in (13) may be due to the different information structure of these sentences. These possibilities are beyond the scope of this paper and will need to be investigated in future studies.

4.2 Resolving prosodic ambiguity in questions with the final *tu*

As reported in the case of final-*tu* in questions, both the particle *tu* in the sentence-final position and the question sentence are expressed with a rising intonation or high F0. As reported in Lee *et al.* (2023: 1492), we also observed that the intonational characteristics of questions with Final-*tu* are not obligatory but optional. Speakers show that they use alternative strategies when resolving prosodic ambiguities.

In sentences without the particle *tu*, questions show the final rising of pitch, so the sentence-final adjective *no̲no* 'fat' with a penult rising is realized as *nono̲* with an ultimate

---

[6] These adjuncts include morphemes for vocatives, locative demonstratives, final tags (such as the exclusive particle *tu*), and words such as *sasa* 'now' (Maw and Kelly 1975: 12).

rising (cf. section 3.2.). When a sentence has a final *tu*, the pitch rising does not occur in the particle itself. Rather, speakers take another strategy. When producing question intonation without the particle (No-*tu* sentences), three speakers (SWH103, 104, 106) placed a high pitch on the ultimate vowel of the final word and produced it as *nono̲* with an ultimate rising. When creating a question, placing the high pitch on *tu* creates ambiguity; these speakers repaired this ambiguity by lengthening the vocalic part of *tu* or by creating a pitch fall in *tu*. At least the speaker (SWH102) used F0 to disambiguate questions from statements in Final-*tu* sentences. The *tu* in question is realized with a much higher F0.

Resolving ambiguity between intonational prosody (statements vs. questions) and the particle *tu* in multiple ways by different individuals demonstrates that this strategy is not yet grammaticalized in Zanzibar Swahili. Using any of the phonetic repairs (vowel lengthening, pitch fall, or pitch raise) will disambiguate statements from questions for these speakers. All of these repair strategies are cross-linguistically attested for various phonological patterns, and Zanzibar Swahili speakers are employing them.

5. Conclusion

This paper has investigated the prosody of Zanzibar Swahili in statements and questions with or without the exclusive particle *tu*. From our data, we conclude first that the intonation of statements has a downstep with a high penultimate syllable in each word. Our new finding is that the downstep is realized by level tones of the ultimate syllable of a word and the first syllable of the following word. Second, the intonation of yes-no questions has a sentence-final high-falling in significantly more speeches than a natural downstep final. However, we find that the higher pitch in a final word may fall on either the penultimate vowel or the ultimate vowel of the final word, while the final vowel of a sentence is described to be the target of question intonation in previous studies. Third, we demonstrate that the intonation of sentences with *tu* is not affected by the high pitch of *tu* in both PV-*tu* and Final-*tu* statements. In Final-*tu*, however, the longer duration, pitch fall or pitch raise of the final *tu* may be used to ensure the question intonation in sentences with the particle *tu*.

References

Ashton, Ethel. O. 1944. *Swahili Grammar (Including Intonation)*. London: Longmans.

Boersma, Paul and David Weenink. 2023. *Praat: Doing Phonetics by Computer*. https://www.fon.hum.uva.nl/praat/.

Clements, Nick and Annie Rialland. 2008. "Africa as a Phonological Area." *A Linguistic Geography of Africa*. (B. Heine and D. Nurse, eds.), 36–85, Cambridge: Cambridge University Press.

Connell, Bruce. 2001. "Downdrift, Downstep, and Declination." *Typology of African Prosodic System (TAPS) Proceedings*. Bielefield University, Germany.

Downing, Laura J. 2004. "What African Languages Tell Us about Accent Typology." *ZAS Papers in Linguistics*, 37: 101–136

Kamano, Shigeto. 2023. "A Downstep Analysis of Swahili Prosody: A Case of the Exclusive Particle

TU." *ICU Working Papers in Linguistics (ICUWPL)*, 24: 93–100.

Lee, Seunghun J., Shigeto Kamano, Yuko Abe, and Kumiko Miyazaki. 2023. "Prosodic Prominence of Swahili TU in Polar Interrogatives". *Proceedings of the 20th International Congress of Phonetic Sciences*, 1513–1517. Prague: Guarant International.

Maddieson, Ian and Bonny Sands. 2019. "The Sounds of the Bantu Languages". *The Bantu Languages*, (M. Van de Velde, K. Bostoen, D. Nurse and G. Philippson, eds.), 79–127. New York: Routledge.

Maw, Joan. 1969. *Sentences in Swahili: A Study of Their Internal Relationships*. London: University of London, School of Oriental and African Studies.

Maw, Joan and John Kelly. 1975. *Intonation in Swahili*. London: University of London, School of Oriental and African Studies.

Meeussen, Achille Emile. 1967. "Bantu Grammatical Reconstructions." *Africana Linguistica*, 3: 79–121.

Nurse, Derek, Thomas J. Hinnebusch and Gérard Philipson. 1993. *Swahili and Sabaki: A Linguistic History*. Berkeley: University of California Press.

Polomé, Edgar C. 1967. *Swahili Language Handbook*. Washington, D.C.: Center for Applied Linguistics.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Shinagawa, Daisuke and Nico Nassenstein. 2020. "On Variation in Swahili: Current Approaches, Trends and Directions." *Swahili Forum*, 26: 1–44.

Wilson, Peter M. 1985. *Simplified Swahili*. Nairobi: Longman.

Journal of Asian and African Studies, Supplement, No. 4, 2024

Prosody of Question Elements in Two Bantu Languages:
Shingazidja and Xitsonga*

Patin, Cédric
Lee, Seunghun J.

This paper reports how question elements (wh-words and question particles) of two distinct Bantu languages are prosodically realized. Shingazidja (the Comorro islands) and Xitsonga (South Africa) are distantly located, but their tonal and prosodic grammars are comparable. Even so, question elements do not behave in a comparable manner. Wh-words in Xitsonga appear in the in-situ position, whereas the wh-words in Shingazidja appear in the clause-final position. Except for wh-words in the subject position, the prosody of wh-words behave as expected. Distribution of other question elements also differs in the two languages: the question element *xana* in Xitsonga has a free placement in a clause, while the question element *yé* in Shingazidja prefers to appear in the clause-final position.

Keywords: prosody, question elements, Bantu, Shingazidja, Xitsonga

DOI: https://doi.org/10.15026/0002000634

[*] This research is written as part of the ILCAA joint research project "Phonetic typology from cross-linguistic perspectives (PhonTyp)" (2021–2023). The research has been funded by a Kakenhi Grant (C) 16K02641 and a Visiting Scholars Program by the Japan ICU Foundation. This work was also supported by JSPS KAKENHI Promotion of Joint International Research (Fostering Joint International Research (B)) Grant Number 21KK0005, and the Kakenhi (B) Grant 23H00622. We thank consultants from the Shingazidja and the Xitsonga communities. Comments from the audience at the 55th Annual Meeting of Japan Association of African Studies and TIE in Göteborg were helpful. Any errors remain our own.

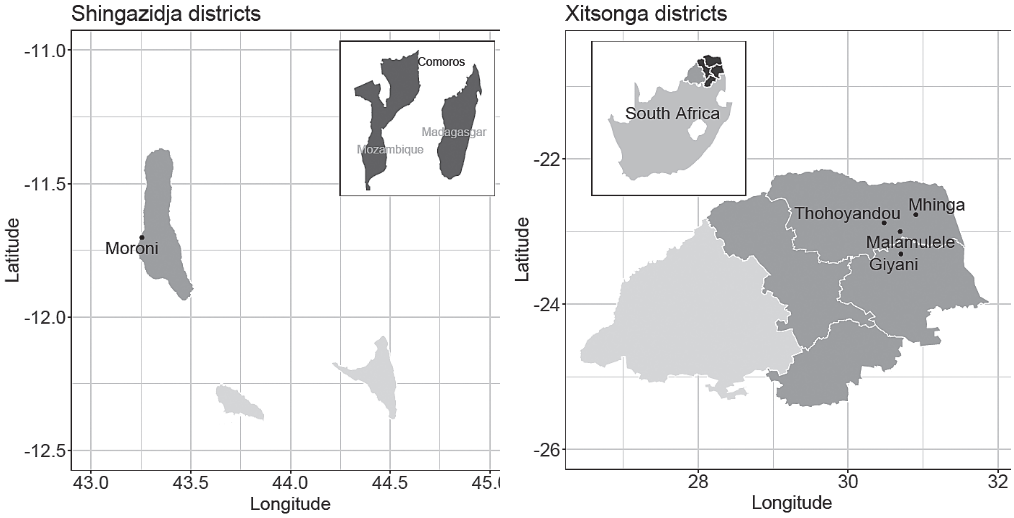


Figure 1 Maps of areas where Shingazidja (the Comorro islands) and Xitsonga ([S53] South Africa) are spoken (created by courtesy of Julián Villegas)

1. Introduction

Questions in African languages show uncommon prosodic status such as lax question prosody in languages spoken in the Sudanic belt (Rialland 2007, 2009). Even so, the question prosody of African languages remains an understudied area. While some recent efforts, e.g. Downing & Rialland (2017a) among others, improved our understanding in this area, we lack more detailed and nuanced view, in particular, concerning the prosodic behavior of question elements (QEs), i.e. wh-words and question particles. The current study aims to address this gap by reporting on the prosody of question elements in two Bantu languages: Shingazidja (G44a, Comoros) and Xitsonga (S53, South Africa), maps in figure 1 show their geographical locations.

Shingazidja and Xitsonga are not closely related, nor are they spoken in the same area. However, the phonology of Shingazidja and Xitsonga share commonalities. Both languages have a privative high (H) tone system where H tone contrasts with tonelessness as in (1). Multiple occurrences of lexical H tone show downstep, where H tones following another H tone is lowered relative to the preceding H tone as in (2). In both languages, the subject with a final H tone on the surface is followed by a H tone verb that is downstepped. Subsequent H tones in the sentences are further downstepped.

(1) Privative H tone

a. **Shingazidja**

i.	/m-ɖó/	mɖó	‘river’ (3-*river*)	vs.
	/m-ɖo/	ḿ̩ɖo	‘fire’ (3-*fire*)	
ii.	/ʦí húsi/	ʦí=husi	‘no summer’ (neg=9.*summer*)	vs.
	/ʦi-hulá/	ʦi-hulú	‘I bought’ (1sg.pfv-*buy*)	

b. **Xitsonga**

i.	/ú/	ú	‘s/he’	vs.
	/u/	u	‘you’	
ii.	/^{m}bila/	^{m}bila	‘xylophone’	vs.
	/^{m}bílá/	^{m}bílá	‘dassie’	vs.
	/^{m}bíla/	^{m}bíla	‘rockrabbit’	

(2) Downstep

a. **Shingazidja**

[(*ye=m-lima̱djí*)$_{\varphi}$ (*ha-ni̱kꜜá*)$_{\varphi}$ (*e=m-levꜜí̱*)$_{\varphi}$ (*e=ɲ-umɓꜜá̱*)$_{\varphi}$]$_{\iota}$
AUG$_1$=1-farmer 1.PFV-give AUG$_1$=1-drunkard AUG$_9$=9-house
‘The farmer gave the house to the drunkard.’ (Patin 2017: 291)

b. **Xitsonga**

[(*ɦo:sí̱*)$_{\varphi}$ ((*wꜜá̱-ɲi̱ká* *mú-lú:ŋgú*)$_{\varphi}$ *ta:ndzꜜá̱*)$_{\varphi}$]$_{\iota}$
chief 1.SM-give 1-European egg
‘The chief gives an egg to a European.’

Both languages have a mobile H tone: H tone shifts in Shingazidja but spreads in Xitsonga. The shifting in Shingazidja as well as the spreading in Xitsonga are both sensitive to phonological phrase boundaries. In (2a), four phonological phrases are present: the subject, the verb, the indirect object and the direct object. Independently a DP with an augment must be preceded by a phonological phrase boundary. The tone of the Shingazidja noun *mlima̱dji* ‘farmer’, which is associated with the penult at the underlying level, shifts to the final syllable of the noun that is followed by a phonological phrase boundary. The tone of the subject NP cannot shift to the first syllable of the verb. The H tone of the verb shifts to the final syllable of the verb, but it cannot shift onto an NP that is introduced by an augment.

H tone spreading in Xitsonga also shows sensitivity to phonological phrase boundaries. In (2b), three phonological phrases are shown: the subject, the verb with the indirect object, and the entire verb phrase. The tone of the verb spreads to the toneless indirect object NP but it does not spread beyond it; the blocking of H tone spreading is due to the presence of a phonological phrase boundary after the indirect object.

The details differ in that Shingazidja targets the end of a maximal phonological phrase for H tone shift while Xitsonga targets the end of minimal phonological phrase for H tone spread. In (3a), different from (2a), Shingazidja has a single phonological phrase because the nouns do not have an augment. The H tone of the verb *tsiní̱ka* ‘I gave’ shifts to the first syllable of the direct object *mapé̱sa̱* ‘money’, throughout the toneless beneficiary *wánḓu* ‘people’. Xitsonga has a binarity requirement for the size of phonological phrases (Lee & Selkirk 2022). The Xitsonga sentence in (3b) shows that the H tone of the verb *ni-ɲí̱ka* spreads to the toneless beneficiary object *βa-n̤u* ‘people’, but not beyond; the direct object *ɲa:ma* ‘meat’ remains toneless. If H tone spread in Xitsonga is sensitive to a maximal phonological phrase, we would expect the H tone spreading onto *ɲa:ma* ‘meat’. H tone spreading in Xitsonga, thus, is sensitive to a minimal phonological phrase in a recursive phi structure.

(3) Mobile H tone

a. H tone shift in **Shingazidja**

[(*ʦi-ni̱ka* *wa-nḓu* *má-pe̱sꜜá̱*)$_\varphi$]$_\iota$
1SG.PFV-give 2-person 6-money
‘I gave money to people.’ (Patin 2017: 288)

b. H tone spread in **Xitsonga**

[((*ni-ɲí̱ká* *βá-n̤ú*)$_\varphi$ *ɲa:ma*)$_\varphi$]$_\iota$
1SG-give 2-person 9.meat
‘I give meat to people.’

Intonation patterns in both languages are similar as well. Both languages have a sentence-final L% boundary tone. Compare the two verbal forms in the Shingazidja example (4a), and Figure 2: the F_0 curve associated with the high tone of the leftward example is flat,[1] in comparison to the high tone of the penult of the rightward form, which is characterized by a rising shape; additionally, the F_0 curve drops sharply on the final vowel of the rightward example, due to the absence of high tone and the presence of an L% boundary tone. The Xitsonga example in (4b) also shows the presence of a L% boundary tone. In Figure 3, the verbal H tone spreads to the penultimate syllable as shown by the F_0 curve that is rising after the subject agreement *ni*. The H tone plateau continues until the penultimate syllable after which the F_0 drops sharply during the ultimate syllable due to the L% boundary tone.

---

[1] This is only true for some dialects, see Patin (2017) for details.

(4) L% boundary tone

a. **Shingazidja**

i. *ʦi-ɾeŋgé̱* vs. ii. *ʦi-i̱-ɾéŋge̱*
1SG.PFV-take 1SG.PFV-OM$_9$-take
‘I took’ ‘I took it (e.g. the ring).’

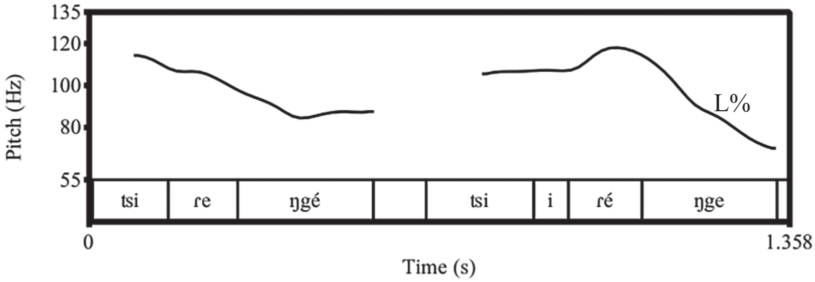


Figure 2 F$_0$ curves of (4ai) – left – and (4aii) – right –, as produced by a male speaker from Moroni.

b. **Xitsonga**

ni lí̱már-í:le
1SG injure-PST
I injured ‘I'm injured’

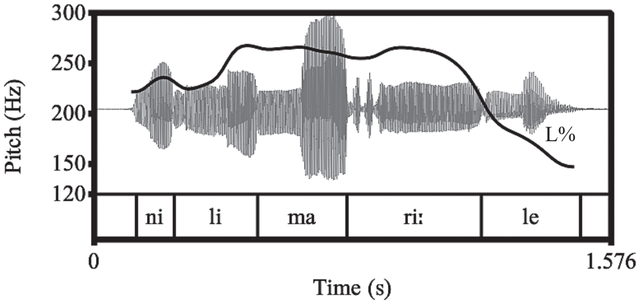


Figure 3 F$_0$ curves of *ni limarile* ‘I'm injured’ by a female speaker of Xitsonga.

Non-finality is another area where the two languages show similarity. In Shingazidja, evidence for non-finality comes from the fact that a tone cannot shift on to the final syllable of an intonational phrase. The H tone of the verb in (5a) shifts to the penultimate syllable of a toneless object. In Xitsonga, the non-finality blocks the spreading of H tone onto the final syllable of an intonational phrase. In (5b), the verbal H tone spreads rightward but not beyond the penultimate syllable of the toneless object noun.

(5) Non-finality
a. **Shingazidja**
[(*ha-wo̲no* *n-dóvu*)$_{\varphi}$]$_{\iota}$ **hawo̲no ndovú*
1.PFV-see 9-elephant
‘He saw an elephant.’
b. **Xitsonga**
[(*ni-βó̲ná* *n-dló:pfu*)$_{\varphi}$]$_{\iota}$ **niβó̲ná ndló:pfú*
1-see 9-elephant
‘He sees an elephant.’

Question elements in this paper refer to both question particles and wh-words. In polar questions, the questions are signaled with the particles *(y)é̲* in Shingazidja and *xana* in Xitsonga. These question elements can occur in the initial position as well as the non-initial position of an interrogative sentence. Although the two languages display similarities in tonal and intonational phonology, an examination of the intonation pattern of these question elements shows that the similarities in basic intonation patterns do not extend to the intonation of sentences with question elements. In Xitsonga, the question element *xana* does not show any marked prosodic effect, but in Shingazidja, the question element *(y)é̲* prosodically affects sentential intonation when it appears in the non-initial position. The other type of question elements are wh-words in argument questions: *má̲ni* ‘who’ or *yiní̲* ‘what’. In both languages, object wh-words appear in-situ with no specific effect on the prosody whereas subject wh-words are placed in a cleft sentence accompanied by a prominence on the wh-words.

The goal of this paper is completing a prosodic typology of question elements based on the question elements from Xitsonga and Shingazidja. Section 2 first reviews studies about the intonation of yes-no questions and wh-questions in African languages. The prosody of wh-words is presented in section 3 and the prosody of question elements in section 4. A discussion about the prosodic typology of question elements appears in section 5.

2. Background: Question Intonation in Bantu Languages and the Prosody of Question Elements

Question intonation in Bantu languages is an understudied field, and descriptions of Bantu languages often do not include sections that explain the intonation patterns. The edited volume by Downing & Rialland (2017a) is one of the first to offer a systematic and comparable overview of the intonation patterns including question intonation. An overview of the question intonation is offered in this section. Intonation of yes-no questions displays prosodic differences from corresponding statements, but intonation of wh-questions does not exhibit such prosodic differences. The rest of this section will review the intonation of yes-no questions in section 2.1 and that of wh-questions in section 2.2.

2.1 Yes-No Question Intonation

The morphosyntactic structure of yes-no questions does not differ from their declarative counterparts in most Bantu languages (Downing & Rialland 2017b), except for the presence of an optional question marker (see section 4). Yes-no questions are prosodically marked with a specific intonation pattern,[2] most commonly a rising(-falling) pattern associated with the final syllable of the utterance (e.g., in Ha [JD66] – Harjula 2004: 150; in Herero [R31], where "*a general slight raising of pitch with a fall on the final syllable yields a question from a statement*" – Elderkin 2003: 607), or the penult (e.g., in Chichewa [N31b] and Tumbuka [N21a] – Downing 2011, 2017; in Kinyarwanda[3] [JD61] – Jarnow 2020; in Xhosa [S41] – Jones, Louw & Roux 2001).

Shingazidja marks yes-no questions with a specific intonation on the penult. As described in detail in Patin (2017: 305–317), polar questions in Shingazija do not differ from their declarative counterparts in their morphosyntactic structure, except for the presence of an initial question marker whose behavior will be extensively discussed in section 4. Prosodically, yes-no questions are characterized by the presence of a superhigh tone that emerges on the penultimate, stressed syllable of the utterance (6b), labeled as LH* in an autosegmental-metrical perspective, and by the suspension of downstep. In addition, the superhigh tone builds a tone plateau with the preceding lexical tones.[4]

(6) **Shingazidja**
a. [(*ha-nik̲á*)$_\varphi$ (*e=ɲ-uŋgu̲* *n-dzꜝíro*)$_\varphi$]$_\iota$
1.PER-give AUG$_9$=9-pot 9-heavy
'He gave the heavy cooking-pot.'
b. [(*hanik̲á*)$_\varphi$ (*é=ɲúŋgú̲* *ndzi̋ro*)$_\varphi$]$_\iota$
'Did he give the heavy cooking-pot?'

A striking property of the intonation of polar questions in Shingazidja is a tone-intonation interface phenomenon whereby the superhigh tone will be retracted on the antepenult if a lexical tone surfaces on the final syllable of the utterance (Patin 2017). Compare (7) with (6), where the superhigh emerges on the antepenult because of the presence of the final high tone (7b).

---

[2] Some non-Bantu African languages do not display specific intonation patterns; no postlexical tone, for instance, signals yes-no questions in Yoruba (Connell and Ladd 1990), or in Mambila (Connell 2005).

[3] When there is no lexical tone in the sentence. Otherwise, the final tone is raised.

[4] See Patin (2017: 305) for additional examples of superhigh. Superhigh tone in Shingazidja also signals alternative questions, where it lands on the penult of the first part of the alternative choices, and biased questions. Questions denoting surprise and rhetorical questions are also signaled by pitch range expansion, whereas echo and surprise questions add a final !H%.

(7) **Shingazidja**

a. [(*ha-ni̱ká*)$_φ$ (*e=ɲ-uŋgꜜú̱*)$_φ$]$_ι$
1.PER-give AUG$_9$=9-pot
'He gave the cooking-pot.'

b. [(*hani̱ká*)$_φ$ (*e̱̋=ɲuŋgꜜú̱*)$_φ$]$_ι$
'Did he give the cooking-pot?'

In addition to a specific melodic shape, Bantu Yes-No questions frequently are associated with an expansion of the pitch range (e.g. in Bemba [M42] – Kula & Hamann 2017; in Chichew̑a – Myers 1996, Downing 2017; in Tswana [S31] – Zerbian 2017), a raised register (e.g. in Sesotho [S33] – Zerbian 2006: 260; Mixdorff *et al.* 2011; in Tswana – Zerbian 2017: 416; in Xhosa – Jones, Louw & Roux 2001), and with a suspension of downstep (e.g. in Bemba – Kula & Hamann 2017; in Chichew̑a – Downing 2017; in Chimiini – Kisseberth 2017: 236; in Shingazidja – Patin 2017: 306) and/or downdrift (e.g. in Ha – Harjula 2004: 150).

Xitsonga is a language that expands its pitch range in yes-no questions. The morphosyntactic structure remains identical to declaratives, and the pitch excursion pattern is also similar between yes-no questions and declaratives. Hlongwani & Lee (2018) report that the overall pitch range of yes-no questions is higher as shown in (8) and Figure 4. The declarative sentence has a pitch plateau around 140 Hz, while the yes-no question has the plateau around 180 Hz. The waveform of the question also shows stronger intensity on the first syllable of the verb *lá̱va* 'to want'.

(8) a. *u lá̱vá ngúlú:ve*
2SG want 9.pig
'You want a pig'.

b. *u lá̱vá ngúlú:ve*
2SG want 9.pig
'Do you want a pig?'

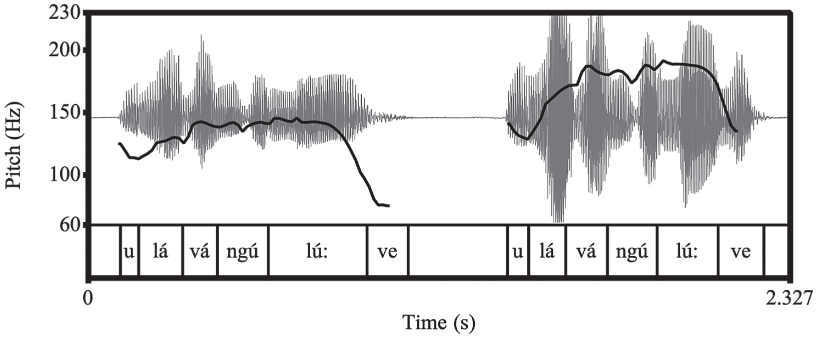


Figure 4 Pitch track of a declarative sentence in (a) and a yes-no question in (b) in Xitsonga

In addition to the pitch manipulation, yes-no questions are also distinguished from declaratives utilizing parameters such as a higher speech rate (e.g. in Sesotho – Mixdorff *et al.* 2011; in Xhosa – Jones, Louw & Roux 2001), a higher intensity (e.g. in Xhosa – Jones, Louw & Roux 2001), or the absence of penultimate lengthening (e.g. in Sesotho – Zerbian 2006, Mixdorff *et al.* 2011; in Tswana – Zerbian 2017).

2.2 WH Questions Intonation

The intonation of wh-questions does not differ from the prosody of statements in a number of Bantu languages. In Chimiini, for instance, "*[t]here are no intonational properties associated uniquely with wh-questions*" (Kisseberth 2017: 257). In Embosi, "*wh-questions [...] do not have any specific intonations[, t]hey display the same intonations as assertive utterances*" (Rialland & Embanga Aborobongui 2017: 214), and the same holds in Sotho (Zerbian 2006: 266).

Yet, a specific set of prosodic clues is frequently associated with content questions in Bantu: pitch-range expansion, or pitch-range rising. Pitch-range expansion is attested in Bemba (Kula & Hamann 2017) or Tswana, especially in wh-questions containing high tones (Zerbian 2017: 420); pitch-range expansion has been reported, e.g. in Chichewa and Tumbuka ("*we find no obligatory question melody in either language, though the overall pitch is raised somewhat compared to statements*" – Downing 2017: 383), and Fipa ([M13], Riedel & Patin 2011). Fipa does not follow the pattern that has been described in this section, since wh-questions are signaled by a final H(L)% melody, as well as the suspension of downdrift and the absence of final devoicing.

In both Shingazidja and Xitsonga, content questions are primarily identified by the presence of an interrogative pronoun. Optional question elements can additionally signal the interrogative of a sentence: *(y)é̱* (or *á̱*) in Shingazija and *xana* in Xitsonga.

As for intonation, a wh-question differs from a statement in Shingazidja in that "*it does not exhibit the L% boundary tone that characterizes the latter*" (Patin 2017: 316), except when the wh-word is dislocated. When a wh-word is dislocated in Xitsonga wh-questions, the dislocated wh-word is realized with pitch boost. Unlike Shingazidja, an L% boundary tone is present in wh-questions.

The overview in this section has shown that the prosody of questions was mainly studied at the utterance level, and prosodic effects of question elements, if there are any, were overlooked. The intonation of wh-words has rarely been explored with an exception of involvement with a focus. The lack of studies on the intonation patterns of question elements may stem from the optionality of these elements. The rest of this paper introduces and contrasts the question elements *yé̱* (or *á̱*) in Shingazija and *xana* in Xitsonga and show how the positional restrictions of these elements interact with the grammar of intonation.

3. The Prosody of Wh-words

This section introduces the prosody of wh-words to show whether these words have any specific contribution to the prosodic patterns. After offering some background on what is known of the prosody of wh-words in Bantu in section 3.1, we describe their characteristics in Shingazidja (Section 3.2) and Xitsonga (Section 3.3).

3.1 Background: The Prosody of wh-words in Bantu

While several studies have been dedicated to the prosody of questions from sentential perspectives, research on the prosodic shape of wh-words remains marginal. What is known is the absence of a specific prosody associated with wh-words in Chicheŵa or Tumbuka (Downing 2017: 384), or Ha (Harjula 2004: 150). In Bemba, for instance, "*the question word does not show specific significant pitch raising [and w]hen the question word occurs in clause-final position it is subject to final lowering*" (Kula & Hamann 2017: 346).

Prosodic variation can occur when wh-words are concerned. Downing (2017: 384) points out that "*there is some tendency to put a phrase break after a wh-word*" in Chicheŵa, and in Tumbuka there is "*an optional raised (↑) register melody on a wh-question word when it appears in sentence-final (Intonational Phrase-final) position, and [that] the final syllable of the wh-question word is also raised*".

When wh-words bear a focus, specific prosodic patterns emerge in several languages: a focused wh-word is followed by a phrase boundary in Chimiini (Kisseberth 2017: 256), and in Fipa its tone shifts to the penult (e.g. *tʃaaní* 'what' > *tʃáani* – Riedel & Patin 2011: 162–163). When focus or contrastive wh-words are fronted as in Makaa (A83, Heath 2003: 340), wh-words often appear with a specific prosody; in Bemba, "*[t]here is some increase in pitch when the question word is non-final*" (Kula & Hamann 2017: 346)

3.2 Prosody of wh-words in Shingazidja

In Shingazidja, as in other Bantu languages (e.g. Ekoti – Schadeberg & Mucanheia 2000; Nkore-Kiga – Taylor 1985: 139; Ha – Harjula 2004: 150), the default position for a wh-word is the utterance-final position (9).[5] In such a context, the prosody associated with the wh-word does not differ from those of any other prosodic word in the language (in Patin (2017: 315): "*the prosody of a wh-question does not seem to differ from that of a statement*"), except for the lack of L% tone (see Section 2.2) – compare for instance (9) with (4a).

(9) [(*ya-ɾe̱má*)$_{\varphi}$ (*nɗo=ꜝβí̱*)$_{\varphi}$]$_{\iota}$
1.REL.PER-hit who=SPEC
'Which one (of you) did hit (it)?'

In particular, the wh-word is subject to downstep, as illustrated in (9) and the left part of Figure 5. In order to emphasize, to express a high degree of concern on the part of the speaker, a wh-word can however be fronted as shown in (10). Compare the pitch track of (10) with that of (9).

---

[5] One speaker, [M], sometimes produced subject wh-words in their in-situ position, e.g. [(*nɗo=βi̱* yámmo̱na)$_{\varphi}$]$_{\iota}$ '*who* has seen him?'.

(10) [(*nɗo=ꜛβí̱*)$_{\varphi}$]$_{\iota}$ (*yaꜝɾ̱é̱ma*)$_{\varphi}$]$_{\iota}$ ‘Which one (of you) did hit (it)?’ [M]

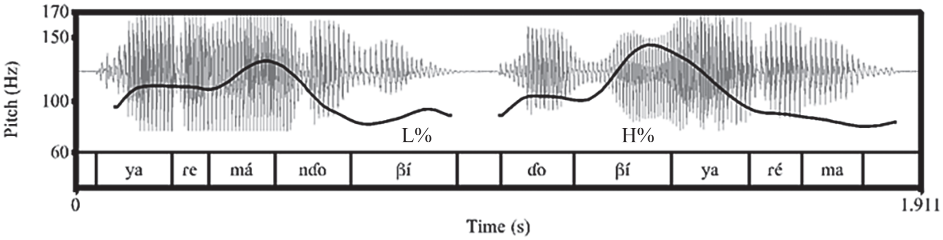


Figure 5 Signal and F_0 curves of (9) – left – and (10) – right –, as produced by a male speaker from Moroni.

In (10), the wh-word is separated from what follows by an Intonation Phrase boundary. This analysis is supported by two main arguments. First, the tone of the wh-word is higher than a corresponding tone that is not followed by an IP boundary. As it has been shown in Figure 5 & O’Connor & Patin (2015) and Patin (2017), an H% boundary tone is associated with the right boundary of an IP in Shingazidja. Second, the comparison of the F_0 curves of the final tones of (9) – rising – and (10) – flat – in Figure 5 reveals that the latter is associated with a L% boundary tone, contrary to the former. Thus, while the wh-word exhibits an unmarked prosody when it emerges in its canonical, utterance-final position, it is clearly strengthened when fronted.

3.3 Prosody of wh-words in Xitsonga

Xitsonga have wh-words *máni* ‘who’, *yiní* ‘what’, *kwíhi* ‘where’, etc., which have varying tonal patterns with an initial H tone or a final H tone. The word *yini* [jiní] ‘what’ appears syntactically in-situ between the verb and the locative adjunct (11a), rather than the final position as it would have been in Shingazidja. The prosody of the wh-word does not differ from that of a sentence where the verb is followed by two constituents as in Figure 6 that represents (11). High tones are downstepped after another high tone. The final syllable of the intonational phrase is subject to non-finality blocking the spreading of H tone on it. Wh-words may appear in the sentence-initial position with a topic marker (11b), which is prosodically accompanied with an F0 boost on the wh-words, but not on the topic marking prefix.

(11) a. (((*mi βón-é ji:nꜜí*)$_{\varphi}$ *ꜜélánda:ni*)$_{\varphi}$)$_{\iota}$
you see-PFV what in.London
'What did you see in London?'

b. ((*í ↑j:íní*)$_{\varphi}$ (*mi ʃí βꜜón-é:ke ꜜélánda:ni*)$_{\varphi}$)$_{\iota}$
TOP what you OM7 see-PFV in.London
'What is it that you saw in London?'

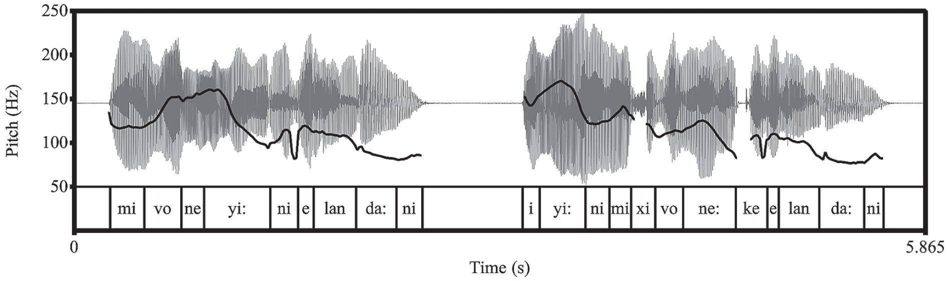


Figure 6 Xitsonga wh-questions: a. wh-in-situ, b. fronted wh-phrase

The default position of wh-words in the two languages syntactically differ from each other: the clause-final position in Shingazidja and in-situ in Xitsonga. In these positions, the prosody of wh-words is akin to the prosody of other types of words, suggesting that wh-words themselves do not contribute any specific question prosody. In both languages, wh-words can be fronted from their original final position in Shingazidja and the in-situ position in Xitsonga. A move of a wh-word to this initial position of a sentence is marked with a sharp boost of F_0 in both languages. In the prosodic structure, a sentence corresponds to an intonational phrase, and the fronted wh-word is prosodically adjoined to the intonational phrase. Although the syntactic origin of the wh-words differs in the two languages, the common property of pitch boost of fronted wh-words suggest that the boost is due to the prosodic adjunction (exemplified in (10) and (11b)); a reason why both languages prosodically pattern in an identical manner.

4. The Prosody of Question Elements

Questions in both Shingazidja and Xitsonga appear with question elements. This section discusses these elements and their intonation. After an introduction to what is known of these elements in Bantu (Section 4.1), we turn to their prosody as it emerges in the two languages of our study. As Xitsonga shows the unmarked patterns, we will first discuss the Xitsonga data before discussing the Shingazidja data, respectively in sections 4.2 and 4.3.

4.1 Background

The scarcity of the research dedicated to the prosody of wh-words in Bantu languages extends to all question elements (QE), of which very little is known. One of the major reasons behind this may be the optional appearances of question elements in Bantu languages. As explained in section 2, yes-no questions do not differ much from declaratives in the Bantu morphosyntax, except for the presence of question elements. Thus, identifying intrinsic prosodic properties of questions are readily done with questions without optional question elements. Another possible explanation is that QE do not display any specific prosodic property compared to other types of words, especially the wh-words, of a given Bantu language; for instance, in Chimiini, interrogative enclitics have the same prosody as wh-words (Kisseberth 2017). QEs also do not display any impact on the prosody of questions; in discussing the particle *a* of Tswana, Zerbian (2017: 416) states that "*a yes/no-question marked with the question particle* a *does not differ in its intonation from yes/no-questions produced without the question particle*". We argue that examining the prosody of QEs offers a fresh look into understanding their prosodic roles; this paper thus examines data from two prosodically similar languages that show QEs that differently behaves in the prosody.

The position of QE displays cross-linguistic variability (cf. Dryer 2013). QE appears in the clause-initial position (cf. !Xóõ (Khoesan), Traill 1994: 18), in the clause-final position (cf. Majang (Nilo-saharan), Unseth 1989: 126) or in the clause-second position (cf. Yurok (Algic), Robins 1958: 139). In some languages QEs appear between the verb and the subject (cf. Hmong Njua (Hmong-Mien), Harriehausen 1990: 205), while QE appear in variable positions in other languages (cf. Hunde (Bantu [JD51]), Kahombo 1992: 171). Close to 40 percent of the languages listed in WALS do not mark polar questions with a QE.

Some descriptions mention the prosody of QE in describing the intonation patterns: *bùshé* in Bemba (Kula & Hamann 2017), *koodí* in Chichewa and *káasi* in Tumbuka (Downing 2017), *mbere* in Ha (Harjula 2004: 150), *ndi* in Ngangela (Zavoni 2003), *kí* in Ekegusii[6] (Cammenga 2002), the *na/naa* – standard yes-no or content questions – and *a/afa* (rhetorical questions) particles in Sotho (Zerbian 2006: 262ff) or the enclitic *-pí* 'which?' in Herero [R31] (Elderkin 2003: 595).

4.2 Question Elements in Xitsonga

Xitsonga question element *xana* represents a language with an unmarked pattern because *xana* does not display any unique prosodic effects. Previous studies on Tswana (Zerbian 2017) report that the presence of a question element may not have any effect on the prosody of question. Xitsonga *xana* behaves akin to the question element in Tswana because no discernible effect on prosody is observed when a question element is used in a sentence.

The occurrence of *xana* is not restricted to a single position either; a question can begin or end with *xana*. While *xana* shows prosodic patterns akin to in-situ wh-words in the non-initial position, the element *xana* has the prosody of an ex-situ wh-words (with an F_0 boost) in the initial position.

(12) a. *u* *lává* *kófi* ***xá:na***?
2SG want coffee xana
'Do you want coffee?'
b. ***xána*** *u* *lává* *kó:fi*?
xana 2SG want coffee

---

[6] This QE also exists in Luganda (Hyman & Katamba 1993).

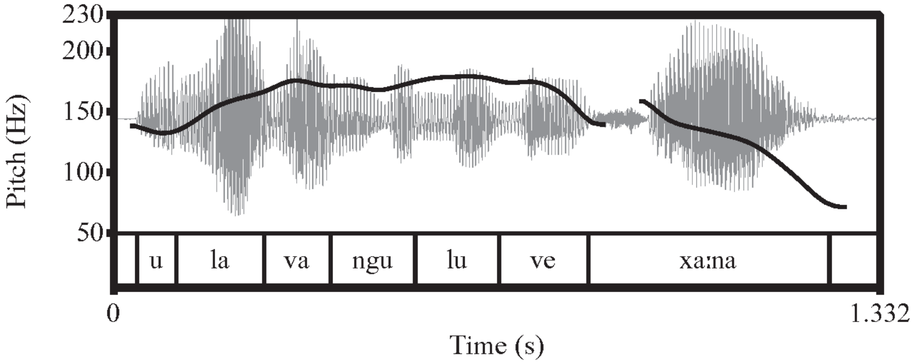


Figure 7 Pitch track of a sentence *u lává ngúlúvé xá:na* 'Do you want a pig?' with the sentence-final *xana*.

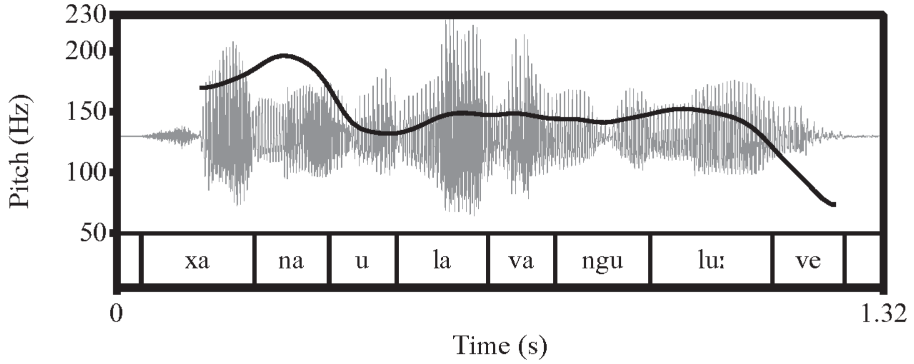


Figure 8 Pitch track of a sentence *xána u* ꜜ*lává ngúlú:ve* 'Do you want a pig?' with the sentence-initial *xana*.

In Figures 7 and 8, a comparison of the prosodic patterns of *xana* is shown. When *xana* appears in the sentence-final position, the f_0 is falling over *xana*, and *xana* itself is produced with high amplitude. Note also that the lexical H tone on *xana* is not realized in the sentence-final position. Unlike the *xana* in Figure 7, the sentence-initial *xana* is realized with a pitch boost with the second syllable with a lexical H tone having a high pitch, but the word itself is produced without strong intensity throughout the word.

Xitsonga *xana* is realized with a prosodic pattern that is expected in questions without *xana* (compared Figure 8 with Figure 4b). This kind of unmarked pattern in prosodic realizations differs from the prosodic realization of the question element *(y)é̱* in Shingazidja, which varies in relation to the position of *(y)é̱* in a sentence.

4.3 Question Elements in Shingazidja

In Shingazidja, a question is usually introduced by the question element *(y)é̱*[7] (13). As explained in section 2, this is true for both the polar/alternative and content questions, with the possible exception of echo or rhetorical questions.

---

[7] According to some speakers, another question particle, *á̱*, is used as an alternative by some speakers, especially the elders. Since none of our informants ever used this particle unprompted, we will not explore its distribution, which seems to be similar to that of *(y)é̱*, in further detail.

(13) a. *ye̱ ŋgó-fiḳiɾi uká̱ ha̋-limꜝí̱* [N]
QE 2SG.IPFV-think.PRS that 1.PFV-cultivate
b. *ŋgofiḳiɾi uká̱ ha̋liꜝmí̱*
‘Do you think he cultivated?’

The factors that determine the presence of *(y)é̱* (13a) or the absence of *(y)é̱* (13b) in the beginning of a question remain unclear at this point of our research. Some of our informants ([M], [N] and [W]) claim that the element is optional, while others ([S] and [B]) consider it to be mandatory, while admitting that it can (‘incorrectly’) be dropped by some speakers. All our informants, however, even those who claim that *(y)é̱* is optional, spontaneously used *(y)é̱* when they were asked to produce a question. Since [N] and [B], who originate from the same area, differ in their appraisal of the distribution of *(y)é̱*, the variability in its emergence can hardly be linked only to dialectal variation.

The question element *(y)é̱* begins with a palatal approximant which is probably epenthetic, and ends with a vowel which is underlyingly associated with a high tone (14).

(14) [(*yé̱*)$_{\varphi}$ (*ha-wo̱nó pa̱̋ha*)$_{\varphi}$]$_{\iota}$ [W]
QE 1.PFV-see 5.cat
‘Did he see a cat?’

The element is separated from what follows by a Phonological Phrase boundary: in (15), the tone of *(y)é̱* does not shift to the first syllable of the verb *hawó̱no* ‘he saw’, even if this syllable is not associated with a high tone at the underlying level. However, the tone of *(y)é̱* is frequently displaced on the following syllable in [M]’s and [S]’s productions, and almost systematically in those of [B], [N] and [W] – see for instance (13), or (14).

(15) [(*ye̱*)$_{\varphi}$ (*há-ɾeŋge̱̋ m-pi̱ꜝɾá̱*)$_{\varphi}$]$_{\iota}$ [S]
QE 1.PFV-take 3-balloon
‘Did he take a balloon?’

This postlexical H tone shift is due to a constraint against the left-alignment of high tones with an Intonational Phrase boundary, especially in utterance-initial position (Patin (2010, 2017) – see Kisseberth (1984) for a similar rule in Digo, an E73 Bantu language of Kenya).

When *(y)é̱* introduces a content question, it usually triggers the downstepping of the following high tone(s) (16).

(16) [(*yé̱*)$_{\varphi}$ (*ya-ɾe̱ꜝmá*)$_{\varphi}$ (*nɗo=ꜝβí̱*)$_{\varphi}$]$_{\iota}$ [M]
QE 1.REL.PFV-hit who=FOC
‘Which one did he hit?’

However, there is considerable variation on this point in our data. The tone of the question element can also be lower than those of the following words (compare for instance the two realizations of (17) in Figure 9).

(17) [(*ye̱*)$_\varphi$ (*zá-zi̱$^{(!)}$ɗí*)$_\varphi$ (*[hi]ꜝndí̱*)$_\varphi$]$_\iota$ [W]
QE 8.REL.PFV-change what
'What happened?'

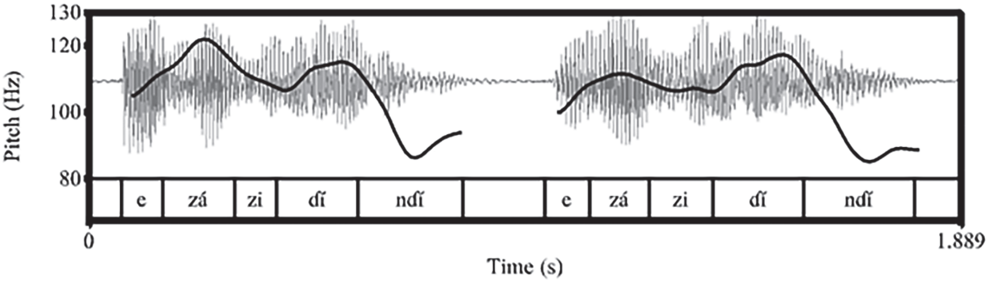


Figure 9 **Signal and F_0 curves of two iterations of (16), as produced by a male speaker from Washili in a single working session.**

The lowering of the tone of *(y)é̱* in regard to the following tones is systematic in Yes-No question, and its presence can be questioned in some cases. Consider for instance the realization of (18) in Figure 10.

(18) [(*yé̱*)$_\varphi$ (*ha-wo̱nó* *pá̱̋ːhḁ*)$_\varphi$]$_\iota$ [N]
QE 1.PFV-see 5.cat
'Did he see a cat?'

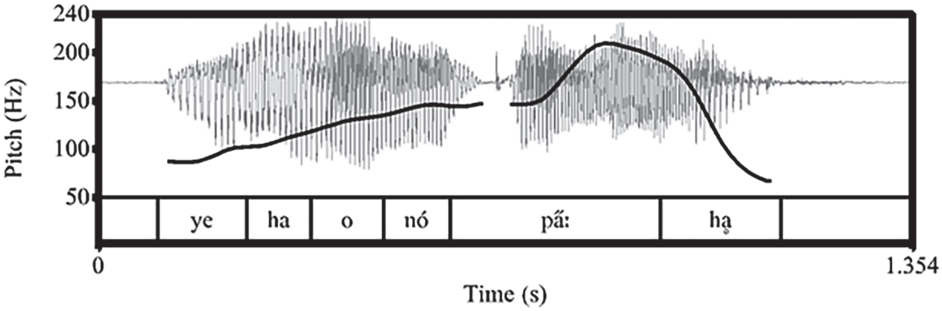


Figure 10 **Signal and F0 curves of (18) as produced by a male speaker from Mbeni.**

Only a slight rise of the F_0 curve on the first syllable may indicate the presence of the tone of the question element in Figure 10. Along with its frequent shift on the following syllable, this prosodic weakness indicates that the question element *(y)é̱* emerges as 'unmarked', not to say 'weak', when it is produced in its canonical initial position.

The question element can also be realized immediately before a wh-word, as in (19b), or a verb, as in (20) – compare with (13). It seems that the word that follows the QE is

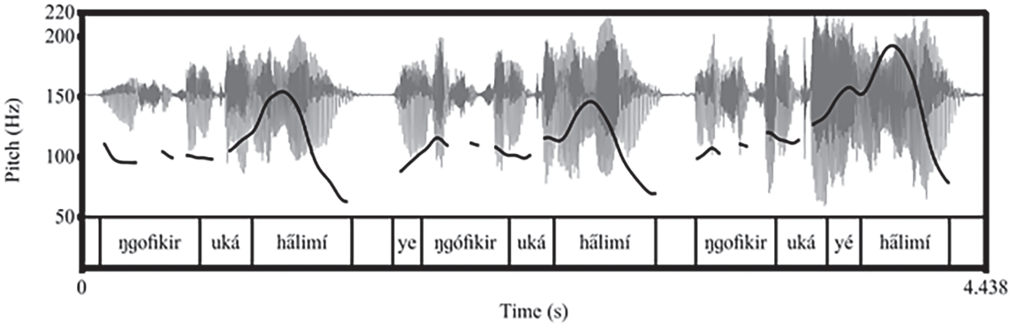


Figure 11 Signal and F_0 curves of (20) – right –, compared to two realizations where *(y)é̱* emerges in initial position (13b) – center – or is absent (13a) – left.

then focalized. However, the element cannot occur sentence-finally (19c) and hardly occurs elsewhere (18d).

(19) a. [(*yé̱*)$_{\varphi}$ (*ya-o̱ná*)$_{\varphi}$ (*e=m-levꜜí̱*)$_{\varphi}$ (*ndo=ꜜβí̱*)$_{\varphi}$]$_{\iota}$ [W]
QE 1.REL.PFV-hit AUG$_1$=1-drunkard who=FOC
‘Who has seen the drunkard?’
b. [(*yao̱ná*)$_{\varphi}$ (*e=mleꜜví*)$_{\varphi}$ (*yé̱*)$_{\varphi}$ (*ndo=βꜜí̱*)$_{\varphi}$]$_{\iota}$
‘Who has seen the drunkard?’
c. *[(*yao̱ná*)$_{\varphi}$ (*e=mleꜜví*)$_{\varphi}$ (*ndo=ꜜβí̱*)$_{\varphi}$ (*yé̱*)$_{\varphi}$]$_{\iota}$
d. $^{??}$[(*yao̱ná*)$_{\varphi}$ (*yé̱*)$_{\varphi}$ (*e=mleꜜví*)$_{\varphi}$ (*ndo=ꜜβí̱*)$_{\varphi}$]$_{\iota}$

(20) [(*ŋgo-fiḳiri* *uká̱*)$_{\varphi}$ (*yé̱*)$_{\varphi}$ (*hã́-liꜜmí̱*)$_{\varphi}$]$_{\iota}$
2SG.IPFV-think.PRS that QE 1.PFV-cultivate
‘Do you think he cultivated (it)?’

Figure 11 compares (20) with two other sentences: *(y)é̱* in initial position or *(y)é̱* preceding the wh-word. When the question element occurs in the non-initial position, F_0 is raised higher than when it is in its canonical position (the initial position). Note that *(y)é̱* in this case is aligned with the peak of intensity.

The Shingazidja question element *(y)é̱* thus differs from its Xitsonga counterpart in that its prosodic behavior varies according to its position. It is weak (or unmarked) in its canonical initial position, but it is highly prominent when it occurs elsewhere. This distribution – unmarked/weak in the initial position, strong elsewhere – strikingly distinguishes *(y)é̱* from the other question elements in the language, the wh-words (see Section 3).

5. Discussion

This paper has introduced two languages Shingazidja and Xitsonga that are neither genetically nor geographically close to each other, but share many prosodic properties.

Both languages have a privative H tone that exhibit long distance phenomena that are constrained by the limits of Phonological phrase boundaries: long-distance H tone spread in the case of Xitsonga, and long-distance H tone shift in Shingazidja. The downstep of successive lexical H tones and non-finality also apply in both the languages, and a L% aligned with the end of the utterance is another shared property.

Building upon these similarities, one may expect that the prosody of the question words or particles in Shingazidja and Xitsonga would be similar in nature, or at least share properties. As for wh-words, this is indeed the case. The default position of wh-words differs in the two languages in that wh-words naturally occur in the final position in Shingazidja, while they remain in situ in Xitsonga. Even so, wh-words behave prosodically similarly, as they do not differ from the other words in the language as far as prosody is involved, except when they are fronted. Fronted wh-words display a F0 boost which signals that they are strengthened. Such a distribution is not uncommon in Bantu: Kula & Hamann (2017: 346) indicate that “*there is some increase in pitch*” when the question word of a constituent question is non-final.

However, the prosodic behavior of the question elements *xana* (Xitsonga) and *(y)é̱* (Shingazidja) clearly differs. While *xana* in Xitsonga is strengthened in the initial position, and thus has a prosodic behavior that is similar in the nature to those of the wh-words in the language, *(y)é̱* in Shingazidja is prosodically weak in its default initial position, and associated with a prominence when it occurs elsewhere. As for the initial position, the reason for this difference may lie in the fact that the question element *(y)é* in Shingazidja is monosyllabic, and thus prosodically weak. As explained in Section 4.3, a postlexical rule typically affects the first syllable in an utterance of Shingazidja, in that a tone that is supposed to be realized on this syllable (if a Phonological phrase boundary follows, or if a lexical tone is associated to the second syllable) is delayed to the following one. The prosodic behavior of the Xitsonga *xana*, which is disyllabic, is expected to be acceptable in the strong initial position.

The motivation for the difference in the prosodic behavior of question elements when they appear elsewhere than in the absolute initial position is less straightforward. The fact that both the question marker and the following lexeme are raised when *(y)é̱* occurs elsewhere as in the initial position in Shingazidja is linked confirms its emphatic role, while so such an effect arises in Xitsonga, where no semantic or pragmatic distinction seems to be associated with any given position where *xana* can emerge.

6. Conclusion

We have shown that *(y)é̱* in Shingazidja has prosodically complementary distribution with wh-words, while *xana* in Xitsonga shares its prosodic distribution with wh-words. We propose that the initial position of a question in these two languages is prosodically prominent (F_0 boost), but this requirement is violable as shown by *(y)é̱* in Shingazidja.

References

Cammenga, Jelle. 2002. *The Phonology and Morphology of Ekegusii, a Bantu Language of Kenya*. Köln: Rudiger Koppe Verlag.

Connell, Bruce. 2005. "Pitch Realization of Questions vs Statements in Mambila." *Proceedings of the 4th World Congress of African Linguistics* (A. Akinlabi ed.), 107–116, Köln: Rüdiger Köppe Verlag.

Connell, Bruce and D. Robert Ladd. 1990. "Aspects of Pitch Realization in Yoruba." *Phonology*, 7(1): 1–29.

Downing, Laura J. 2011. "Wh-questions in Chewa and Tumbuka: Positions and Prosodies." *Questions in Bantu Languages: Prosodies and Positions* (L. J. Downing, ed.), *ZASPiL*, 55: 33–46.

Downing, Laura. J. 2017. "Tone and Intonation in Chichewa and Tumbuka." (L. Downing and A. Rialland, eds.), 365–391.

Downing, Laura J. and Annie Rialland (eds.) 2017a. *Intonation in African Tone Languages*. Berlin/Boston: De Gruyter Mouton.

Downing, Laura J. and Annie Rialland. 2017b. "Introduction." (L. Downing and A. Rialland, eds.), 1–16.

Dryer, Matthew S. 2013. "Position of Polar Question Particles." *WALS Online (v2020.3)* [Data set] (M. S. Dryer & M. Haspelmath, eds.). Zenodo. https://doi.org/10.5281/zenodo.7385533

Elderkin, Edward D. 2003. "Herero." *The Bantu Languages* (D. Nurse and G. Philippson, eds.), 581–608, London: Routledge.

Harjula, Lotta. 2004. *The Ha Language of Tanzania: Grammar, Texts and Vocabulary*. Köln: Köppe.

Harriehausen, Bettina. 1990. *Hmong Njua: Syntaktische Analyse einer gesprochenen Sprache mithilfe datenverarbeitungstechnischer Mittel und sprachvergleichende Beschreibung des südostasiatischen Sprachraumes*. Tübingen: Max Niemeyer Verlag.

Heath, Teresa. 2003. "Makaa (A83)." *The Bantu Languages* (D. Nurse and G. Philippson, eds.), 335-348. London: Routledge.

Hlongwani, Crous M. and Seunghun J. Lee. 2018. "Yes-No Question Intonation in Xitsonga." A Paper Presented at the 20th International Congress of Linguists. Cape Town, South Africa.

Hyman, Larry M. and Francis X. Katamba. 1993. "The Augment in Luganda: Syntax or Pragmatics?" *Theoretical Aspects of Bantu Grammar* (Sam Mchombo, ed.), 209–256, Stanford: C.S.L.I.

Jarlow, Alexander. 2020. "Making Questions with Tone: Polar Question Formation in Kinyarwanda". *Proceedings of the Linguistic Society of America*, 5(1): 213–227.

Jones, Jacquelynn, Japie A. Louw, and Justus C Roux. 2001. "Queclaratives in Xhosa: An Acoustic Analysis." *SAJAL, supplement*, 36: 3–18.

Kahombo, Mateene. 1992. *Essai de grammaire du Kihunde*. Münster: LIT Verlag.

Kisseberth, Charles W. 1984. "Digo Tonology". *Autosegmental Studies in Bantu Tone* (G. N. Clements and J. Goldsmith, eds.), 105–182, Berlin, New York: De Gruyter Mouton.

Kisseberth, Charles W. 2017. "Chimiini Intonation." (L. Downing and A. Rialland, eds.), 225–284.

Kula, Nancy C and Silke Hamann. 2017. "Intonation in Bemba." (L. Downing and A. Rialland, eds.), 323–364.

Lee, Seunghun J. and Elisabeth Selkirk. 2022. "Xitsonga Tone: The Syntax-Phonology Interface." *Prosody and Prosodic Interfaces* (Haruo Kubozono, Junko Ito and Armin Mester, eds.), 337–373, Oxford: Oxford University Press.

Mixdorff, Hansjörg, Lehlohonolo Mohasi, Malillo Machobane, and Thomas Niesler. 2011. "A Study on the Perception of Tone and Intonation in Sesotho." *Proceedings of Interspeech 2011*: 3181–3184. https://pdfs.semanticscholar.org/2264/f8fc5e9d871789e2461a5a7cf3c1e60a09b4.pdf

Myers, Scott P. 1996. "Boundary Tones and the Phonetic Implementation of Tone in Chichewa." *Studies in African Linguistics*, 25(1): 29–60.

O'Connor, Kathleen M. and Cédric Patin. 2015. "The Syntax and Prosody of Apposition in Shingazidja." *Phonology* 32.1 (S. Lee and E. Selkirk, eds.), Special issue 'Constituent Structure in Sentence Phonology': 111–145.

Patin, Cédric. 2010. "The Prosody of Shingazidja Relatives". *ZAS papers in Linguistics* 53 (L. Downing *et al.*, eds.), *Papers from the Workshop on Bantu Relative Clauses*: 187–210.

Patin, Cédric. 2017. "Tone and Intonation in Shingazidja." (L. Downing and A. Rialland, eds.), 285–320.

Rialland, Annie. 2007. “Question Prosody: An African Perspective”. *Tones and Tunes: Studies in Word and Sentence Prosody* (C. Gussenhoven and T. Riad, eds.), 35–62, Berlin: Mouton de Gruyter.

Rialland, Annie. 2009. “The African Lax Question Prosody: Its Realisation and Geographical Distribution.” *Lingua*, 1119: 928–949.

Rialland, Annie and Martial Embanga Aborobongui. 2017. “How Intonations Interact with Tones in Embosi (Bantu C25), a Two-tone Language without Downdrift.” (L. Downing and A. Rialland, eds.), 195–222.

Riedel, Kristina and Cédric Patin. 2011. “Question Structure and Intonation in Fipa.” *ZAS papers in Linguistics* 55 (L. Downing dir.), *Questions in Bantu Languages: Prosodies and Positions*: 141–160.

Robins, R. H. 1958. *The Yurok Language, Grammar, Texts, Lexicon.* Berkeley / Los Angeles: University of California Press.

Schadeberg, Thilo and Francisco U. Mucanheia. 2000. *Ekoti: The Maka or Swahili Language of Angoche*. Köln: Köppe.

Taylor, Charles. 1985. *Nkore-Kiga*. London: Croom Helm.

Traill, Anthony. 1994. *A !Xóõ Dictionary*. Köln: Rüdiger Köppe Verlag.

Unseth, Peter. 1989. Sketch of Majang Syntax. Topics in Nilo-Saharan Linguistics (Bender, M. Lionel, ed.) 97–127. Hamburg: Helmut Buske Verlag.

Zavoni, Ntondo. 2003. Éléments de description du ngangela. PhD Thesis, Université Lumière Lyon 2.

Zerbian, Sabine. 2006. “Questions in Northern Sotho.” *Linguistische Berichte*, 208: 385–405.

Zerbian, Sabine. 2007. “A first approach to information structuring in Xitsonga/Xichangana.” *SOAS Working Papers in Linguistics*, 15: 65–78.

Zerbian, Sabine. 2017. “Sentence intonation in Tswana (Sotho-Tswana Group).” (L. Downing and A. Rialland, eds.), 393–434.

Zerbian, Sabine and Etienne Barnard. 2008. “Phonetics of Intonation in South African Bantu Languages.” *Southern African Linguistics and Applied Language Studies*, 26(2): 235–250.

Journal of Asian and African Studies, Supplement, No. 4, 2024

A Phonetic Typology of Laryngeal Contrasts:

Data from *Illustrations of the IPA*[*]

LEE, Seunghun J.
SUZUKI, Michinori

Phonetic typology of laryngeal contrasts in plosives is explored. A corpus based on recordings of plosives in 103 languages are analyzed based on the number of laryngeal contrast in plosives: languages show a two-way, a three-way, a four-way contrast, or no laryngeal contrast. The recordings come from Illustrations of the International Phonetic Alphabet, of the Journal of International Phonetic Association. Ninety languages display a laryngeal contrast, from which half of the languages showed a two-way contrast. The analysis based on voice onset time (VOT) suggests the following. First, phonetic realizations of plosives have a wider range of VOT values in languages with less laryngeal contrast. Second, more variability of VOT values is observed in the voiced and voiceless aspirated categories than in the voiceless unaspirated category. Third, languages with a four-way laryngeal contrast may utilize phonetic cues other than VOT in the voiced aspirates category.

Keywords: laryngeal contrast, plosives, voice onset time, closure duration, IPA

DOI: https://doi.org/10.15026/0002000635

[*] This research is written as part of the ILCAA joint research project “Phonetic typology from cross-linguistic perspectives (PhonTyp)” (2021–2023). The research has been funded by a Kakenhi (C) Grant 16K02641. This work was also supported by JSPS KAKENHI Promotion of Joint International Research (Fostering Joint International Research (B)) Grant Number 21KK0005, the Kakenhi (B) Grant 23H00622, and Kakenhi 23KJ1921 for JSPS Fellows. Any errors remain our own.

1. Introduction

This paper examines the VOT-based phonetic typology of laryngeal contrast using the data from the Journal of International Phonetic Association (JIPA) illustrations. The illustrations contain a description of phonetic aspects of a language with an emphasis of demonstrating the use of the International Phonetic Alphabet (IPA). As a requirement, illustrations submit sound files accompanying examples in the main text. Most illustrations report phonetics of consonants, vowels, suprasegmentals (stress or tone), syllable structure as well as a translation of the story "The North Wind and the Sun" that is transcribed into IPA. The sound files are downloadable from the JIPA website[1] with a membership, and we utilized this resource to investigate the phonetic typology of plosives from available languages. We zoom into voice onset time (VOT) when exploring the phonetic typology of laryngeal contrasts.

Lisker & Abramson 1964 report that the laryngeal contrast in plosives of 11 languages is best to be explained by VOT. Since then VOT was used as a primary acoustic measurement for studying voicing distinction as well as other laryngeal contrast in plosives in various languages (Abramson & Whalen 2017). However, Cho & Ladefoged 1999, after analyzing data from 18 languages, report that VOT values may show more variation than expected in predicting the laryngeal contrast of a language. Data in their study suggests that phonetic variation due to different places of articulation can be predicted by universal implementation rules, and that the use of VOT in realizing the laryngeal contrast is parametrized in each language (cf. Keating 1984, 1990). Shimizu 1989 also suggests language-specific phonetic variability in the laryngeal categories. Shimizu's conclusion is based on an examination of various acoustic and articulatory cues in five languages: Japanese with a two-way laryngeal contrast, Korean, Burmese and Thai with a three-way laryngeal contrast, and Hindi with a four-way laryngeal contrast. Later on, Shimizu 2018 argues that VOT functions as a primary cue for identifying laryngeal contrasts in the examined languages. In this paper, we analyze our dataset only with VOT echoing Abramson & Whalen 2017, who argue that other phonetic cues for the laryngeal contrasts (F0, amplitude, duration of closure etc.) may add complexity with less benefit for understanding cross-linguistic laryngeal contrasts.

Nasals are phonetically stops that also involve closures by articulators. As such, a full phonetic typology of stops would need to include both oral and nasal stops as suggested by Shinagawa & Komori 2020 in their discussion about voicing contrasts in Niger-Congo languages. We will leave an expansion to include nasal stops to future studies.

In section 2, methods of creating the corpus and annotation scheme are presented. Phonetic typology of laryngeal contrasts by the number of laryngeal contrast shows that languages mainly have two-, three-, or four-way laryngeal contrasts in section 3, where

---

[1] The website URL is https://www.internationalphoneticassociation.org/news/201609/free-illustrations-ipa-online-jipa

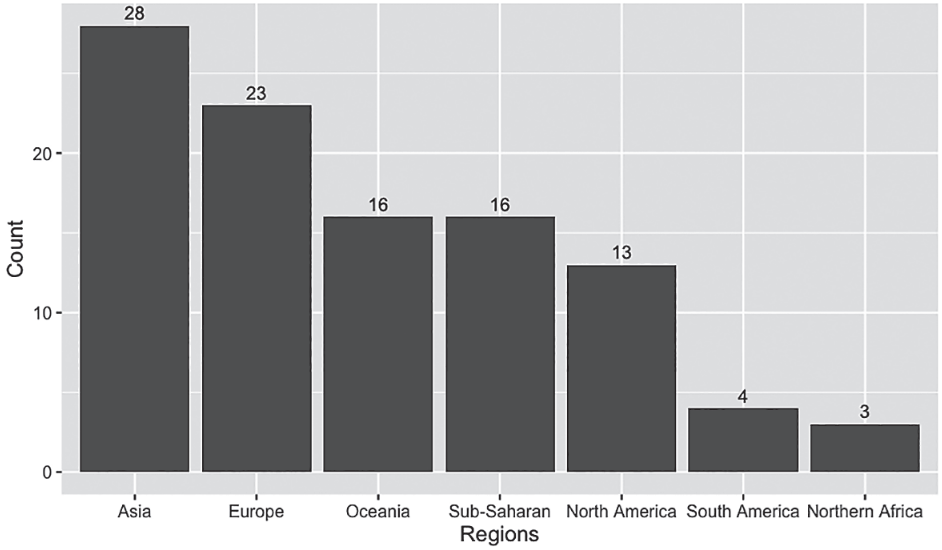


Figure 1 Distribution of languages by geographical regions

we also devote a subsection to the phonetics of plosives in languages with the laryngeal contrast. Section 4 discusses the implications of our findings in section 3.

2. Data and Analyses

The illustrations of the Journal of International Phonetic Association are short articles that share basic phonetic aspects of a language, or a dialect of a language, accompanied by mandatory submission of sound files of words or phrases that appear in an illustration. These sound files are available upon becoming a member of the association, which can then be used for further analysis.

At the time of this study, we identified 119 languages described in the illustrations. Out of these languages, 103 languages had sound files accompanying the illustrations. All of these languages had recordings from a single speaker who produced words that appear in each of the articles. As shown in Figure 1, the languages come from diverse geographical regions, and therefore it is not balanced across language families.

Data points were selected based on the following criteria. First, the type of laryngeal contrast: no contrast, two-way, three-way, or four-way contrast. Second, the place of articulation of the laryngeal contrast, focusing on labials, coronals (combined dentals and alveolars) and velars. Most consonants appear in the word-initial position, and VOT was annotated on a tier in the TextGrid file.

A sample of an annotation is shown in Figure 2. Positive VOT was marked from the burst to the beginning of the vocalic signal, and negative VOT was marked from the beginning of voicing to the burst. Audio and visual cues were used to mark VOT. The number of tokens is 655 (see Appendix 1 for detailed information). The distribution of plosives based on the number of laryngeal contrast is shown in Table 1.

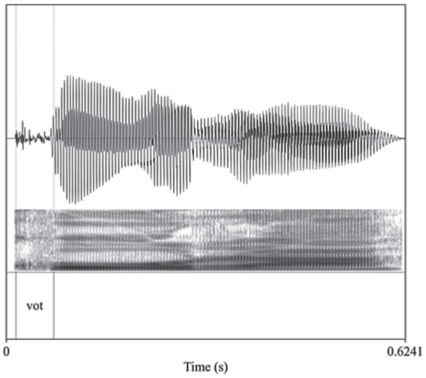


Figure 2a An example of positive VOT in Polish /k/ in *kura* 'hen'.

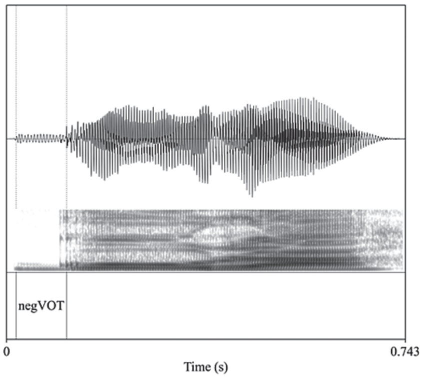


Figure 2b An example of negative VOT in Polish /g/ in *góra* 'mountain'.

Table 1 Distribution of plosives

	Aspirated	Breathy	Voiced	Voiceless
two-way	6	0	170	188
three-way	47	0	51	53
four-way	13	26	16	15
No contrast	2	0	4	36

3. Results

Results of the laryngeal analysis are reported in this section. As languages differ in how they utilize VOT information based on the absence and presence of voicing contrast, this section divides the results into subsections based on the number of laryngeal contrast. Although data from 103 languages are analyzed, the geographical distribution of these languages show a bias in the type of languages; it is weighted more heavily towards Europe and Asia. We first begin with languages that have a two-way laryngeal contrast.

3.1 Languages with a Two-way Voicing Contrast

In our corpus, 66 languages have a two-way voicing contrast; a majority of the languages belong to this group. In 64 languages, there is a contrast between voiced and voiceless plosives (see Appendix 1) based on VOT: negative VOT for voiced plosives and positive VOT for voiceless plosives. In Figure 3, voiceless plosives consistently take positive VOT values, while voiced plosives are realized with both negative and short-positive VOT values. The VOT values between the two categories are significantly different ($t(230) = -19.729$, $p < 0.001$).

Since the voiced category can be realized with both negative and positive VOT, examining the correlation of the VOT values in corresponding voice category is informative. In Figure 4, the voiced-voiceless pair for a place of articulation in a language is plotted to explore whether a VOT value in the voiced sound is associated with a larger VOT value in the corresponding voiceless sound. The correlation is positive ($R = 0.44$, $p < 0.001$)

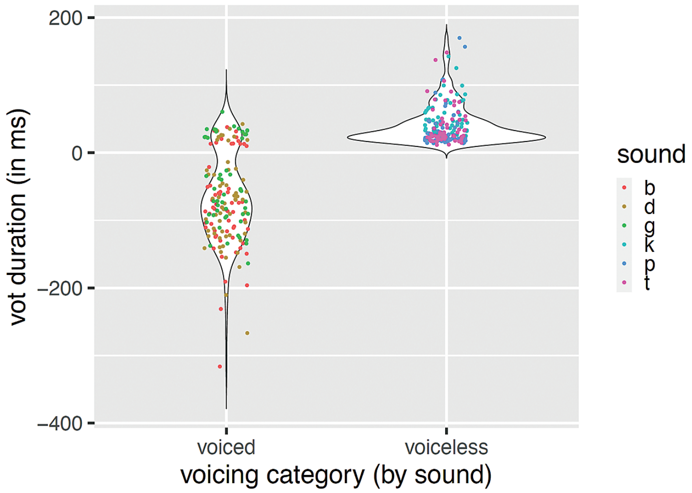


Figure 3 VOT of languages with a two-way voicing contrast by underlying voicing categories (voiced vs. voiceless). Each phoneme is color-coded. The y-axis indicates VOT duration (ms).

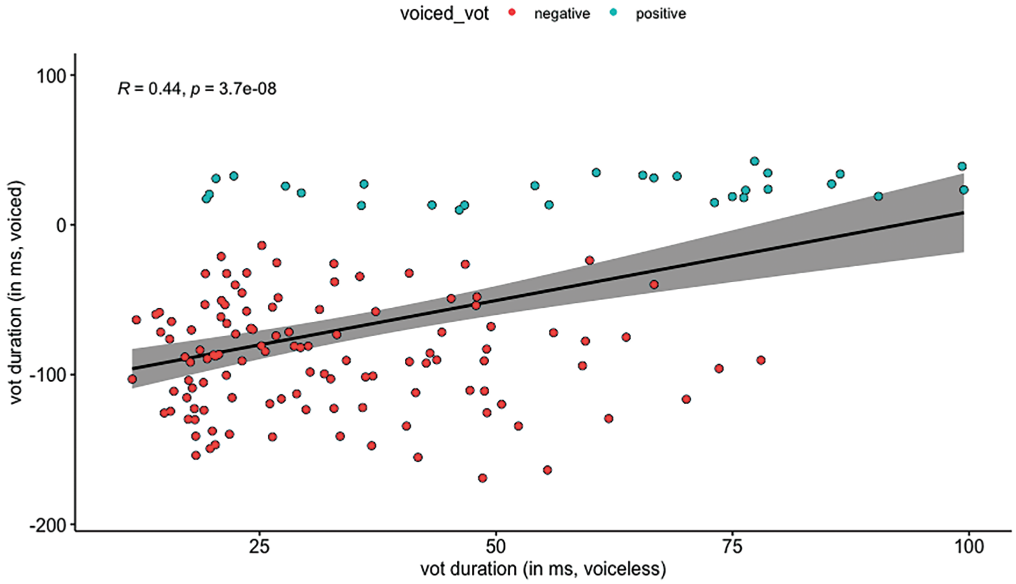


Figure 4 The correlation of VOT in languages with a two-way laryngeal contrast. The x-axis indicates voiceless plosives and the y-axis indicates voiced plosives. Each dot denotes a voiced-voiceless pair with the same place of articulation in a language.

suggesting that a language that has a short VOT value for voiceless plosives tends to have negative VOT for voiced plosives. When a language has long VOT in voiceless plosives, the voiced category is more likely to be realized with a positive VOT.

Two languages, Hakka Chinese and Standard Chinese, were explicitly described as languages that have an aspiration-based two-way laryngeal contrast. These two languages display a contrast between unaspirated and aspirated plosives, which is distinguished by

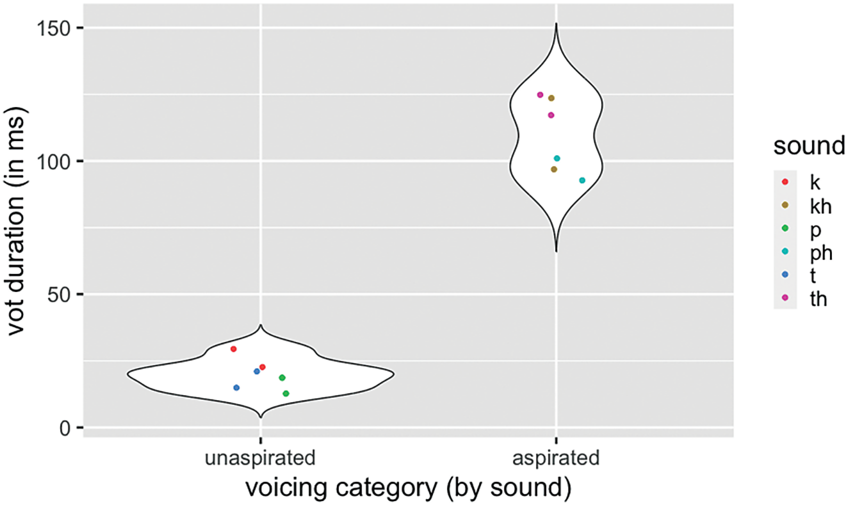


Figure 5 VOT of languages with a two-way voicing contrast by aspiration categories (unaspirated vs. aspirated). Each phoneme is color-coded.

short-positive VOT for voiceless unaspirated and long positive VOT for aspirated. All plosives take a positive VOT, but unaspirated categories have a relatively short VOT (less than 50 ms) as shown in Figure 5. It is possible that some languages plotted in Figure 4 may be classified as languages with an aspiration-based contrast because both categories have a positive VOT, but we leave the exploration of that possibility to a future study.

No language in our corpus displays a contrast by maximizing the difference in VOT: negative VOT for voiced plosives and long positive VOT for voiceless plosives. In dispersion theory (Flemming 2017), it is argued that languages aim to create contrast that maximally use an acoustic context. For example, the number of vowels in a language correlates with the amount of acoustic space that can be used by a single vowel. A language with three vowels has a more dispersed vowel space than languages with five vowels. If dispersion applies to a two-way laryngeal contrast, a language that utilizes negative VOT and long positive VOT would be expected, but our data suggests that it was not the case.

Our corpus has only two languages that have an aspiration-based voicing contrast. This number does not reflect any tendency in phonetic typology, but rather it comes from the nature of the sample. Sinitic languages are underrepresented in our corpus, and if there were more languages in this group, we expect an increase of languages with aspiration-based contrast.

3.2 Languages with a Three-way Contrast

Nineteen languages in our corpus display a three-way contrast: voiced, voiceless, and aspirated. Voiced plosives have negative VOT, voiceless unaspirated plosives have a short positive VOT, and aspirated plosives have a long positive VOT. Figure 6 shows the distribution of these three laryngeal categories.

When the data points are divided by place of articulation as in Figure 7, VOT patterns

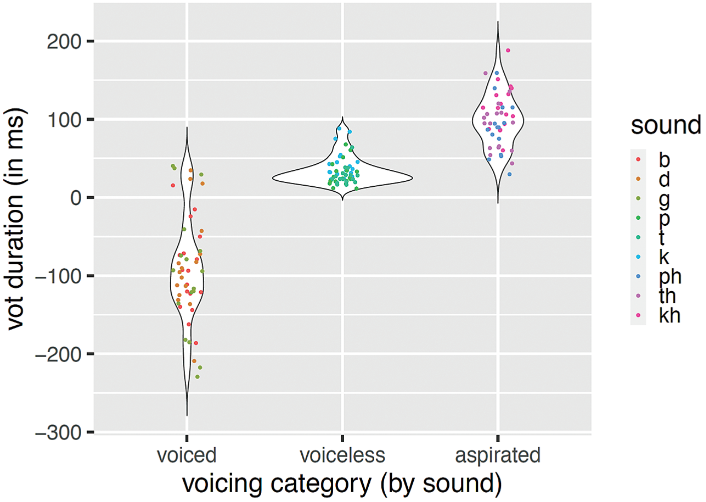


Figure 6 VOT of languages with a three-way voicing contrast by underlying voicing categories (voiced vs. voiceless vs. aspirated). Each phoneme is color-coded.

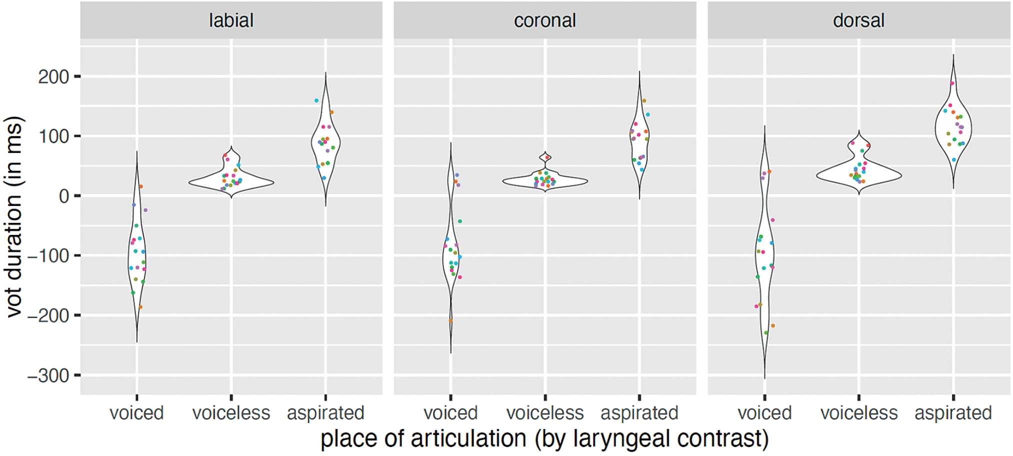


Figure 7 VOT of languages with a three-way laryngeal contrast by underlying voicing categories (voiced vs. voiceless vs. aspirated), by places of articulation. Each language is color-coded.

of the three laryngeal categories remain the same. A linear mixed model analysis was performed with laryngeal category as a fixed effect, and place of articulation and language as random effects. With VOT as the dependent variable, the model that fits the most shows both the voiceless category and the aspirated category have significantly longer VOT than the voiced category ($p < 0.001$). A separate t-test showed that the VOT difference between voiceless and aspirated is also significant ($t(66.77) = -11.73$, $p < 0.001$).

The VOT values of voiced and unaspirated voiceless categories as well as the VOT values of unaspirated and aspirated voiceless categories overlap. In Figure 8, we plotted the VOT values of each triplet in a given language by place of articulation. The data demonstrates that most triplets distinguish the laryngeal contrast by the VOT values: the voiced plosive has negative VOT, the unaspirated voiceless plosive shows short positive

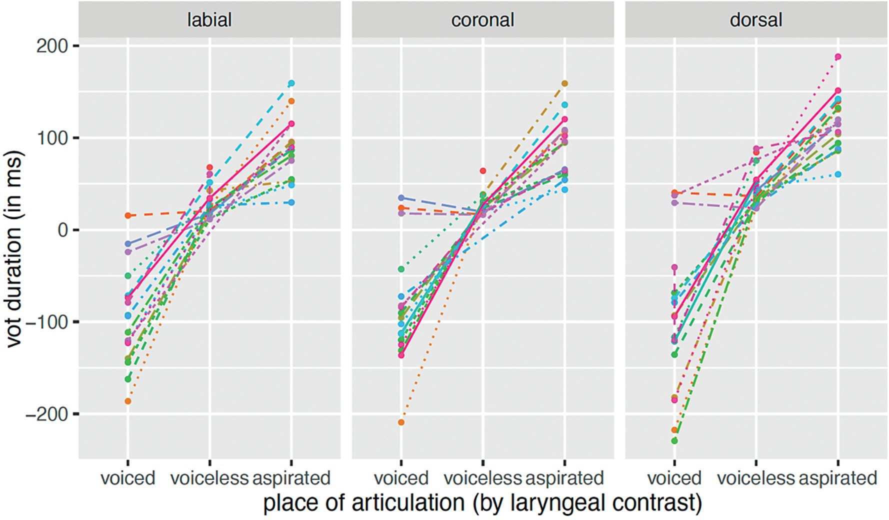


Figure 8 VOT of languages with a three-way laryngeal contrast by underlying voicing categories (voiced vs. voiceless vs. aspirated), by places of articulation. Each dotted line indicates a language.

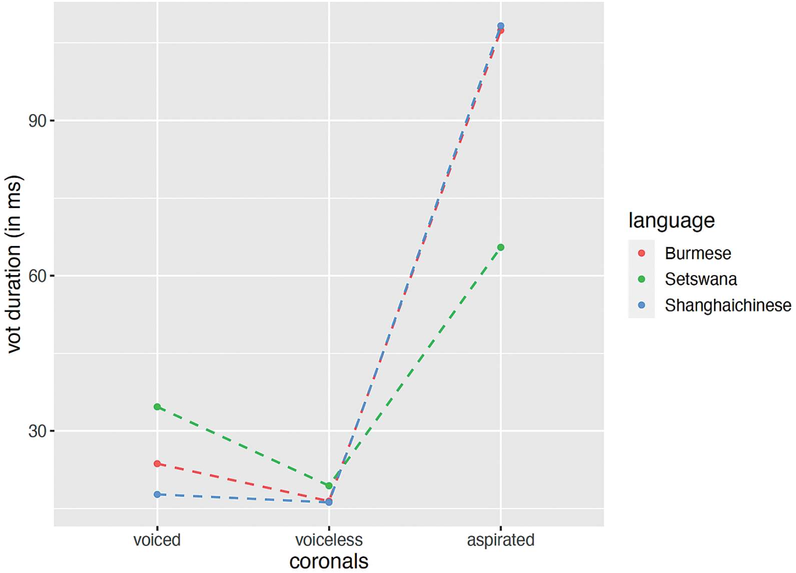


Figure 9 A part of Figure 8 focusing on the three languages with an exceptional pattern regarding VOT. Each dotted line indicates a language.

VOT, and the aspirated plosive has long positive VOT.

From Figure 8, languages that utilize short positive VOT in both voiced and voiceless unaspirated categories are separated in Figure 9 (coronal data only). The three languages, Burmese, Setswana and Shanghai Chinese, suggest that VOT may not be a salient cue for distinguishing the two categories.[2]

---

[2] Other possible phonetic cues may be fundamental frequency (F0) or phonation of the following vowel, or closure duration. Examining those cues are beyond the scope of this article, and we will leave it for a future study.

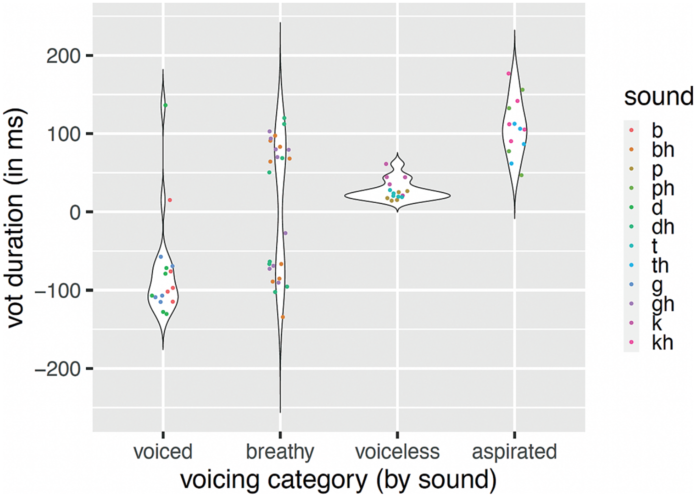


Figure 10 **VOT of languages with a four-way laryngeal contrast by underlying voicing categories (voiced vs. breathy vs. voiceless vs. aspirated). Each phoneme is color-coded.**

3.3 Languages with a Four-way Contrast

The corpus includes five languages with a four-way laryngeal contrast: Assamese, Bengali, Malayalam, Nepali, and Telugu. In addition to the voiced, unaspirated and aspirated voiceless plosives, these languages have a fourth laryngeal category. This fourth category is usually characterized as voiced aspirates (Schwarz *et al.* 2019), but it is also accompanied by breathy voice in the following vowel (for an example of a phonetic study on Drenjongke, see Lee *et al.*, 2019).

The VOT of the four categories are illustrated in Figure 10. While VOT distinguishes three out of four categories (voiced, voiceless, aspirated). The voiced aspirated (or breathy) category shows a wide range of VOT values that are distributed from negative to positive. As discussed in Schwarz *et al.* 2019 and Lee *et al.* 2019, the results in Figure 10 show that VOT is not a salient cue for identifying voiced aspirates.

Among the breathy category, the breathy dorsal /g^h/ shows different phonetic quality between languages. The dorsal /g^h/ in Nepali and Bengali has breathy phonation, while that in Assamese, Malayalam, Telugu is voiceless and takes positive VOT.

Again, relationships in VOT among the four categories are examined in Figure 11 to examine the overlap between voiced and breathy categories. Figure 11 indicates that the VOT values of breathy category overlap with either voiced or aspirated categories in each language, suggesting that this category cannot be distinguished solely by using VOT. In Drenjongke, for example, this category is always followed by a vowel that begins with low f0 (Kunzang Namgyal *et al.* 2020). Although adding too many phonetic cues may add complexity, the necessity of other acoustic cues to distinguish this category is also pointed out by Abramson & Whalen 2017.

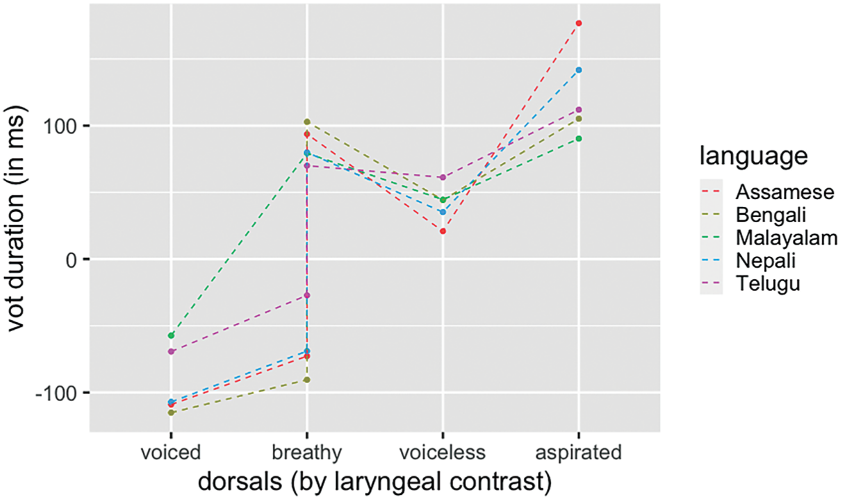


Figure 11 VOT of languages with a four-way laryngeal contrast by underlying voicing categories (voiced vs. voiceless vs. aspirated), by places of articulation. Each dotted line indicates a language.

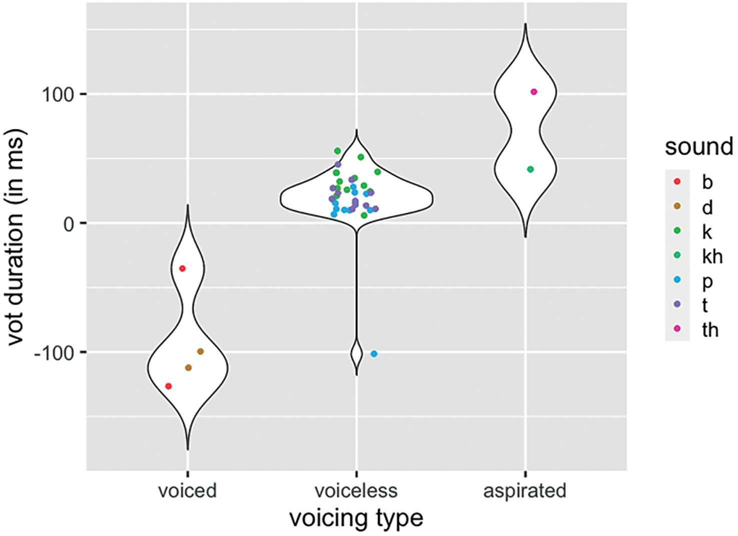


Figure 12 VOT of languages with no voicing contrast by underlying voicing categories. Each phoneme is color-coded. The y-axis indicates VOT duration (ms).

3.4 Language with No Voicing Contrast

Out of the 103 languages in the corpus, thirteen languages have no voicing contrast. However, the plosives that they have do not behave as a single group in terms of VOT. The languages generally use voiceless unaspirated, but some exceptions are realized as either voiced with a negative VOT value, or voiceless aspirated with a long positive VOT. Figure 12 displays VOT durations of these thirteen languages.

Although most of the voiceless unaspirated plosives take short-lag VOT, Bardi has an unaspirated plosive with a negative VOT, as shown in Figure 13. It should be noted,

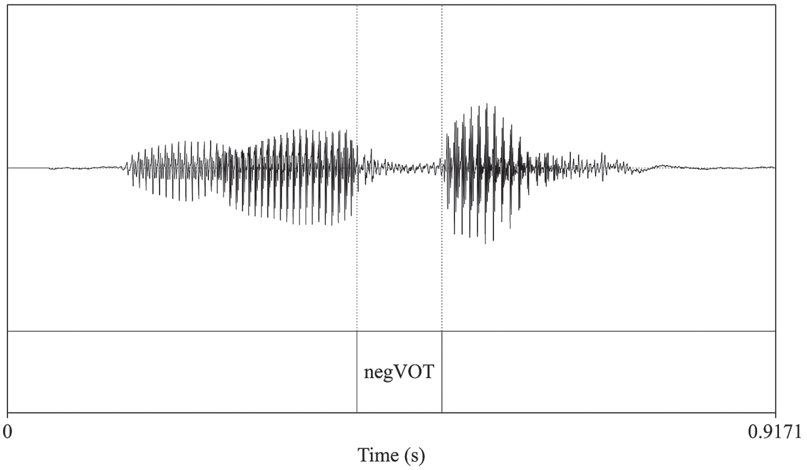


Figure 13 The outlier in the voiceless category Bardi (token has [t] in the word-medial position)

Table 2 Unmarked and marked patterns of languages with only one voicing categories. The underlines indicate exceptional cases: voiced or aspirated.

a. Unmarked pattern		b. Marked pattern	
Bardi	p, t, k	Ibibio	$\underline{b}$, $\underline{d}$, k
Central Arrernte	p, t, k	Vietnamese	$\underline{b}$, $\underline{d}$, k, t, $\underline{t^h}$
Estonian	p, t, k	Upper Saxon	p, t, k, $\underline{k^h}$
Itunyoso Trique[3]	p, t, k		
Mapudungun	p, t, k		
Mavea	p, t, k		
Pitjantjatjara	p, t, k		
Seri	p, t, k		
Shiwilu	p, t, k		
Yine	p, t, k		

however, that this token is in an intervocalic position and the negative VOT could be an example of passive voicing.

Table 2 lists the plosive inventory of languages with no voicing contrast; most of them have a voiceless unaspirated plosive. This distribution implies that voiceless unaspirated /p, t, k/ is the unmarked plosive (as many theories of markedness assume, Vaux and Samuels 2005, Brown 2016), whereas voiced /b, d, g/ and voiceless aspirated /p^h, t^h, k^h/ are marked plosives. Languages with the marked patterns display gaps in the laryngeal contrast. In Ibibio, labial and coronal plosives are voiced, but the velar plosive is voiceless. In Vietnamese, coronals have a three-way laryngeal contrast, but other places do not show laryngeal contrast: the labial is voiced and the velar is voiceless. In Upper Saxon, only the velar has a two-way laryngeal contrast between unaspirated and aspirated (akin to Chinese section 3.1), but the labial or the alveolar is not aspirated.

---

[3] Itunyoso Trique also has prenasalized plosives.

3.5 Summary

This section has reported cross-linguistic patterns of how laryngeal contrast is phonetically distributed. Following the suggestion by Abramson & Whalen 2017, our cross-linguistic analysis on the systems of laryngeal contrast is based on VOT. Though the results show sampling bias because of overrepresentation of certain language families, we have shown that the VOT-based phonetic typology is informative in recognizing phonetic aspects of the laryngeal typology. In the next section, three points of the results are discussed further.

4. Discussion

4.1 Phonetic Realizations in Laryngeal-contrast Systems

Phonetic variation found in individual languages is argued to be the nature of the phonetic realizations (Cho & Ladefoged 1999). We found such variation in our corpus, but we also observe that the variation is constrained by the type of laryngeal contrast. In languages with a two-way contrast and also languages with no voicing contrast, VOT values varied greater compared to languages with more laryngeal contrasts. The number of contrast and the degree of phonetic variability could be an instance of the use of acoustic dimension in a manner proposed in the dispersion theory. Languages with no voicing contrast have the entire VOT continuum available for the realization of plosives. Similarly, languages with a two-way laryngeal contrast can utilize various dimensions of the VOT continuum.

The relatively free availability of the VOT continuum appears to be constrained in the two-way laryngeal contrast system. As discussed in section 3.1, the voiceless category displays a more restrictive use of VOT compared to the voiced category. We interpret this pattern resulting from a constraint that prefers the inclusion of unmarked category whose phonetic target is the unaspirated voiceless plosive. Once a language has a phonetic target based on VOT of unaspirated voiceless plosives, the other category can utilize the complementary part of the VOT continuum. Languages that have voiced category (negative VOT) are at the liberty of utilizing negative to positive VOT for the category as long as it is different from the unaspirated voiceless category of that language. Languages with an aspiration contrast have the unaspirated voiceless category with a short positive VOT; the aspirated plosive in those languages is realized with a long VOT (more than 50 ms).

The use of VOT continuum becomes restricted in the voiceless categories in languages with a three-way laryngeal contrast. The voiced category in these languages utilize negative VOT, and the voiceless unaspirated plosives have positive VOT between 0 and 50 ms, but never negative VOT. Aspirated plosives in general have longer than 50 ms VOT, which sometimes can be as long as 200 ms. Our observation is that voiceless unaspirated plosives have the most constrained VOT range out of the three laryngeal categories. Shimizu 1989 states in his conclusion that when the glottal and supralaryngeal timing events are considered, variability in the laryngeal categories is mostly found in voiceless unaspirated stops, but not in aspirated plosives. The findings in this paper are still in line with what is reported (Shimizu 1989), because the VOT distribution in the five languages reported in Shimizu 1989:3 shows similarities between the two studies. For example, Shimizu presents that a language with a two-way contrast (Japanese) shows a wider range of VOT for voiceless unaspirated (15–100 ms), while languages with a three-way contrast (Burmese and Thai) demonstrate narrower ranges (0–35 ms or 5–30 ms). These reports agree with our observations where the use of VOT continuum is more restricted in a three-way contrast than in a two-way contrast.

An additional laryngeal category, voiced aspirates, demonstrates that VOT itself may not be enough for fully exploring phonetic typology of laryngeal systems when a large number of contrasts are involved. The three other categories (voiced, voiceless unaspirated, voiceless aspirated) show VOT patterns akin to languages with the three-way laryngeal contrast. The voiced aspirates, however, may utilize the entire VOT continuum: some languages with negative VOT, and other languages with a positive VOT (Lee *et al.* 2019, Schwarz *et al.* 2019).

It would be useful to think about the VOT distribution from the dispersion theory. VOT values exhibit more dispersed values between laryngeal categories with less variation within the categories when the number of laryngeal contrast is higher. Since no known languages have a five-way or six-way laryngeal contrast in plosives, testing the dispersion theory in the VOT continuum is challenging. Even so, we find an interesting instance of the dispersion principle applied to achieving contrast in laryngeal activities. Languages with a two-way contrast in our corpus do not show a contrast, in which one category is realized with negative VOT, and the other category with a long positive VOT (i.e. aspiration). The gap or absence of this pattern seems to be counter to the dispersion theory.

We suggest that the dispersion theory that accounts for the distribution of vowels within the vowel space (Liljencrants & Lindblom 1972, Fleming 2017) may not directly extend to an explanation of the distribution of laryngeal contrast. Markedness may also be at work. Articulatorily, voicing that requires vocal fold vibration or aspiration that requires lag of voicing are produced with more effort compared with unmarked voiceless unaspirated plosives. When we couple the dispersion theory with the theory of markedness, languages with a laryngeal contrast can be argued to have an unmarked laryngeal category (i.e., voiceless unaspirated) and marked laryngeal categories. In a language with a two-way contrast, a language would have a marked category with negative VOT or a marked category with a long positive VOT. No language may show contrasts by employing two marked categories, even though such a language satisfies the tenets of the dispersion theory better. In our corpus of 103 languages, the observed patterns imply that the articulatory constraints to minimize efforts cross-linguistically outrank dispersion constraints that prefer more distinct contrast to less distinct ones.

4.2 Variability in VOT Values

What drew our attention is the variability of VOT values. It is well known that languages differ in their organization of VOT for the voicing category (Lisker & Abramson

1964, Zsiga 2012). In English, voiceless plosives have a long positive VOT (over 50 ms), and voiced plosives have short positive VOT or negative VOT. In Spanish, however, voiceless plosives have a short VOT, while voiced plosives have negative VOT. In Thai, voiced plosives have a negative VOT, voiceless unaspirated plosives have a short positive VOT, and aspirated plosives have a long positive VOT. These examples demonstrate that phonetic realization of an acoustic measure is language specific.

The degree of variability of VOT in each laryngeal category is not identical. Across the examined languages with three-way or four-way contrasts, the VOT range of voiceless unaspirated plosive is constrained between 0 and 50 ms. This relatively narrow range of VOT in this plosive stands in contrast with the other types of plosives that take a wider range of VOT values. This VOT distribution may suggest that this category serves as an anchor point (i.e., unmarked realization of a plosive) for other laryngealized plosives that deviate from the point. The data points show languages are content with longer positive VOT for aspirated plosives, and also longer negative VOT for voiced plosives.

4.3 Acoustic Cues beyond VOT

By examining a relatively large group of languages, this paper has shown that VOT is useful in understanding the phonetic typology of laryngeal contrasts if a language has up to three types of plosives. No language in our corpus exhibited a five-way laryngeal contrast, but once a language has a four-way laryngeal contrast, phonetic cue(s) other than VOT would be needed to fully contrast the plosives. The fourth category in these languages is the voiced aspirates, and this section reviews some of the studies that discuss details about them.

In Drenjongke (Lee *et al.* 2019), these voiced aspirates display variable VOT (from negative to positive) in all 12 speakers who took part in the study. What differed in Drenjongke voiced aspirates was the consistently low F0 in the vowel immediately following voiced aspirates, and high F1 in the beginning of the vowel, presumably due to more jaw opening at the beginning of the vowel.

Faytak & Steffman 2023 offer a useful overview about the phonetics of voiced aspirates. In Clements & Khatiwada 2007, VOT was not a sufficient indicator but a period of breathy phonation after the consonant with secondary cue of lowered f0 was. Based on the laryngeal contrast observed in Nepali spoken in Sikkim, India, Schwarz *et al.* 2019 report that VOT of voiced aspirates mainly overlap with voiced plosives, but prevoicing of voiced aspirates is long as in voiced plosives, and post-vocalic interval duration in voiced aspirates is longer than voiced plosives; further suggesting that VOT may not be sufficient to distinguish voiced aspirates from other plosives.

The few languages with a four-way laryngeal contrast in our corpus come from the Indian subcontinent, and thus show similarly distributed contrast: voiced, voiceless unaspirated, voiceless aspirated, and voiced aspirates. Different types of four-way laryngeal contrast system is also found in other languages. For example, in Xitsonga, a southern Bantu language, the four-way laryngeal contrast consists of the following four plosives:

unaspirated [p], aspirated [pʰ], implosive [ɓ], prenasalized [ᵐb].

Non-VOT cues are also utilized even in languages with less than a four-way laryngeal contrast. In Tokyo Japanese (Gao & Arai 2019), which has a two-way voicing contrast, a combination of VOT and the onset f0 of the following vowel are used to enhance the voicing contrast in the word-initial position when the effect of VOT is insufficient, but not in word-medial position. Korean three-way laryngeal contrast is also distinguished by VOT and f0 (Shimizu 1989).

Languages that contain non-pulmonic sounds in the realization of the plosives need to undergo closer examination. Voiced plosives may be realized as implosives or it may be prenasalized. Voiceless unaspirated stops may be ejectives. Comparing the number of place contrasts in oral plosives and nasal plosives can additionally elucidate our understanding of laryngeal contrast in the context of phonetic typology.

5. Conclusion

This paper has explored the phonetic typology of laryngeal contrast focusing on 103 languages of the JIPA illustrations. Cross-linguistic data indicate that the use of VOT continuum has different patterns based on the number of laryngeal categories. Comparing languages with a two-way and ones with a three-way laryngeal contrast, voiceless unaspirated plosives in languages with a two-way contrast show wider ranges of VOT than in languages with a three-way contrast. Simultaneously, both types of languages display a wide range of VOT for voiced unaspirated plosives. However, in languages with a four-way laryngeal contrast, VOT ranges for the voiced aspirated plosives show either large positive or large negative values. These VOT distributions of voiced aspirated plosives are overlapping with ones of voiceless aspirated or voiced unaspirated plosives. As the VOT continuum can be divided into three regions (voicing lead, short-lag, and long-lag), languages with more than three-way laryngeal contrasts might utilize non-VOT cues to differentiate the laryngeal categories. This paper also investigates VOT of languages with no laryngeal contrasts and demonstrates that most of the languages produce voiceless unaspirated plosives /p, t, k/ with short-lag VOT. This observation implies that voiceless unaspirated is the cross-linguistically unmarked category.

Appendix 1 The list of languages in our corpus. Type column is based on the JIPA illustration.

No	Language	Plosive	Type
L001	Bardi	p, t, k	no contrast
L002	Central Arrernte	p, t, k	no contrast
L003	Estonian	p, t, k	no contrast
L004	Ibibio	b, t̪, d, k	no contrast
L005	Itunyoso Trique	p, t, k, nd, ng	no contrast
L006	Mapudungun	p, t, k	no contrast
L007	Mavea	p, t, k	no contrast

L008	Pitjantjatjara	p, t, k	no contrast
L009	Seri	p, t, k	no contrast
L010	Shiwilu	p, t, k	no contrast
L011	Upper Saxon	p, t, k, kh	no contrast
L012	Vietnamese	t, th, k	no contrast
L013	Yine	p, t, k	no contrast
L014	Amarasi	p, b, t, k	two-way
L015	Azerbaijani	p, b, t, d, k, g	two-way
L016	Basaá	p, mb, t, nd, k, ŋg	two-way
L017	Bearnais	p, b, t, d, k, g	two-way
L018	Belgian Standard Dutch	p, b, t, d, k	two-way
L019	Brazilian Portugues	p, b, t, d, k, g	two-way
L020	Brunei Malay	p, b, t, d, k, g	two-way
L021	Bukharan Tajik	p, b, t, d, k, g, q	two-way
L022	Castillain Spanish	p, b, t, d, k, g	two-way
L023	Central Sama	p, b, t̪, d, k, g	two-way
L024	Chicahuaxtla Triqui	p, b, t̪, d, k, g	two-way
L025	Chickasaw	p, b, t, k	two-way
L026	Chistabino	p, b, t, d, k, g	two-way
L027	Cicipu	p, b, t, d, k, g	two-way
L028	Cocos Malay	p, b, t̪, d, k, g	two-way
L029	Czech	p, b, t, d, k, g	two-way
L030	Dari	p, b, t, d, k, g, q	two-way
L031	Ega	p, b, t, d, k, g	two-way
L032	Friulian	p, b, t, d, k, g	two-way
L033	Gayo	p, b, t, d, k, g	two-way
L034	Gitksan	p, p’, t, t’, kw, k’w, q, q’, b, d, gw	two-way
L035	Hakka Chinese	p, ph, t, th, k, kh	two-way
L036	Ika Igbo	p, b, t, d, k, g	two-way
L037	Indonesian	p, b, t̪, d, k, g	two-way
L038	Isthmus Zapotec	p, b, t, d, k, g	two-way
L039	Italian	p, b, t, d, k, g	two-way
L040	JamaicanCreole	p, b, t, d, k, g	two-way
L041	Jicarilla Apache	p, th, t, d, k, kh	two-way
L042	Kabiye	p, b, t, d, k, g	two-way
L043	Kalabari Ijo	p, b, t, d, k, g	two-way
L044	Kedayan	p, b, t̪, d, k, g	two-way
L045	Kera	p, b, t, d, k, g	two-way
L046	Kumzari	p, b, t, d, k, g, q	two-way
L047	Lusoga	p, b, t, d, k, g, mp, mb, nt, nd, nk, ng	two-way
L048	Luxembourgish	p, b, t, d, k, g	two-way
L049	Lyonnais	p, b, t, d, k, g	two-way
L050	Makasar	p, b, t, d, k, g	two-way
L051	Mambay	p, b, t, d, k, g	two-way

L052	Mennonite Plautdietsch	p, b, t, d, k, g	two-way
L053	Mono	p, b, t, d, k, g	two-way
L054	Munji	p, b, t, d, k, g	two-way
L055	Murcian Spanish	p, b, t, d, k, g	two-way
L056	NewZealandEnglish	p, b, t, d, k, g	two-way
L057	Northwest Sahaptin	p, p’, t, t’, k, k’	two-way
L058	Nuuchahnulth	p, p’, t, t’, k, k’	two-way
L059	Palula	p, b, t, d, k, g	two-way
L060	Polish	p, b, t, d, k, g	two-way
L061	Russian	p, b, t, d, k, g	two-way
L062	Slovak	p, b, t, d, k, g	two-way
L063	Southeastern Pashayi	p, b, t, d, k, g	two-way
L064	Standard Austrian German	p, b, t, d, k, g	two-way
L065	Standard Chinese	p, ph, t, th, k, kh	two-way
L066	Standard Malay	p, b, t, d, k, g	two-way
L067	Tamil	p, b, t, d, k, g, mb, nd, ng	two-way
L068	Tashlhiyt Berber	b, t, d, k, g	two-way
L069	Tausug	p, b, t, d, k, g	two-way
L070	Temne	p, b, t, d, k	two-way
L071	Tena Quichua	p, b*, t, d*, k, g*	two-way
L072	Tera	p, b, t, d, k, g, mb, nd, ng	two-way
L073	Hasselt	p, b, t, d, k	two-way
L074	Tilquiapan Zapotec	p, b, t, d, k, g	two-way
L075	Tyneside English	p, b, t, d, k, g	two-way
L076	Ukranian	p, b, t, d, k, g	two-way
L077	Upper Sorbian	p, b, t, d, k, g	two-way
L078	Zurich German	p, b, t, d, k, g	two-way
L079	English (British RP, Australian, South Michigan)	p, b, t, d, k, g	two-way
L080	Bemba	p, mp, mb, t, nt, nd, k, ŋk, ŋg	three-way
L081	Burmese	p, ph, b, t, th, d, k, kh, g	three-way
L082	Ersu	p, ph, b, t, th, d, k, kh, g	three-way
L083	Goemai	p, ph, b, t, th, d, k, kh, g	three-way
L084	Greek Thrace Xoraxane Romane	p, ph, b, t, th, d, k, kh, g	three-way
L085	Khowar	p, ph, b, t, th, d, k, kh, g	three-way
L086	Lizu	p, ph, b, t, th, d, k, kh, g	three-way
L087	Mah Meri	p, ph, b, t, th, d, k, kh, g	three-way
L088	Mono Lake Northern Paiute	p, b, bb, t, d, dd, k, g, gg	three-way
L089	Nen	p, b, mb, t, d, nd, k, g, ng	three-way
L090	Nuosu Yi	p, ph, b, mb, t, th, d, nd, k, kh, g, ng	three-way
L091	Salasaca Quichua	p, ph, b, t, th, d, k, kh, g	three-way
L092	Sandawe	p, ph, b, t, th, d, k, kh, g	three-way
L093	Setswana	p, ph, b, t, th, d, k, kh	three-way

L094	Shanghai Chinese	p, ph, b, t, th, d, k, kh, g	three-way
L095	Standard Georgian	ph, p’, b, th, t’, d, kh, k’, g	three-way
L096	Sumi	p, ph, b, t, th, d, k, kh, g	three-way
L097	Lower Xumi	p, ph, b, t, th, d, k, kh, g	three-way
L098	Upper Xumi	p, ph, b, t, th, d, k, kh, g	three-way
L099	Assamese	p, ph, b, bh, t, th, d, dh, k, kh, g, gh	four-way
L100	Bengali	p, b, bh, t, th, d, dh, k, kh, g, gh	four-way
L101	Malayalam	p, ph, b, bh, t, th, d, dh, k, kh, g, gh	four-way
L102	Nepali	p, ph, b, bh, t, th, d, dh, k, kh, g, gh	four-way
L103	Telugu	p, ph, b, mb, t, th, d, nd, k, kh, g, ng	four-way

References

Abramson, Arthur and Douglas Whalen. 2017. “Voice Onset Time (VOT) at 50: Theoretical and Practical Issues in Measuring Voicing Distinctions.” *Journal of Phonetics*, 63: 75–86.

Brown, Jason. 2016. “Laryngeal Assimilation, Markedness and Typology.” *Phonology*, 33(3): 393–423.

Cho, Taehong and Peter Ladefoged. 1999. “Variation and Universals in VOT: Evidence from 18 Languages.” *Journal of Phonetics*, 27(2): 207–229.

Clements, George N. and Rajesh Khatiwada. 2007. “Phonetic Realization of Contrastively Aspirated Affricates in Nepali.” *Proceedings of the 16th International Congress of Phonetic Sciences* (Jürgen Trouvain & William John Barry, eds.), 629–632, Saarbrücken: Universität des Saarlandes.

Faytak, Matt, and Jeremy Steffman. 2023. “Voiced Aspirates with Mixed Voicing in Yemba, a Grassfields Bantu Language of Cameroon.” *Journal of the International Phonetic Association*, 1–38.

Flemming, Edward. 2017. *Dispersion Theory and Phonology*. Oxford Research Encyclopedia of Linguistics. Oxfordre.com.

Gao, Jiayin, and Takayuki Arai. 2019. “Plosive (de-)voicing and f0 Perturbations in Tokyo Japanese: Positional Variation, Cue Enhancement, and Contrast Recovery.” *Journal of Phonetics*, 77: 100932.

Keating, Patricia A. 1984. “Phonetic and Phonological Representation of Stop Consonant voicing.” *Language*, 60: 286–319.

Keating, Patricia A. 1990. “Phonetic Representations in a Generative Grammar.” *Journal of Phonetics*, 18: 321–334.

Kunzang Namgyal *et al.* 2020. “A Phonetic Analysis of Drenjongke: A First Critical Assessment.” *Indian Linguistics*, 81(1-2): 1–14.

Liljencrants, Johan and Bjorn Lindblom. 1972. “Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast.” *Language*, 48(4): 839–862.

Lisker, Leigh and Arthur S. Abramson. 1964. “A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements.” *Word*, 20: 384–422.

Lee, Seunghun J., Shigeto Kawaharam, Céleste Guillemot, and Tomoko Monou. 2019. “Acoustics of the Four-way Laryngeal Contrast in Drenjongke (Bhutia): Observations and Implications.” *Journal of Phonetic Society of Japan*, 23: 65–75.

Schwarz, Martha, Morgan Sonderegger, and Heather Goad. 2019. “Realization and Representation of Nepali Laryngeal Contrasts: Voiced Aspirates and Laryngeal Realism.” *Journal of Phonetics*, 73: 113–127.

Shinagawa, Daisuke and Junko Komori. 2020. “Stop Series in Niger-Congo.” Presentation at the First Meeting, ILCAA Joint Research Project on ‘*Studies in Asian and African Geolinguistics*’ (2020–2022).

Shimizu, Katsumasa 1989. “A Cross-Language Study of Voicing Contrasts of Stops.” *[Studies on Phonetic Science] Onsei Kagaku Kenkyu*, 23: 1–12.

Vaux, Bert and Bridget Samuels. 2005. “Laryngeal Markedness and Aspiration.” *Phonology*, 22(3): 395–436.

Zsiga, Elizabeth C. 2012. *The Sounds of Language: An Introduction to Phonetics and Phonology*. Chichester: John Wiley & Sons.

CONTRIBUTORS

Abe, Yuko
Professor
Center for Studies of Ethnic Groups in Northwest China
Lanzhou University (China)

Aoi, Hayato
Assistant Professor
World Language and Society Education Centre
Tokyo University of Foreign Studies (Japan)

Guillemot, Céleste
Lecturer
World Language and Society Education Centre
Tokyo University of Foreign Studies (Japan)

Kamano, Shigeto
Ph.D. Student
Department of Linguistics, College of Arts and Sciences
Stony Brook University (USA)

Kurabe, Keita
Associate Professor
Research Institute for Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies (Japan)

Lalhminghlui, Wendy
Ph.D. Scholar
Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati (India)

Lee, Seunghun J.
Senior Associate Professor of Linguistics
College of Liberal Arts
International Christian University (Japan)

Adjunct Professor
Department of African Language
University of Venda (South Africa)

Honorary Associate Professor
Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati (India)

Miyazaki, Kumiko
Visiting researcher
Global Center for Swahili Studies and Advancement
State University of Zanzibar (Tanzania)

Patin, Cédric
Maître de conférences
CNRS, UMR 8163 - STL - Savoirs Textes Langage
Université de Lille & CNRS (France)

Sarmah, Priyankoo
Professor
Department of Humanities and Social Sciences, Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati (India)

SHINAGAWA, Daisuke
Associate Professor
Research Institute for Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies (Japan)

SUZUKI, Michinori
Ph.D. Candidate
Graduate School of Arts and Sciences
International Christian University (Japan)

TERHIIJA, Viyazonuo
Ph.D. Candidate
Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati (India)

UCHIHARA, Hiroto
Associate Professor
Institute of Global Studies
Tokyo University of Foreign Studies (Japan)

UETA, Naoki
Assistant Professor
Research Institute for Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies (Japan)

YAMAMOTO, Kyosuke
Associate professor
Institute of Global Studies
Tokyo University of Foreign Studies (Japan)

ISSN 2436-9179

Crossroads for Phonetic Typology

Edited by
Seunghun J. Lee, Daisuke Shinagawa & Keita Kurabe

JOURNAL OF ASIAN AND AFRICAN STUDIES
SUPPLEMENT

アジア・アフリカ言語文化研究　別冊

No.04

Research Institute for Languages and Cultures of Asia and Africa (ILCAA)