ILLUSTRATIONS OF THE IPA

Baima

Katia Chirkova
Centre National de la Recherche Scientifique, Centre de Recherches Linguistiques sur l'Asie Orientale
& Nankai University, School of Literature
katia.chirkova@gmail.com

Tanja Kocjančič Antolík
Charles University, Faculty of Arts, Institute of Phonetics
tkocjancic@gmail.com

Angélique Amelot
Centre National de la Recherche Scientifique, Laboratoire de Phonétique et Phonologie
angelique.amelot@gmail.com

The Baima language (/pêkê/, Chinese 白马语 *báimǎyǔ*, ISO-639 code *bqh*) is a little-studied Tibeto-Burman (Bodic or Himalayish) language spoken in the southwestern part of the People's Republic of China. Approximately 10,000 people, who traditionally reside in three counties in Sichuan Province (Pingwu 平武, Songpan 松潘 (in Written Tibetan, hereafter WT, *zung chu*), and Jiuzhaigou 九寨沟 (WT *gzi rtsa sde dgu*), and in one county in Gansu Province (Wenxian 文县), speak the Baima language (see Sichuan Sheng Minzu Yanjiusuo 1980, Zeng & Xiao 1987), see Figure 1.[1] The largest concentrations of Baima speakers are in Baima Township (白马乡, Baima /tôpû/) of Pingwu County, and in Tielou Township (铁楼乡) of Wenxian County.

The area of distribution of the Baima language lies at the historical Sino-Tibetan border, in a multi-ethnic and multi-lingual region. In all counties of its present distribution, immediate linguistic neighbors of Baima include varieties of Mandarin (mostly Southwestern Mandarin), and varieties of Tibetan (or Tibetic) languages.[2] In Songpan County, Baima is additionally neighbored by the local variety of the Rma (or Qiang) language (see Evans & Sun 2015).

The basic vocabulary of Baima is mostly of Tibetan origin (Zhang 1994a, b; Huang & Zhang 1995), and Baima also exhibits characteristic phonological features of Tibetic

[1] According to Suzuki (2015: 120), Baima is also spoken in a few villages in Boyu township (博峪, WT *bod yul*) of Zhouqu County (舟曲县, WT *'brug chu rdzong*), which borders on Jiuzhaigou County. He notes, however, that there are no more speakers of Baima in the younger generation in Zhouqu.

[2] The term 'Tibetic' refers to a family of languages derived from Old Tibetan (hereafter OT), which was originally spoken in the Yarlung valley at the time of the Tibetan empire (7th–9th centuries) (Tournadre 2014). Tibetic languages spoken in China are conventionally divided into three groups: Central (or Dbus-gtsang), Kham, and Amdo, which correspond to the three namesake historical provinces of Tibet (Hu 1991, Gesang Jumian & Gesang Yangjing 2002). Among the three groups, Central Tibetan and Amdo groups are characterized by linguistic homogeneity, whereas the Kham group is internally more diverse. The border areas of the historical Amdo and Kham provinces in the present-day Northern Sichuan and Southern Gansu, where the Baima language is spoken, are particularly diverse linguistically and are home to many language-like Tibetic varieties (see J. Sun 2014).

Journal of the International Phonetic Association (2023) 53/2
doi:10.1017/S0025100321000219

First published online 18 November 2021

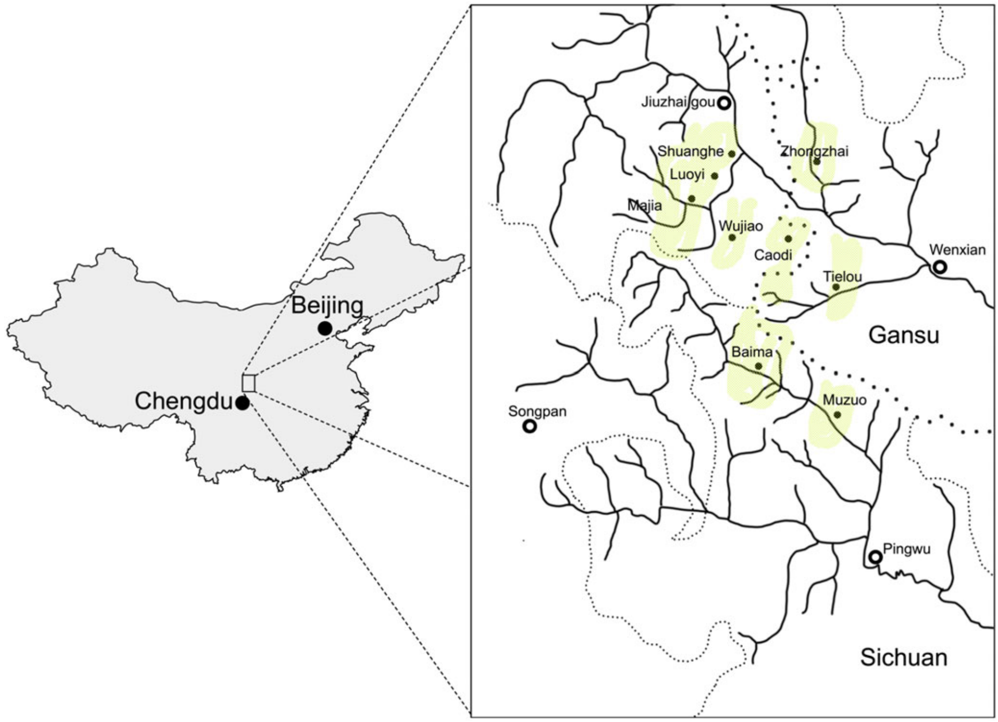

Figure 1 (Colour online) Distribution of the Baima language (adapted from Sichuan Sheng Minzu Yanjiusuo 1980). Black dots with white centers indicate county seats, black dots indicate townships, shaded areas indicate Baima villages.

languages (see Tournadre 2005, 2014).[3] At the same time, Baima is not mutually intelligible with other Tibetic varieties in its neighborhood, and it is considered a distinct language by its speakers. The linguistic affiliation of Baima – as a Tibetic language (Zhang 1994a, b; Huang & Zhang 1995) or a Bodic language distinct from Tibetic languages (H. Sun 1980, 2003; Nishida & Sun 1990 ; H. Sun et al. 2007) – has long been disputed, mainly due to the controversy surrounding the ethnic classification of the Baima group (see Chirkova 2007 for an overview).

Previous work on Baima has mostly concentrated on the variety spoken in Baima Township of Pingwu County. In Baima, this variety is called /tôpû k^{h}â̱kê/. It has been described in Nishida & Sun (1990), Huang & Zhang (1995), H. Sun, Chirkova & Liu (2007), and Feng (2017) (all in Chinese), and it is also the focus of the present Illustration. In Pingwu County, Baima speakers are a minority, Chinese speakers making up the majority. Pingwu Baima is in contact with the local variety of Mandarin Chinese (which belongs to the Chengyu cluster 成渝小片 of the Chuanqin group 川黔片 of Southwestern Mandarin dialects, Li 2012: 85). To our knowledge, there are no longer any monolingual speakers of Pingwu Baima, as all age groups are bilingual in the local variety of Mandarin. Mandarin (both the local variety and the closely related Standard Mandarin, the official language of the People's Republic of China) also dominates the education system and work in public domains. As a result, Baima is mostly restricted to family and community events, and is therefore endangered.

[3] Some common phonological innovations of Tibetic languages include palatalization of *ty, *ly, *sy, *tsy, e.g. WT *gcig* 'one', Baima /t͡ʃî/ [t͡ʃʒə̂]; WT *chi* 'what', Baima /t͡ʃ^{h}ə̂/; WT *shing* 'tree, wood', Baima /ʃ^{h}è/; WT *bzhi* 'four', Baima /ʒə̂/. A commonly cited example of a lexical innovation in Tibetan is the form *bdun* for 'seven', Baima /dè/.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Baima is remarkable for its phonological complexity, and for a number of features that are typologically uncommon. These include non-modal phonation type contrasts in both consonants and vowels, four series of affricates contrasting in place of articulation, a three-way manner contrast in fricatives, a large vowel inventory with eleven monophthongs and four diphthongs, and a tonal system characterized by syllable-level contrasts, involving the redundant use of pitch, voice quality, and vowel length. This Illustration aims to (i) present the exceptionally rich phonetic and phonological system of Baima in a language that makes it accessible also to a non-Chinese speaking audience, and (ii) provide an updated phonological analysis of Baima. The main points of difference from previous analyses include (i) epiglottalization as secondary articulation in Baima consonants, and (ii) voice quality of the vowel as one of the phonetic correlates of tone.[4] Of these, epiglottalization as secondary articulation in consonants has not been previously reported for Tibetic languages. Its development in Baima may possibly be attributable to language contact. Notably, distinctive secondary articulations (such as velarization or uvularization) are being progressively discovered and explored in northern Rma varieties (e.g. Evans 2006, Evans & J. Sun 2015, Evans et al. 2016), which are immediate linguistic neighbors of Baima.

The sample words in this Illustration come from two sets of field recordings in Pingwu made in 2015 and 2019. Both sets contain words spoken in isolation (a basic vocabulary list of 2,000 words, with an average of three repetitions per word) and a few short texts by one male speaker, Mr. Li Degui 李德贵. Born in 1939 in Luotongba 罗通坝 village (Baima /jɑ̀rūk^{h}uɛ̂/), Mr. Li has been bilingual in Baima and Mandarin since childhood. Prior to retirement, Mr. Li worked as an elementary school teacher. He continues to speak Baima in his everyday life with family and friends.

Phonation types and the voice quality of the syllable

The three contrastive tones in Baima – high falling, mid, low – are related to different phonation types in vowels, respectively, tense, modal, and lax. In addition, in the high falling tone, there is a binary phonation contrast in consonants: epiglottalized and non-epiglottalized. Epiglottalized consonants have a harsh voice quality. Epiglottalized consonants include (i) non-nasalized voiced stops, affricates, and fricatives; (ii) prenasalized stops and affricates; and (iii) nasals, and lateral and palatal approximants. Non-epiglottalized consonants include (i) non-nasalized voiceless stops, affricates and fricatives; (ii) prenasalized (voiced) stops and affricates; and (iii) nasals, and lateral and palatal approximants.

Non-modal vowels and vowels adjacent to epiglottalized consonants share some acoustic features (such as a lower harmonics-to-noise ratio, a higher spectral slope, a lower f0 relative to modal vowels). In addition, vowels adjacent to epiglottalized consonants are characterized by higher F1, which indexes laryngeal raising, and lower strength of excitation (SoE), which indexes a greater constriction in the larynx or vocal tract and weaker voicing (see Garellek 2020 and references therein). Non-modal vowels and vowels adjacent to epiglottalized consonants also display differences in timing. Non-modal vowel phonation (tense, lax) persists throughout the entire vowel, whereas non-modal phonation due to consonantal epiglottalization typically lasts for less than half of the vowel duration.

The voice quality of syllables beginning with non-epiglottalized consonants depends on the default voice quality of the tone. The voice quality of syllables beginning with epiglottalized consonants, on the other hand, results from the interaction of the default voice quality of the tone and the harsh voice quality due to consonantal epiglottalization. The consonant-derived harshness typically does, but need not, replace the tone's default voice quality. Table 1 provides an overview of phonation types associated with different syllable onsets and tones

[4] We are grateful to the editors for prompting us to describe different types of phonation found in Baima, and for suggesting the helpful references cited in the text.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Table 1 Phonation types associated with different syllable onsets and tones. T = voiceless unaspirated obstruent, T^{h} = voiceless aspirated obstruent, D = (non-nasalized) voiced obstruent, $^{\text{N}}$D = prenasalized stop or affricate, S = sonorant, V = vowel. Shaded cells show non-occurring combinations.

Tone / Syllable onset	High falling: Epiglottalized	High falling: Non-epiglottalized	Mid	Low
TV		tense	modal	
T^{h}V				
DV				
$^{\text{N}}$DV		tense	modal	
SV	tense or harsh			lax

in our data. Quantitative descriptions of various phonation types are provided in sections on consonants and tones.

A combination of a binary phonation contrast on consonants and a three-way phonation contrast on vowels, as described in this Illustration, is unusual for Tibetic languages and, more broadly, for other Tibeto-Burman languages of Southwestern and Southern China, many of which, similar to Baima, contrast tone and phonation (such as Yi or Bai languages). Given the complex interactions between syllable onsets and tones (as in Table 1) and variability of in the realization of consonant-derived harshness, the Baima data at our disposal do not lend themselves straightforwardly to a reanalysis in terms of syllable-level contrasts in register that is separate from tone. For that reason, this Illustration focuses more on the phonetic details of phonation contrasts in this language. The unusualness of the Baima system naturally warrants further investigation, based on larger data sets and more speakers (a follow-up study is in preparation). The historical development of the Baima tonal system from Old Tibetan is discussed in Chirkova (2021).

Consonants

The Baima consonant inventory consists of 57 phonemes, presented in the consonant table below. Sequences of a nasal and a voiced stop or a voiced affricate (e.g. $^{\text{m}}$b, $^{\text{n}}$d, $^{\text{n}}$d͡ʒ), which are treated as nasal–stop and nasal–affricate clusters in previous descriptions of Baima, are here analyzed as unitary segments (prenasalized stops and prenasalized affricates). Arguments for our analysis include: (i) in such sequences, the place of articulation of the nasal is always homorganic with that of the stop or the affricate, (ii) such sequences only occur in syllable-initial position, (iii) such sequences have a duration comparable to that of other consonants in syllable-initial position. Conversely, phonological evidence that would argue for a cluster analysis (such as compensatory lengthening, resyllabification, or independent tone specification) is unavailable in Baima.

Contrastive epiglottalized prenasalized stops and affricates, nasals, and lateral and palatal approximants are notated with the IPA superscript voiced epiglottal fricative (or trill) (e.g. $^{\text{m}}$b$^{\text{ʢ}}$). In our corpus, contrastive epiglottalized vs. non-epiglottalized pairs with prenasalized stops and affricates are only attested at the bilabial and velar places of articulation for prenasalized stops, and at the alveolar and alveolo-palatal places of articulation for prenasalized affricates. All words with prenasalized postalveolar affricates in the high falling tone are epiglottalized and marked for epiglottalization in transcriptions. All words with prenasalized alveolar stops and prenasalized retroflex affricates in the high falling tone, on the other hand, are non-epiglottalized. Examples include [$^{\text{n}}$d͡ʒ$^{\text{ʢ}}$â̰] 'to eat (IPFV)', [$^{\text{n}}$d͡ʒ$^{\text{ʢ}}$ṵ̂] 'snout', [$^{\text{n}}$də̂] 'this', [$^{\text{n}}$dâ̰] 'to tremble'; [$^{\text{n}}$ɖ͡ɽɛ̂] 'ghost', [$^{\text{n}}$ɖ͡ɽə̂] 'female yak'.

/ʈ͡ɽ ʈ͡ɽ$^{\text{h}}$ ɖ͡ɽ/, which are treated as retroflex affricates in previous descriptions of Baima (/ʈ͡ʂ ʈ͡ʂ$^{\text{h}}$ ɖ͡ʐ/), are in our data, rather sequences of a stop and a trill, which is raised and fricated. We

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

analyze /ʈ͡ɽ ʈ͡ɽʰ ɖ͡ɽ/ as unitary segments, for there are no phonological arguments to treat them as clusters. Note that in our notation, /ɽ/ represents not a voiced retroflex flap (its IPA official meaning), but a fricated postalveolar trill, for which there is currently no IPA symbol. In our choice of the symbol /ɽ/, we follow conventions in Spajić et al. (1996). In their work on trills in Toda, they use /ɽ/ for the retroflex trill in that language.

In our corpus, all consonant phonemes in the consonant table can occur at the onset position of monosyllabic roots in the language. The following minimal and near-minimal sets illustrate consonant phonemes in Baima.

	Bilabial	Alveolar	Post-alveolar	Palatal or Alveolo-palatal	Retroflex	Velar
Plosive	p pʰ b ᵐb ᵐbˤ	t tʰ d ⁿd				k kʰ g ᵑg ᵑgˤ
Affricate		t͡s t͡sʰ d͡z ⁿd͡z ⁿd͡zˤ	t͡ʃ t͡ʃʰ d͡ʒ ⁿd͡ʒˤ	t͡ɕ t͡ɕʰ d͡ʑ ᶮd͡ʑ ᶮd͡ʑˤ	ʈ͡ɽ ʈ͡ɽʰ ɖ͡ɽ ⁿɖ͡ɽ	
Nasal	m mˤ	n nˤ		ɲ ɲˤ		ŋ ŋˤ
Trill		r				
Fricative		s sʰ z	ʃ ʃʰ ʒ	ɕ ɕʰ ʑ		x ɣ
Approximant	w			j jˤ		
Lateral approximant		l lˤ				

p	pê	‘Tibetan’[5]	ⁿd͡ʒˤ	ⁿd͡ʒˤâ (də)	‘to (be) solidify(ing)’
pʰ	pʰè	‘to level, flatten’	ʃ	ʃê	‘to wipe’
b	bù dzù	‘fist’	ʃʰ	ʃʰè	‘tree’
ᵐb	ᵐbû	‘insect’	ʒ	ʒè	‘trace’
ᵐbˤ	ᵐbˤû	‘to fly’	t͡ɕ	t͡ɕâ	‘to weigh (PFV/IMP)’
m	mē	‘to plough (PFV/IMP)’[6]	t͡ɕʰ	t͡ɕʰâ	‘blood’
mˤ	mˤê	‘feces; fertilizer’	d͡ʑ	d͡ʑâ	‘to mend (IPFV)’
w	wô	‘rhododendron’	ᶮd͡ʑ	ᶮd͡ʑɛ̂	‘to win’
t	tê	‘to push’	ᶮd͡ʑˤ	ᶮd͡ʑˤâ	‘to be cold’
tʰ	tʰê	‘to pull open’	ɲ	ɲɛ̂	‘fire’
d	dê	‘to ruminate’	ɲˤ	ɲˤê	‘to find’

5 /e/ may be realized as diphthongized ([ei]). This realization is non-contrastive in our corpus.

6 Baima verbs can have one to three different stems (imperfective, perfective, and imperative). The type of stem is indicated in brackets after the verb gloss. If no indication is given, the verb has one single stem.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

ⁿd	ⁿdē	‘to sunbathe’	j	jyə̂	‘to yawn’
n	nâ	‘forest’	jˤ	jˤyə̂	‘to be lax’
nˤ	nˤâ	‘pus’	ç	çâ̠	‘bird; chicken’
r	rē	‘cloth’	çʰ	çʰâ̠	‘to open (door) (PFV/IMP)’
l	lâ	‘to lick’	ʑ	ʑâ̠	‘to borrow (tools)’
lˤ	lˤâ̠	‘deity’	ʈ͡ɽ	ʈ͡ɽê	‘to cover, enmesh’
t͡s	t͡sê	‘to arrive’	ʈ͡ɽʰ	ʈ͡ɽʰê	‘to meet (PFV/IMP)’
t͡sʰ	t͡sʰê	‘winnowing fork’	ɖ͡ɽ	ɖ͡ɽì [ɖ͡r̩r̩ə̀]	‘to roll’
d͡z	d͡zê	‘to pour, drench’	ⁿɖ͡ɽ	ⁿɖ͡ɽē	‘rice’
ⁿd͡z	ⁿd͡zɛ̂	‘to abstain from’	k	kê	‘sound’
ⁿd͡zˤ	ⁿd͡zˤɛ̂	‘night’	kʰ	kʰè	‘official’
s	sê	‘to kill’	g	gè	‘to be old’
sʰ	sʰê	‘to tear, rip open’	ŋ	ŋâ̠	‘drum’
z	zî [zz̩ə̂]	‘leopard’	ŋˤ	ŋˤâ̠	‘five’
t͡ʃ	t͡ʃê	‘to chop; break’	ᵑg	ᵑgâ̠	‘to open (flowers)’
	t͡ʃâ	‘iron’	ᵑgˤ	ᵑgˤâ̠	‘city’
t͡ʃʰ	t͡ʃʰê	‘to be tired’	x	xuâ̠suɛ̂	‘to think, reflect’
	t͡ʃʰâ	‘to break (INTR)’	ɣ	ɣâ̠	‘fox’
d͡ʒ	d͡ʒâ	‘tongue’			

Voicing and epiglottalization in non-nasalized obstruents

Baima non-nasalized stops, affricates, and fricatives contrast in voicing; voiceless stops, affricates, and fricatives are further differentiated in terms of aspiration. Voiced obstruents are epiglottalized. Degrees of both voicing and epiglottalization in voiced obstruents may vary.

In word-initial position, Baima (non-nasalized) voiced stops, affricates, and fricatives may be realized (i) as voiced (that is, showing voicing lead through much or all of the closure and through release), (ii) as prevoiced (with voicing diminishing throughout the closure and/or non-existent during release), and also as (iii) devoiced. In addition, voiced stops, affricates, and fricatives may be realized as prenasalized. This is illustrated in Figure 2, which shows waveforms, spectrograms, and pitch tracks for two repetitions of the words /gâ̠/ ‘horn’ (Figure 2a), /d͡zâ̠/ ‘mountain peak’ (Figure 2b), and /zâ̠/ ‘to filter’ (Figure 2c).

As shown in Figure 2a, the first of the two repetitions of the word /gâ̠/ is voiced, [ga̠]; whereas the second repetition is devoiced ([g̊a̠]). Of the two repetitions of the word /d͡zâ̠/

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

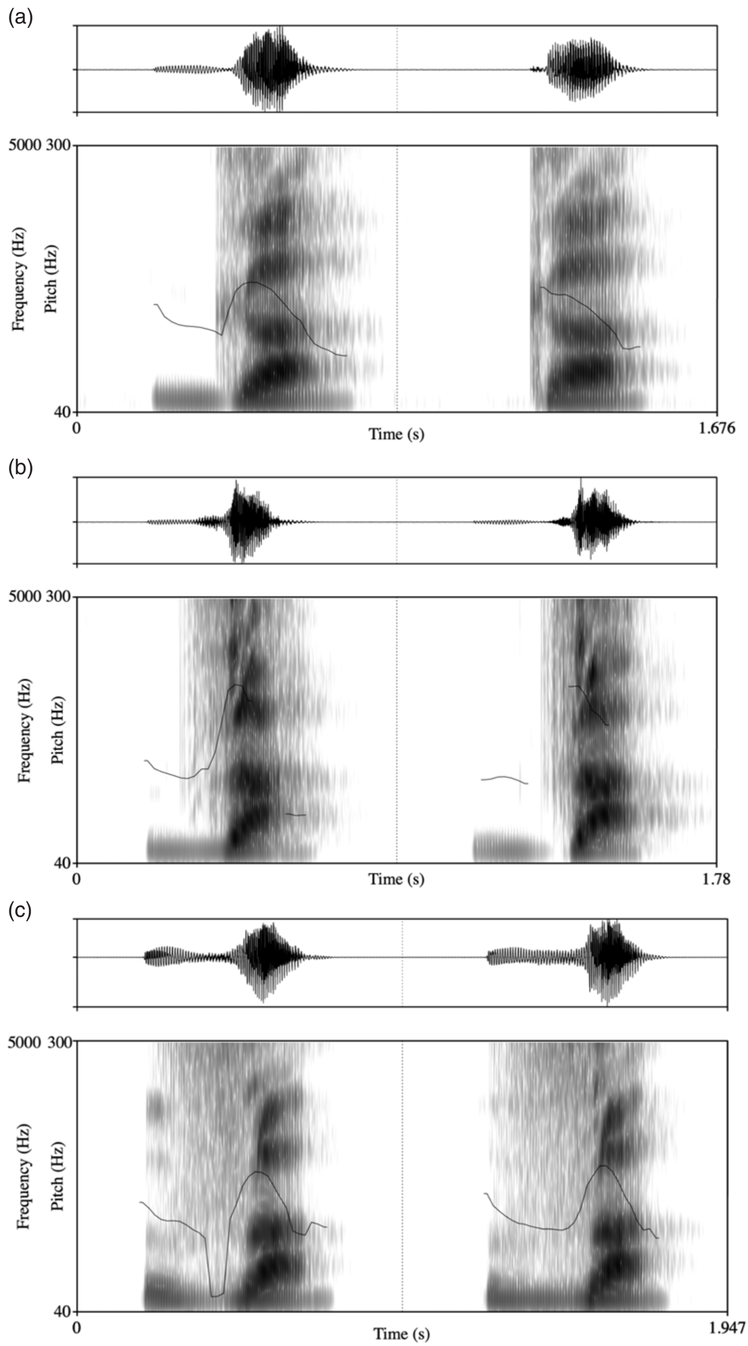

Figure 2 Waveforms, spectrograms, and pitch tracks for the two repetitions of the words /g̱â/ 'horn' (2a); /d͡ẕâ/ 'mountain peak' (2b), and /ẕâ/ 'to filter' (2c).

'mountain peak' in Figure 2b, the first is again voiced, whereas the second is prevoiced. Finally, the first repetition of the word /zâ̱/ 'to filter' in Figure 2c is prenasalized at the beginning and then shows diminishing voicing halfway through the frication period. The second repetition is fully voiced. Voiced obstruents are likely undergoing devoicing in Baima (see below), and prenasalization may be used as a strategy to facilitate voicing (due to the phonetic affinity between prenasalization and voiced consonants, see Ohala 1983: 194–201). In those words in our corpus, where voiced obstruents occur intervocalically, voiced obstruents are more consistently voiced (as in /nōgâ̱/ 'cow horn').

Voiced stops, affricates, and fricatives are characterized by a distinctive harsh voice quality, which is clearly observable over approximately half of the vowel duration, but may also persist throughout the entire vowel. Relative to vowels adjacent to non-epiglottalized (voiceless) stops, affricates, and fricatives, vowels following epiglottalized stops, affricates, and fricatives are characterized by (i) a lower harmonics-to-noise ratio (measured here using harmonics-to-noise ratio below 500 Hz [HNR05]), (ii) a higher harmonic ratio (difference between the amplitude of the first and second harmonics, corrected for the effects of formants and bandwidths [H1*–H2*]), and (iii) high to low SoE, (iv) higher F1, and (v) lower f0 values.

The acoustic characteristics of vowels adjacent to voiced stops, affricates, and fricatives are illustrated in Figure 3 on the basis of three minimal pairs at the alveolar and velar places of articulation: /dâ̱/ 'to grind (PFV/IMP)' vs. /tâ̱/ 'to look (IPFV)'; /gâ̱/ 'horn' vs. /kâ̱/ 'to be bright'; /gɑ̂/ 'to bind' vs. /kɑ̂/ 'hoe' (three repetitions per word, a total of 18 tokens). We segmented the vowel and further divided it into five regions: 0–20%, 20–40%, 40–60%, 60–80% and 80–100% of the vowel duration. We then used VoiceSauce (Shue et al. 2011) to measure (i) HNR05, (ii) H1*–H2*, (iii) SoE, (iv) F1, and (v) f0 at these regions over the segmented intervals. Figure 3 shows the time courses of these measures over the vowel.

As shown in Figure 3, vowels following voiced stops are characterized by (i) higher H1*–H2* values throughout approximately half of the vowel duration, and (ii) lower HNR05, (iii) high-to-low SoE, (iv) higher F1, (v) lower f0 throughout the entire vowel duration, as relative to vowels following voiceless stops. Hence, vowels adjacent to voiced stops generally conform to the characteristics of harsh vowels, as defined in Garellek (2020) (low HNR05, high-to-low H1*–H2*, low SoE, high F1, low f0).[7] We note that f0 differences on vowels adjacent to voiced and voiceless obstruents is one major phonetic correlate to the voicing and phonation contrast in obstruents in Baima. The magnitude of f0 perturbations (which in Figure 3, ranges from over 40 Hz to almost 70 Hz) is considerably greater than what is common for languages with voicing contrasts (ranging from 8 to 16 Hz, as reported for English, Swedish, German, Dutch, Italian, and French in Coetzee et al. 2018: 187), and similar to f0 differences found in languages analyzed as undergoing incipient tonogenesis (such as Seoul Korean or Afrikaans). Large f0 differences between Baima voiced and voiceless obstruents may be suggestive of a change in progress, whereby f0 is overtaking other multiple covarying cues as the primary cue for the voicing contrast in obstruents. (F0 differences of similar

[7] Baima voiced obstruents develop from OT initial clusters consisting of voiced obstruents with preradicals (*r*-, *l*-, *s*-, *b*-, *d*-, *g*-) (Huang & Zhang 1995). Consider the following examples from the Consonant table: /dê/ 'to ruminate', WT *ldad*; /zî/ 'leopard', WT *gzig*; /dʒɑ̂/ 'tongue', WT *ljags*; /ʒè/ 'trace', WT *rjes*. In some Tibetic varieties neighboring Baima, these OT clusters have breathy reflexes (see J. Sun 2003, Suzuki 2015). Voiced obstruents in Baima have likely gone through a similar, breathy stage of development. The characteristic harsh quality associated with words with voiced obstruents in the high falling tone (as in our audio files, see examples above) can be tentatively explained as a by-product of (historically) breathy voice followed by tense voice (the default voice quality of the high falling tone).

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

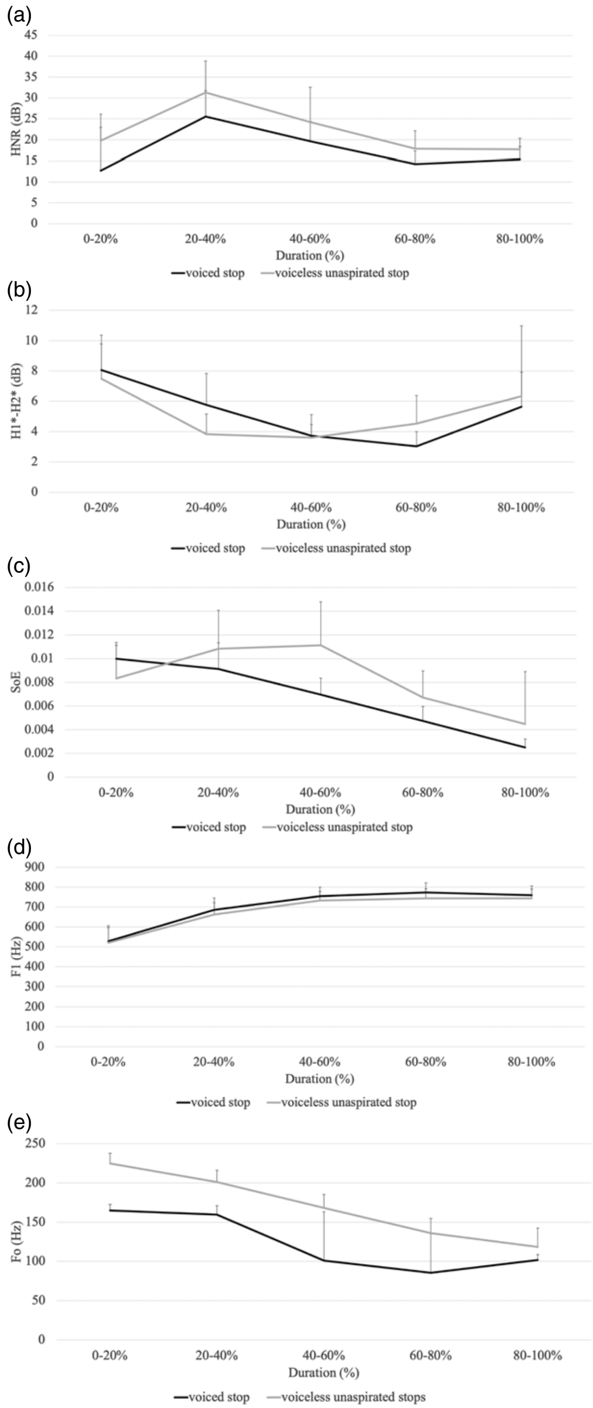

Figure 3 Time courses of HNR05 (dB) (3a), H1*–H2* (dB) (3b), SoE (3c), F1(Hz) (3d), and f0 (Hz) (3e) for vowels in the three minimal pairs with voiced stop onsets (in black) and voiceless unaspirated stop onsets (in gray) at the alveolar and velar places of articulation before the vowels /a̱/ and /ɑ/, uttered in isolation. The x-axis represents vowel duration divided into five regions: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%. The number of analyzed tokens is 18.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

magnitude are also observed in contrastive pairs with epiglottalized and non-epiglottalized prenasalized obstruent and sonorant onsets, see below.[8])

While vowels adjacent to the sample words with voiced stops in Figure 3 are markedly harsh, the degree of harshness in vowels following voiced obstruents may vary. This can be illustrated with the two repetitions of the words /dî/ 'ogress, a man-eating female demon' and /dɜ̂/ 'stone'. For each word, the first repetition is characterized by higher levels of noise and weaker voicing, as compared to the second repetition.

Contrastive epiglottalization in prenasalized obstruents and sonorants

In the high falling tone, we observe a contrast between epiglottalized and non-epiglottalized prenasalized stops and affricates (/ᵐbˤ ᵑgˤ ⁿd͡zˤ ᶮd͡ʑˤ/ – /ᵐb ᵑg ⁿd͡z ᶮd͡ʑ/) and epiglottalized and non-epiglottalized nasals (/mˤ nˤ ɲˤ ŋˤ/ – /m n ɲ ŋ/). (Near) minimal pairs include: /ᵐbˤû/ 'to fly' vs. /ᵐbû/ 'insect'; /ᵑgˤâ̱/ 'city' vs. /ᵑgâ̱/ 'to open (flowers)'; /ⁿd͡zˤô/ 'lake' vs. /ⁿd͡zô/ 'dzo (a hybrid of yak and domestic cattle)'; /ᶮd͡ʑˤɑ̂/ 'to be cold' vs. /ᶮd͡ʑɛ̂/ 'to win'; /mˤɑ̂/ 'army' vs. /mâ̱/ 'wound'; /nˤɑ̂/ 'pus' vs. /nɑ̂/ 'forest'; /ɲˤô/ 'to be mad' vs. /ɲô/ 'to buy' (IPFV); /ŋˤâ̱/ 'five' vs. /ŋâ̱/ 'drum'. This contrast is also exceptionally observed in one near-minimal pair with lateral approximant initial consonants (/lˤâ̱/ 'deity' vs. /lɑ̂/ 'to lick') and two (near) minimal pairs with palatal approximant initial consonants (/jˤyə̂/ 'to be lax' vs. /jyə̂/ 'to yawn'; /jˤyɛ̂/ 'to collapse, fall down' vs. /jyɛ̄/ 'lung').

Epiglottalized prenasalized obstruents and sonorants are characterized by a harsh quality observed both on the initial consonant and at the onset of the following vowel. This is illustrated in Figure 4 on the basis of three (near) minimal pairs with epiglottalized and non-epiglottalized nasals (/mˤɑ̂/ 'army' vs. /mâ̱/ 'wound', /nˤɑ̂/ 'pus' vs. /nɑ̂/ 'forest', /ŋˤâ̱/ 'five' vs. /ŋâ̱/ 'drum'), measured over the entire syllable (onset and vowel). The difference in amplitude between the first and second harmonics is corrected for the effects of formants and bandwidths for vowels and palatal approximants (H1*–H2*) and uncorrected for consonants (H1–H2) (because the correction is only tested for vowels, see discussion in Garellek 2019: 91).

As can be seen in Figure 4, epiglottalized nasals are characterized by (i) low HNR05, (ii) high H1–H2, (iii) low SoE, and (iv) low f0. Vowels adjacent to epiglottalized prenasalized stops and affricates, on the other hand, are characterized by (i) low HNR05, (ii) high H1*–H2*, (iii) high to low SoE, (iv) high F1, and (v) low f0 values throughout approximately half of the vowel duration.[9] Hence, compared to voiced stops, the non-modal quality in epiglottalized nasals is phased earlier in the voiced portion of the syllable.

[8] Note that the contrast between voiced, epiglottalized obstruents and voiceless non-epiglottalized obstruents in the high falling tone, as described in this section, is uniformly described in previous analyses of Baima as that of voicing and tone (voiced obstruents having a lower falling tone relative to voiceless obstruents). The contrast between epiglottalized and non-epiglottalized prenasalized obstruents and sonorants, on the other hand, is variably described in previous analyses as that of tone (e.g. Nishida & Sun 1990, H. Sun et al. 2007) and as that of consonant phonation type (breathy vs. non-breathy) and tone (Huang & Zhang 1995). In both cases, epiglottalized prenasalized obstruents and sonorants are analysed as having a lower falling tone relative to their non-epiglottalized counterparts.

[9] Baima epiglottalized sonorants regularly develop from OT sonorant initial consonants with preradicals. Epiglottalized nasals develop from OT sonorant initial consonants with the preradicals *r*-, *l*-, *s*-, *g*-, *d*-, *b*-, and *m*-. Epiglottalized lateral and palatal approximants develop from the proto-sonorant cluster *lh*. Baima epiglottalized prenasalized stops and affricates, on the other hand, regularly develop from OT voiceless aspirated initial consonants with the OT nasal preradicals *m* and ˈ-. Examples include: /nˤɑ̂/ 'pus', WT *rnag*; /jˤyə̂/ 'to be lax', WT *lhod*; /ᵐbˤû/ 'to fly', WT ˈ*phur*; [ⁿd͡ʒˤâ̱] 'to eat (IPFV)', WT ˈ*cha*ˈ; [ⁿd͡ʒˤû] 'snout', WT *mchu* 'lips'. A few exceptions can be noted. For instance, in the pair /ŋâ̱/ 'drum', WT *rnga* vs. /ŋˤâ̱/ 'five', WT *lnga*; both initial consonants in WT are preceded by preradicals, however, in Baima, the initial of the word 'drum' is non-epiglottalized. In many Tibetic varieties spoken

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

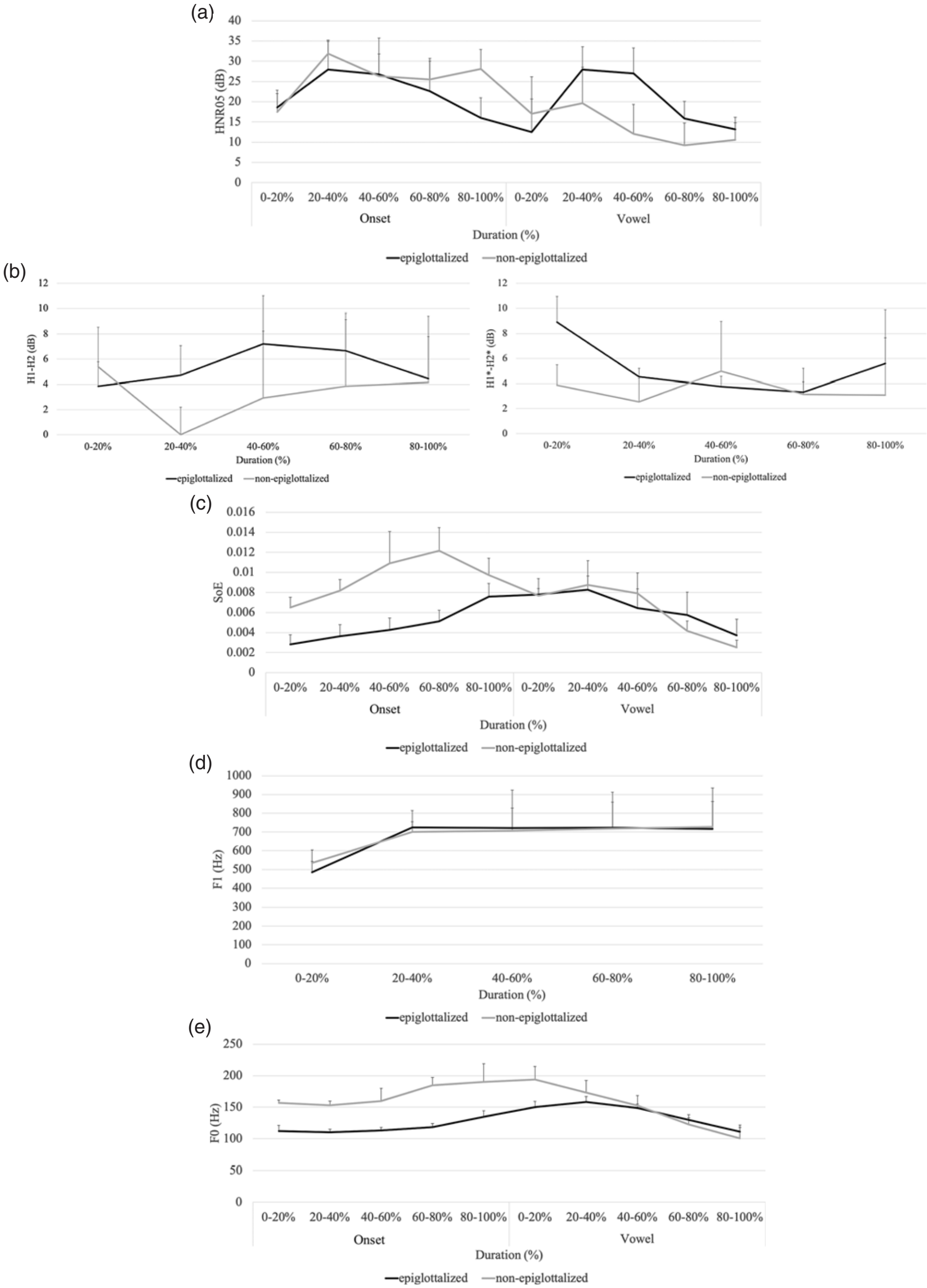

Figure 4 Time courses of HNR05 (dB) (4a), H1–H2 (dB) (4b left), H1*–H2* (dB) (4b right), SoE (4c), F1 (Hz) (d), and f0 (Hz) (4e) in three (near) minimal pairs with epiglottalized nasals (in black) and non-epiglottalized nasals (in gray). The x-axis represents onset and vowel duration divided, respectively, into five regions: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%. Results for HNR05 (dB), SoE, and f0 (Hz) are presented in the same plot. Results for the difference in amplitude between the first and second harmonics are presented separately for the onset (uncorrected measures, H1–H2) and for the vowel (corrected measures, H1*–H2*). F1 (Hz) measures are only taken over the vowel. Based on three (near) minimal pairs at the bilabial, alveolar, and velar places of articulation before the vowels /â̰/ and /ɑ/, uttered in isolation. The total number of tokens is 18.

Affricates

In previous analyses, Baima has been described as having four series of affricates, contrasting in place of articulation: alveolar (/t͡s t͡sʰ d͡z/), postalveolar (/t͡ʃ t͡ʃʰ d͡ʒ/), alveolo-palatal (/t͡ɕ t͡ɕʰ d͡ʑ/), and retroflex (/t͡ʂ t͡ʂʰ d͡ʐ/). We note that in our data, the latter series are rather sequences of a stop and a trill, which is raised and fricated (/ʈ͡ɽ ʈ͡ɽʰ ɖ͡ɽ/), than those of a stop and a fricative. Examples include:

/t͡sâ̰/ 'grass' – /t͡ʃâ̰/ 'tea' – /t͡ɕâ̰/ 'hair' – /ʈ͡ɽâ̰/ 'to sever, cut'

/t͡sʰâ̰/ 'salt'; 'to ache' – /t͡ʃʰâ̰/ 'pair' – /t͡ɕʰɑ̂/ 'blood' – /ʈ͡ɽʰâ̰/ 'to be fine'

/d͡zɑ̀/ 'moon; month' – /d͡ʒɑ̂/ 'tongue' – /d͡ʑɑ̂/ 'to mend (IPFV)' – /ⁿɖ͡ɽɑ̄/ 'to be good'

/ⁿd͡zɛ̂/ 'to abstain from' – /ⁿd͡ʒɛ̀/ 'fang, sharp protruding teeth' – /ⁿd͡ʑɛ̂/ 'to win' – /ⁿɖ͡ɽɛ̂/ 'ghost'

The stop release and the voicing of the trill with a periodical waveform in /ʈ͡ɽ/ and /ɖ͡ɽ/ are clearly visible in waveforms and spectrograms. The trill part in /ʈ͡ɽʰ/ is more fricated and devoiced. Furthermore, /ʈ͡ɽʰ/ has a clear period of aspiration between the frication portion and the vowel. This is illustrated in Figure 5, which provides waveforms and spectrograms of the near-minimal set /ʈ͡ɽē/ 'mule', /ʈ͡ɽʰê/ 'to meet (PFV/IMP)', and /ⁿɖ͡ɽē/ 'rice'.

The stop–trill segments may have a trill release, involving one to three contacts. Figure 6 illustrates a production of /ɖ͡ɽì/ 'to roll', which has two full contacts (about 17 ms apart), marked by two arrows.

After the two full contacts in /ɖ͡ɽì/ 'to roll', we observe several more periods, visible as alternating presence and absence of energy on the spectrogram. These alternations suggest aerodynamically induced movements of the tongue tip, with the tongue tip vibrating and approaching a closure. Note also that the trill part of the segment in Figure 6 is accompanied by fricative noise: the high frequency energy noise in the 3000–5000 Hz region. In voiceless aspirated fricatives, the number of contacts is more difficult to see (as in the word /ʈ͡ɽʰê/ 'to meet (PFV/IMP)' in Figure 5).

In line with previous analyses we treat /ʈ͡ɽ ʈ͡ɽʰ ɖ͡ɽ/ as retroflexes. This preliminary conclusion is based on kinesthetic sensations during the imitation of these sounds.[10] For this study, we were not able to collect palatographic and linguagraphic data that would provide reliable

in Sichuan, Gansu, and Qinghai provinces, OT sonorant initials with preradicals and the proto-sonorant cluster *lh* develop into breathy sonorants (see Renzeng 1987; Gesang Jumian & Gesang Yangjing 2002: 178–179). Notably, epiglottalized sonorants in our data are explicitly described as 'voiced aspirated' (*zhuó sòngqì* 浊送气) in the phonetic and phonological sketch of Baima by Huang & Zhang (1995). In view of their distribution (contrastive epiglottalized vs. non-epiglottalized obstruent and sonorant pairs are only found in the high falling tone, the default voice quality of which is tense), it is conceivable that the characteristic harsh quality associated with these segments in our data is a by-product of (historically) breathy phonation followed by tense phonation.

[10] These sounds are auditorily and kinesthetically similar to retroflex affricates with a trill release, as described in our previous work on the Ersu language (Chirkova et al. 2015). In the case of Ersu, we had palatographic data of three speakers showing that retroflex affricates in that language are produced with the point of contact on the roof of the mouth, with the contact made with the underside of the tongue. Compared to other affricates (dental, alveolar, alveolo-palatal), Ersu retroflex affricates are articulated furthest back in the mouth.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

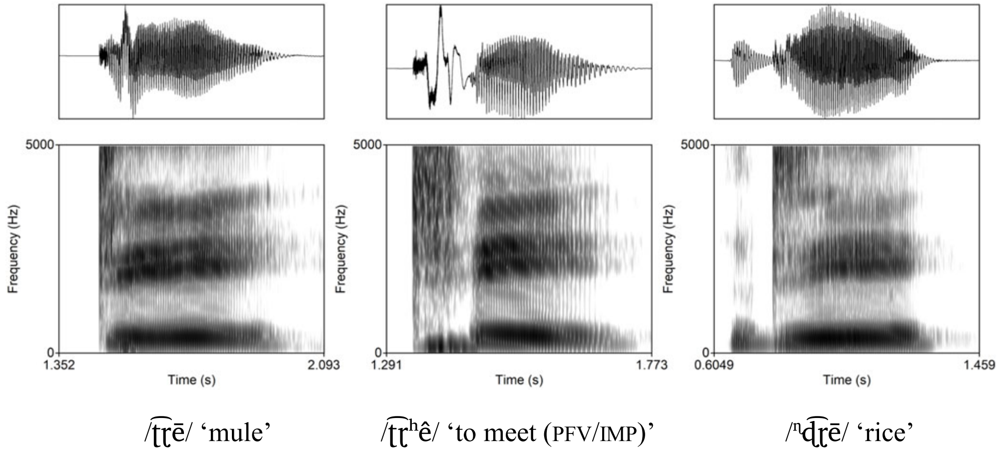

Figure 5 Waveforms and spectrograms for the words /ʈ͡ɽē/ ‘mule’, /ʈ͡ɽ^{h}ê/ ‘to meet (PFV/IMP)’, and /^{n}ɖ͡ɽē/ ‘rice’.

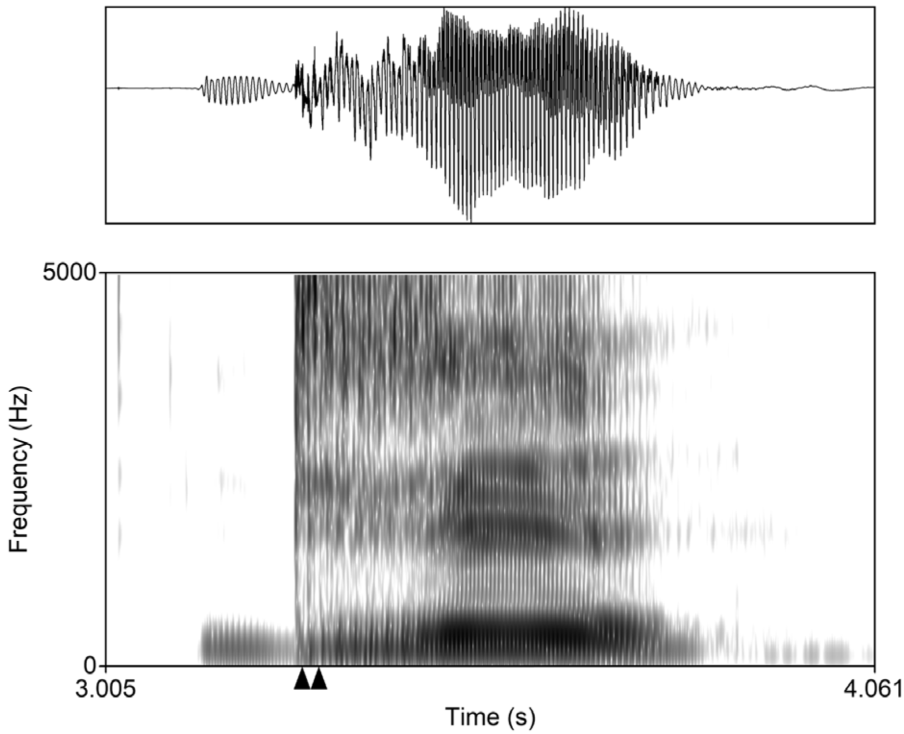

Figure 6 Waveform and spectrogram for the word /ɖ͡ɽì / ‘to roll’. Arrows indicate contacts between the tip of the tongue and the roof of the mouth.

information on the contact areas on both the tongue and the palate, and ascertain the place and manner of articulation for the four contrastive affricate series in Baima. For the present analysis, we examined center of gravity (COG) values, which in studies on perceptual cues for differentiating place of articulation for fricatives and affricates (e.g. Jongman, Wayland

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Table 2 COG values (Hz) for the spectral slices taken from the middle of the frication portion of [t͡s], [t͡ʃ], [t͡ɕ], [ʈ͡ʂ]. Based on 89 tokens.

Place of articulation	Affricate	No of tokens	COG (Hz)	SD (Hz)
Alveolar	d͡z	6	7464	1778
	t͡s	9	7491	2041
	t͡sʰ	10	7975	1647
Postalveolar	d͡ʒ	6	5376	1846
	t͡ʃ	6	5594	1718
	t͡ʃʰ	12	5151	1756
Alveolo-palatal	d͡ʑ	12	3858	1950
	t͡ɕ	6	5916	1820
	t͡ɕʰ	7	4081	1743
Retroflex	ɖ͡ʐ	3	4593	1854
	ʈ͡ʂ	6	4809	1977
	ʈ͡ʂʰ	6	4586	2428

& Wong 2000, Gordon, Barthmaier & Sands 2002, Lee, Zhang & Li 2014), have been suggested as providing salient acoustic cues for the place of articulation. COG and its standard deviation (SD) were measured in Praat (Boersma & Weenink 2020) over a 25.6 ms window centered at the middle of the frication part of the affricate. All recordings were first filtered under 500 Hz and above 10,000 Hz in order to remove ambient noise and the effect of voicing. The number of analyzed tokens ranged between three (for /ɖ͡ʐ/)[11] and twelve per affricate (89 tokens in total), see Table 2.

Our results show the highest COG values for alveolar affricates, followed by postalveolar affricates. Alveolo–palatals and retroflex affricates appear to be articulated further back in the mouth than the former two series. Similar values in alveolo-palatals and retroflexes potentially suggest a similar place of articulation. (We note that the observed COG pattern for the alveolo-palatal voiceless unaspirated affricate /t͡ɕ/ has higher COG than the postalveolar counterpart /t͡ʃ/. The unexpectedly high values may be due to palatalization, which in /t͡ɕ/ raises formants and the COG in the frication part. We suggest that the exact degree of palatalization varies for alveolo-palatals, /t͡ɕ/ being produced with a higher degree of palatalization than its voiced and voiceless unaspirated counterparts.)

An important characteristic of retroflex consonants, distinguishing them from other non-anterior coronals (i.e. postalveolars and alveolo-palatals) is a lowered third formant (see Hamann (2003: 27, 53–64) for detailed discussion, and references therein). Language-specific investigations of retroflexes furthermore suggest that the transitions from a vowel into a consonant (or VC transitions) show a more distinct lowering of the third formant for retroflexes than the transitions from a consonant into a vowel (or the CV transitions). The CV transitions from retroflexes are further closer to those from other coronals, especially apicals (ibid.). In our corpus consisting of words with the CV structure spoken in isolation, we could only investigate the CV transitions from retroflexes. We used two near-minimal sets involving the four affricate series in two vowel environments: (1) followed by /a̠/ (that is, /t͡sâ̠/ 'grass' – /t͡ʃâ̠/ 'tea' – /t͡ɕâ̠/ 'hair' – /ʈ͡ʂâ̠/ 'to sever, cut') and (2) followed by /e/ (that is, /t͡sʰê/ 'winnowing fork' – /t͡ʃʰê/ 'to be tired' – /ʈ͡ʂʰê/ 'to meet (PFV/IMP)'). We segmented the vowel, divided it into five regions (as above), and extracted F3 values for the vowels in the analyzed words. Figure 7 shows the mean F3 values for the two vowels in the data set.

[11] /ɖ͡ʐ/ only occurs in one word in our corpus, /ɖ͡ʐì/ 'to roll'.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

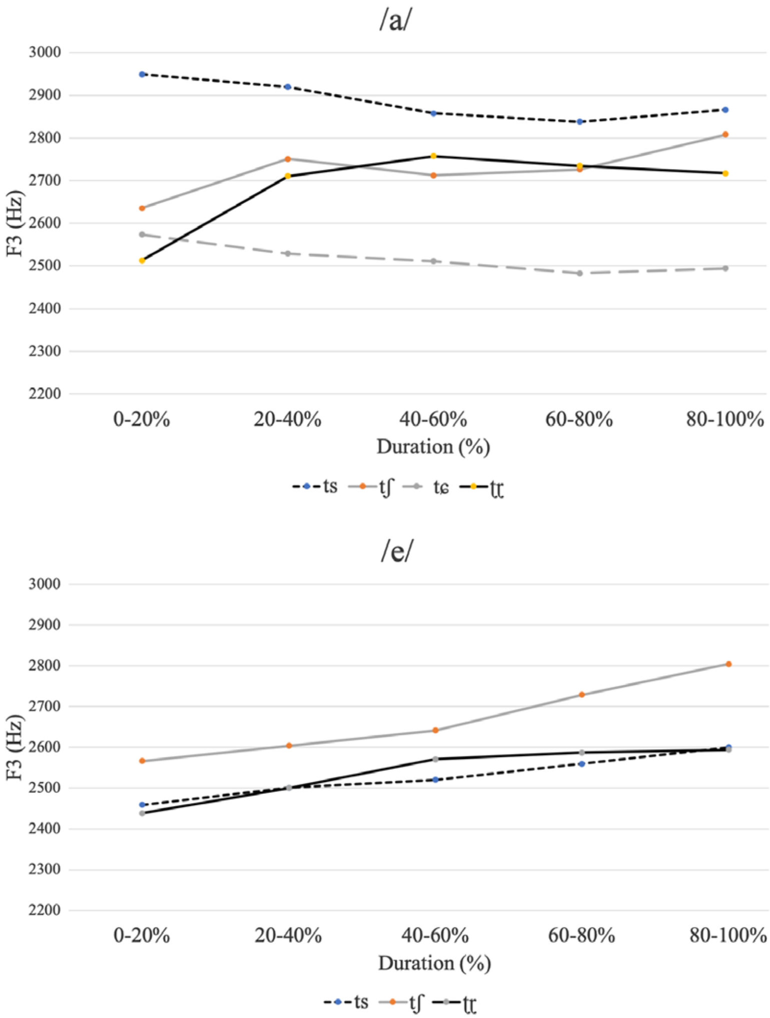

Figure 7 Mean F3 values of the vowel /a̱/, following voiceless unaspirated alveolar, postalveolar, alveolo-palatal, and retroflex affricates (left), and mean F3 values of the vowel /e/, following voiceless aspirated alveolar, postalveolar, alveolo-palatal, and retroflex affricates (right). Based on two near-minimal sets: (1) /t͡sâ̱/ 'grass' – /t͡ʃâ̱/ 'tea' – /t͡ɕâ̱/ 'hair' – /ʈ͡ɽâ̱/ 'to sever, cut' (three repetitions per word, a total of 12 tokens), and (2) /t͡sʰê/ 'winnowing fork' – /t͡ʃʰê/ 'to be tired' – /ʈ͡ɽʰê/ 'to meet (PFV/IMP)' (three repetitions per word, a total of 9 tokens). The x-axis represents vowel duration divided into five regions: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%. /t͡s/ in black dotted line, /t͡ʃ/ in gray solid line, /t͡ɕ/ in gray dashed line, /ʈ͡ɽ/ in black solid line.

Figure 7 shows that the retroflex affricates /ʈ͡ɽ/ and /ʈ͡ɽʰ/ affect the following vowels, lowering F3 in the first half of the vowel's duration. In the second half of the vowel, F3 raises to its average values for /a̱/ and /e/.

Results in this section provisionally support the previous analysis of /ʈ͡ɽ ʈ͡ɽʰ ɖ͡ɽ/ as retroflexes. However, this preliminary conclusion needs to be confirmed in a larger and more diverse corpus, and it also remains to be substantiated by instrumental investigation. In particular, palatographic and linguagraphic data would be essential to further explore this interesting type of sound.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Aspiration contrast in voiceless fricatives

Similar to stops and affricates, Baima fricatives distinguish between voiceless unaspirated, voiceless aspirated, and voiced counterparts. Cross-linguistically, such a distinction is more common in stops and affricates, and rarer in fricatives. Despite its rarity cross-linguistically, the three-way manner contrast in fricatives is commonly attested in Tibetic languages of the Amdo and Kham groups, at the border of which Baima is spoken (see X. Wang 1988, Sh. Wang & Chen 2010 for instrumental studies). In Baima, the three-way contrast between voiceless unaspirated, voiceless aspirated, and voiced fricatives is observed at the alveolar, postalveolar, and alveolo-palatal places of articulation. Figure 8 illustrates the contrast with a minimal set at the alveolar place of articulation: /sâ̠/ 'to itch', /sʰâ̠/ 'earth, soil', /zâ̠/ 'to filter'.

In our corpus, voiceless aspirated fricatives have longer duration than their unaspirated and voiced counterparts. This is shown in Table 3, which compares the mean duration of voiceless unaspirated, voiceless aspirated, and voiced fricatives, as measured in 12 (near) minimal sets of fricatives at the alveolar, postalveolar, and alveolo-palatal places of articulation, together with all remaining examples of words with voiceless aspirated fricatives in our corpus. Duration measures for each fricative were made from the waveform and spectrogram. The onset of noise in the waveform and the presence of high resonant frequencies in the spectrogram were taken as benchmarks for determining the beginning of the fricative. The offset of noise in the waveform and the appearance of vowel formants in the

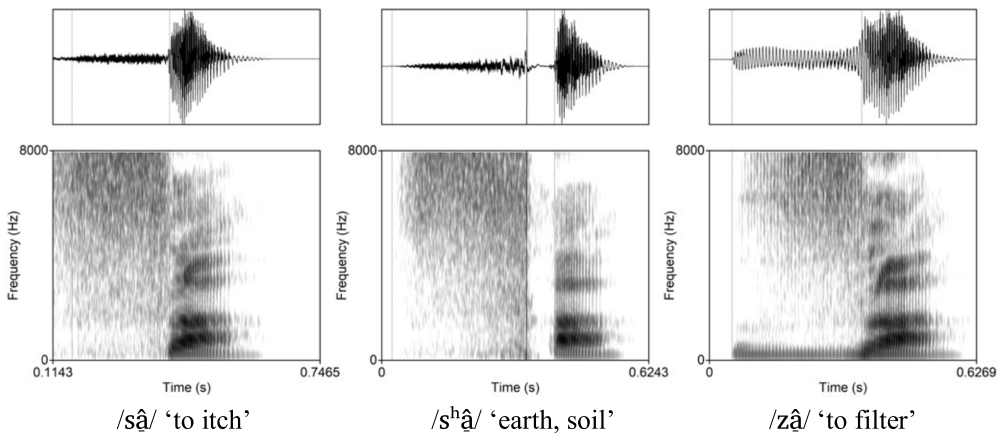

Figure 8 Waveforms and spectrograms for the minimal set /sâ̠/ 'to itch', /sʰâ̠/ 'earth, soil', /zâ̠/ 'to filter'. The dashed lines mark the onset of frication noise and the onset of voicing, the solid line marks the beginning of the aspiration period in the voiceless aspirated fricative /sʰ/.

Table 3 Number of tokens analyzed, mean duration and mean standard deviation (in ms) of voiceless unaspirated, voiceless aspirated, and voiced fricatives at the alveolar, postalveolar, and alveolo-palatal places of articulation.

Fricative	Voiceless unaspirated	Voiceless aspirated	Voiced
N of tokens	49	36	29
Duration (ms)	269	287	271
SD (ms)	43	47	54

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

spectrogram were taken to indicate the end of the fricative and the beginning of the vowel, see Figure 8.

The mean duration of the aspiration period in aspirated fricatives (across all possible vowel contexts) is just 7 ms (SD 9 ms). (In measuring the aspiration duration in aspirated fricatives, the visible decrease in amplitude of the aperiodic waveform, along with the appearance of vowel formants in the spectrogram was taken as benchmark for determining the beginning of the aspiration period, whereas the onset of vowel periodicity was taken to indicate the end of the aspiration period, see Figure 8.) That is considerably shorter than the duration of the aspiration period in aspirated stops and affricates (respectively, 102 ms, SD 17 ms; and 46 ms, SD 12 ms, as based on a small sample of 20 aspirated stops and 20 aspirated affricates).

In our corpus, voiceless aspirated fricatives do not occur before the close vowels /i ʉ ʊ/ (and there is only one example before the close back vowel /u/, /ʃʰù/ 'mushroom'), whereas voiceless unaspirated and voiced fricatives all occur in this environment (e.g. /ʑî/ 'writing, letter, character', /sʉ̄/ 'who', /zʊ̀/ 'to close (door) (PFV/IMP)'). This is noteworthy because the same co-occurrence patterns have been reported for some other Tibetic and non-Tibetic Tibeto-Burman languages that have a similar three-way manner contrast in fricatives (e.g. Sgaw Karen, Salgado, Slavic & Ye 2013; Kami Tibetan, Chirkova 2014).

Compared to fricatives at other places of articulation, Baima velar fricatives /x/ and /ɣ/ have a more restricted distribution. Furthermore, word-initially, /x/ is only found in loanwords from Mandarin. Examples include: /xuâ̠suɛ̂/ 'to think, reflect' (Standard Mandarin, hereafter Mandarin, *huásuàn* 划算 'consider, deliberate'), /xuā̠ᵑguā̠/ 'minority official (in imperial China)' (Mandarin *fānguān* 番官). /x/ also occurs word-medially between two vowels in a few native Baima words. In such cases, it likely represents a historical lenition of /p/, as in /ʃòxù/ [ʃofv̩̀] 'wealthy, rich' (the Written Tibetan etymology for this word is *phyug po*). Realization of /x/ as [f] before the rhyme /u/ is a characteristic feature of the Mandarin dialect of Pingwu (Pingwuxian Xianzhi Bianzuan Weiyuanhui 1997: 900), as well as of Southwestern Mandarin dialects in general (e.g. Yuan 2001: 29). The voiced velar fricative /ɣ/ only occurs before /a̠/, /ɣâ̠/ 'fox'.

Trill

Baima trill /r/ is produced with a series of rapid aerodynamically induced movements of the tongue tip that involve contact of the tip of the tongue against the alveolar ridge. In our data there is a general tendency for the alveolar trill to involve four to five contacts, as shown in Figure 9. A contact, that is, a moment of closure of the oral cavity, is reflected on the spectrograms by reduction of energy.

Syllabic consonants

Some syllables in the corpus have a nucleus that is a voiced prolongation of the initial consonant. These include syllables that begin with the alveolar trill, alveolar, postalveolar, retroflex affricates, and alveolar and postalveolar fricatives. The latter portion of such syllabic consonants tends to be realized as a vowel with a formant structure that is similar to that of the preceding consonant. This generally represents a centralized close-mid vowel. Examples include: [rr̩ə̂] 'snake', [t͡sʰz̩ə̂] 'joint', [ɖ͡rr̩ə̀] 'to roll', [zz̩ə̂] 'leopard', [ʃʒ̩ə̂] 'louse'. Figure 10 provides waveforms and spectrograms of the words [rr̩ə̂] 'snake' and [zz̩ə̂] 'leopard' (see also the word /ɖ͡ɽì/ / [ɖ͡rr̩ə̀] 'to roll' in Figure 6).

Syllabic consonants in Baima are contrastive with syllables with alveolar trill, alveolar, postalveolar, retroflex affricates, and alveolar and postalveolar fricatives onsets followed by

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

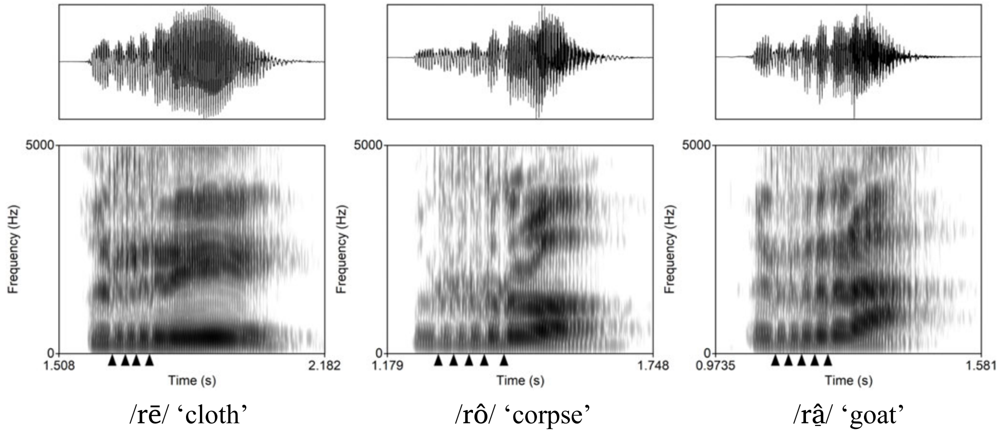

Figure 9 Waveforms and spectrograms for the words /rē/ 'cloth', /rô/ 'corpse', and /rə̂/ 'goat' with arrows indicating contacts between the tip of the tongue and the roof of the mouth.

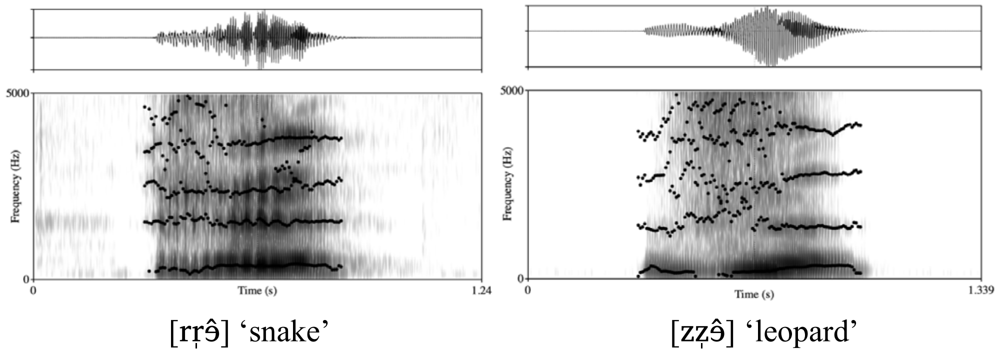

Figure 10 Waveforms, spectrograms, and formant tracks for the words [rr̩ə̂] 'snake' and [zz̩ə̂] 'leopard'.

schwa. Examples include: [rr̩ə̂] 'snake' vs. [rə̂] 'mountain range', [ʃʒ̍ə̂] 'louse' vs. [ʃə̂] 'to die (PFV)', compare also [t͡ʃʒ̍ə̂] 'one' vs. [t͡ʃʰə̂] 'what'.

The exact nature of the nucleus – a genuine vowel, a syllabic approximant (after alveolar trill and retroflex affricate onsets), a syllabic fricative (after alveolar, postalveolar, and alveolo-palatal affricates and fricatives onsets) – is open to varied interpretation. In our analysis, we follow the tradition in Sinitic languages (e.g. Chao 1972 [1948]) to analyze such consonant-like syllabic segments as conditioned variants of high vowels, in the case of Baima, the vowel /i/.

Vowels

Baima has a large vowel inventory consisting of 11 monophthongs, three native diphthongs, and one diphthong that only occurs in loanwords from Mandarin (/ua̠/).

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Monophthongs

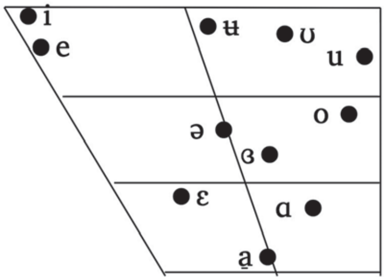

Figure 11 shows the mean F1 and F2 (11a), and F2 and F3 (11b) values for Baima monophthongs, as produced by our speaker. F1 values demonstrate the closed–open domain (from /i/ to /a̠/), F2 the front–back domain (from /i/ to /u/), and F3 the presence of lip rounding (highest degree of lip rounding for /ʉ/ and /ʊ/).

/ʊ/ has two allophones in complementary distribution: [y] and [ʊ]. [y] occurs after alveolar onsets. Examples include: [tỳ] 'poison', [dỳ] 'bracelet', [tʰỳ] 'rice gruel; to be watery', [lỳ] 'to leak, overflow'. [ʊ] occurs after bilabial stops and nasals, and after postalveolar, and retroflex onsets. Examples include [ᵐbʊ̀] 'awl', [ʈʂʰʊ̀] 'woolen cloth', [gʊ̀] 'to wait', [tɕêʃʊ̀] 'knife sheath'.

The open-mid central rounded vowel /ɞ/ mostly appears in the words with the high falling tone (the default voice quality of which is tense), and /o/ mostly appears in the words with the mid and low tones (the default voice quality of which is respectively, modal and lax). However, in our data we have minimal pairs, in which the two vowels are contrasted in words with the high falling tone. Examples include: /kɞ̂/ 'to dig' – /kô/ 'to be full'; /jɞ̂/ 'year' – /jô/ 'road'. For that reason, /ɞ/ and /o/ are analyzed as separate phonemes.

The difference between the open vowels /a̠/ and /ɑ/ is, for many non-native ears, rather subtle. Minimal pairs include /t͡ʃâ̠/ 'tea' vs. /t͡ʃɑ̂/ 'iron', /t͡ʃʰâ̠/ 'pair' vs. /t͡ʃʰɑ̂/ 'to break (INTR)',

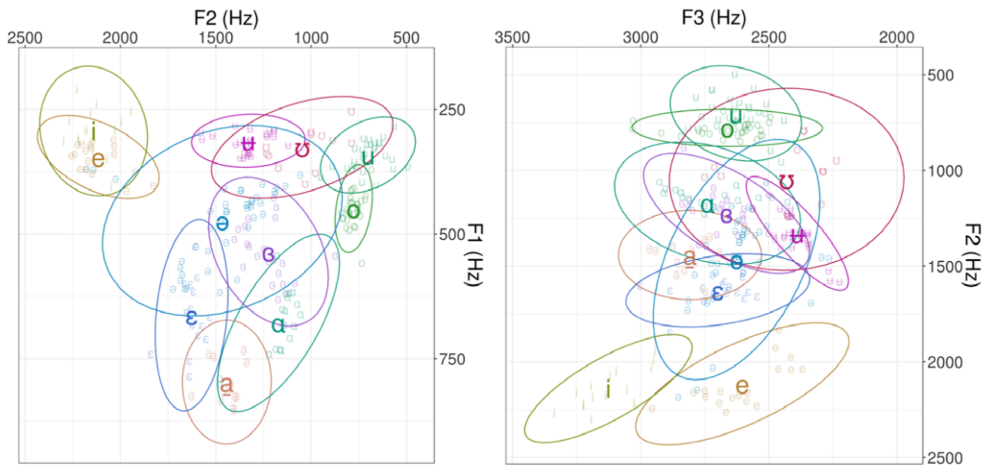

Figure 11 (Colour online) Acoustic vowel plots showing vowel phonemes with ellipses fit to the data and average F1–F2 values (left). Acoustic vowel plots showing vowel phonemes with ellipses fit to the data and average F2–F3 values (right). Formants were sampled at the token midpoint using the website Visible Vowels (https://www.visiblevowels.org). Based on 245 tokens across different initial consonants and tones from the Baima recordings of one male speaker in this Illustration.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

/t͡ɕâ̠/ ‘hair’ vs. /t͡ɕɑ̂/ ‘to weigh (PFV/IMP)’, /d͡ʑâ̠/ ‘hundred’ vs. /d͡ʑɑ̂/ ‘to mend (IPFV)’, /kâ̠/ ‘to be bright’ vs. /kɑ̂/ ‘hoe’, /ʑâ̠/ ‘to borrow (tools)’ vs. /ʑɑ̂/ ‘male yak’.

Table 4 illustrates Baima vowels phonemes by minimal sets of contrasts in the three contrastive tones after alveolar stop initial consonants.

Table 4 Examples of Baima monophthongs in words with alveolar stop initial consonants in the three tones.

	High falling		Mid		Low	
i	dî	‘ogress, a man-eating female demon’			dì	‘to grow’
e	dê	‘to ruminate’	$^{\text{n}}$dē	‘to sunbathe’	dè	‘seven’
ɛ	dɛ̂	‘village’	tɛ̄	‘cushion’	dɛ̀	‘bottom part’
ʊ	dʊ̂	‘bracelet’			dʊ̀	‘to collect’
ə	$^{\text{n}}$də̂	‘this’				
ʉ	dʉ̂	‘to insert’	tʉ̄tʉ̄	‘orphan’		
u	dû	‘hill’	tū	‘to wrap’	dù	‘to pound’
ɜ	dɜ̂ [dʌ̂]	‘stone’				
o			tō	‘weaving knife’	dò	‘to be tense’
â̠	dâ̠	‘to grind (PFV/IMP)’				
ɑ	tɑ̂	‘mark, trace’	tɑ̄	‘banner’	$^{\text{n}}$dɑ̀	‘to grind (IPFV)’

Diphthongs

The three native Baima diphthongs (/iɛ yʉ uɛ/) are plotted in Figure 12 in an F1–F2 space.

Of the three diphthongs that occur in Baima native vocabulary, /uɛ/ and /yʉ/ have broad distribution; they may occur after alveolar stops, affricates, fricatives; postalveolar affricates and fricatives; alveolo-palatal affricates, fricatives, nasal, palatal approximant; and velar stops. In addition, /uɛ/ also occurs as an onsetless syllable, /uɛ̀/ ‘can’ (which word we did not record in isolation, see the appended text). /iɛ/ only occurs after bilabial, alveolar, and velar stops, and only in the mid and low tones.

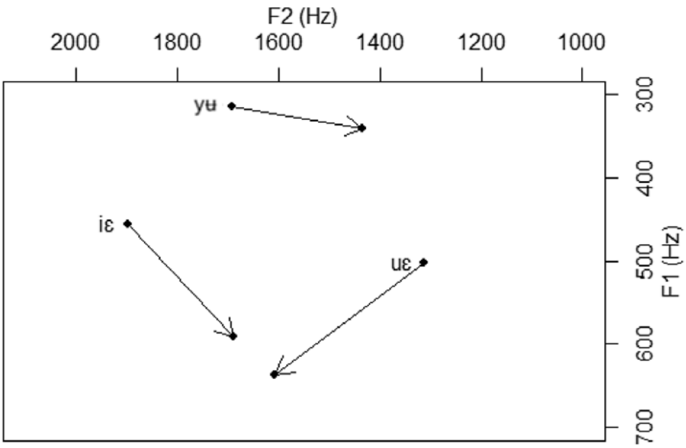

Figure 12 Average F1–F2 values for native Baima diphthongs. Based on 24 tokens.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Table 5 Examples of diphthongs in words with alveolar onsets in the three tones.

	High falling		Mid		Low	
iɛ			tiɛ̄	'patch (of land)'	t^{h}iɛ̀	'ashes'
yʉ	dyʉ̀	'to live, sit (IMP)'	tyʉ̀	'to look (PFV/IMP)'		
uɛ	tuɛ̂	'to be hungry'	tuɛ̄	'smoke'	duɛ̀	'deer'
					t^{h}uɛ̀	'hammer'

The diphthong /uɛ/ is realized as [yɛ] after alveolopalatal onsets, as in [ɕyɛ̄] 'mouse', [ʑyɛ̀] 'knife handle'; and as [uɛ] before other onsets, as in [suɛ̄] 'hemp', [ʒuɛ̀] 'flea'.

Examples of diphthongs in words with alveolar onsets are presented in Table 5. Examples of the diphthong /ua̱/ include: /xuâ̱suɛ̂/ 'to think, reflect' (Mandarin *huásuàn* 划算 'consider, deliberate'), /ʃuâ̱tsî/ 'brush' (Mandarin *shuāzi* 刷子).

Syllable structure

Baima has a simple syllable structure of the (C)V type, where C (which is optional) can be any consonant in the consonant table, and V stands for vowel. Examples include:

V /ò/ 'shoes, boots'

CV /dò/ 'to be tense', /^{n}dò/ 'yesterday', /^{n}d͡ɽò/ 'necklace'

In onsetless syllables with open and open-mid vowels in the low tone, which is correlated with lax (or harsh) phonation (see below), we observe weak voiced frication (variously heard as the phonetic segment [ɦ], [ʕ], or [ʢ]). Examples include: /ɑ̀ò/ [ɦɑ̀ò] 'temple', /ɑ̀pâ̱/ [ʕɑ̀pâ̱] 'wind', /ɛ̀^{n}dʊ̂/ [ʢɛ̀^{n}dŷ] 'to gather, assemble'. Onsetless syllables with the vowels /ə/ and /a̱/ in the high falling tone, which is correlated with tense (or harsh) phonation, may be preceded by a glottal (or pharyngeal, or epiglottal) stop (in some words we also observe creaky phonation on the vowel). Examples include: /â̱pâ̱/ [ʔâ̱pâ̱] 'father', /â̱î/ [ʡâ̱î] 'old women', /ə̂ə̂/ [ʡə̂ʡə̂] 'expression of agreement'. The occurrence of weak voiced frication, on the one hand, and the glottal (pharyngeal or epiglottal) stop, on the other hand, is therefore predictable, and the two are not analyzed as phonemes.[12]

The two Baima approximants, /j/ and /w/, are clearly phonemic before non-close vowels. Examples include: /ɛ̀^{n}dʊ̀/ 'to gather, assemble' – /jɛ̀/ 'to weave'; /ɑ̀ò/ 'temple' – /tʃ^{h}ûjô/ 'buffalo, water ox'; /â̱pâ̱/ 'father' – /jâ̱^{m}bâ̱/ [ȷ̂â̱^{m}pâ̱] 'fool'; /jɑ̄/ 'hand', /wɑ̄/ 'goiter', /ɑ̀ɑ̀/ 'expression of disagreement'.

The distribution of the two approximants in relation to onsetless syllables with close vowels is more difficult to analyze, given the limited number of relevant examples in our corpus. We have a few clear examples of onsetless syllables with close front vowels in monosyllabic and disyllabic words, as in /â̱î/ 'old woman'. We also have examples, where close front vowels are preceded by the approximant /j/, with a clear segment boundary (similar to its realization before close-mid, open-mid, and open vowels). Examples include /jyə̂/ 'to yawn', /jyɛ̄/ 'lung'. In such words, /j/ is characterized by lesser energy than that of the following vowel, and it is marked by frication in addition to voicing. Overall, words with the onset /j/

[12] We are grateful to the editors for their comments and references to comparable distributional patterns in other languages (Shanghai Chinese, Chen & Gussenhoven 2015; Goemai, Tabain & Hellwig 2015; Tongan, Garellek & Tabain 2020) that prompted us to re-examine our original analysis of the syllable-initial weak voiced frication and glottal, pharyngeal or epiglottal stop.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

tend to have a greater degree of frication in the high falling tone than in the mid and low tones (compare the words /jyə̂/ 'to yawn'; /jê/, /â̠jê/ 'to be bad'; /jyɛ̄/ 'lung'; /jɛ̀/ 'to weave').

In the case of monosyllabic morphemes with close back vowels, in the vast majority of cases, /ʊ/ and /u/ in monosyllabic and disyllabic words are preceded by an approximant-like element. (One exception is the word /uɛ̀/ [uɛ̀] 'can' in the appended text.) Here again, in words in the high falling tone, the approximant-like element tends to have a greater degree of frication than in words with the mid and low tones. Examples include /wû/ [vû] 'rice polishings', /wʊ̄pâ̠/ [ʋʊ̄pâ̠] 'bellows'. In our corpus, we also have examples of onsetless [u], as in /mōû/ [mʌ̄v̩̂] 'much, many'.

In sum, the data that we have at our disposal are inconclusive as to the distribution of glides in relation to onsetless syllables with close vowels. It is possible that some glides are phonetically-conditioned variants, preceding close vowels in vowel-initial syllables, but more data are necessary to support this conclusion.

Prosodic organization

Monosyllabic words

Baima is a tone language with a complex interrelation between (a) pitch height, (b) voice quality of the vowel, and (c) vowel length.

Baima has three contrastive tones, which in a prepausal position all end in a falling pitch contour. The high falling tone has a falling contour throughout the vowel duration. The mid tone is level throughout most of its duration, with a small fall at the end. The low tone is low rising. This is illustrated in Figure 13 with the minimal triplet /ɲɛ̂/ 'fire', /ɲɛ̄/ 'to sleep', and /ɲɛ̀/ 'mountain god', uttered in isolation by our male speaker (three repetitions per word). The figure presents mean f0 and duration across three repetitions. The f0 values were extracted at every ten percent of the duration, from 10% to 90%, and at 2% and 98%. The latter was selected in order to exclude the extremes time points of 0% and 100%, which were preceded or followed by silence.

When followed by another syllable, the mid and low tones are realized as mid and low level, respectively. Examples include: /tʉ̄tʉ̄/ 'orphan', /xuā̠ᵑguā̠/ 'minority official (in imperial China)'; /ʃòxù/ 'wealthy, rich', /pʰɛ nô/ 'in the belly' (from /pʰɛ̀/ 'belly; stomach', /nô/ 'inside'). The high falling tone consistently has a falling pitch contour, both

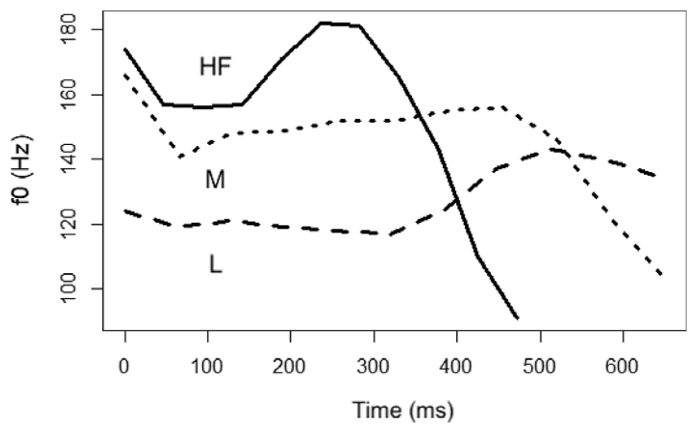

Figure 13 Pitch contours and vowel duration in the three tones. Based on the minimal triplet /ɲɛ̂/ 'fire', /ɲɛ̄/ 'to sleep', and /ɲɛ̀/ 'mountain god' (mean of three repetitions per word, a total of 9 tokens), uttered in isolation by our male speaker. High falling tone (HF) in solid line, mid tone (M) in dotted line, low tone (L) in dashed line.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

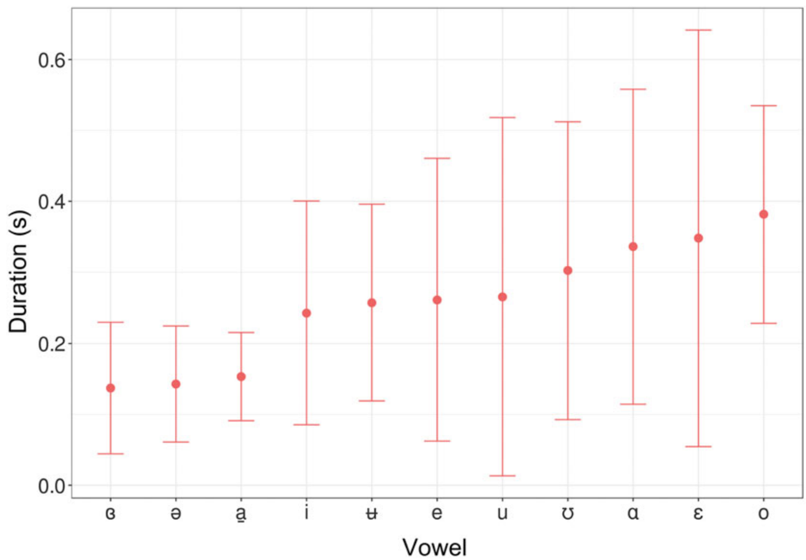

Figure 14 (Colour online) Mean duration and mean standard deviation (in s) of Baima monophthongs, based on 245 tokens across different initial consonants and tones from the Baima recordings of one male speaker in this Illustration.

in a prepausal position and when followed by another syllable. Examples include: /tôpû/ 'Baima township'; /jâ̠ᵐbâ̠/ 'fool'. When followed by another syllable, the high falling tone is frequently realized as mid falling. Examples include:

/pêkê/ [pe³¹ke⁵³] 'Baima, Tibetan language', from /pê/ 'Tibet' + /kê/ 'sound'

/t͡ʃʰûŋgô/ [t͡ʃʰu³¹ŋgo⁵³] 'spring, source of a river', from /t͡ʃʰû/ 'water' + /ᵑgô/ 'head'

As can be seen in the minimal set in Figure 13, words in the high falling tone are shorter, whereas words in the mid and low tones are longer. Consecutive temporal analysis revealed that the mean vowel duration (mean of three repetitions) is 225 ms in the high falling tone, 503 ms in the mid tone, and 511 ms in the low tone. Relative vowel duration hence cross-cuts the three contrastive pitch categories. Differences in vowel duration are further illustrated in Figure 14, which shows mean duration of Baima monophthongs in the same sample of 245 tokens, as used in the plots in Figure 11 above.

The three shortest vowels in Figure 14 (/ə ɞ a̠/) only occur in the high falling tone.[13] The longest vowel in Figure 14 (/o/) mainly occurs in the mid and low tones, but it is also contrastive with the short vowel /ɞ/ in the high falling tone (see examples in the section on vowels). Overall, the eight vowels that occur in all three tones (that is, /i e ɛ ʉ ʊ u o ɑ/) are shorter in the high falling tone, and longer in the mid and low tones.[14] The eight vowels that occur in all three tones do not differ in vowel quality.

[13] Written Tibetan etymologies of Baima words with these three vowels suggest that they have developed from OT syllables with zero codas. Examples include: 'four': Baima /ʒə̂/, WT *bzhi*; 'male': Baima /pʰɞ̂/, WT *pho*; 'father': Baima /â̠pâ̠/, WT *a p*(*h*)*a*.

[14] This reflects the origin of all vowels in the high falling tone but /ə ɞ a̠/ in Old Tibetan syllables with stop codas (*-b*, *-d*, *-g*); and the origin of the vowels in the mid and low tones in OT syllables with the codas *-m*, *-n*, *-ng*, *-s*, *-l*, and from coalesced syllables. Examples include: 'mark, trace': Baima /tɑ̂/, WT *rtags*; 'food, cooked rice': Baima /sō/, WT *za ma*; 'to pound': Baima /dù/, WT *rdung*.

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

Table 6 Examples of tone contrasts on words with sonorant and prenasalized stop and affricate onsets.

Tone / Initial	High falling: Non-epiglottalized	High falling: Epiglottalized	Mid	Low
m		mˤɑ̂ 'army'	mɑ̄ 'butter'	mɑ̀ 'son-in-law'
n	nô 'inside'		nō 'heaven, sky'	nò 'to exist'
ᵑg	ᵑguɛ̂ 'to bark'		ᵑguɛ̄ 'pan'	ᵑguɛ̀ 'downstairs'
ⁿd͡ʒ		ⁿd͡ʒˤû 'snout'	ⁿd͡ʒū 'thrush'	ⁿd͡ʒù 'to jump'

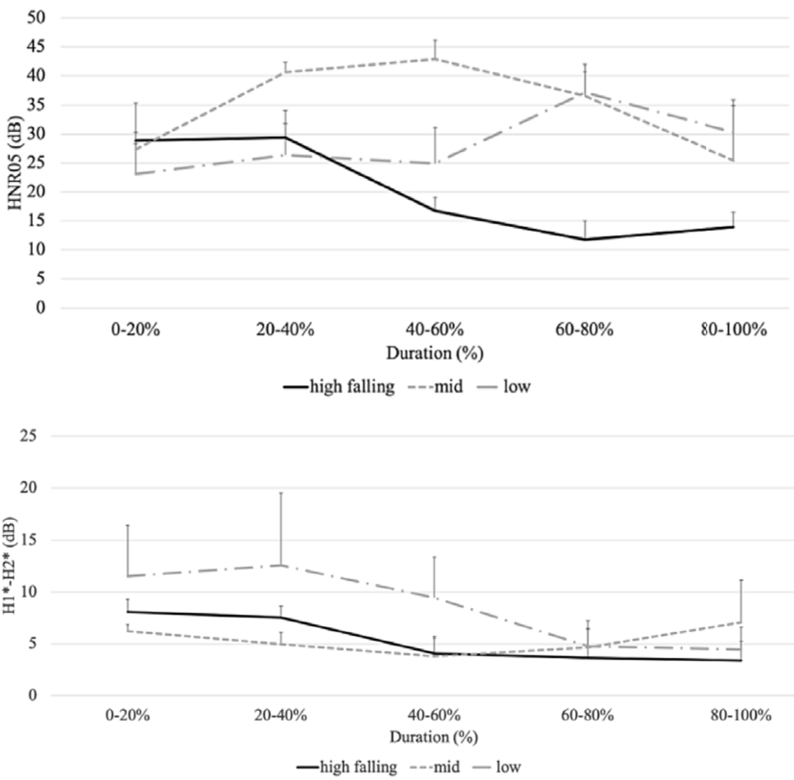

Figure 15 Time courses of HNR05 (dB) (left) and H1*–H2* (dB) (right) in two minimal triplets for tone: (i) /ɲɛ̂/ 'fire', /ɲɛ̄/ 'to sleep', /ɲɛ̀/ 'mountain god', and (ii) /nô/ 'inside', /nō/ 'heaven, sky', /nò/ 'to exist' (with three repetitions per word, a total of 18 tokens). The x-axis represents vowel duration divided into five regions: 0–20%, 20–40%, 40–60%, 60–80%, 80–100%. High falling tone in black solid line, mid tone in light gray dotted line, low tone in gray dashed line.

Syllables with sonorant and prenasalized stops and affricate onsets allow the full set of tonal contrasts. Table 6 provides additional examples of minimal triplets for tone in our corpus.

Phonation differences in the three tones are illustrated in Figure 15 on the basis of two minimal triplets for tone: (i) /ɲɛ̂/ 'fire', /ɲɛ̄/ 'to sleep', /ɲɛ̀/ 'mountain god', and (ii) /nô/ 'inside', /nō/ 'heaven, sky', /nò/ 'to exist' (with three repetitions per word, a total of 18 tokens).

Figure 15 shows that the mid tone is characterized by the highest HNR05 and the lowest H1*–H2* values in the set, suggesting periodic voicing, low level of noise, and therefore, modal voice quality. Vowels in the low tone are characterized by the highest H1*–H2* values

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

in the set, and HNR05 values that are lower than in the vowels in the mid tone, but higher than in the vowels in the high falling tone (suggesting some noise and somewhat irregular voicing). Given these characteristics, vowels in the low tone can be defined as having breathy-like, lax voice quality. Finally, vowels in the high falling tone are characterized by the lowest HNR05 values in the set, suggesting weak voicing and high levels of noise. H1*–H2* values in vowels in the high falling tone are higher than in modal vowels in the mid tone, but lower than in lax vowels in the low tone. They are most tense-like in the set.

The non-modal phonation type that co-occurs with the low tone can vary from lax (as in /lʊ̀/ 'to leak, overflow', /gʊ̀/ 'to wait') to harsh (as in /jɛ̀/ 'to weave' or /dɛ̀/ 'bottom part'). Alternation between lax and harsh phonation is likely a by-product of lax (breathy) phonation with a low tone, in which a speaker tends to lower his larynx (see Liu et al. 2020: 13 for detailed explanation and references therein). Overall, the association of low tone with breathy phonation is commonly observed in Tibetic languages of China (e.g. J. Sun 1997, 2001; Huber 2005), Bodic languages outside China (e.g. Mazaudon & Michaud 2008) as well as, more broadly, in languages of other subgroups of the Tibeto-Burman language family (e.g. Jianchuan Bai, Edmondson et al., published online 27 April 2020) and in languages of other language families of Southeast Asia (e.g. Hmu, Hmong-Mien, Liu et al. 2020).

Polysyllabic words and compounds

Baima morphemes (lexical and grammatical) are generally monosyllabic, but words are generally disyllabic. Polysyllabic words are mostly composite, formed through compounding. Content morphemes in Baima are specified for lexical tone (e.g. /rē/ 'cloth'; /ᵐbû/ 'insect'; /kʰâ̰/ 'mouth'; /kʰɑ̀/ 'snow'). Grammatical morphemes (such as the indefinite marker /ʃi/ or the conjunctions /rɛ/ and /ɲi/ in the appended text) do not occur in isolation, and their lexical tone cannot be determined. In connected speech, grammatical morphemes are prosodically weak, they usually have shorter duration than content morphemes and no fixed tone value (tone on grammatical morphemes is not marked in the appended text).

In polysyllabic words, each syllable bears its own (etymological) tone (correlated with a specific phonation type). Table 7 provides examples of all nine possible combinations of the three contrastive tones in disyllabic words.

Table 7 Examples of disyllabic words with nine possible combinations of tones.

Tone	High falling	Mid	Low
High falling	/jâ̰ᵐbâ̰/ 'fool'	/ⁿd͡ʒˤûpɑ̄/ 'lips'	/sâ̰jò/ 'early'
Mid	/wʊ̄pâ̰/ 'bellows'	/t͡ʈʂōᶯd͡ɖʐɑ̄/ 'fragrant'	/rūwɑ̀/ 'one household'
Low	/ɑ̀pâ̰/ 'wind'	/ʃʰuɛ̀rō/ 'dung fly'	/ⁿdɑ̀ⁿdɑ̀/ 'persimmon'

When words combine into compounds, no tonal change occurs, each word preserves its etymological tone. Consider the following examples: /t͡ɕʰĩ̂ŋuɛ̄ tsâ̰/ 'foxtail grass', from /t͡ɕʰĩ̂/ 'dog', /ŋuɛ̄/ 'tail', /tsâ̰/ 'grass'; /â̰î nɑ̄/ 'agaric, wood ear mushroom', from /â̰î/ 'old woman', /nɑ̄/ 'ear'; /ɕâ̰ rɛ ʃʰɑ̀/ 'game and wild animals', from /ɕâ̰/ 'bird; chicken', /ʃʰɑ̀/ 'wild animals' (/rɛ/ conjunction, does not occur in isolation).

The appended text in this Illustration provides several examples of words 'in focus' position (sentence-level stress). The phonetic correlates of phrasal prominence in Baima are: (i) a distinctive rising pitch contour (pitch accent), overlaid over the etymological tone of the word in focus, and (ii) longer duration of the vowel. Take as an example the sentence [ʑè sʉ̌ː ᶮd͡ʑɛ̂ ʃi tɛ̂ tâ̰ ʃâ̰ t͡ʈʂû dzɛ̀ ɲi] 'let's see which one of the two of us will win', where the word is focus is /sʉ̄/. This word is lengthened and carries a rising pitch contour, which is distinct from its

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

mid-level pitch contour in isolation (/sʉ̄/ 'who'). In the appended text, the intonational pitch accent is marked by a rising tone and the length diacritic on the vowel ('ǎː').

Transcription of the recorded text 'The North Wind and the Sun'

The original audio and video recordings (made with a Zoom H6 Handy Recorder, an AKG C520L headset microphone) have been made available to the *JIPA* along with this analysis.

Transcription of recorded passage with interlinear morphemic glossing

This passage is transcribed phonemically, using the symbols presented in the consonant and vowel charts. Abbreviations used in interlinear glossing follow the Leipzig Glossing Rules (LGR, http://www.eva.mpg.de/lingua/resources/glossing-rules.php). Non-standard abbreviations (those not included in the LGR) are: CONJ = conjunction, EGO = egophoric, HOR = hortative, ITRJ = interjection, ITSF = intensifying verbal marker.[15]

tôᵑgǒː |	â̠kû	ʃi	tə |	ɲyɛ̄	rɛ	ɑ̀pâ̠
in.the.past	time	IND	one?	sun	CONJ	wind

ɲˤî	tɛ̂ |	ʈ͡ɽʰê	ɕɛ̄	ɲi |	kʰâ̠sʰɑ̂ ||
two	DEF	meet.PFV/IMP	do.PFV	CONJ	dispute

ɲˤî-ᵑgə̂	dzɛ̀	ɲi |	ʑè	sʉ̌ː |	ᶮd͡ʑɛ̂	ʃi	tɛ̂
two-head.CLF	say.PFV	CONJ	1DU	who	win	PFV	DEF

tâ̠	ʃâ̠	ʈ͡ɽû	d͡zɛ̀	ɲi |	ɑ̀pâ̠…		ɑ̀pâ̠	i
look.IPFV	HOR	that.way	say.PFV		CONJ	wind	wind	ERG

d͡zɛ̀	ə̂ə̂	d͡zɛ̀	ʃi ||
say.PFV	expression.of.agreement	say.PFV	PFV

tò	ɑ̀pâ̠	tə |	kʰʉ̂	ŋˤʉ̌ː	zô	ʃâ̠ |	ə̂ = uɛ̂ ||
then	wind	DEF	LOG	front	make.IPFV	HOR	Q=can

ɑ̀pâ̠	ʃǐː-tɕō	ɕɛ̄ |	ɲə̂	tɛ̂	ʈ͡ɽʰê
wind	ITSF-blow	do.PFV	person	DEF	meet.PFV/IMP

ɕɛ̄ |	ɲə̂	ʃi	nɑ̂	wē	də	ɲi |
do.PFV	person	IND	here	come.IPFV	PROG	CONJ

ɲə̂	tɛ̂	ʈ͡ɽʰê	ɕɛ̄	ɲi |	ɑ̀pâ̠	ʃìː-tɕō
person	DEF	meet.PFV/IMP	do.PFV	CONJ	wind	ITSF-blow

[15] Analysis of evidentiality markers is based on Chirkova (2017).

https://doi.org/10.1017/S0025100321000219 Published online by Cambridge University Press

ɕɛ̄ | ɲə̂ tɛ̂ ə̂wʊ̂p^{h}ûtsə̂ tə ʃĭː-^{n}dû ɕɛ̄ ^{n}dǔː |
do.PFV person DEF cloak DEF SUP-wrap do.PFV wrap

ɑ̀pâ̠ tɕō tɕō | k^{h}ʉ̂ tâ̠ ɲə̂ lɛ̂ kʉ̄ lɛ̂
wind blow blow LOG now person that clothes that

^{m}bì gâ̠rɛ̀ mâ̠ = rè | k^{h}ʉ̂ tâ̠ lɘ̂ mû =
take.off.IPFV a.little? NEG=COP LOG now idea NEG

nò | ^{n}də̂ rè d͡zɛ̀ ʃi || ɲyɛ̄ i d͡zɛ̀ | tâ̠
exist.N-EGO that COP say.PFV PFV sun ERG say.PFV now

k^{h}ʉ̂ mə k^{h}ʉ̂ zɘ̂ gô-rè tî || ɔ̂̀-tɕâ̠ tō
LOG ITRJ LOG do.IPFV want-COP DEF.GEN again-? then

ɲyɛ̆ː ʃĭː-mbɛ̂ ɕɛ̄ | ɲə̂ tə̂ ŋă̠ː i tə̂ wē |
sun ITSF-blow do.PFV person DEF scald GEN DEF come.PFV

ɲə̂ tə̂ ə̂wʊ̂ tə̂ mbì-s^{h}â̠ re ^{m}bì ɲi |
person DEF cloak DEF take.off-quick? COP take.off CONJ

tò ɑ̀pâ̠ d͡zɛ̀ | ej ɲyɛ̄ t͡ɕ^{h}ʉ̂ uɛ̀ t͡ɕ^{h}ɛ̀ ʃi |
then wind say.IPFV ITRJ sun 2SG can go.PFV PFV

t͡ɕ^{h}ʉ̂ uɛ̀ t͡ɕ^{h}ɛ̀ ʃi | t͡ɕ^{h}ʉ̂ ^{n}d͡ʑɛ̂ ^{m}bo ʃi | ŋā̠
2SG be.capable go.PFV PFV 2SG win COMPL PFV 1SG

k^{h}ô ^{m}bo ʃi | t͡ʈû d͡zɛ̀ ʃi ||
lose call.IPFV PFV that.way say.PFV PFV

Acknowledgments

We would like to thank our Baima language consultant, Mr. Li Degui for his patience, assistance, and continuous support of our work. We are indebted to the journal's editors Marc Garellek and Marija Tabain and the three anonymous reviewers for their detailed reviews and insightful suggestions, which helped to improve both the analysis and the presentation of the Illustration. Thanks are also due to Radek Skarnitzl for his helpful suggestions and to Paul van Els for creating Figure 1 in this paper. We gratefully acknowledge the financial support of the OP VVV project no. CZ.02.2.69/0.0/0.0/17_050/0008466 for Tanja Kocjančič Antolík. This work is also partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the program 'Investissements d'Avenir' (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris - ANR-18-IDEX-0001.

Supplementary material

To view supplementary material for this article (including audio files to accompany the 752 language examples), please visit https://doi.org/10.1017/S0025100321000219.

References

Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer. Version 6.1.28, retrieved 20 October 2020 from http://www.praat.org/.

Chao, Yuen Ren. 1972 [1948]. *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press & London: Oxford University Press. [First published in 1948 by Harvard University Press.]

Chen, Yiya & Carlos Gussenhoven. 2015. Shanghai Chinese. *Journal of the International Phonetic Association* 45(3), 321–337.

Chirkova, Katia. 2007. Between Tibetan and Chinese: Identity and language in Chinese South-West. *Journal of South Asian Studies* 30(3), 405–417.

Chirkova, Katia. 2014. Kami. In Jackson T.-S. Sun (ed.), *Phonological profiles of little-studied Tibetic varieties* (Language and Linguistics Monograph Series 55), 1–75. Taipei: Academia Sinica, Institute of Linguistics.

Chirkova, Katia. 2017. Evidentials in Pingwu Baima. In Lauren Gawne & Nathan W. Hill (eds.), *Evidential systems of Tibetan languages*, 445–459. Berlin & Boston, MA: De Gruyter Mouton.

Chirkova, Katia. 2021. Tonal developments in Baima. Presented at the Sixth Workshop on Sino-Tibetan Languages of Southwest China, Kobe City University of Foreign Studies, Kobe, Japan, 7–11 September 2021.

Chirkova, Katia, Wang Dehe, Yiya Chen, Angélique Amelot & Tanja Kocjančič Antolík. 2015. Ersu. *Journal of the International Phonetic Association* 45(2), 187–211.

Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185–216.

Edmondson, Jerold A., John H. Esling & Li Shaoni. Jianchuan Bai. *Journal of the International Phonetic Association*, doi: 10.1017/S0025100319000379. Published online by Cambridge University Press, 27 April 2020.

Evans, Jonathan P. 2006. Vowel quality in Hongyan Qiang. *Language and Linguistics* 7(4), 937–960.

Evans, Jonathan P. & Jackson T.-S. Sun. 2015. Qiāng 羌 language. In Rint Sybesma (editor-in-chief), *Encyclopedia of Chinese language and linguistics*. doi:http://dx.doi.org/10.1163/2210-7363_ecll_COM_00000347 (accessed on 25 May 2021).

Evans, Jonathan P., Jackson T.-S. Sun, Chenhao Chiu & Michelle Liou. 2016. Uvular approximation as an articulatory vowel feature. *Journal of the International Phonetic Association* 46, 1–31.

Feng, Yun. 2017. *The matching rules of the late Chinese loanwords in Pingwu Baima language*. Singapore: National University of Singapore, ProQuest Dissertations Publishing.

Garellek, Marc. 2019. The phonetics of voice. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics*, 75–106. New York: Routledge.

Garellek, Marc. 2020. Acoustic discriminability of the complex phonation system in !Xóõ. *Phonetica* 77, 131–160.

Garellek, Marc & Marija Tabain. 2020. Tongan. *Journal of the International Phonetic Association* 50, 1–11.

Gesang Jumian (skal bzang 'gyur med) 格桑居冕 & Gesang Yangjing 格桑央京 (skal bzang dbyangs chan). 2002. *Zangyu Fangyan Gailun*《藏语方言概论》*/Bod kyi yul skad rnamb shad/An introduction to Tibetan dialects*. Beijing 北京: Nationalities Press 民族出版社.
Gordon, Matthew, Paul Barthmaier & Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32(2), 141–174.
Hamann, Silke. 2003. *The phonetics and phonology of retroflexes*. Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).
Hu, Tan 胡坦. 1991. Zangyu《藏语》[The Tibetan language]. In Ma Xueliang 马学良 (ed.), *Han-Zang Yu Gailun* 《汉藏语概论》[A general introduction to Sino-Tibetan languages], vol. 2, 109–207. Beijing: Peking University Press 北京大学出版社.
Huang, Bufan 黄布凡 & Minghui Zhang 张明慧. 1995. Baima zhishu wenti yanjiu 白马话支属问题研究 [A study of the genetic affiliation of Baima]. *Zhongguo Zangxue* 《中国藏学》 [Tibetology in China] 2, 79–118.
Huber, Brigitte. 2005. *The Tibetan dialect of Lende (Kyirong): A grammatical description with historical annotations* (Beiträge zur tibetischen Erzählforschung 15). Bonn: VGH Wissenschaftsverlag. [Revised version of Ph.D. thesis, Universität Bern, 2002.]
Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108, 1252–1263.
Lee, Chao-Yang, Yu Zhang & Ximing Li. 2014. Acoustic characteristics of voiceless fricatives in Mandarin Chinese. *Journal of Chinese Linguistics* 42, 150–171.
Li, Lan 李蓝. 2012. B1-10, 11, 12 Guanhua zhi shiyi shi er Xinan Guanhua (Hubei, Hunan, Sichuan, Chongqing, Shanxi, Yunnan, Guizhou, Guangxi Zhuangzu Zizhiqu) B1-10-11-12 官话之十一十二 西南官话（湖北、湖南、四川、重庆、陕西、云南、贵 州、广西壮族自治区）[Mandan dialects Maps 11-12. Southwestern Mandarin (Hubei, Hunan, Sichuan, Chongqing, Shaanxi, Yunnan, Guizhou, Guangxi Zhuang Autonomous District)]. In Zhongguo Shehui Kexueyuan Yuyan Yanjiusuo 中国社会 科学院语言研究所, Xianggang Chengshi Daxue Yuyan Zixun Kexue Yanjiu Zhongxin 香港城市大学语言资讯科学研究中心 (eds.), *Zhongguo Yuyan Dituji* 《中 国语言地图集》[The language atlas of China, 2nd edn.], 82–91. Beijing: Commerical Press 商务印书馆.
Liu, Wen, You-Jing Lin, Zhenghui Yang & Jianping Kong. 2020. Hmu (Xinzhai variety). *Journal of the International Phonetic Association* 50(2), 240–257.
Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: A case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65(4), 231–256.
Nishida, Tatsuo 西田龍雄 & Hongkai Sun 孙宏开. 1990. *Hakuba yakugo no kenkyū: Hakuba no kōzō to keitō* 《白馬譯語の研究:白馬語の構造と系统》 [A study of the Baima–Chinese vocabulary 'Baima yiyu': The structure and affiliation of the Baima language]. Kyoto 京都: Shokado 松香堂.
Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer.
Pingwuxian Xianzhi Bianzuan Weiyuanhui 平武县县志编纂委员会 (eds.). 1997. *Pingwuxian Zhi* 平武县志 [Pingwu Gazetteers]. Chengdu 成都: Sichuan Kexue Jishu Chubanshe 四川科学技术出版社.
Renzeng, Wangmu 仁增旺姆. 1987. Zangyu Xiahehua yuyin yu Zangwen de yuyin duiying 藏语夏河话语音与藏文的语音对应 [On sound correspondences between Xiahe (bLa.brang) and written Tibetan]. *Minzu Yuwen* 《民族语文》 [Minority languages of China] 6, 46–53.
Salgado, Hugo, Jessica Slavic & Zhao Ye. 2013. The production of aspirated fricatives in Sgaw Karen. *Studies in the linguistic sciences* (Illinois Working Papers 2013), 148–161.
Shue, Yen-Liang, Patricia A. Keating, Chad Vicenik & Kristine Yu. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of the Phonetic Sciences* (ICPhS XVII), 1846–1849. Available at: http://www.ee.ucla.edu/~spapl/voicesauce/.
Sichuan Sheng Minzu Yanjiusuo 四川省民族研究所 [Sichuan Nationalities Research Institute]. 1980. *Baima Zangren zushu wenti taolunji* 《白马藏人族属问题讨论集》 [A collection of discussions on the question of the ethnic classification of Baima Tibetans]. Chengdu: Sichuan Sheng Minzu Yanjiusuo.
Spajić, Siniša, Peter Ladefoged & P. Bhaskararao. 1996. The trills of Toda. *Journal of the International Phonetic Association* 26(1), 1–21.

Sun, Hongkai 孙宏开. 1980. Baima ren de yuyan 白马人的语言 [The language of the Baima people]. In Sichuan Institute of Minority Nationalities 四川民族研究所编 (eds.), Baima Zangren zushu wenti taolunji 白马藏人族属问题讨论集 [A collection of discussions on the question of the ethnic classification of Baima Tibetans], 15–25. Chengdu: Sichuan Minzu Yanjiusuo.

Sun, Hongkai 孙宏开. 2003. Baimayu shi Zangyu de yige fangyan huo tuyu ma? 白马语是藏语的一个方言或土语吗? [Is Baima a dialect or vernacular of Tibetan?]. *Yuyan Kexue* 《语言科学》 [Linguistic Sciences] 1(2), 65–75.

Sun, Hongkai, 孙宏开, Katia Chirkova (Qi Kajia) 齐卡佳 Liu Guangkun 刘光坤. 2007. *Baimayu Yanjiu* 《白马语研究》 [A grammar of Baima]. Beijing: Nationalities Press 民族出版社.

Sun, Jackson T.-S. 孫天心. 1997. The typology of tone in Tibetan. *Proceedings of the symposium Chinese Languages and Linguistics IV: Typological studies of languages in China* (Symposium Series of the Institute of History and Philology–Academia Sinica), 485–521. Taipei: Taiwan.

Sun, Jackson T.-S. 孫天心. 2001. Variegated tonal developments in Tibetan. In David Bradley, Randy LaPolla, Boyd Michailovsky & Graham Thurgood (eds.), *Language variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 35–51. Canberra: Pacific Linguistics.

Sun, Jackson T.-S. 孫天心. 2003. Qiuji Zangyu de yuyin tezheng 求吉藏语的语音特征 [Phonological profile of Chos-rje Tibetan]. *Minzu Yuwen* 《民族语文》 [Minority languages] 6, 1–6.

Sun, Jackson T.-S. 孫天心 (ed.). 2014. *Phonological profiles of little-studied Tibetic varieties* (Language and Linguistics Monograph Series 55). Taipei: Academia Sinica, Institute of Linguistics.

Suzuki, Hiroyuki. 2015. New perspectives on the suprasegmentals in mBrugchu Tibetan: An introduction to the tonogenesis triggered by breathy voice. *Bulletin of Chinese Linguistics* 8(1), 117–139.

Tabain, Marija & Birgit Hellwig. 2015. Goemai. *Journal of the International Phonetic Association* 45(1), 81–104.

Tournadre, Nicolas. 2005. L'aire linguistique tibétaine. *Lalies* 25, 7–56. Paris: Presses de l'Ecole Normale Supérieure.

Tournadre, Nicolas. 2014. The Tibetic languages and their classification. In Thomas Owen-Smith & Nathan W. Hill (eds.), *Trans-Himalayan linguistics: Historical and descriptive linguistics of the Himalayan area* (Trends in Linguistics Studies and Monographs 266), 105–129. Berlin: Mouton de Gruyter.

Yuan, Jiahua 袁家骅. 2001. *Hanyu Fangyan Gaiyao* 《汉语方言概要》 [Outline of Chinese dialects]. Beijing: Yuwen Chubanshe 语文出版社. [First published in 1960.]

Wang, Shuangcheng 王双成 & Zhongmin Chen 陈忠敏. 2010. 安多藏语送气擦音的实验研究 [An experimental study of aspirated fricatives in Amdo Tibetan]. *Minzu Yuwen* 《民族语文》 [Minority languages of China] 2, 7–15.

Wang, Xianhai 王贤海. 1988. Guonei jizhong sahoshu minzu yuyan songqi cayin shiyan yanjiu 国内集中少数民族语言送气擦音实验研究 [An experimental study of aspirated fricatives in several minority languages]. *Minzu Yuwen* 《民族语文》 [Minority languages of China] 1, 38–46.

Zeng, Weiyi 曾维益 & Youyuan Xiao 萧猷源 (eds.). 1987. *Baimaren zushu yanjiu wenji* 《白马人族属研究文集》 [A collection of research papers on the genetic affiliation of the Baima]. Pingwu: Pingwu Xian Baimaren zushu yanjiuhui 平武县白马人族属研究会.

Zhang, Jichuan 张济川. 1994a. Baimahua yu Zangyu, shang 白马话与藏语 （上） [Baima and Tibetan, part 1]. *Minzu Yuwen* 《民族语文》 [Minority languages of China] 2, 11–24.

Zhang, Jichuan 张济川. 1994b. Baimahua yu Zangyu, xia 白马话与藏语 （下） [Baima and Tibetan, part 1]. *Minzu Yuwen* 《民族语文》 [Minority languages of China] 3, 58–67.