Open Library of Humanities

Development of a new vowel feature from coarticulation: Biomechanical modeling of rhotic vowels in Kalasha

Jeff Mielke*, Department of English, North Carolina State University, Raleigh, NC, USA, jimielke@ncsu.edu

Qandeel Hussain, Department of English, North Carolina State University, Raleigh, NC, USA, qandeel.hussain@students.mq.edu.au

Scott R. Moisik, Nanyang Technological University, Singapore, scott.moisik@ntu.edu.sg

*Corresponding author.

Coarticulation is an important source of new phonological contrasts. When speakers interpret effects such as nasalization, glottalization, and rhoticization as an inherent property of a vowel, a new phonological contrast is born. Studying this process directly is challenging because most vowel systems are stable and phonological change likely follows a long transitional period in which coarticulation is conventionalized beyond its mechanical basis. We examine the development of a new vowel feature by focusing on the emergence of rhotic vowels in Kalasha, an endangered Dardic (Indo-Aryan) language, using biomechanical and acoustic modeling to provide a baseline of pure rhotic coarticulation.

Several features of the Kalasha rhotic vowel system are not predicted from combining muscle activation for non-rhotic vowels and bunched and retroflex approximants, including that rhotic back vowels are produced with tongue body fronting (shifting the backness contrast to principally a rounding contrast). We find that synthesized vowels that are about 30% plain vowel and 70% rhotic are optimal (i.e., they best approximate observed rhotic vowels and also balance the acoustic separation among rhotic vowels with the separation from their non-rhotic counterparts). Otherwise, dispersion is not generally observed, but the vowel that is most vulnerable to merger differs most from what would be expected from coarticulation alone.

1. Introduction

This is an investigation of what happens when a language develops an entirely new vowel feature. It is difficult to observe how various articulatory parameters such as height, backness, or rounding are added to a vowel system, because the vowel systems of most languages already employ them. Kalasha, an endangered Dardic (Indo-Aryan) language has five contrastive pairs of plain (/i e a o u/) and rhotic vowels (/ɨ˞ e˞ a˞ o˞ u˞/), as well as nasalized counterparts of all of these (/ĩ ẽ ã õ ũ ɨ̃˞ ẽ˞ ã˞ õ˞ ũ˞/) (Cooper, 2005; Hussain & Mielke, 2020;. Kochetov, Arsenault, Petersen, Kalas, & Kalash, 2021). The rhotic vowels are thought to have developed recently from the combination of plain vowels and a source of retroflexion (e.g., /ʈ ɭ ɳ/, Heegård & Mørch, 2004), as schematized in **Figure 1**.

	rhoticity	plain vs.
plain vowels	**source**	**rhotic vowels**

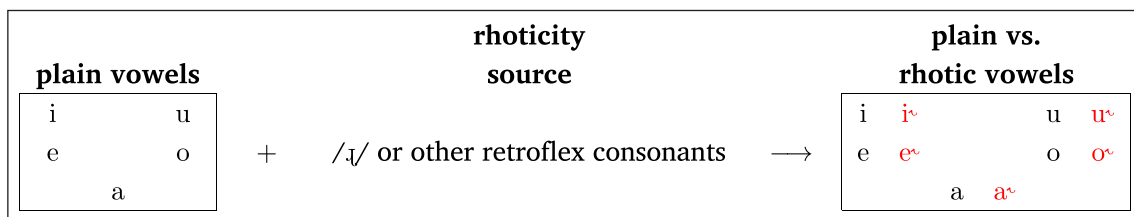| i	u			i	i˞	u	u˞
e	o	$+$	/ɭ/ or other retroflex consonants $\longrightarrow$	e	e˞	o	o˞
	a				a	a˞	

Figure 1: Schematic representation of the development of rhotic vowels in Kalasha.

This recent appearance of rhotic vowels in Kalasha provides an opportunity to explore the development of a new vowel feature (i.e., how does an articulatory gesture get combined with all the vowel qualities in a vowel system?) The genesis of rhotic vowels strongly points to retroflexion, given that it was the retroflex class of consonants that triggered the change (Heegård & Mørch, 2004). Apart from the approximant, retroflex consonants are very likely to have a tip-up or true retroflex configuration. However, Hussain and Mielke (2021) found that present-day Kalasha rhotic vowels are produced with tongue bunching rather than retroflexion. Clearly the vowel system has undergone reorganization from its articulatory basis.

Height, backness, and rounding are ubiquitous vowel quality features. We are interested in how other features interact with vowel quality when they are introduced into a vowel system, as in the case of rhoticity being added to Kalasha's vowel system. While rhotic vowels in Kalasha and other languages have lower F3 than the corresponding plain vowels (Hussain & Mielke, 2021; Kochetov et al., 2021, see section 2.3), making a plain vowel rhotic is not simple, because the gestures involved in lowering F3 are complex and interact with the other aspects of vowel quality such as F1 and F2. To combine plain vowel qualities with rhoticity, it is necessary to consider the gestures involved in producing vowels and rhotics as well as the acoustic consequences of combining these gestures. In this paper we use biomechanical modeling to compare rhotic vowels with the result of applying retroflex and bunched tongue gestures to plain vowels. More generally, this is an attempt to isolate the phonetic bases of phonological patterns. While phonetic

explanations are widely invoked to account for phonological observations, the arguments often take the form of a typological observation concerning the distribution of a particular pattern and a plausible phonetic explanation for it. It is often difficult to study the phonetic motivation directly, because its phonologized result and/or its conventionalized phonetic precursor may already be present in a language under investigation. Here we attempt to produce a realistic vowel + rhotic coarticulation baseline to compare with Kalasha's rhotic vowels.

Previous studies have investigated the phonetics of rhotic vowels and vowel + rhotic sequences in the Dardic languages Kalasha and Dameli, and the Nuristani languages Eastern Kataviri and Kamviri (Hussain & Mielke, 2022). The pairs of rhotic and non-rhotic vowels differ considerably in their lingual articulation (Hussain & Mielke, 2021), but it is not known how much of this difference is directly attributable to retroflex coarticulation and how much is due to subsequent changes as the new vowels have been incorporated into the sound system of Kalasha. In this paper we seek to explore the apparent coarticulatory basis for a new vowel feature through biomechanical modeling of the effects of combining plain vowels with a coarticulatory source derived from a different language. We are interested in whether the development of rhotic vowels from plain vowel + retroflex is basically additive from an articulatory standpoint. We use the rhotic approximant /ɻ/ produced by contemporary speakers of Kamviri as representative of the articulation that conditioned the rhotic vowels of Kalasha. Kamviri speakers have retroflex and bunched versions of rhotic approximant /ɻ/, which provides us the opportunity to investigate how coarticulation of vowels and retroflex/bunched approximants results in the development of rhotic vowels in modern Kalasha.

Kalasha rhotic vowels are predominantly bunched and they retain the lip rounding gestures of the corresponding non-rhotic vowels, but rhotic vowels differ considerably from the corresponding plain vowels in tongue posture and acoustic vowel quality (Hussain & Mielke, 2021, see section 2.3). These differences are difficult to interpret without a realistic model of coarticulation between plain vowels and rhotic consonants. Thus, we use biomechanical and acoustic modeling to address the following questions:

(a) **Do the rhotic vowels retain the lingual articulation of the corresponding non-rhotic vowels?** This will be assessed by combining the corresponding plain vowels with both retroflex and bunched approximants and comparing them with the rhotic vowels and determining whether the differences in tongue height and backness observed within each plain-rhotic pair are accounted for by adding retroflexion or bunching to the plain vowel.

(b) **Do the rhotic vowels retain any signs of retroflexion of the historically present retroflex consonants?** It is already known that Kalasha rhotic vowels are predominantly bunched, but it is unknown how much other aspects of their articulation might be more

similar to the retroflex consonants that provided the original coarticulation leading to their development.

(c) **Do the formant frequencies of rhotic vowels differ from what would be expected from adding retroflexion or bunching to their non-rhotic counterparts?** Although the modern Kalasha vowels are bunched, they could have acoustic properties that are attributable to their previous existence as plain vowels coarticulated with retroflexion in particular. Furthermore, since five-way rhotic vowel quality contrasts are unusual, we wonder if additional quality adjustments are required to keep them perceptually distinct.

2. Background

2.1 Typology of vowel distinctions

Retroflex or rhotic vowels are found in fewer than 1% of the world's languages (Maddieson, 1984; Moran, McCloy, & Wright, 2014). While vowel rhoticity may be considered marginal from a broad crosslinguistic perspective, it is a basic vowel feature in Kalasha. To help contextualize the interaction of rhoticity and vowel quality in Kalasha, we begin by surveying various phonetic vowel distinctions and their interaction with the lingual articulation of vowel quality. **Table 1** shows the proportion of languages (defined as distinct ISO 693-3 codes) in the PHOIBLE database (Moran et al., 2014) that employ various phonetic distinctions in their vowel systems, as well as vowel quality dimensions that they are particularly likely to interact with.[1] We did not require these distinctions to be minimal, only to be present (i.e., /i/ vs. /u/ counts as both backness and rounding).

While lip rounding is utilized in nearly all vowel systems, it is worth considering how it interacts with tongue posture. Although the lips and tongue are articulatorily independent, they may interact through the trading relations involving their effects on formant frequencies. Vowel pairs such as /i y/ that ostensibly differ only in lip rounding often differ in tongue position as well (Jackson & McGowan, 2012). Length and nasalization are the most frequent additional features that are at least partially independent of vowel quality, followed by various forms of laryngealization (creakiness/glottalization/ejective), breathiness, pharyngealization, and rhoticity. Each of these semi-independent vowel features has opportunities to interact with vowel quality.

---

[1] This classification does not distinguish tense/lax from vowel height, and many vowel systems that are analyzed as having tongue root contrasts but transcribed using base symbols are also treated as basically vowel height. The tongue root advancement or retraction category here only includes languages that are listed with diacritics indicating tongue root position.

Distinction	Example	Count	Percentage	Interacts with
height	i and a	2101	100.00%	[see below]
backness	i and u	2095	99.71%	[see below]
lip rounding	i and u	2088	99.38%	backness
length	aː	883	42.03%	height
nasalization	ã	500	23.80%	height
creakiness (or glottalization, ejective)	a̰ a' or aˀ	30	1.43%	height
breathiness	a̤	28	1.33%	height
pharyngealization	aˤ	10	0.48%	height & backness
rhoticity	a˞	9	0.43%	height & backness
voicing	ḁ	7	0.33%	height
tongue root advancement or retraction	a̘ or a̙	6	0.29%	height
velarization	aˠ	2	0.10%	height & backness
epilaryngeal source	aᴱ	1	0.05%	backness

Table 1: Occurrence of vowel distinctions in the PHOIBLE database (Moran et al., 2014).

A prototypical example of length-quality interaction is the development of Persian vowels: Classical Persian has been analyzed as having a three-way quality contrast combined with a two-way length contrast /a aː: i iː u uː/ (Krámskỳ, 1939), but in Modern Persian this is a six-way quality contrast /a ɑ e i o u/ (Nye, 1955) or /a ɑː e iː o uː/ (Toosarvandani, 2004). In this case, the Classical Persian long vowels have developed higher and/or more peripheral qualities in Modern Persian than the corresponding short vowels. This is consistent with the idea that short vowels are more vulnerable to undershoot (Lindblom, 1963), which can be reinterpreted as a basic property of a vowel. The quality differences among the non-low vowels are also consistent with the idea that higher vowels sound longer than lower vowels and speakers may use lowering to signal short duration and raising to signal long duration (Gussenhoven, 2007).

Vowel nasalization is achieved by velum lowering, which is also independent of tongue position but has overlapping acoustic consequences that may lead to lingual differences among ostensibly similar oral-nasal pairs. Nasalization leads to F1 raising in high vowels and F1 lowering in low vowels due to an additional vocal tract resonance in the vicinity of an F1 frequency that is typical for a low-mid vowel (Diehl, Kluender, Walsh, & Parker, 1991; Feng & Castelli, 1996; Fujimura & Lindqvist, 1971; Serrurier & Badin, 2008). This acoustic effect of nasalization may lead to enhancement or compensation (Beddor, 1982; Krakow, Beddor, Goldstein, & Fowler, 1988). The acoustic effects of nasalization are enhanced with a more neutral tongue height in Northern Metropolitan French (Carignan, 2014) and Brazilian Portuguese (Barlaz et al., 2015), and also in Kalasha (Hussain & Mielke, 2021).

Creaky voice quality and glottalization are associated with larynx raising, which shortens the vocal tract and raises F1 and other formants (Laver, 1980), making vowels sound lower, whereas breathy vowels are often produced with a lowered larynx, which has the opposite effect (Esposito, Sleeper, & Schäfer, 2021), making vowels sound higher. Indeed, listeners perceive creaky vowels as sounding lower (Brunner & Zygis, 2011) and breathy vowels as sounding higher (Lotto, Holt, & Kluender, 1997). It is reasonable to expect phonation differences to be enhanced by vowel height, much like length and nasalization differences (see Esposito et al., 2021 for more discussion of the interaction of voice quality and vowel quality).

Pharyngealized vowels are typically produced with centralization of F1 and F2, and they may or may not show signs of rhoticity such as low F3 (Catford, 1983). Pharyngealization makes the back cavity smaller and limits the tongue's freedom of movement to produce extreme vowel postures. Hussain and Mielke (2021) concluded that there is probably considerable overlap between vowels described as pharyngealized and vowels described as rhotic, but that these terms are not equivalent.

Most of the vowel types described so far have acoustic or perceptual effects that may lead to changes in vowel quality. By contrast, rhotic vowels directly interact with the tongue movements used to produce vowels, and accordingly rhoticity is expected to be particularly aggressive at rearranging vowel systems it is introduced to. Rhotic vowels such as /ɚ/ and rhotic approximants such as retroflex (tip-up) [ɻ] and bunched (tip-down) [ɹ] are produced with a wide range of tongue shapes often grouped into retroflex and bunched categories (Delattre & Freeman, 1968; Hussain & Mielke, 2021; Mielke, 2015; Zhou et al., 2008). Retroflexion is achieved by raising and retracting the tongue tip (possibly involving subapical or sublaminal contact with the alveolar ridge). Bunching can be achieved by lowering and retracting the tongue tip while depressing the medial portion of the tongue dorsum, resulting in a concavity that is prominent in the mid-sagittal plane (Hussain & Mielke, 2021; Moisik, 2013; Stavness, Gick, Derrick, & Fels, 2012). Tongue dorsum concavity is a key characteristic of bunched [ɹ] and has also been observed

in retroflex tongue shapes in English and Canadian French (Mielke, 2015; Zhou et al., 2008). Retroflexion, in addition to raising of the tongue tip towards hard palate, may also include a dip in between the tongue tip and dorsum (Hamann, 2003).

Low F3 is a hallmark of bunching and retroflexion (Lehiste, 1962). Tongue tip raising and tongue bunching both give rise to a large sublingual cavity, which results in lowering of F3 (Zhou et al., 2008). Delattre and Freeman (1968) reported a wide range of articulatory to acoustic correlations for the American English [ɹ]. (1) A narrow palato-velar constriction lowers the F3 or brings F2 and F3 close to each other. (2) A dip in the tongue dorsum lowers F3. (3) A wider pharyngeal constriction increases the distance between F2 and F3 but a narrow pharyngeal constriction brings the two formants closer. (4) Lip rounding lowers all the formants.

Table 2 organizes the vowel distinctions that are at least as frequent as rhoticity in descending order of their expected effect on the lingual articulation of vowels. Rhoticity directly affects most aspects of tongue posture, and pharyngeal constriction and tongue root movements affect the posterior tongue. The others have effects that are mediated primarily by acoustics, by affecting the end of the vocal tract as in lip rounding and laryngeal gestures, or the effect of velopharyngeal coupling on F1, or primarily by perception, as in the case of length.

Distinction	Impact
rhoticity	directly affects tongue posture
pharyngealization, tongue root	directly affects pharynx volume
lip rounding	affects front cavity
creakiness, breathinesss	affects pharynx length
nasalization	affects acoustics of F1
length	affects undershoot and perception of height

Table 2: Vowel distinctions by hypothesized hierarchy of impact on lingual articulation of vowels.

2.2. Development of Kalasha rhotic vowels

The question of how rhotic vowels emerged in Kalasha is a major issue in the current Dardic literature (Cooper, 2005; Di Carlo, 2016; Heegård & Mørch, 2004; Hussain & Mielke, 2022; Kochetov et al., 2021). There are reasons to attribute the development of rhotic vowels to internal sound change as well as to external factors. The internal source of rhoticity is the occurrence of retroflex consonants in the environment of a vowel that developed rhoticity. The oral plain

vowels of Northern (Birir, Bumburet, and Rumbur) and Southern (Urtsun and Jinjiret) dialects of Kalasha underwent rhoticization due to the presence of liquids and retroflex consonants in a word. For instance, the rhotic-nasal vowels of modern Kalasha can be reconstructed from the Old Indo-Aryan (Sanskrit) retroflex nasal /ɳ/ (Sanskrit /paɳi/ 'hand' → Kalasha /pẽˑ/ 'palm of the hand': Heegård & Mørch, 2004).

Some evidence about earlier stages of Kalasha comes from fieldwork by Leitner in 1866–72 (Leitner, 1880) and Morgenstierne in the late 1920s (Morgenstierne, 1973). In Leitner and Morgenstierne's wordlists, Kalasha rhotic vowels were transcribed as sequences of plain vowels + palatal fricative /ř/ or rhotics /r ṛ/ (an underdot is generally used to denote retroflexion in Indo-Aryan and Nuristani literature). For example, the word for 'heart' was transcribed as /héra/ by Leitner and /hïːřa/ by Morgenstierne; in modern Bumburet Kalasha, the same word is produced as /hiˑa/, with a rhotic vowel. Morgenstierne (1973) originally used the term *palatal fricative* to refer to a class of r-colored speech sounds in Dardic and Nuristani languages, which may affect the quality of the neighboring vowels. This resembles modern descriptions of Dameli and Nuristani languages (Eastern Kataviri and Kamviri), which have non-phonemic and r-colored vowels in the vicinity of /ɻ/ (Perder, 2013; Strand, 2011).

We infer from these descriptions that the development of rhotic vowels in Kalasha is the result of vowels coalescing with following liquids and retroflex consonants and that researchers 95–160 years ago heard the consonantal portions as palatal fricative /ř/ or rhotics /r ṛ/. During the time period that Kalasha has been developing rhoticity in vowels that are followed by retroflex consonants, it has been in contact with Nuristani languages that have abundant phonetic vowel rhoticity. This vowel rhoticity takes the form of a contrastive retroflex approximant /ɻ/ and non-phonemic retroflex vowels (Heegård & Mørch, 2004; Morgenstierne, 1954), as well as r-coloring of vowels in the vicinity of the retroflex flap /ɽ/ (Strand, 2011).

In summary, Kalasha rhotic vowels are believed to have emerged via loss of retroflex consonants (e.g., /ʈ ɻ ɳ/). The development of phonemic rhotic vowels may have also been encouraged by intensive contact with the Nuristani languages with a contrastive retroflex approximant /ɻ/ (Di Carlo, 2016; Heegård & Mørch, 2004; Hussain & Mielke, 2021, 2022).

2.3 Rhotic vowels in modern Kalasha

A handful of studies have investigated the phonetic correlates of rhotic vowels in Kalasha (Hussain & Mielke, 2021, 2022; Kochetov et al., 2021). **Figure 2** shows mean formant frequencies for the ten plain and rhotic oral vowels of four male speakers reported by Hussain and Mielke (2021) for the tokens included in this study. It can be observed that the rhotic vowels are generally more centralized in F2 relative to their non-rhotic counterparts, their F1 is centralized

or raised, and they have much lower F3 (as also shown by Kochetov et al., 2021).[2] The rhotic vowels are close to steady-state (i.e., they are not generally more rhotic at the end than at the beginning; Hussain & Mielke, 2021).
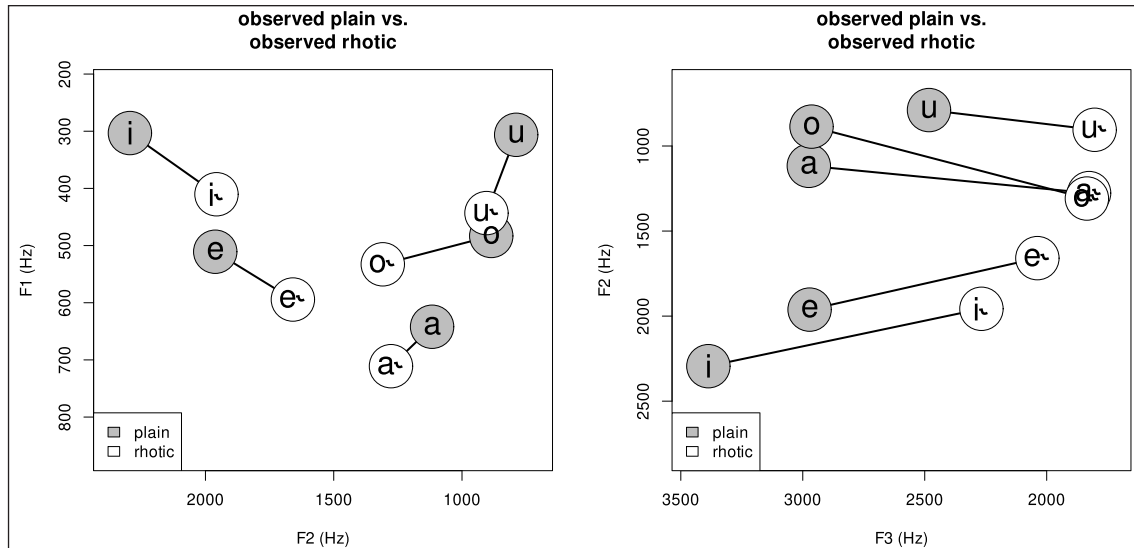

Figure 2: F1, F2, and F3 of observed plain and rhotic oral vowels of Kalasha (Hussain & Mielke, 2021).

Figure 3 shows Smoothing-Spline ANOVA (SSANOVA) comparisons of tongue shapes used to produce these ten vowels. The five non-rhotic vowels are articulated as expected (e.g., on the basis of Wood (1979): /i/ and /e/ have constrictions made toward the hard palate, /u/ has tongue raising toward the velum, and /o/ and /a/ have constrictions in the upper and lower pharynx, respectively). The five rhotic vowels are produced with tongue bunching (rather than retroflexion) by all speakers investigated by Hussain and Mielke (2021). In addition, all five rhotic vowels, including /ɚ ɔ˞ ʊ˞/, are produced with relatively front tongue body position.

While the rhotic vowel tongue shapes are quite different from the tongue shapes used to produce the corresponding plain vowels, Hussain and Mielke (2021) found no significant differences between the lip postures for rhotic vowels and the corresponding non-rhotic vowels. /u/, /ʊ˞/, /o/, and /ɔ˞/ and their nasal counterparts are all rounded and all the other vowels are not. /u/ and /ʊ˞/ are produced with a slightly smaller lip opening than /o/ and /ɔ˞/.

---

[2] There is some ambiguity in how the rhotic vowels differ in F1. Kochetov et al. (2021) found centralized F1 for all of the rhotic vowels, and Hussain & Mielke (2021) found similar differences in nonlow vowels but little difference in F1 between /a/ and /ɚ/, and in the data used for this paper, /ɚ/ has higher F1 than /a/. We note also that we have observed signs of larynx raising in Kalasha rhotic vowels, which could account for general F1 raising in some or all speakers. This is discussed more in the discussion.
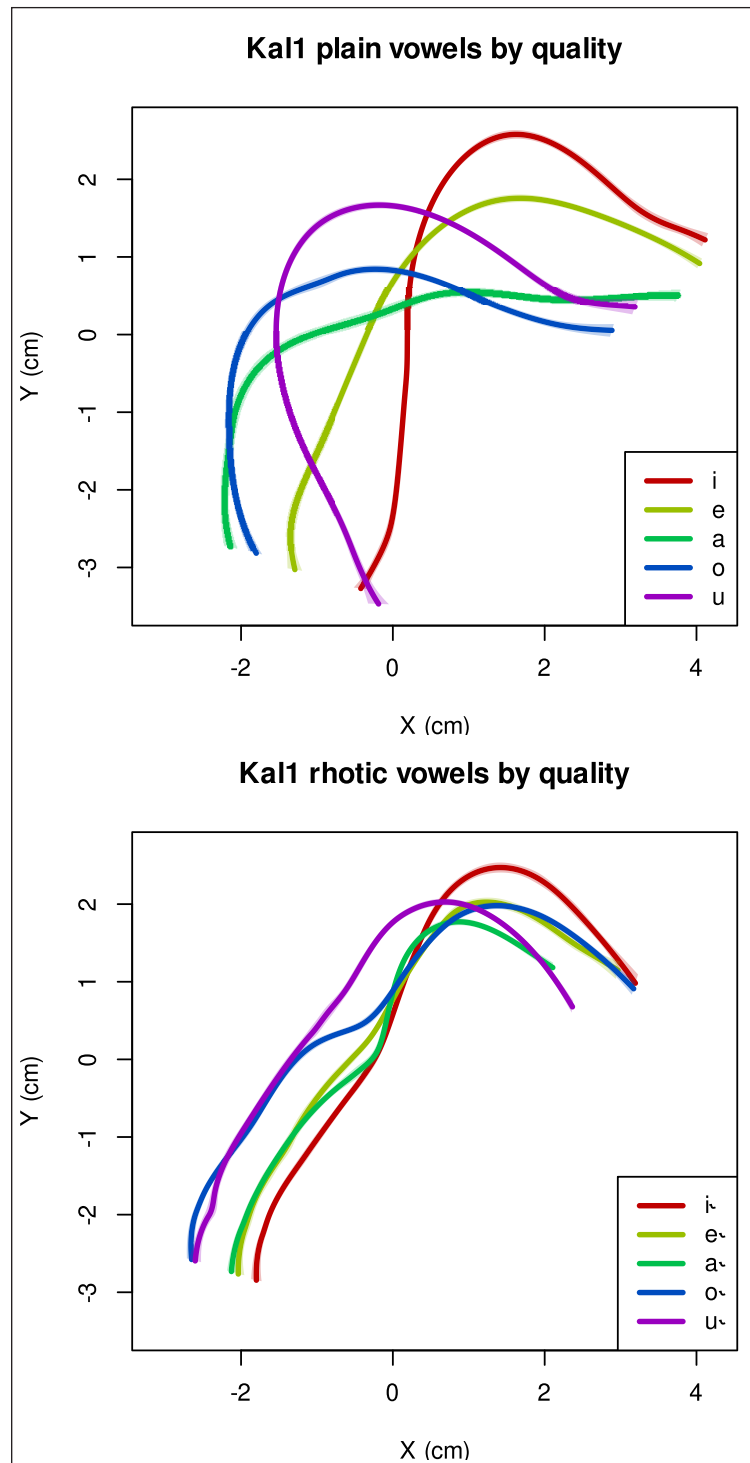
Figure 3: SSANOVA figures showing tongue shapes used to produce plain (top) and rhotic (bottom) vowels by one Kalasha speaker (Kal1). Tongue tip is to the right. X and Y axes are in centimeters (Hussain & Mielke, 2021).

2.4. Tongue muscle activations during vowel and rhotic production

In this paper we model the coarticulation between non-rhotic vowels and rhotic consonants. The tongue shapes involved in coarticulated vowel + rhotic sequences are determined by the combination of muscle activations involved in vowels and rhotic consonants. We begin by reviewing the muscle activations involved in producing vowels and rhotic approximants. The human tongue is controlled by intrinsic and extrinsic sets of muscles (see Sanders & Mu, 2013; chapters 8–9 of Gick, Wilson, & Derrick, 2013; and **Figures 1–4** in Jang, 2022 and references therein for more details). Minor contractions in intrinsic muscles change the shape of the tongue from the inside. The intrinsic tongue muscles consist of inferior longitudinal, which lowers and retracts the tongue tip; superior longitudinal, which raises and retracts the tongue tip; verticalis, which flattens the tongue body; and transversus, which narrows the tongue laterally and causes a sagittal expansion, enlarging the tongue along both anteroposterior and inferosuperior axes due to the muscular hydrostatic nature of the tongue (Smith & Kier, 1989).

The extrinsic tongue muscles connect the tongue to other parts of the body (see Honda, 1996; Takano & Honda, 2007). The genioglossus courses from the superior mental spines inside the mandible to the full length of the tongue. Contracting the anterior fibers of the genioglossus lowers the front of the tongue and contracting middle fibers lowers the tongue dorsum. Contracting posterior fibers of the genioglossus moves the whole tongue forward and raises it toward the palate, as in the production of high and front vowels. The hyoglossus courses from the hyoid bone at the root of the tongue up to the sides of the tongue and pulls the tongue down and back when contracted (as in low and back vowels). The palatoglossus courses from the soft palate to the sides of the tongue and pulls the soft palate down or the tongue up depending on the state of other muscles attached to these structures. The styloglossus courses from the styloid processes in the skull below the ears forward to the sides of the tongue, and has the potential to retract and stabilize the tongue in back vowels. In addition to these intrinsic and extrinsic tongue muscles, the geniohyoid and mylohyoid are muscles that do not directly insert into the tongue but form the floor of the mouth and accordingly support the base of the tongue and facilitate tongue raising.

Stavness, Gick, et al. (2012) modeled the articulation of retroflex and bunched English /r/ variants using the ArtiSynth biomechanical modeling toolkit (www.artisynth.org; Lloyd, Stavness, & Fels, 2012). Their bunched /r/ involved the contraction of the superior longitudinal, inferior longitudinal, and anterior genioglossus (to retract the tongue tip without raising it), middle genioglossus (to lower the tongue dorsum medially), along with some retraction of the transversus and verticalis to bunch the tongue body. Their tip-up (retroflex) /r/ involved greater contraction of the superior longitudinal without contraction of the inferior longitudinal (to retract and raise the tongue tip), plus some contraction of the middle genioglossus (but less than in the bunched variant), and optionally contraction of the hyoglossus to do some of the work of retracting the tongue without depressing the tip. Stavness, Gick, et al. (2012) showed

that their bunched /r/ is very compatible with their simulation of [i], which differs from it by principally having contraction of the posterior genioglossus to front and raise the tongue and less contraction of the middle genioglossus. Their retroflex /r/ is compatible with their simulation of [a], which similarly involves middle genioglossus and hyoglossus to lower the tongue body, and differs from retroflex /r/ in having contraction of the verticalis to further depress the tongue body and lacking contraction of the superior longitudinal (for no retroflexion). These results helped account for the observation that English /r/ is typically bunched in the context of /i/ and that retroflex /r/ is particularly frequent in the environment of /a/ (Mielke, Baker, & Archangeli, 2016; Ong & Stone, 1998).

The aim of the current study is to investigate the similarity between Kalasha rhotic vowels and the superposition of plain vowels and rhotic approximants (i.e., additive combination of the articulatory gestures of plain vowels and retroflex /ɻ/ and bunched /ɹ/ approximants). We use the phonetic data presented in Hussain and Mielke (2021, 2022) as the basis for the current investigation of the biomechanics involved in the production of Kalasha rhotic vowels and combine Kalasha plain vowels with Kamviri-style rhotic approximants /ɻ ɹ/. Moreover, we also examine the role of different tongue muscles in the production of plain and rhotic vowels of Kalasha and compare them with the retroflex and bunched approximants of Kamviri.

3. Methods

3.1. Languages and speakers

Acoustic and articulatory data used as the basis for simulations were described in more detail in Hussain and Mielke (2021, 2022). Recordings were made of four Kalasha (Bumburet dialect) and two Kamviri speakers (all males in their 20s or 30s). All six speakers were from Bumburet valley, Chitral, northern Pakistan. In addition to their native languages, all the speakers could speak Khowar, Pashto, Urdu, and/or English.

3.2. Speech materials

The modeling described in this paper is based on the rhotic vowels in the Kalasha words /piˑ˞ː/ 'press', /heˑ˞/ 'theft', /baˑ˞/ 'lazy', /tʃoˑ˞i/ 'parasite', and /kʰuˑ˞/ 'hat', the non-rhotic vowels in the Kalasha words /pi/ 'drink (verb); from', /pe/ 'if', /paː/ 'go', /po/ 'footprint', and /tu/ 'you', and the word-final rhotic approximant /ɻ/ (retroflex) or /ɹ/ (bunched) in the Kamviri word /parmaɻ/ 'child.'

3.3. Recording procedure

The participants were invited into a quiet room at a hotel in Bumburet valley, Chitral, Pakistan. A Terason t3000 ultrasound machine with Ultraspeech 1.3 software (Hueber, Chollet, Denby, &

Stone, 2008) was used for recording the ultrasound data. The tongue ultrasound and lip video recordings were made in direct-to-disk mode, generating 640 × 480 pixel bitmap images at 60 frames per second. A Terason 8MC3 3–8 MHz ultrasound transducer was positioned underneath each participant's chin, stabilized with an Articulate Instruments aluminum Probe Stabilisation Headset (Scobbie, Wrench, & van der Linden, 2008). A frontal view of the lips was captured using a board camera (The Imaging Source DFM 22BUC03-ML with a 12 × 0.5 mm lens) mounted on the headset about five centimeters in front of each participant's lips using two clip-on LED book lights, which also illuminated the participants' lips.

Simultaneous audio recordings were made with a Shure Beta 53 head-mounted omnidirectional condenser microphone (44.1 kHz, 16-bit). Before the recordings, all the participants were familiarized with the task and went through the wordlists. After audio and ultrasound recording commenced, the participants were asked to hold a mouthful of water to generate ultrasound images of the palate (not used in this analysis) and then held a tongue depressor between their teeth and pressed their tongue against it in order to generate ultrasound images of the occlusal plane. The target wordlists were presented to the participants on a computer screen or they were described to them in Urdu, which is a lingua franca of Pakistan. All the words were elicited in citation form. Each target word was repeated five times.

3.4. Acoustic and articulatory analyses

The ultrasound frames were selected from the midpoints of the Kalasha vowel intervals and the Kamviri rhotic approximant intervals. The tongue contours and the lip opening were manually traced in Palatoglossatron (Baker, 2005). The lip data used to inform the modeling were two points placed mid-sagittally on the edge of the upper and lower lips. The frequencies of the first three formants were extracted at the same time points using Praat (Boersma & Weenink, 2007).

3.5. ArtiSynth modeling

Our goal was to test whether Kalasha rhotic vowels are similar articulatorily and acoustically to the superposition of plain vowels and rhotic approximant articulations. Thus, we have taken plain vowels coarticulated with a following rhotic approximant to be the initial state for Kalasha rhotic vowels. To model this coarticulation, we used the ArtiSynth biomechanical modeling toolkit (www.artisynth.org; Lloyd et al., 2012). ArtiSynth is a free, open-source computational platform for simulating multibody systems comprising rigid and deformable bodies (the latter implemented as finite-element models or FEMs). These bodies can be made to interact via unilateral (e.g., contact/collision) and bilateral (e.g., joint) constraints and manipulated with force effectors of various kinds, such as musculature based on realistic mathematical muscle models. We performed two types of simulations: (i) Inverse simulations of the Kalasha vowel shapes, both plain and rhotic, and Kamviri rhotic approximants (both retroflex /ɻ/ and bunched

/ɹ/ variants); and (ii) forward simulations, which combine the excitations computed during the inverse simulations of the Kalasha plain vowels with excitations of either of the two types of Kamviri rhotic approximant. Along with the biomechanical simulation, we also simulated 1-dimensional acoustics of these vowels, employing frequency-domain acoustic simulation (Birkholz, 2005; Birkholz & Jackel, 2004) and making use of the airway skin mesh (Anderson et al., 2017) to allow for estimation of the vocal tract area function. More details about the biomechanical modeling are included in the Supplementary Materials.

Biomechanical modeling approaches are generally characterized as either forward or inverse. Forward modeling (aka forward-dynamics simulation), is a process of controlling a biomechanical model with given muscle activation signals. Inverse modeling (aka inverse-dynamics simulation), in contrast, is a process which estimates the underlying muscle activations from previously obtained kinematic measurements by using a biomechanical model (Eskes et al., 2017). The ArtiSynth forward models (retroflex and bunched versions) represent the predicted vocal tract shapes due to superposing a bunched or retroflex approximant on plain vowels. The ArtiSynth inverse model of Kalasha vowels is meant to be a close approximation to the actual observed Kalasha vocal tract shapes. When the inverse simulation tongue shapes are quite different from the forward simulation tongue shapes, it suggests that Kalasha speakers are doing something different (such as a more optimal articulatory strategy to achieve a particular perceptual target). If the acoustic output is also different, this suggests that Kalasha has phonologized new perceptual targets for rhotic vowels (i.e., the goal for the phonological category does not even sound like the vowel + rhotic superposition that it is thought to have originated from).

We next describe the inverse simulations, which form the basis of the forward simulations. Inverse models in ArtiSynth take time-varying target points — or trajectories — as input, and they output a set of muscle activations that minimize trajectory-tracking errors while being subject to additional terms for resolving muscle redundancy (Anderson et al., 2017; Stavness, Lloyd, & Fels, 2012) and ensuring smooth activations (as rapid changes can lead to model instability). In our simulations, we used the empirical data from the lips and the tongue to define the trajectories. The final set of inverse target points were two targets for the face FEM (on the midline of the upper and lower lips) and eleven target points along the contour of the tongue, running from the tongue tip to the tongue root. We also developed target points for the mandible (one on the central incisors and one on the pogonion) and hyoid bone (one point on the anterosuperior most point of the body) giving optional parameters to use in cases where model stability was an issue. In all cases, the inverse trajectories started at 0.0 s from their initial configuration and ended at their target location at 0.2 s.

We used MATLAB (MATLAB, 2019) to register the participants' data (lip points and tongue contours) against corresponding details of the ArtiSynth model (**Figure 4**). While it

was straightforward to identify points on the lips that match the flesh points tracked on the participants' lips, it was less clear how to map the ultrasound tongue contour data into the ArtiSynth model. This is because we have no information in the ultrasound image as to what part of the tongue exactly is being imaged, and the imaged portion of the tongue also changes from frame to frame. Thus, homology cannot be guaranteed in the registration, and it was therefore necessary to make assumptions about what the typical visible portion of the tongue was and select reference points on the ArtiSynth tongue model (from tongue tip to root) that matched this. With this in mind, we extracted the location of the lower lip, upper lip, mouth corners, and midsagittal contour of the tongue from tip to root from the ArtiSynth model in its neutral configuration and imported these landmarks into MATLAB for further processing alongside the empirical articulatory data. For each participant, the lip and tongue data were independently registered (using Procrustes superimposition) against the ArtiSynth landmarks. In the case of the tongue, we registered the (participant-wise) mean observed tongue contour against the ArtiSynth tongue contour, using resampling to ensure all contours had the same number of sample points. The registered empirical data were then brought into ArtiSynth and visually examined for how well they fit within the ArtiSynth vocal tract, with the possibility of making slight adjustments to the overall scaling and translation of the registered data once it was in the ArtiSynth environment. The entire process was iterated upon using (most notably) slightly different selections for tongue tip and root positions on the ArtiSynth model to try to best fit the tongue contours while also being articulatorily reasonable. The task proved difficult. Ideally, if more participant data were available for other vocal tract structures, the fit could be improved; it would even be possible to register the ArtiSynth model itself to the participant (if, for example, structural MRI data were available for the participant). In practice, the inverse simulation does not always manage to match the target points exactly and thus no severe acoustic issues arose from the articulation occasionally treading slightly outside of the airway skin.[3] The appearance of the tongue contour data within the ArtiSynth vocal tract setting is illustrated in **Figure 4**.

While alternative reference points on the ArtiSynth tongue model could have been explored systematically, small differences in reference points would not result in large differences with the current findings (see, e.g., Howson, Moisik, & Żygis, 2022). This is in part because there is always some amount of error that the inverse simulation makes in hitting the targets (particularly when there are many of them for a given contour and many contours are being used across a large set of simulations). Systematic exploration of this choice is infeasible, especially given the many other assumptions we have made alongside this that could also arguably be deserving of similar attention (such as the number of inverse targets to employ). The best we can do then is to present

---

[3] As a precaution to prevent collapse of the area function (and thus zero acoustic output), we clamp the minimum cross-sectional area of the airway skin to a value of 1.0×10^{-5} m^2 (0.1 cm^2).

our findings in the light of the model design choices that we have made with the knowledge that small deviations (such as shifting the reference nodes by one node forward or backward or by using more or less inverse target nodes) would lead to similarly small changes to the results.
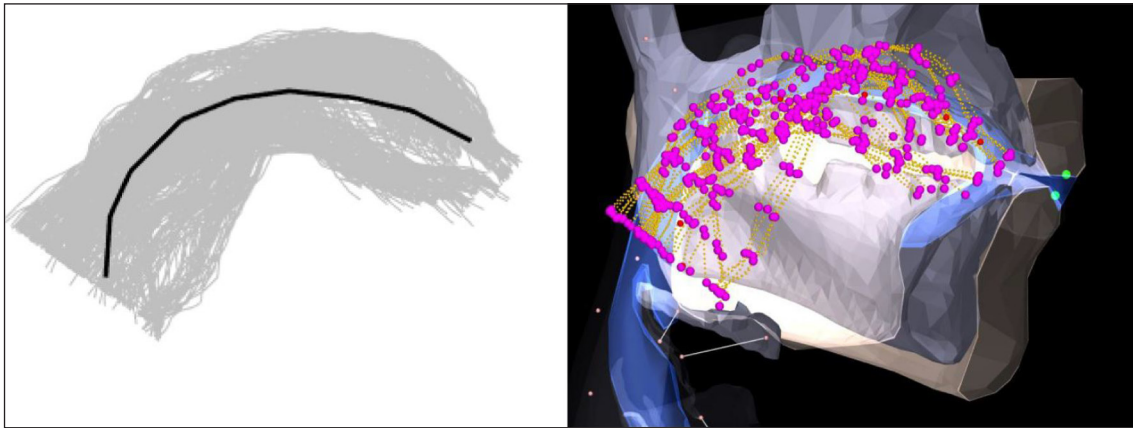

Figure 4: Registration of one participant's tongue contours (gray lines) against the ArtiSynth tongue contour in black line (left) and a selection of these tongue contours (representing the all types of plain and rhotic vowels) within the ArtiSynth vocal tract (right). The magenta circles are the inverse target locations that define the inverse simulation trajectory.

Participant	i	e	a	o	u	ɨ˞	e˞	a˞	o˞	u˞	Total	
Kal1	5	5	5	5	4	4	5	5	5	5	48	
Kal4	5	6	5*	6	5	5	6*	6**	5	5	54	****
Kal5	5	5	5	3	5	6	6	6*	6	9	56	*
Kal8	4	5	5	5*	6	5	5	5	5	5	50	*
All	19	21	20*	19*	20	20	22*	22***	21	24	208	******

Table 3: Number of simulated tokens by participant and vowel. The number of failed simulation runs included in the count are indicated with a corresponding number of asterisks.

For the inverse simulation, we simulated selected productions from four Kalasha participants. We ran simulations covering five basic vowel qualities /i e a o u/ within two vowel types (plain and rhotic), each with approximately five tokens, as shown in **Table 3**, giving us 208 simulations. We also simulated five tokens each of the Kamviri rhotic approximant variants (retroflex and bunched), using data from two different participants who produced the sound differently. The process of bringing these data into ArtiSynth followed the same procedure as

that used for the Kalasha simulations outlined above. **Figure 5** illustrates inverse simulations of /o/ and /ɔ˞/.

Once the inverse simulations were complete, we proceeded with the second type of modeling by creating forward simulations of the Kalasha rhotic vowels using the muscle activations estimated in the inverse simulations. Specifically, for both the retroflex and bunched versions of the Kamviri rhotic approximants, we combined the token-wise means of the muscle activations of these sounds with the token-wise means of the activations that were estimated for the Kalasha plain vowels. This resulted in a further 10 sets of simulations of "pseudo" Kalasha rhotic vowels – five with a retroflex basis and five with a bunched basis – that were intended to serve as a point of comparison against the inverse simulations of the Kalasha rhotic vowels. To accomplish the superposition of the muscle activations, we used a simple additive combination rule, adding activations from each component articulation (plain vowel and rhotic approximant variant) for each muscle exciter. To model a range of possible coarticulatory blends of plain vowels and rhotic approximants, each rhotic vowel was simulated using 11 different mixtures of plain vowel and rhotic approximant activations, essentially a "crossfade" from 100% plain vowel to 100% rhotic approximant at 10% increments. This resulted in 11 steps of rhotic mix proportion for each of the five forward simulation rhotic vowels that could be compared with the inverse simulation of the same rhotic vowel.
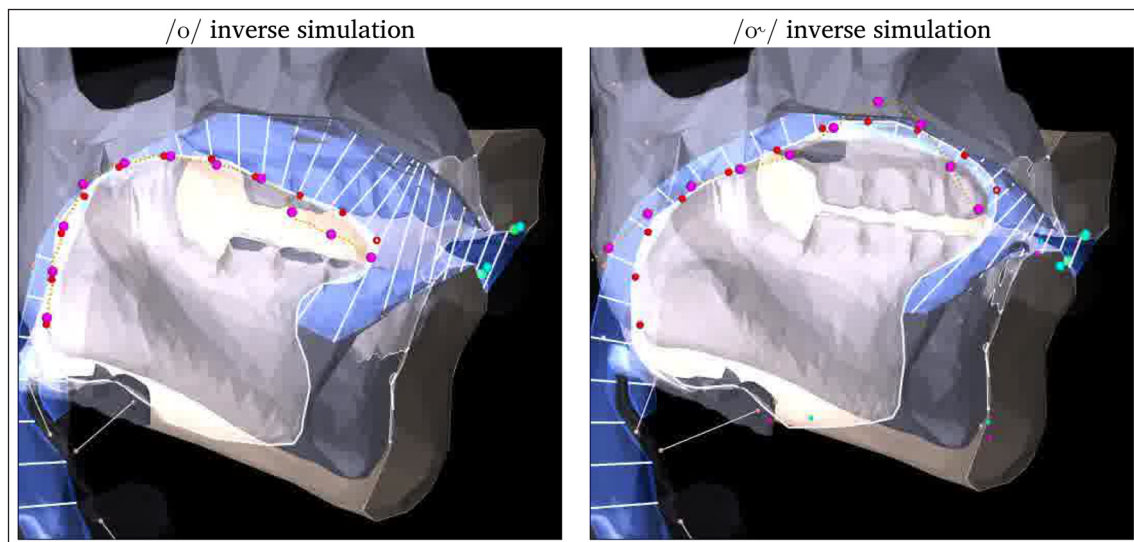

Figure 5: Sample tongue shapes of inverse simulations of plain /o/ in /po/ 'footprint' and rhotic /ɔ˞/ in /krɔ˞/ 'chest' (orange points = detailed ultrasound samples; purple points = inverse targets).

We applied several measures to the simulated tongue postures in order to compare them, as illustrated in **Figure 6**. Each tongue shape is represented by a polygon with 50 vertices.

The centroid of the polygon is indicated by a dot in the middle of the polygon. Its *x* and *y* coordinates represent the overall advancement and height of the tongue body, respectively. The most posterior point on each polygon (at a point indicated by another dot) represents tongue root advancement. The tongue tip is defined as the most anterior point on the tongue polygon (indicated by a third dot). The *x* value of this point represents tongue tip advancement. The angle above the horizontal from the tongue centroid to the tongue tip (represented by a line segment) is an indicator of retroflexion. The tongue blade is taken to be the portion of the tongue that is 2–5 points (out of the 50 points) posterior to the tip. The angle between the two points defining the tongue blade (represented by a thick line between these two points) is another indicator of retroflexion.
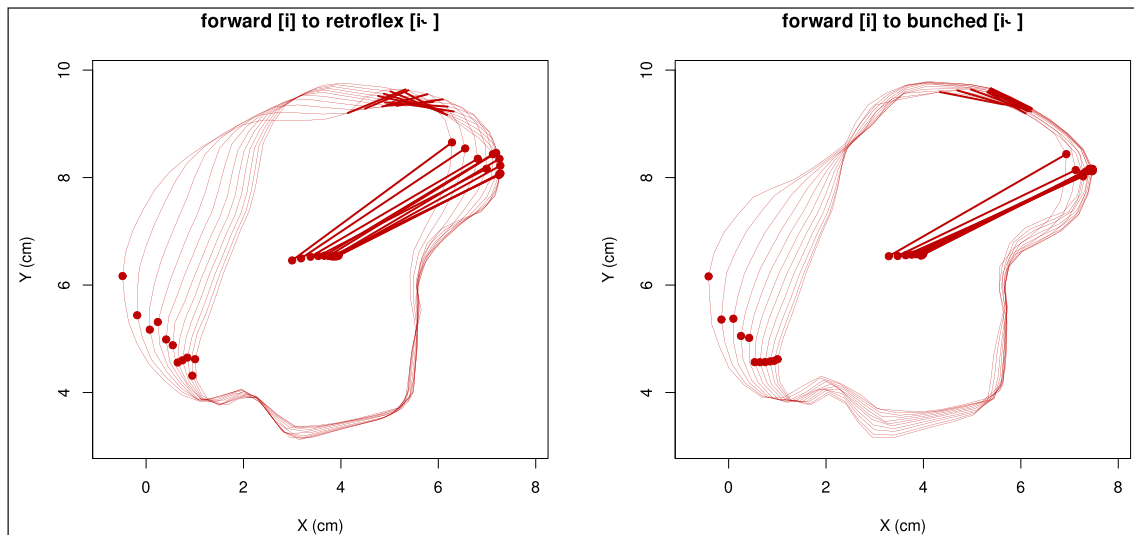
Figure 6: Illustration of measurements applied to the simulated tongue shapes. Dots indicate tongue root, tongue centroid, and tongue tip. Lines indicate angle measure (tongue blade angle and angle from tongue centroid to tongue tip).

Here we reprise the research questions introduced at the end of §1 in terms of the inverse and forward simulations:

(a) **Do the rhotic vowels retain the lingual articulation of the corresponding non-rhotic vowels?** This is a tongue shape comparison between the inverse simulation rhotic vowels (representing modern Kalasha rhotic vowels) and the retroflex and bunched forward simulations (representing an earlier coarticulatory stage). Articulatory properties of the inverse simulation rhotic vowels that are not found in the forward simulations suggest an articulatory reorganization of the Kalasha vowel system.

(b) **Do the rhotic vowels retain any signs of retroflexion of the historically present retroflex consonants?** This is also a tongue shape comparison between the inverse simulation rhotic vowels and the retroflex and bunched forward simulations. Articulatory properties of the inverse simulation rhotic vowels that resemble the retroflex forward simulations in particular suggest that the modern bunched rhotic vowels retain articulatory signs of their retroflex origins.

(c) **Do the formant frequencies of rhotic vowels differ from what would be expected from adding retroflexion or bunching to their non-rhotic counterparts?** This is a formant comparison (based on the acoustic synthesis derived from the biomechanical models), comparing the inverse simulation rhotic vowels with the retroflex and bunched forward simulations, particularly looking for acoustic similarities between the inverse simulations and retroflex forward simulations (suggesting that the modern Kalasha vowels are preserving acoustic details attributable to the original coarticulatory basis) and looking for signs that the inverse simulation vowels are acoustically more distinct than would be expected from the forward simulation vowels (suggesting compensation for the centralizing effects of rhoticity).

To address these questions, we have compared a set of inverse simulations based on actual productions by four Kalasha speakers (representing our best biomechanical models of actual Kalasha vowels) to 22 different forward simulations for each rhotic vowel, representing 11 different degrees of overlap between averaged inverse simulation Kalasha plain vowels and averaged inverse simulation Kamviri retroflex and bunched approximants (the rhotic mix proportion). Any point along the continuum from 100% plain vowel to 100% rhotic could have formed the basis for modern Kalasha rhotic vowels, so we consider these possibilities as a group, asking, for example, does the inverse simulation /o˞/ resemble *any* of the mixtures of /o/ and /ɻ/ or /ɹ/ among the forward simulations?

4. Results

Figure 7 shows the tongue body shape in the inverse simulations of the five non-rhotic and five rhotic vowels, and retroflex/bunched approximants of Kamviri. They are broadly consistent with what is shown for observed Kalasha vowels above in **Figure 3**. The rhotic vowels involve more tongue front bunching than their non-rhotic counterparts, and for front vowels and /ʊ˞/, they involve more tongue root retraction. The Kamviri retroflex approximant is characterized by a slightly raised (tip-up) tongue posture, whereas the bunched approximant exhibits a bunched (tip-down) tongue gesture. It can also be observed that the the Kamviri bunched approximant resembles the rhotic vowels of Kalasha.
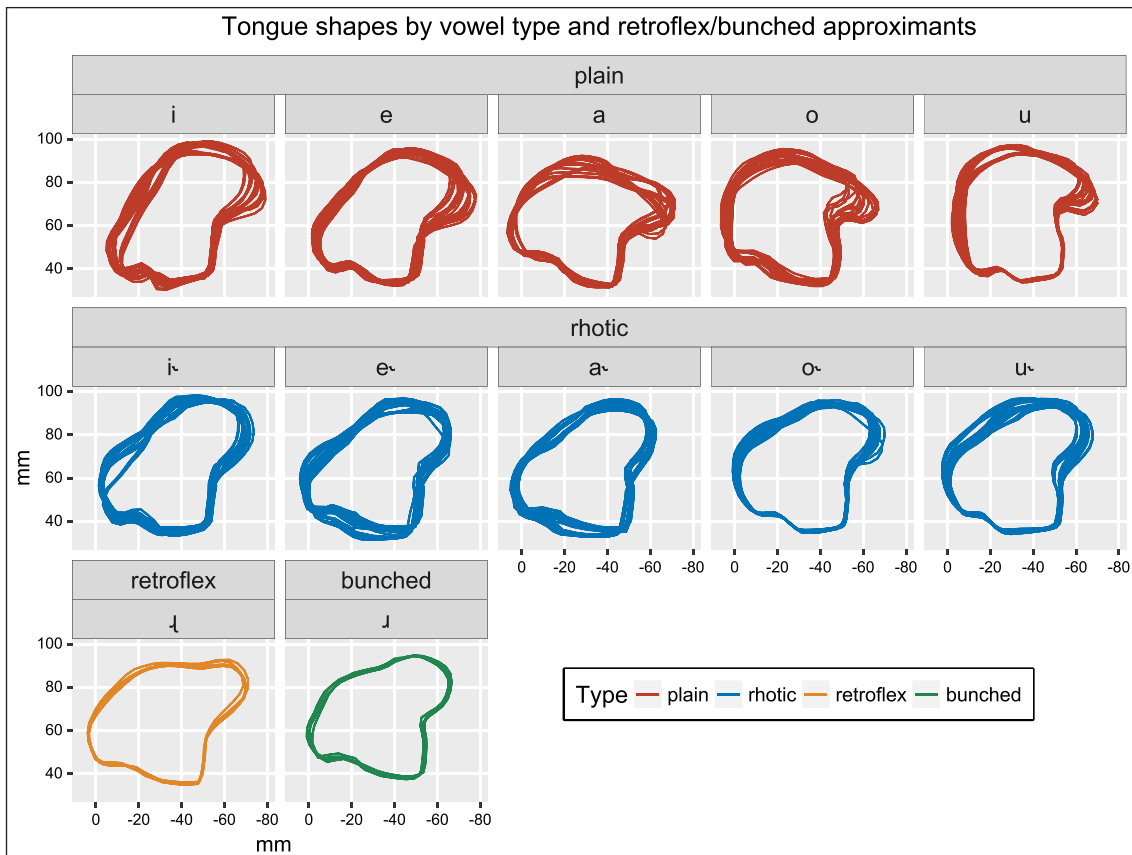
Figure 7: Tongue shape comparisons in inverse simulation plain and rhotic vowels, and retroflex and bunched approximants. Tongue tip points to the right.

Figure 8 shows the muscle activity used in these inverse simulations. Anterior fibers of genioglossus are used to produce /a/ and /ɚ/. Medial genioglossus fibers are used to produce all four front vowels and rhotic /ɚ/. Consistent with how vowels are observed to be produced in Kalasha, the greatest differences in overall tongue shape are observed between the five non-rhotic vowels. Consistent with this, extreme contraction of the posterior genioglossus fibers is seen only in /i/, and contraction of the hyoglossus is seen only in /a/. Inferior longitudinal is more active in rhotic vowels (retracting the tongue tip) than in corresponding non-rhotic vowels for all pairs except for /o ɚ/, where the plain vowel involves more tongue retraction than the rhotic vowel. Styloglossus and superior longitudinal muscles are also generally more active in rhotic vowels than their non-rhotic counterparts. Unexpectedly, transversus is more active in /ɚ ɚ/ than in their non-rhotic counterparts, and verticalis is more active in /o u/ than in their rhotic counterparts. This may be because the inverse simulation is fed only information about the mid-sagittal plane. Geniohyoid is more active in /e a ɚ o/. Kamviri's bunched approximant is consistently characterized by higher muscle activation in genioglossus anterior, medial, and

posterior, inferior longitudinal, styloglossus, and verticalis. Hyoglossus, superior longitudinal, and transversus muscles are actively involved in the production of retroflex approximant.
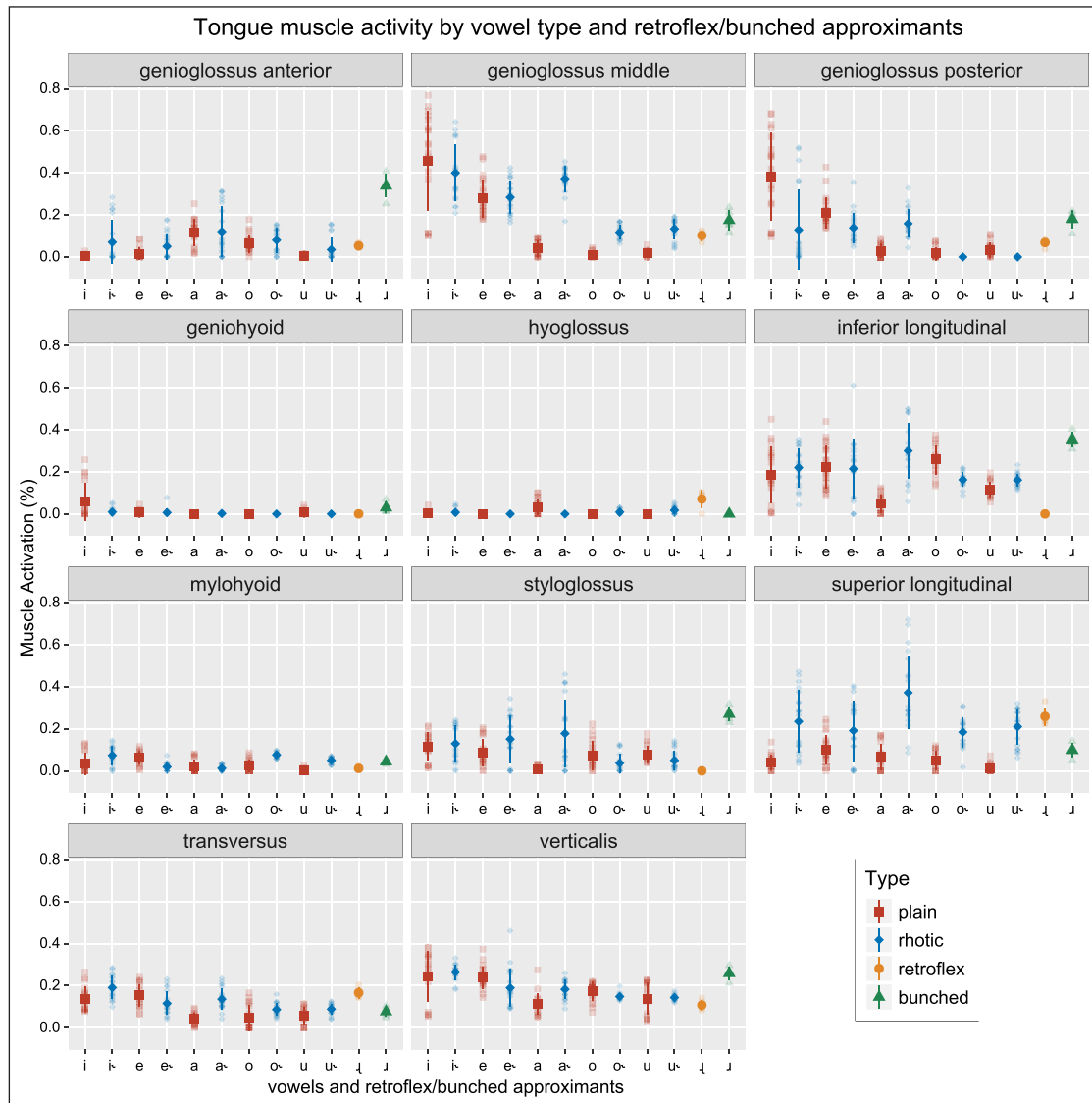

Figure 8: Tongue muscle activations in plain and rhotic vowels, and retroflex and bunched approximants.

There are some differences between our inverse simulation Kamviri rhotic approximants and Stavness et al.'s (2012) English /ɹ/ simulations. Their bunched /ɹ/ does not involve posterior genioglossus or styloglossus but it does involve a small contraction of transversus and it involves equal contraction of the inferior longitudinal and superior longitudinal. Their retroflex /ɻ/ involves medial genioglossus but not transversus, and it optionally involves hyoglossus.

Recall that the forward simulations were designed to simulate coarticulation of vowels and retroflex or bunched approximants. They were produced at 11 rhotic mix proportion steps ranging from 100% vowel and 0% rhotic (similar to the inverse simulation plain vowels) to 0% vowel and 100% rhotic (where all five rhotic vowels are identical because there is no vowel information included in the simulation). We are interested in whether any of the intermediate points resemble the inverse simulation rhotic vowels, which would support the idea that rhotic vowels originated from coarticulation between plain vowels and retroflex or bunched consonants without much further phonetic development. The forward simulated combinations of plain vowels and rhoticity are considered to be vowel qualities that naturally occur in the event of coarticulation.[4] We are interested in how they differ from inverse simulation Kalasha rhotic vowels, because these differences point to ways in which the actual Kalasha rhotic vowels have been established as separate vowel categories that are distinct from vowel + rhotic superposition.

In each panel of **Figure 9**, thin horizontal lines indicate the formant frequencies of the inverse simulation plain and rhotic vowels for one vowel quality. In all cases the rhotic vowel has lower F3 frequency, and in most cases the rhotic vowel has less extreme F1 and F2 frequencies. From left to right within each panel, the thick solid contour shows the formant frequencies of the forward simulation as more and more retroflex approximant (and less and less plain vowel) is mixed in. The thick dashed contour shows the formant frequencies of the forward simulation as more and more bunched approximant is mixed in. The thick vertical lines indicate the step at which the retroflex and bunched forward simulation vowels are acoustically most similar to the inverse simulation rhotic vowels according to the Root Mean Squared Error (RMSE) for the bark-scaled formant frequencies. In some cases, like /a/ and /u/, the 0% rhotic mix steps are fairly close to the inverse simulation non-rhotic vowel formants. In other cases, there are some differences, attributable to the fact that the inverse simulation formant values are based on multiple tokens averaged in acoustic space, whereas the forward simulation formant values are those produced by an average articulatory configuration.

For the /i ɨ˞/ and /e e˞/ simulations, as more of either type of rhotic is mixed in, F1 mostly increases and F2 and F3 mostly decrease. The step at which the forward simulation of bunched /ɨ˞/ most resembles the inverse simulation rhotic vowel is 50%, an equal mix of plain /i/ and

---

[4] It should be noted that articulatory-target timing is introduced into the simulation to allow the model to attain the target posture from its resting position at a pace commensurate with what might be observed in speech (200 ms for the inverse simulation) while also providing ample timing for stability purposes. However, the simulations are idealized in all other respects with respect to timing and should be viewed as canonical steady-state productions (i.e., free of context and hence coarticulatory influence on muscle activity patterns). We did not make any attempt to simulate muscle activation dynamics that might arise from effects such as coarticulation, which are empirically documented for speech (Leidner, 1976; MacNeilage & DeClerk, 1969; Sussman, MacNeilage, & Hanson, 1973) and other complex motor-control tasks (e.g., Winges, Furuya, Faber, and Flanders 2013).
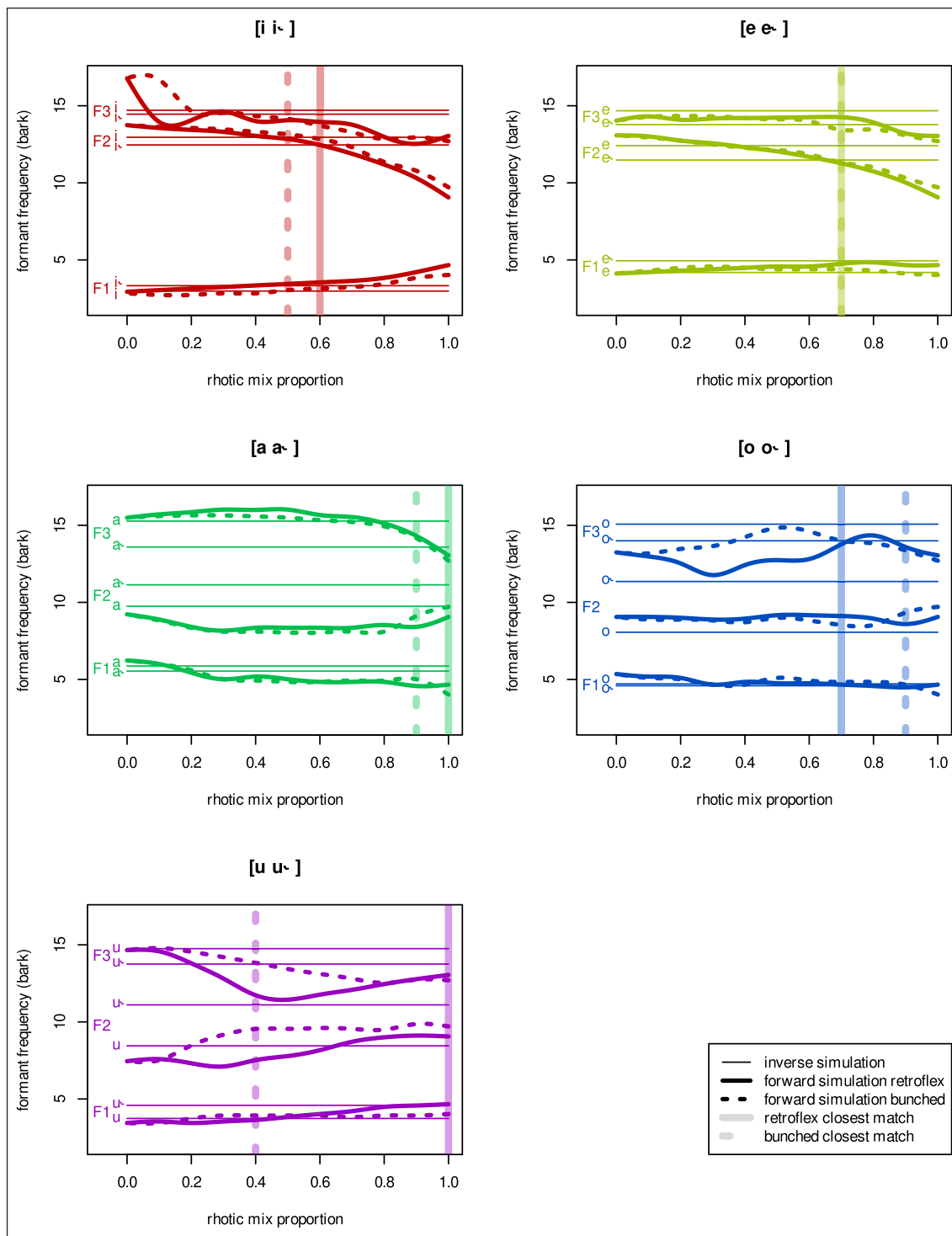
Figure 9: Formant frequencies of forward simulation rhotic vowels ranging from 0 to 100% rhotic mix proportion, with reference formant values for inverse simulation plain and rhotic vowels.

a rhotic approximant. For retroflex /ɨ˞/, 60% rhotic most resembles the acoustics of the inverse simulation rhotic vowel. The /e e˞/ simulations most closely resemble the inverse simulation [e˞] when 70% rhotic is mixed in for both retroflex and bunched versions.

For the back vowels, the picture is a bit different. For /a a˞/, the most similar steps for retroflex and bunched /a˞/ are 90% rhotic for bunched and 100% rhotic for retroflex. This is consistent with the fact that the model rhotics used for these simulations were produced in an /a/ context, so the rhotic approximant itself is already similar to a blend of /a/ and a rhotic approximant.[5] Forward simulation /o˞/ is acoustically most similar to inverse simulation /o˞/ at 70% retroflex rhotic and 90% bunched rhotic. Forward simulation /u˞/ is acoustically most similar to inverse simulation /u˞/ at 100% retroflex rhotic (similar to the other back vowels) and 40% bunched rhotic (similar to the other high vowel). Forward simulation back vowels tend to resemble the inverse simulation rhotic vowels near the 100% rhotic end of the scale. This is consistent with the fact that the Kalasha back rhotic vowels are articulatorily quite different from the corresponding plain vowels (as shown in **Figure 2**), and thus the plain vowel portion of the mixture is not as helpful as it is for the front rhotic vowels. The fact that the plain back vowels do not help much is also a clue that some of the Kalasha rhotic vowels are not articulated in a way that can be predicted from combining the corresponding plain vowel with a gesture for rhoticity.

Figure 10 shows how the forward simulation vowels move through F1–F2 and F2–F3 space as they become increasingly rhotic. The forward simulation 100% plain vowels are indicated by IPA symbols /i e a o u/, and the 100% retroflex and bunched rhotic ends of the scales are indicated by "ʐ" and "ɹ". The inverse simulation formant frequencies are indicated by small black IPA symbols, and the observed formant frequencies are indicated by small gray IPA symbols. Even though they differ from the observed formant frequencies, the inverse simulation formant frequencies are a more appropriate point of comparison for the forward simulation formant frequencies. Comparing the forward simulations with the observed Kalasha vowels would conflate effects of coarticulation with effects of biomechanical modeling, but comparing the forward simulations with the inverse simulations isolates the effects of coarticulation. The paths from 0% rhotic to 100% rhotic are indicated by curves (solid for retroflex and dashed for bunched), and each step along the way is indicated by a dot. The front vowels pass by the inverse simulation rhotic versions as they become more rhotic, and the back vowels generally do not. This suggests that the present-day /ɨ˞ e˞/ are similar acoustically to coarticulated /i e/ and a rhotic consonant,

---

[5] In our earlier exploration of the methods for performing the blending simulations, we attempted to use idealized rhotics as targets (based in part on the work of Stavness, Gick, et al., 2012). We abandoned this approach in favor of one that makes use of exemplars drawn from our ultrasound data of the Kamviri-style rhotic approximants to improve the connection of our simulation materials to our research question. This decision, however, comes with the drawback that the use of natural speech data involving real-word productions makes coarticulatory contamination unavoidable. We opted to use the rhotic low-vowel context because this is the only context where we have articulatory data from the relevant languages (Hussain and Mielke, 2022). Thus our findings must be interpreted with this in mind.
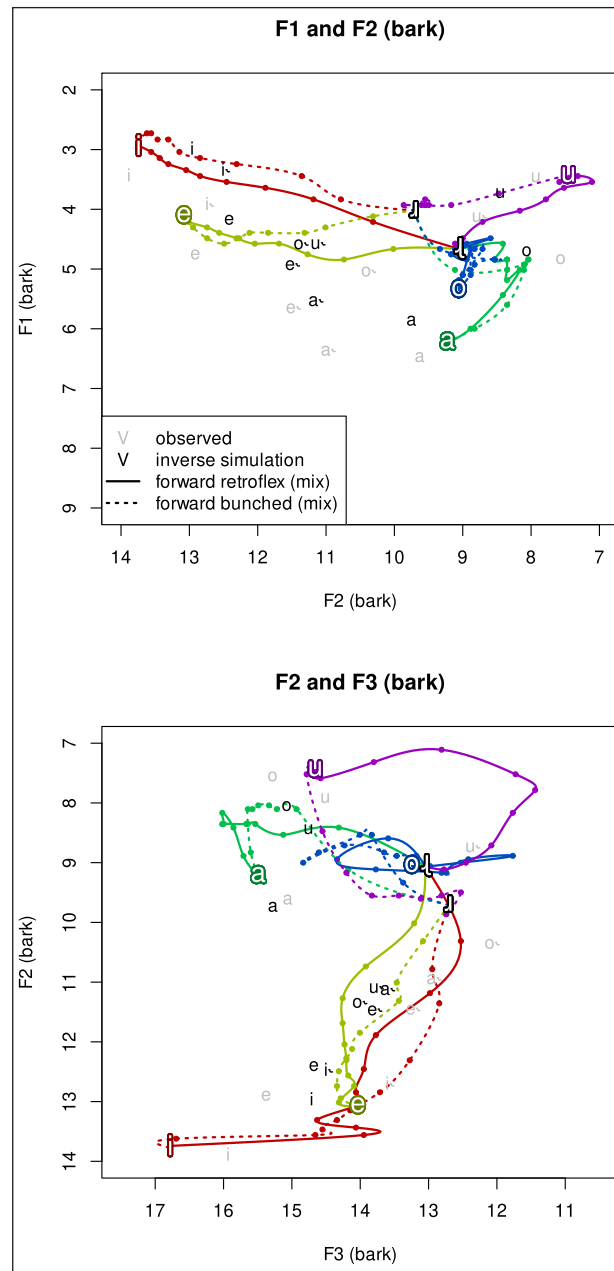
Figure 10: Paths through formant space as vowels become more rhotic: Forward simulation rhotic vowels ranging from 0 to 100% rhotic mix proportion in F1–F2 space (top) and F2–F3 space (bottom). Plain vowel IPA symbol endpoint is 0% rhotic and [ɻ] and [ɹ] endpoints are 100% rhotic (retroflex and bunched, respectively). Steps are at 10% increments.

either bunched or retroflex, while the back vowels generally are not similar. As /a/ becomes more rhotic, it moves away from /ɚ/ in F1–F2 space, and becomes more similar to it mainly by dropping F3 as it approaches the rhotic end of the scale. /o/ becomes more similar to /ɚ/ only in

F1. /u/ approaches /ʊ˞/, but the inverse simulation version of the rhotic vowel is located beyond the 100% rhotic endpoints in terms of F2. One thing that these back vowel mismatches have in common is that the inverse simulation rhotic vowels (and for /ɑ˞ ɔ˞/ also the observed vowels) have high F2 frequencies that are not accounted for by the superposition of rhoticity. This is likely because the rhotic counterparts of back vowels are produced with tongue fronting that is not accounted for by the gestures used to produce plain back vowels or rhotics.

Figures 11–15 show the shape of the simulated tongue for all the steps of the retroflex and bunched forward simulations and the acoustic distance between each of these steps and the corresponding inverse simulation rhotic vowel. Each of the five figures contains four panels showing the same types of information for each of the five rhotic vowels. The top two panels show tongue shapes and the bottom two panels show acoustic distance. The left two panels show retroflex forward simulations and the right two panels show bunched forward simulations. In the top tongue trace panels, all 11 steps of the forward simulation (from 0% rhotic to 100% rhotic) are shown with solid lines. The step that is acoustically most similar to the rhotic vowel is indicated by a heavier contour.
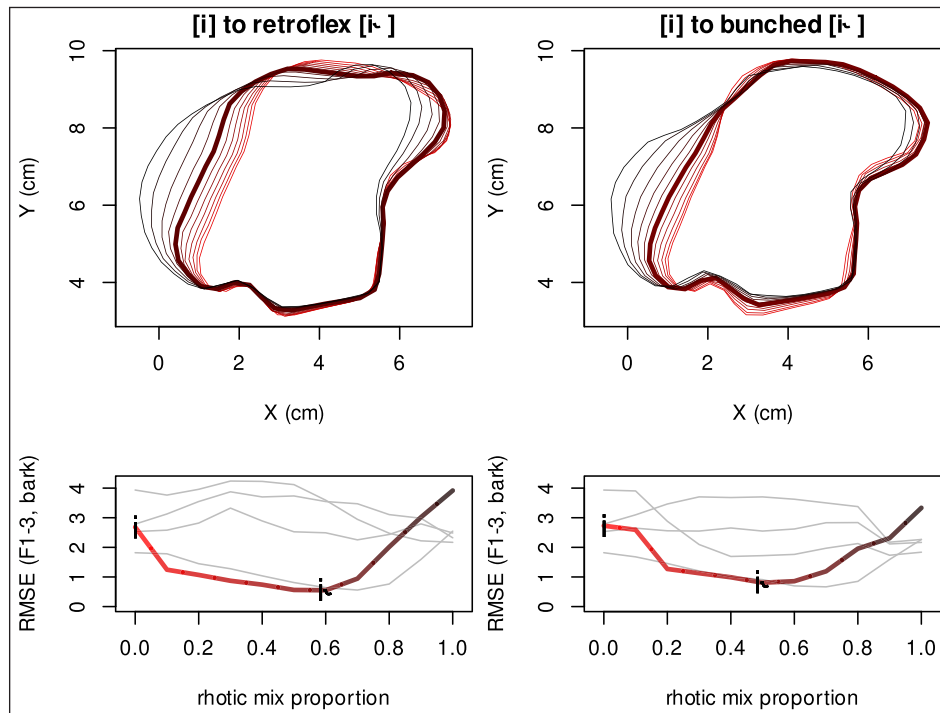

Figure 11: Mid-sagittal tongue shapes (top) and acoustic distances (bottom) for all steps of the retroflex (left) and bunched (right) forward simulations of /ɨ˞/. Tongue tip is to the right in tongue traces. The step that is acoustically most similar to the inverse simulation rhotic vowel is indicated by a heavy contour in the top panel and a rhotic vowel IPA symbol in the bottom panel. The least rhotic step has the brightest shading and the most rhotic step has the darkest shading, in both the top and bottom panels.
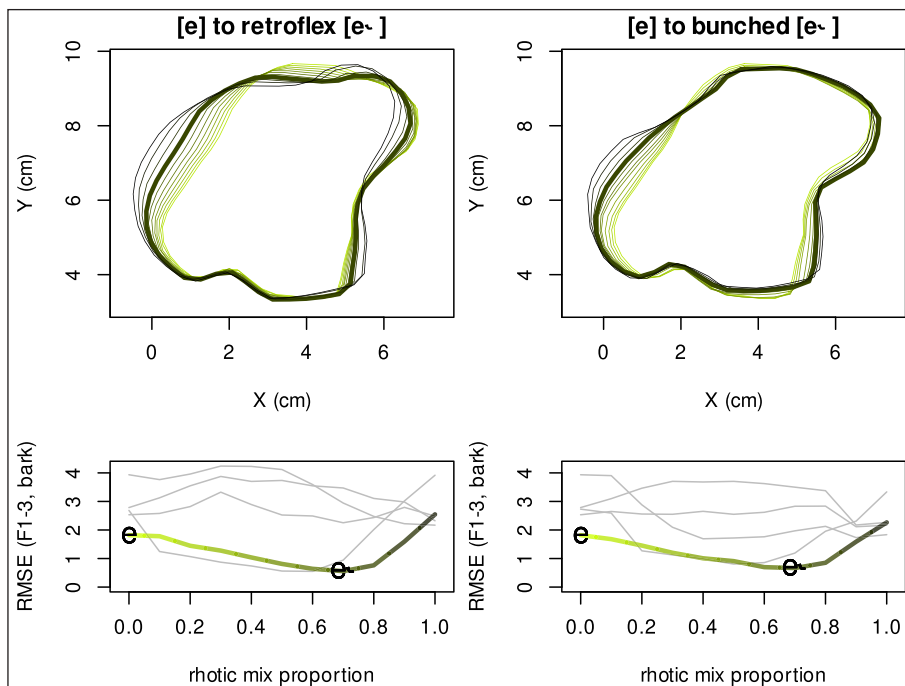
Figure 12: Mid-sagittal tongue shapes (top) and acoustic distances (bottom) for all steps of the retroflex (left) and bunched (right) forward simulations of /e˞/.
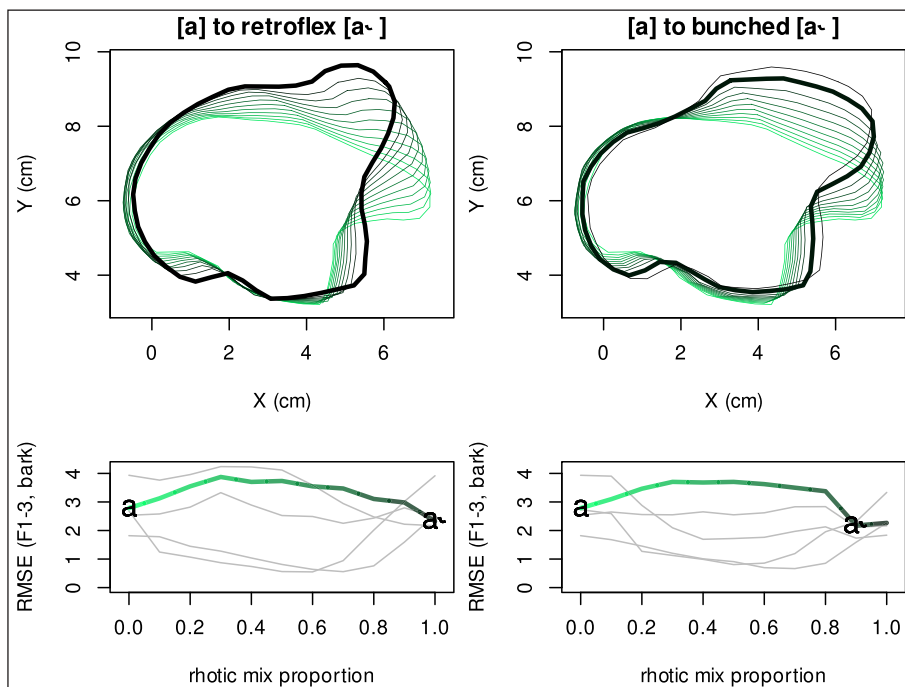

Figure 13: Mid-sagittal tongue shapes (top) and acoustic distances (bottom) for all steps of the retroflex (left) and bunched (right) forward simulations of /a˞/.
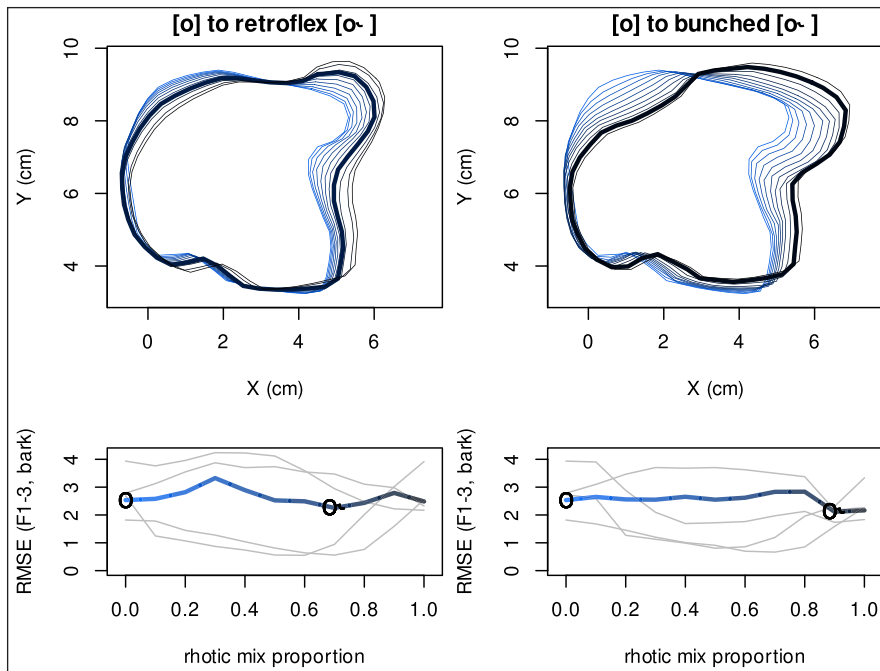
Figure 14: Mid-sagittal tongue shapes (top) and acoustic distances (bottom) for all steps of the retroflex (left) and bunched (right) forward simulations of /ɔ˞/.
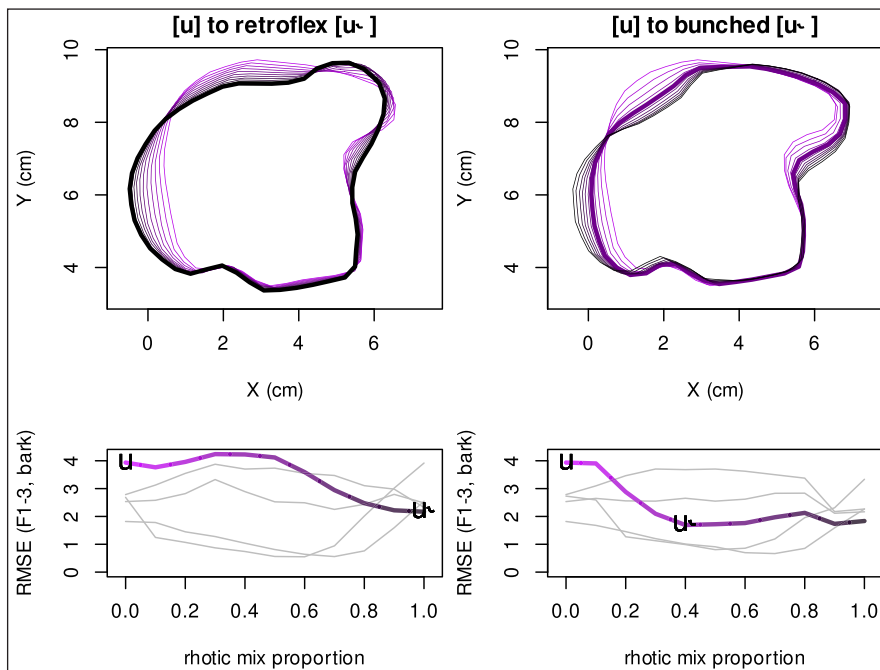

Figure 15: Mid-sagittal tongue shapes (top) and acoustic distances (bottom) for all steps of the retroflex (left) and bunched (right) forward simulations of /ʊ˞/.

The bottom acoustic distance panels show the Root Mean Squared Error (RMSE) for the bark-scaled formant frequencies of all the forward simulation rhotic vowels, with the same *x*-axis scale as **Figure 9**. Non-rhotic IPA symbols are always at the first (0% rhotic) step and rhotic IPA symbols are located at the step of each series (retroflex and bunched) that is most similar acoustically to the inverse simulation rhotic vowel (corresponding to the heavy tongue trace in the panel above it and the vertical lines in **Figure 9**).

The acoustic distances in the bottom panels of these figures show that the best matches for the front vowels are closer matches than the best matches for the back vowels, which mostly occur close to the 100% rhotic end of the scale. In other words, the inverse simulation rhotic front vowels /ɨ˞ e˞/ are intermediate between the corresponding plain front vowel and a rhotic approximant, while the inverse simulation rhotic non-high back vowels /a˞ ɔ˞/ are not very similar to any of the mixture proportions, but most similar to the rhotic end of the scale. The bunched rhotic high back vowel /u˞/ is more similar to the front vowels, and the retroflex rhotic /ʊ˞/ is more similar to the back vowels.

Figures 16–17 show how the various simulations compare according to the articulatory parameters illustrated in **Figure 6**. The distributions of inverse simulation plain and rhotic vowels are represented by boxes, and the individual steps of the forward simulations are represented by circles in between them, with the 11 retroflex simulations (labeled with ɻ) always to the left of the 11 bunched simulations (labeled with ɹ), with each group getting more rhotic from left to right. The circle for the step that is acoustically most similar to the inverse simulation rhotic is filled. Note that for every vowel, there is no difference between the first step of the retroflex and bunched forward simulation series, because no rhotic is mixed in to the plain vowel, and that the final step for every retroflex (or bunched) simulation is the same as all the other retroflex (or bunched) simulations, because no vowel is mixed in.

In terms of tongue body height, all of the forward simulations are close to the inverse simulation rhotic vowel distribution at the step that is acoustically most similar to it (the filled circle), with the exception of retroflex /o˞/. Adding retroflexion to /o/ does not provide any of the tongue body raising that is observed in Kalasha /o˞/.

In terms of tongue body advancement, the front and back vowels react differently to the addition of rhoticity. The retroflex /ɨ˞/ and /e˞/ match the inverse simulation rhotic vowels at the acoustically most similar step, but the corresponding bunched simulations do not match the inverse simulation rhotic vowels until additional bunching is added. Retroflex /a˞/ has the right amount of tongue body advancement, but the acoustically most similar bunched step overshoots it by a little. The other back vowels have an appropriate amount of tongue body advancement only in the bunched simulations. The forward simulation retroflex versions of /o˞/ and /ʊ˞/ have insufficient tongue body advancement.
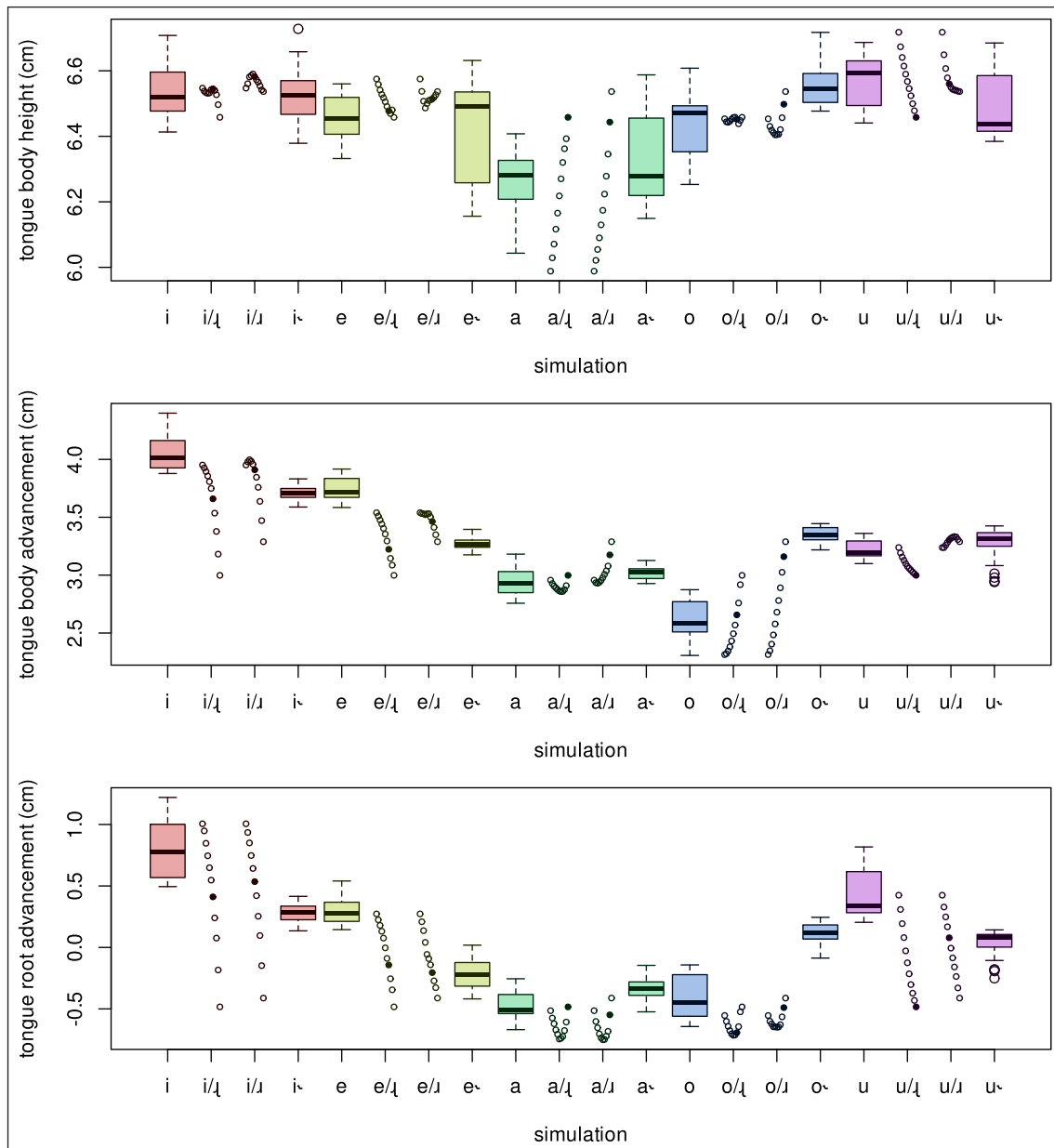
Figure 16: Tongue body and tongue root measures for all vowel simulations. Distributions of simulated plain and rhotic vowels are represented by boxes. Each series of 11 forward simulations is represented by circles, 0–100% retroflex first, followed by 0–100% bunched, with a filled circle for the forward simulation acoustically closest to the inverse simulation rhotic vowel.

Figure 17: Tongue tip and tongue blade measures for all vowel simulations. Distributions of simulated plain and rhotic vowels are represented by boxes. Each series of 11 forward simulations is represented by circles, 0–100% retroflex first, followed by 0–100% bunched, with a filled circle for the forward simulation acoustically closest to the inverse simulation rhotic vowel.
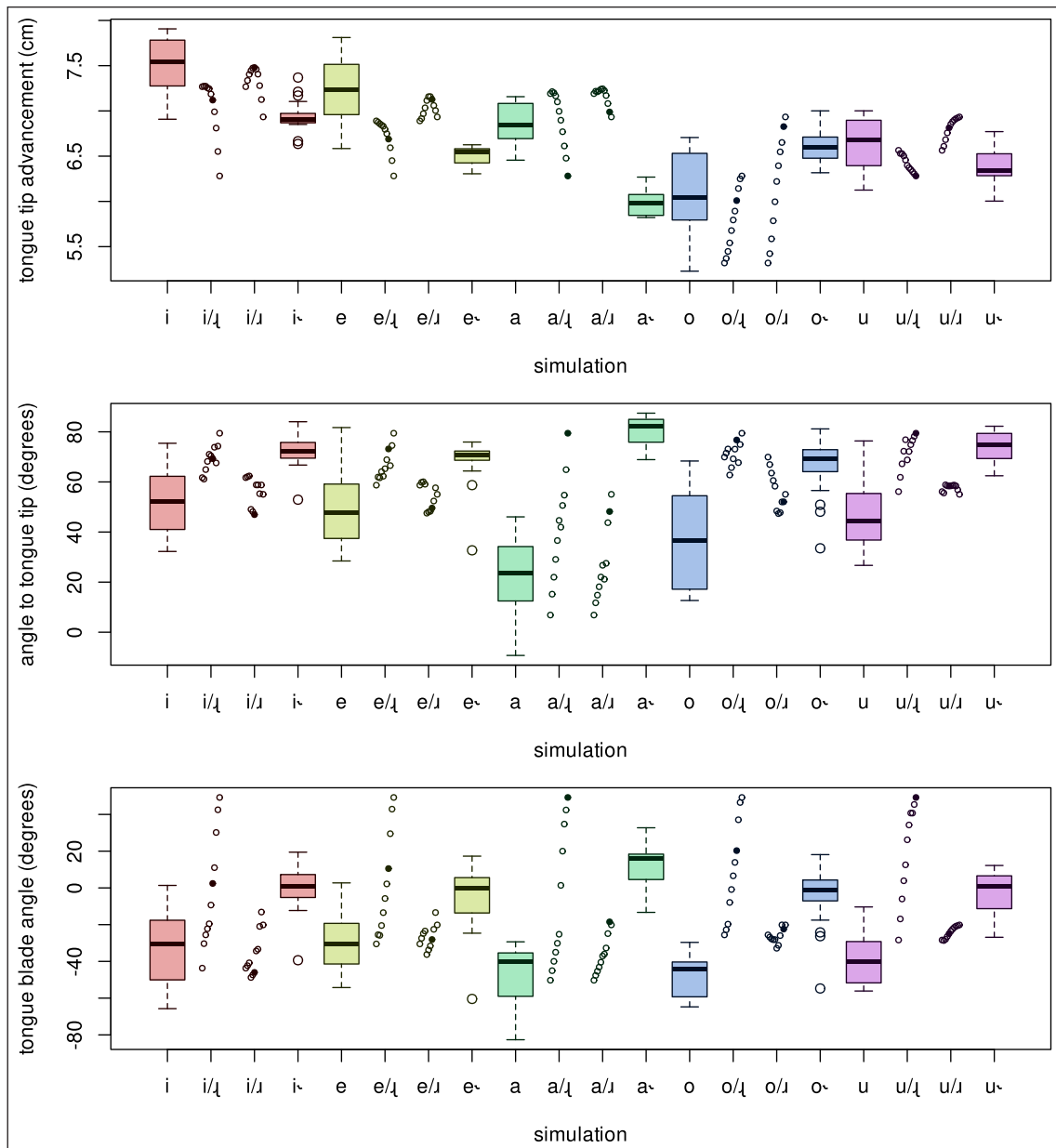
For tongue root advancement, all /ɨ˞/ and /e˞/ simulations get close to the corresponding inverse simulations. The /a˞/ simulations do not show any tongue root advancement relative to the plain category until the last step of each series. None of the steps of the /o˞/ forward simulations have the tongue root advancement that is seen in /o˞/. Like the front vowel simulations, the /u˞/ simulations span a wide range of tongue root positions that include the degree of retraction of the inverse simulation /u˞/, but resemblance to the inverse simulation occurs at the acoustically most similar step only for the bunched simulation.

Moving to the tongue tip and blade (**Figure 17**), the bunched and retroflex simulations produce too little tongue tip retraction for /ɨ˞ e˞ a˞/ at the step that is closest acoustically to the inverse model, although most of them reach the appropriate amount of retraction at a more extreme step. The forward simulation bunched /o˞/ has a good amount of tip advancement, but the retroflex simulation is too retracted. The forward simulation retroflex /u˞/ has a good amount of tip retraction, but the bunched simulation is too advanced. The angle to the tongue tip is too low for all of the bunched simulations and too high for the retroflex /o˞ u˞/ simulations. Tongue blade angle is too high for all retroflex simulations except /ɨ˞/, and too low for all bunched rhotic simulations.

Another way that the forward simulation rhotic vowels differ from observed Kalasha rhotic vowels is that their lip posture is interpolated between the plain vowel and the rhotic approximant, but observed Kalasha rhotic vowels have virtually the same lip posture as the corresponding non-rhotic vowels. The rhotic approximants used for the forward simulations are produced with a small lip opening that is closer to /u/ than any of the other non-rhotic vowels, but it is achieved by a relatively closed jaw and relatively relaxed facial muscles rather than by orbicularis oris contraction as in /u/. So most of the forward simulation rhotic vowels generally have a smaller lip opening than the observed Kalasha rhotic vowels, but less lip protrusion in /u˞/. This lack of variation in lip posture is likely to be one reason for the lack of acoustic differentiation between the forward simulation rhotic vowels.

We also note that all of the observed Kalasha rhotic vowels have higher F1 than their non-rhotic counterparts, as seen in **Figure 2**, but in the inverse and forward simulations all the rhotic vowels are closer to the middle of the F1 range than their non-rhotic counterparts. This may be accounted for by an additional phonetic feature of Kalasha rhotic vowels that has not been included in the modeling: Kalasha rhotic vowels appear to be produced with a raised larynx voice quality. We have recognized this as an auditory feature of rhotic vowels and we have noticed large changes in the angle of the hyoid bone shadow in our ultrasound images. Larynx raising shortens the vocal tract, raising the frequency of F1 in particular. If larynx raising were included in the inverse simulation rhotic vowels, their F1 values would be higher, more in line with the observed rhotic vowels. In any case, larynx raising appears to be an additional feature of Kalasha rhotic vowels that is not accounted for by coarticulation to rhotic approximants.

5. Discussion

This study investigated the development of rhoticity in a vowel system using biomechanical modeling. The likely sources of Kalasha rhotic vowels are deleted postvocalic retroflex consonants (by sound change) and Nuristani rhotic approximants (by contact). It is reasonable to consider the initial state of Kalasha rhotic vowels to be a plain vowel coarticulated with a rhotic approximant. So we have sought to find out how similar present-day Kalasha rhotic vowels are to a rhotic approximant blended with the corresponding plain vowel. Here we revisit the questions from the introduction, where we asked whether the development of rhotic vowels from plain vowel + rhoticity is basically additive from an articulatory standpoint (i.e., whether the rhotic vowels retain the lingual articulation of the corresponding non-rhotic vowels, whether they retain any signs of retroflexion from the historically present retroflex consonants, and whether the acoustic properties of rhotic vowels resemble what would be expected from adding retroflexion or bunching to the corresponding non-rhotic vowels). The genesis of rhotic vowels strongly points to retroflexion (given that it was the retroflex class of consonants that triggered the change).

5.1. Answers to research questions

Do the rhotic vowels retain the lingual articulation of the corresponding non-rhotic vowels? The Kalasha plain vowels /i e a o u/ differ from each other in tongue height and advancement in the expected ways, as seen above in **Figure 16**: /i e/ have a relatively advanced tongue body, /u o a/ have a relatively retracted tongue body, and within those front and back groups the vowels are distinguished by tongue body height. /i/ has the most advanced tongue root, followed by /e u/ and then /a o/. For most of these vowels, adding either bunching or retroflexion is expected to increase tongue body height and neutralize advancement while also increasing tongue root retraction.

The inverse simulation rhotic vowels /ɪ˞ e˞ o˞ u˞/ all have similar tongue body height, and /a˞/ is lower. The retroflex and bunched forward simulations all generally capture these tongue height effects, indicating that tongue height in most rhotic vowels is attributable to the results of coarticulation. The exception is retroflex /o˞/, which shows none of the tongue body raising observed in /o˞/.

Turning to tongue body advancement, the inverse simulation /e˞ o˞ u˞/ have advancement similar to each other and similar to plain /u/. /ɪ˞/ is more advanced and /a˞/ is less advanced. The retroflex forward simulations capture the tongue body advancement of /ɪ˞ e˞ a˞/ well, and the bunched forward simulations have steps that match the degree of tongue body advancement seen in the inverse simulation, but the step that is most similar acoustically has too much tongue body advancement for these three vowels. For /o˞ u˞/, the bunched forward simulations are closer. Retroflex inverse simulation /o˞/ has insufficient advancement and retroflex inverse simulation

/ʊ˞/ introduces unnecessary retraction. As with tongue body height, introducing retroflexion to plain /o/ does not yield appropriate tongue body advancement for /o˞/.

Turning to tongue root advancement, the inverse simulation /ɨ˞ o˞ ʊ˞/ have similar tongue root position, while /e˞ a˞/ show tongue root retraction similar to plain /a o/. The bunched and retroflex forward simulations of /ɨ˞ e˞ ʊ˞/ all include steps with tongue root retraction similar to the inverse simulation rhotic vowels, although the retroflex /ʊ˞/'s tongue root retraction match occurs at a step that is not similar acoustically to inverse simulation /ʊ˞/. None of the forward simulations of /a˞ o˞/ approach the tongue root advancement that is seen in these vowels, which is particularly large for /o˞/.

In summary, the advancement and raising of /o˞/ is not accounted for by coarticulation to any type of rhotic. It is similar in tongue posture to /e˞/, from which it is distinguished by lip rounding. Hussain and Mielke (2021) concluded that /o˞/ might better be classified as a front vowel, and we have shown that this fronting is not accounted for by coarticulation to a rhotic consonant. Beyond this fact about /o˞/, bunching does the best job of accounting for the tongue position of back rhotic vowels and retroflexion does the best job of accounting for the tongue position of front rhotic vowels.[6]

The addition of rhoticity to plain vowels does a good job of capturing the tongue body and tongue root position found in the front rhotic vowels /ɨ˞ e˞/, and the retroflex version is more accurate, particularly for tongue body advancement. On the other hand, adding bunching to the back vowels yields a closer match to the rhotic vowels than adding retroflexion. Retroflexion misses the tongue body and tongue root advancement of /o˞ ʊ˞/ and the tongue body raising of /o˞/. Bunching provides a much closer approximation of /o˞ ʊ˞/ but misses the tongue root advancement of /o˞/ and underestimates its tongue body advancement. The bunched and retroflex forward simulations of /a˞/ are similar, and the main miss is the slight tongue root advancement.

A possible interpretation of these facts is that the front vowels straightforwardly reflect the effects of coarticulation to retroflex consonants, and the tongue body fronting observed in back vowels may have been introduced at a stage when the dominant articulatory strategy for rhotic vowels shifted from retroflexion to bunching. This shift seems likely to have occurred once the vowel + retroflex consonant sequences were reinterpreted as vowels. Rhotic vowels have been found to be predominantly bunched (more so than rhotic approximants) in languages where they have been studied articulatorily (Jiang, Chang, & Hsieh, 2019; Mielke, 2015; Mielke et al., 2016).

---

[6] Note that this is unlikely to be accounted for bunched/retroflex allophony in rhotic approximants. Where vowel-conditioned bunched/retroflex allophony has been observed in English, the distribution of bunched and retroflex allophones is the opposite of what would account for this pattern: Retroflexion is compatible with back vowels and bunching is compatible with front vowels (Mielke et al., 2016; Ong & Stone, 1998; Stavness, Gick, et al., 2012)

Do Kalasha rhotic vowels retain any signs of retroflexion of the historically present retroflex consonants? Although the direct source of Kalasha rhotic vowels is believed to be coarticulation to a following retroflex (not bunched) consonant, Hussain and Mielke (2022) found only bunched variants of Kalasha rhotic vowels, for all vowel qualities. Here we simulated coarticulation to both retroflex and bunched rhotic approximants. As seen above in **Figure 17**, the inverse simulation rhotic vowels all have higher tongue blade angle and a higher angle to the tongue tip than their non-rhotic counterparts, and all but /o˞/ have a more retracted tongue tip. The bunched forward simulation rhotic vowels achieve a much better match to the inverse simulation rhotic vowels in terms of tongue blade angle, and they are comparable in terms of tongue tip advancement and the angle to the tongue tip. Coarticulation to the Kamviri bunched approximant provides a clearly better match to Kalasha rhotic vowels than coarticulation to the Kamviri retroflex approximant does, and this difference is most apparent in tongue blade angle. Kalasha /o˞/ has a more advanced tongue tip than would be expected for a retroflex version of /o/ because it is a more advanced vowel than /o/, but the bunched simulation nevertheless does a good job of approximating the observed tongue tip advancement.

Do the formant frequencies of rhotic vowels differ from what would be expected from adding retroflexion or bunching to their non-rhotic counterparts? Nearly all of the ways that the observed Kalasha rhotic vowels differ from their plain counterparts are replicated in the inverse simulation vowels, although the magnitudes of the differences vary a lot. This is shown above in **Figure 10**. The rhotic vowels all have lower F3 than their plain counterparts. F2 is lower in front rhotic vowels /i˞ e˞/ and higher in back rhotic vowels /a˞ o˞ u˞/ relative to their plain counterparts. F1 is higher in /i˞ u˞ e˞/ and lower in /a˞/, and it is similar between /o˞/ and its non-rhotic counterpart. This differs somewhat from the wholesale F1 increase shown in **Figure 2**, and we have suggested that the F1 increase in Kalasha rhotic vowels could be due to larynx raising, which is not included in any of the simulations.

The bunched and retroflex forward simulations both do a good job of capturing the F1 increase and F2 and F3 reduction observed in /i˞ e˞/. Both forward simulations of /a˞/ have F1 and F2 that are too low. The bunched and retroflex forward simulations differ the most for the two back rounded rhotic vowels /o˞ u˞/, with higher F3 for bunched /o˞/ and higher F2 for bunched /u˞/, relative to the retroflex versions of these vowels. The bunched models are somewhat better at capturing the F3 values (the retroflex models yield too much F3 decrease), but all of the models of these vowels are too low in F2. They do not account for the acoustic centralization or fronting of these vowels. In summary, there does not seem to be any evidence that any acoustic details of the Kalasha rhotic vowels are attributable specifically to the retroflexion in their history rather than rhoticity in general.

An important possibility to consider is that Kalasha rhotic vowels are somehow optimized to maintain perceptual distinctiveness within the entire vowel system. This is explored in the next subsection.

5.2. Vowel dispersion

Kalasha appears to have once had a system of five oral vowels distinguished primarily by F1 and F2, and now it has a system of ten oral vowels distinguished by F1, F2, and F3.[7] **Figure 18** shows how adding rhoticity affects the dispersion of vowels in F1–F2–F3 space. The top panel shows that at a rhotic proportion of zero, all rhotic vowels are identical to their non-rhotic counterparts, and at a rhotic proportion of one, all rhotic vowels are identical to each other. As the rhotic proportion is increased from zero to one, the difference within rhotic-non-rhotic pairs generally increases, and the difference between rhotic vowels generally decreased. The curves for rhotic vowels and rhotic-non-rhotic pairs cross at a rhotic proportion between 0.7 and 0.8, where the average distances between vowels are closely matched by the average distances between inverse simulation vowels. This is also a rhotic mix proportion that results in vowels that most closely match the inverse simulation acoustically. At low rhotic mix proportions, bunched rhotic vowels are somewhat less distinct from each other than similar retroflex rhotic vowels, but these differences disappear before the rhotic mix proportion reaches 0.7. We note that for /ɨ˞ e˞/, the two vowels with a good acoustic match between inverse and forward simulations, the best match was found at a rhotic mix proportion of 0.5–0.7 (a mixture of 50–70% bunched or retroflex approximant and 30–50% plain vowel). For the vowels with a relatively poor acoustic match, the closest step tended to be closer to the 1.0 rhotic approximant end of the scale. Thus, maximizing the acoustic dispersion of the forward simulation vowels and maximizing their acoustic similarity to inverse simulation vowels both point to 70% rhotic approximant and 30% plain vowel as a reasonable mixture, at least for rhotic front vowels, or at least for vowels that resemble rhotic versions of plain vowels.

The middle and bottom panels show all rhotic-non-rhotic pairs, and all pairs of rhotic vowels, respectively. Liu and Kewley-Port (2004) identify 0.37 barks as a threshold for listeners to distinguish vowels.[8] All pairs of vowels at rhotic mix proportions from 0.2 to 0.9 exceed this threshold. The most indistinct rhotic-non-rhotic pair is /o o˞/. The most indistinct pairs of rhotic vowels are bunched /a˞ o˞/, retroflex /e˞ ɨ˞/, and bunched /e˞ ɨ˞/. Pairs of vowels differing in both quality and rhoticity are not depicted here, but for all rhotic mix proportions from 0.2 to 0.7, and for both retroflex and bunched simulations, the most indistinct among these pairs is /e ɨ˞/.

---

[7] It also has the ten nasalized counterparts of these vowels, which are not addressed here.

[8] We note that Liu and Kewley-Port (2004) compared pairs of vowels differing in one formant, and here we are measuring Euclidean distance based on three formants.
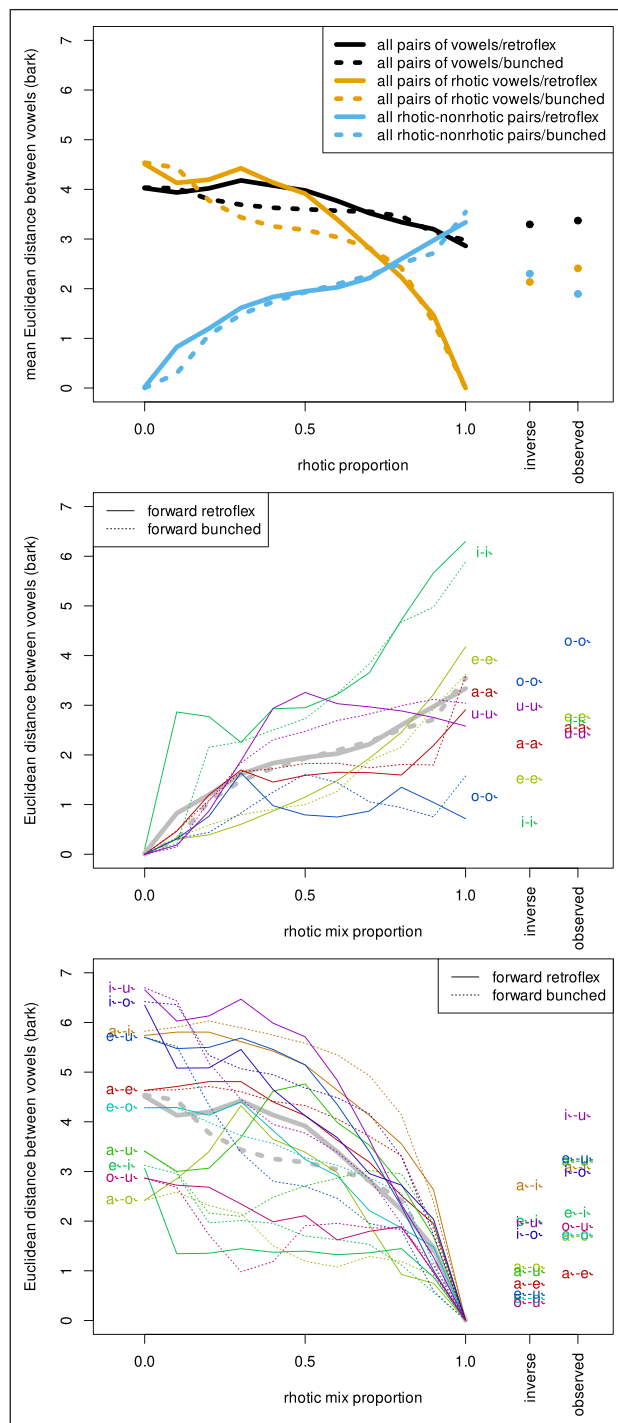
Figure 18: Vowel dispersion by rhotic mix proportion and simulation type. Mean Euclidean distance (based on bark-scaled F1, F2, and F3) among all pairs of vowels, rhotic vowels only, and rhotic-non-rhotic pairs only (top); all rhotic-non-rhotic pairs (middle); all pairs of rhotic vowels (bottom).

If Kalasha vowels are more dispersed than would be expected based on coarticulation, we expect the inverse simulation vowels to be more dispersed than the forward simulation vowels. Looking at the whole system, the inverse simulation vowels are no more dispersed than any step of rhotic proportion, and bunching does not make for a more dispersed vowel system than retroflexion. The mean dispersion within the most vulnerable vowels (the rhotic-non-rhotic pairs and the pairs of rhotic vowels) is very similar between the inverse simulation vowels and the forward simulation vowels at a rhotic mix proportion of 0.7, which is also the point where the forward simulation rhotic tended to match the inverse simulation vowels most closely. There does appear to be some exaggeration of the differences in particular vulnerable pairs of vowels. /o ɔ˞/ is the least distinct pair of vowels in the forward simulations, but it is the most distinct pair of vowels in the inverse simulations, and in the observed Kalasha tongue shapes. As previously discussed, /ɔ˞/ is considerably different from /o/ in articulation, and probably better described as a front vowel with a tongue shape similar to /e˞/ combined with lip rounding similar to /o/.

Some Kalasha vowels are more dispersed acoustically than they would be if they were simply plain vowels combined with either type of a rhotic approximant. This is consistent with adaptive dispersion (de Boer, 2000; Flemming, 2002; Lindblom, 1986, 1990; Padgett & Tabain, 2005) manifesting early in the development of a new vowel sub-system. The observed acoustic dispersion involves articulatory gestures not explained by the rhotic approximants or the plain vowels believed to form the basis for the new rhotic vowels. However, we do not observe dispersion generally across the Kalasha vowel system. If it is an active force here, it appears to be limited to the most vulnerable pairs of vowels that are created by adding rhoticity to the system.

5.3. Comparison to nasal vowel subsystems

The most likely sequence of events in the development of Kalasha vowels seems to be that coarticulatory retroflexion was exaggerated and extended to more of the vowel interval, introducing new vowel qualities that were achieved through tongue bunching by later generations. When the vowels coalesced with rhoticity, they retained their characteristic lip postures. In the course of developing new vowel categories, the back rhotic vowels, and /ɔ˞/ in particular, were established as front rounded vowels. The development of rhotic vowels in Canadian French (e.g., [pnø] ~ [pnɚ] 'tire') has shown a similar reorganization from a different starting point: rhotic vowels developed from front rounded vowels, which were undergoing backing over time (Mielke, 2013). The modern rhotic vowels are produced as retroflex by some speakers (Mielke, 2015). The Kalasha /ɔ˞/ is phonetically similar to the Canadian French rhotic /ø/, both produced with lip rounding and a bunched tongue in the front of the oral cavity, despite one originating from a back vowel and retroflexion and the other originating from a front vowel and no articulatory source of retroflexion. See Hussain and Mielke (2022) for further discussion of these two cases.

The reorganization of the rhotic vowel subsystem recalls the reorganization that is seen in nasalized vowel subsystems. Vowel nasalization changes the acoustic output of the vocal tract in ways that interact with the perception of vowel quality, such as by making high vowels sound lower and low vowels sound higher (Diehl et al., 1991; Feng & Castelli, 1996; Fujimura & Lindqvist, 1971; Serrurier & Badin, 2008). As such, nasalization is articulatorily independent but acoustically integrated with vowel quality. The acoustic effects of vowel nasalization may be enhanced or compensated for through direct changes to tongue or pharynx posture (Barlaz et al., 2015; Beddor, 1982; Carignan, 2014, 2018; Carignan, Shosted, Fu, Liang, & Sutton, 2015; Krakow et al., 1988). Articulatory enhancement of vowel nasalization is also seen in Kalasha /ã õ ũ/ and /æ̃˞/, which appears to be merged with /ẽ/ in some speakers (Hussain and Mielke, 2021).

Since rhoticity is achieved largely with the tongue, it is less independent of vowel quality than nasalization is. The articulatory gestures that help achieve acoustic characteristics of rhoticity such as low F3 also directly affect the quality of vowels as realized through F1 and F2 by changing the shape of the oral cavity and pharynx. We have seen that the present-day Kalasha rhotic vowel subsystem has apparently compensated for some of these effects through enhancement of F1 and F2 differences among the rhotic vowels. It would not have been possible to observe this as enhancement without the reference point provided by the biomechanical and acoustic modeling.

5.4. Comparison to pre-rhotic vowel subsystems

It seems very likely that Kalasha went through a stage with distinct pre-rhotic vowel allophones prior to developing monophthongal rhotic vowels from coarticulated vowel-rhotic sequences. North American English has a bunched/retroflex rhotic approximant that affects the quality of vowels around it. We can examine the pre-rhotic vowel system of North American English for possible clues about the earlier development of Kalasha rhotic vowels before they became established as monophthongs.

North American English is typically described as having one rhotic vowel quality /ɚ/ (which may be transcribed /ɝ/ or /ɻ/), but it has many distinct pre-rhotic vowel variants. Thomas (2001, p. 44) summarizes the effects of following /ɹ/ on English vowels as follows:

> Coarticulation with /r/ shrinks the vowel space of pre-/r/ vowels and obliterates certain cues such as gliding that are used to distinguish vowels. These effects, in turn, lead to difficulty by speakers in identifying pre-/r/ vowels with particular vowel phonemes…. They also lead to mergers…

Kalasha vowels are generally monophthongal (Hussain & Mielke, 2021), so they are less susceptible to changes in gliding, and Kalasha has fewer non-rhotic oral vowel qualities than

English, so mergers may be less likely. The syllabic R of North American English is the result of a merger of Middle English /ɛɹ ɪɹ ʊɹ ɜɹ əɹ/ (Wells, 1982). None of these vowel qualities are found in Kalasha. The remaining English pre-rhotic vowels are more similar to the Kalasha vowels that merged with rhotics.

Thomas (2001, pp. 44–48) describes North American English monophthongs before /ɹ/ as follows. /iɹ/ has retracted and sporadically merged with /eɹ/. /eɹ/ lacks the upgliding found in non-pre-rhotic contexts and it is lowered and retracted (and sporadically merged with /iɹ/). /ɑɹ/ varies considerably in the front-back dimension ([ɑɹ] to [æɹ]) and also may be rounded. /oɹ/ has mostly merged with /ɔɹ/ and is pronounced as [oɹ] (without the upgliding found in non-pre-rhotic contexts). In words which historically had /uɹ/, this sequence has merged with /ɝ/ or /oɹ/, or it has been reinterpreted as bisyllabic /uɚ/. These patterns in English are largely consistent with Kalasha rhotic vowels. We have seen that rhoticity causes F2 decrease in Kalasha front vowels, and /e˞ i˞/ are very close together in the simulations. Kalasha /a˞/ is acoustically fronted, much like English /ɑɹ/. Kalasha /o˞/ is quite different from English /oɹ/, and much more like English /ɝ/. Kalasha /u˞/ is acoustically lowered, which could have caused it to merge with /o˞/ if /o˞/ had not moved toward the front of the vowel space. Indeed, /u˞/ and /o˞/ are very close in the forward simulations, which do not take into account the fronting of Kalasha /o˞/. The fronting of /oɹ/ to avoid a merger with /uɹ/ would have been more problematic in North American English, with /ɚ/ sitting in front of /oɹ/ in the vowel space.

Thomas (2001, p. 44) notes that the distinct pre-rhotic variants of English vowels are difficult for speakers to connect to vowel categories occurring in non-pre-rhotic contexts. While the conventional IPA transcription /i˞ e˞ a˞ o˞ u˞/ suggests a close relationship between rhotic vowels and their non-rhotic counterparts, we have seen that some pairs, most notably /o o˞/, bear very little resemblance to each other in the way of lingual articulation. The rhotic vowels are spelled <a' e' i' o' u'>, transparently relating them their non-rhotic counterparts <a e i o u> in Cooper's (2005) orthography. This may encourage association between historically related vowel pairs among speakers who read and write using this orthography. We have not collected metalinguistic judgments about associations between rhotic and non-rhotic vowels, and we are not aware of any phonological patterns supporting phonological relationships between particular vowels (see Kochetov et al. 2021), but we note one similarity (lip rounding) which may support the continued connection between rhotic vowels and their historically related non-rhotic vowels. Rhotic vowels are produced with virtually identical lip postures to their non-rhotic counterparts (Hussain & Mielke, 2021). Even though lip rounding would be an excellent way to help achieve the low F3 targets for these vowels, speakers maintain lip postures that match the vowels' historical non-rhotic counterparts.

Walker and Proctor (2019) show that American English /ɑɹ/ and /oɹ/ involve little movement of the posterior tongue compared to /iɹ/ and /eɹ/, in which the tongue retracts considerably going from the vowel to the rhotic. They interpret this as a reason why /ɑɹ/ and /oɹ/ appear to only have a single mora each and may be followed by coda consonants as in *dark* and *fork*, whereas other vowel + /ɹ/ sequences appear to be bimoraic and cannot be followed by coda consonants. Walker and Proctor (2019) suggest that /ɑɹ/ and /oɹ/ might be particularly well-suited to being monophthongal rhotic vowels resembling coarticulated sequences of plain vowels and rhotic approximants. However, we have seen that the anterior tongue position is quite different between rhotic and non-rhotic /a/ and /o/ in Kalasha, and the rhotic vowels are rather different from what is predicted based on coarticulation. In summary, what makes /ɑɹ/ and /oɹ/ particularly good sequences in English does not extend well to monophthongal rhotic vowels with a single tongue-front posture throughout the vowel.

6. Conclusion

Despite the historical source of rhotic vowels being *retroflex*, Kalasha rhotic vowels are (at least predominantly) *bunched*. To explore what kind of vowels would result from coarticulation between Kalasha's plain vowels and a source of rhoticity, we used biomechanical modeling to superimpose the gestures and acoustic modeling to examine the resulting formant frequencies. We found that present-day Kalasha rhotic vowels are not simply plain vowels with retroflexion or bunching added. The introduction of rhoticity has caused a reorganization of the vowel space, most notably in the form of the fronting of back vowels, especially /ɚ/, which is much more acoustically distinct from /o/ than would be predicted based on coarticulation alone. We did not observe wholesale dispersion of Kalasha's ten oral vowels, but we observed that /ɚ/, the rhotic vowel most vulnerable to merger in its apparent original coarticulated form, is the vowel that differs the most from what would be expected based on coarticulation. It has shifted to being a front rounded vowel, potentially avoiding the mergers that affected pre-rhotic /u/ and /o/ in North American English.

We found some converging evidence that ideal rhotic vowels are 70% rhotic and 30% vowel in terms of muscle activation. The best balance of acoustic dispersions within the rhotic vowel subsystem and within the pairs of rhotic and non-rhotic vowels happens at rhotic mix proportions of 0.7 and 0.8. Also, the rhotic vowels whose forward simulations best matched their inverse simulations matched the best at rhotic mix proportions of 0.6 or 0.7.

The present study shows how a language can exploit already-available consonantal features to transform a simple vowel inventory into a crowded vowel space. Compared to more frequent features such as nasality and length, rhoticity directly interacts with the tongue postures used to achieve vowel quality and necessitates reorganization of the articulatory realization of vowel

contrasts. The introduction of rhoticity has transformed the lingual articulation of rhotic vowels, especially /ɔ˞/. This involves changes in its position in the vowel space relative to what would be expected from coarticulation, and it also involves new lingual means of achieving the low F3 of rhotic vowels (bunching rather than retroflexion). Through all this, lip rounding stands as a conspicuously unexploited means of lowering F3 in rhotic vowels. Despite the fact that lip rounding could aid in producing low F3 in rhotic vowels that are not already rounded, the rhotic vowels maintain the lip rounding specifications of their corresponding non-rhotic vowels, suggesting that they are still related in the minds of Kalasha speakers.

We end with a brief history of technology and our understanding of rhotic vowels. Trail and Cooper (1985), Heegård and Mørch (2004), and Cooper (2005) described the 20 vowels of modern Kalasha in terms of vowel quality, nasality, and retroflexion, and Heegård and Mørch (2004) and Di Carlo (2016) described how they likely originated from coarticulation and contact. Hussain and Mielke (2022) postulated that the emergence of Kalasha rhotic vowels was probably further amplified by retroflex approximant /ɻ/, which is widely found in neighboring Nuristani languages. Kochetov et al. (2021) provided an acoustic description of the Kalasha vowel system confirming the low F3 of rhotic vowels and the effects of rhoticity on F1 and F2. Hussain and Mielke (2021) used ultrasound and lip video data to show that the rhotic vowels are bunched and their lip postures closely match their non-rhotic counterparts, but could only speculate about how the observed tongue postures relate to the apparent origins of rhoticity in the Kalasha vowel system. Present-day Kalasha rhotic vowels probably owe their existence to coarticulation but they are many decades removed from it and even the rawest phonetic data incorporates the phonological consequences of communicating with rhotic vowels for multiple generations of speakers. Biomechanical modeling has enabled us to construct a plausible coarticulation-only simulation of the Kalasha rhotic vowel system to compare with our simulation of the actual rhotic vowels, and learn about how they are different, and find out what happens when a language develops a new vowel feature.

Supplementary materials

The overall construction of the model made use of pre-existing face (Nazari, Perrier, Chabanas, & Payan, 2010), tongue (Stavness, Lloyd, Payan, & Fels, 2011), and rigid structure model components available within ArtiSynth (Anderson et al., 2017). Collisions are used very sparingly since they can greatly reduce the stability of the model and were not deemed essential here. Collision processing was only used between the lips and the teeth. The design closely (although not exactly) follows the material properties and coupling specified in the original model sources. However, modifications were made to several model components to improve the speed of model development and provide greater model flexibility deemed necessary to achieve the widely varying set of empirical tongue shapes observed in the ultrasound data. The most important changes were converting the face and tongue FEMs to tetrahedral meshes, trimming of the face FEM, and elaborating the musculature of the face and tongue models. Concerning the first change, using meshes dominantly comprised of linear tetrahedral elements is less preferred compared to other types (such as quadratic tetrahedral or linear hexahedral topologies) because of issues such as mesh locking that can arise (Benzley, Perry, Merkley, Clark, & Sjaardama, 1995; Hughes, 1987). This issue was not found to be a major impediment to simulation here, but rather the gains in design flexibility and speed of model prototyping were deemed to be worthy trade-offs to make.

Figure 19 shows original face FEM and the trimmed version used in the present simulations. Concerning trimming of the face FEM, the original face model has an inferosuperior extent from below the chin (at roughly the level of the vocal folds) to above the brow ridge and an anteroposterior extent from the tip of the nose to roughly a coronal plane set just in front of the ears. While this expansive amount of facial structure could allow for simulation of details of facial expression, it is unnecessary for speech models and its presence only adds to the computational burden of the simulation. The trimming, which was conducted in Blender (www.blender.org), reduced the face to just the lower orofacial area, with a superior, transverse planar-border set immediately below the nose, an oblique plane oriented roughly parallel to the inferior border of the body of the mandible, and a posterior, roughly coronal planar-border aligned to the posterior margin of the mandibular rami.

Figures 20 and 21 show the tongue and face muscles. Elaboration of the tongue and face musculature was performed to increase the density of muscle fiber representation in both models, which are, in the original versions, quite sparse. This was done to improve the distribution of the musculature within a material-based modeling approach (and specification of muscle fiber directions at element integration points). Redesign of the muscles was performed in Blender and a flexible system was developed along with an export script to facilitate the process of making small changes to the musculature and facilitate further refinements in future models. Both sets of musculature are based on available anatomical resources (e.g., Zemlin, 1998), but the tongue
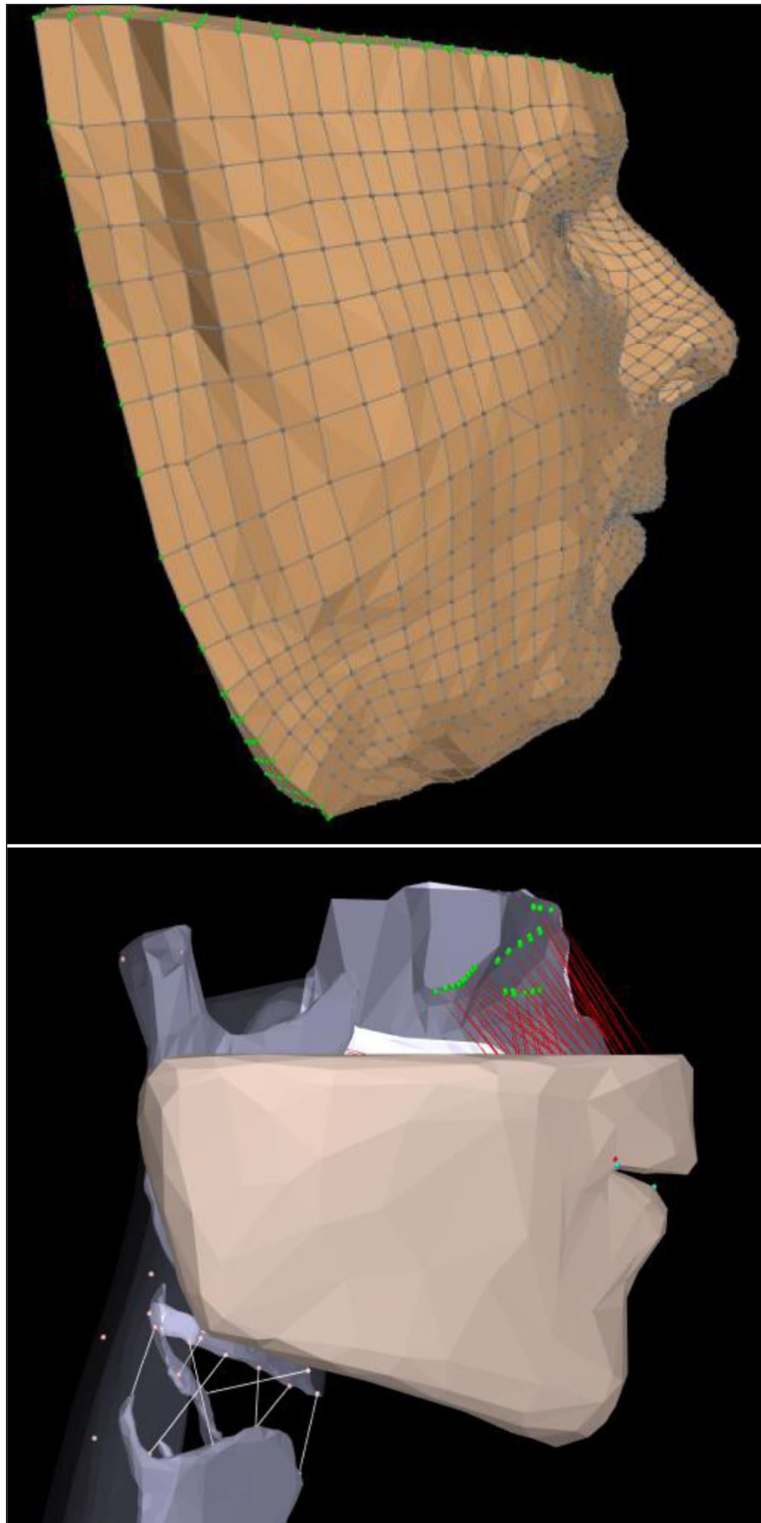
Figure 19: Original face FEM (top) and the trimmed version (bottom) used in the present simulations.
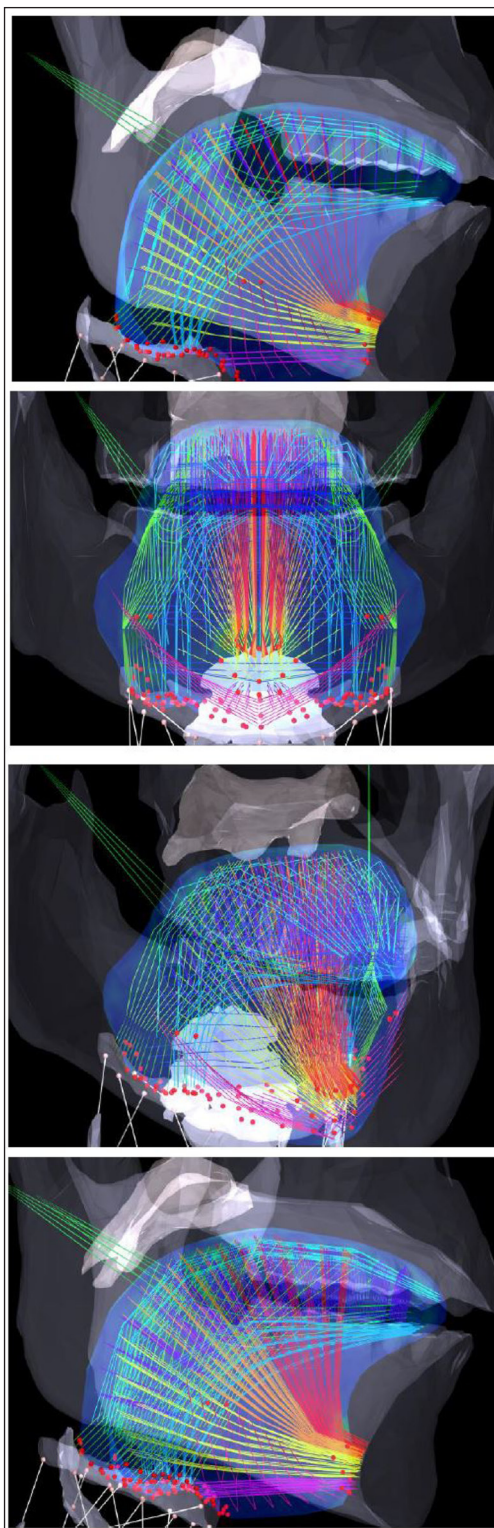
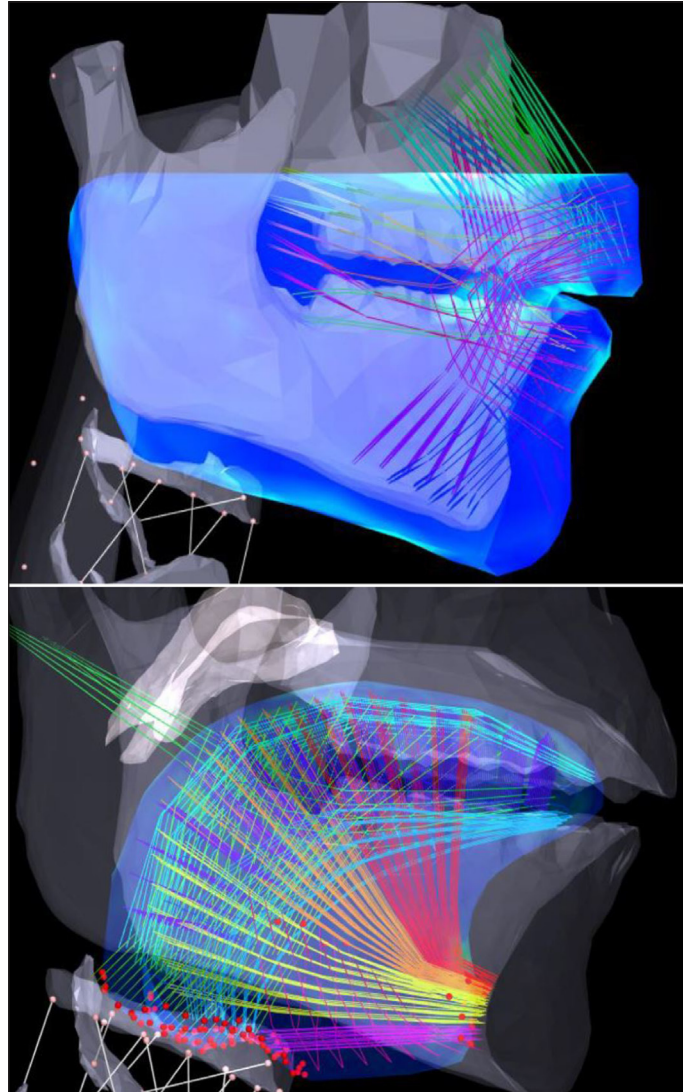**Figure 20:** Tongue muscles for ArtiSynth modeling.

Figure 21: Refined muscles of the face (top) and tongue (bottom).

muscles are also closely based upon details presented in Sanders and Mu (2013), with the images therein having been imported into Blender and used to guide the layout of the muscle fibers. Muscles requiring extrinsic attachment to skeletal structure (outside of the bounding FEMs) were supported by axial muscles partially embedded in the FEM. A tendinous origin of the genioglossus muscle fibers was also developed.

All inverse simulations used a damping term of 0.5 and an l2-normalization term of 1.0. Target weighting across the different articulators was freely adjusted to facilitate finding a balance between accuracy of the tracking and stability of the simulation (some empirical configurations were difficult to match with the model and we needed to relax the target weighting where this occurred).

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Authors' contributions

JM: Conceptualization, Investigation, Methodology, Data analysis and visualization, Writing

QH: Conceptualization, Investigation, Methodology, Data collection and visualization, Writing

SRM: Investigation, Methodology, Data analysis and visualization, Writing

References

Anderson, P., Fels, S., Harandi, N. M., Ho, A., Moisik, S., Sánchez, C. A., … Tang, K. (2017). Frank: A hybrid 3d biomechanical model of the head and neck. In *Biomechanics of living organs* (pp. 413–447). Academic Press. DOI: https://doi.org/10.1016/B978-0-12-804009-6.00020-1

Baker, A. (2005). Palatoglossatron 1.0 [Computer software manual]. Tucson, Arizona. (http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf)

Barlaz, M. S., Fu, M., Dubin, J., Liang, Z.-P., Shosted, R., & Sutton, B. P. (2015). Lingual differences in Brazilian Portuguese oral and nasal vowels: An MRI study. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). Glasgow, UK: University of Glasgow. (Paper number 819)

Beddor, P. S. (1982). *Phonological and phonetic effects of nasalization on vowel height* (Unpublished doctoral dissertation). University of Minnesota. (Bloomington, IN: Indiana University Linguistics Club).

Benzley, S. E., Perry, E., Merkley, K., Clark, B., & Sjaardama, G. (1995). A comparison of all hexagonal and all tetrahedral finite element meshes for elastic and elasto-plastic analysis. In *Proceedings of the 4th international meshing roundtable* (pp. 179–191).

Birkholz, P. (2005). *3d-artikulatorische sprachsynthese* (Unpublished doctoral dissertation). Universität Rostock, Germany.

Birkholz, P., & Jackel, D. (2004). Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system. In *Proceedings of Interspeech 2004* (pp. 1125–1128). DOI: https://doi.org/10.21437/Interspeech.2004-409

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer [Computer program]. (Version 6.0.30, http://www.praat.org)

Brunner, J., & Zygis, M. (2011). Why do glottal stops and low vowels like each other? In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 376–379). City University of Hong Kong, Hong Kong.

Carignan, C. (2014). An acoustic and articulatory examination of the "oral" in "nasal": The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics*, *46*, 23–33. DOI: https://doi.org/10.1016/j.wocn.2014.05.001

Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America*, *143*(5), 2588–2601. DOI: https://doi.org/10.1121/1.5034760

Carignan, C., Shosted, R. K., Fu, M., Liang, Z.-P., & Sutton, B. P. (2015). A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French. *Journal of Phonetics*, *50*, 34–51. DOI: https://doi.org/10.1016/j.wocn.2015.01.001

Catford, J. C. (1983). Pharyngeal and laryngeal sounds in Caucasian languages. In D. M. Bless & J. H. Abbs (Eds.), (pp. 344–350). College-Hill Press.

Cooper, G. (2005). *Issues in the development of a writing system for the Kalasha language* (Unpublished doctoral dissertation). Macquarie University, Sydney.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465. DOI: https://doi.org/10.1006/jpho.2000.0125

Delattre, P., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, *6*(44), 29–68. DOI: https://doi.org/10.1515/ling.1968.6.44.29

Di Carlo, P. (2016). Retroflex vowels? phonetics, phonology, and history of unusual sounds in Kalasha and other languages of the Hindu Kush region. *Archivio per L'Antropologia e la Etnologia*, *CXLVI*, 103–121.

Diehl, R. L., Kluender, K. R., Walsh, M. A., & Parker, E. M. (1991). Auditory enhancement in speech perception and phonology. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes, vol 3: Applied and ecological perspectives* (pp. 59–76). Hillsdale, NJ: Erlbaum.

Eskes, M., Balm, A. J., Van Alphen, M. J., Smeele, L. E., Stavness, I., & Van Der Heijden, F. (2017). sEMG-assisted inverse modelling of 3D lip movement: a feasibility study towards person-specific modelling. *Scientific Reports*, *7*(1), 1–14. DOI: https://doi.org/10.1038/s41598-017-17790-4

Esposito, C. M., Sleeper, M., & Schäfer, K. (2021). Examining the relationship between vowel quality and voice quality. *Journal of the International Phonetic Association*, *51*(3), 361–392. DOI: https://doi.org/10.1017/S0025100319000094

Feng, G., & Castelli, E. (1996). Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization. *The Journal of the Acoustical Society of America*, *99*(6), 3694–3706. DOI: https://doi.org/10.1121/1.414967

Flemming, E. S. (2002). *Auditory representations in phonology*. New York: Routledge.

Fujimura, O., & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. *The Journal of the Acoustical Society of America, 49,* 541–558. DOI: https://doi.org/10.1121/1.1912385

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics.* Malden, MA: Wiley-Blackwell.

Gussenhoven, C. (2007). A vowel height split explained: Compensatory listening and speaker control. In J. Cole & J. Hualde (Eds.), *Papers in laboratory phonology 9: Change in phonology* (pp. 145–172). Berlin: Mouton de Gruyter.

Hamann, S. (2003). *The phonetics and phonology of retroflexes.* Utrecht, The Netherlands: LOT.

Heegård, J., & Mørch, I. E. (2004). Retroflex vowels and other peculiarities in the Kalasha sound system. In A. Saxena (Ed.), *Himalayan Languages: Past and Present* (pp. 57–76). Berlin: De Gruyter.

Honda, K. (1996). Organization of tongue articulation for vowels. *Journal of Phonetics, 24*(1), 39–52. DOI: https://doi.org/10.1006/jpho.1996.0004

Howson, P. J., Moisik, S., & Żygis, M. (2022). Lateral vocalization in Brazilian Portuguese. *The Journal of the Acoustical Society of America, 152*(1), 281–294. DOI: https://doi.org/10.1121/10.0012186

Hueber, T., Chollet, G., Denby, B., & Stone, M. (2008). Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In *Proceedings of the Eighth International Seminar on Speech Production* (pp. 365–369). Strasbourg, France.

Hughes, T. J. (1987). *The finite element method: Linear static and dynamic finite element analysis.* USA: Prentice Hall.

Hussain, Q., & Mielke, J. (2020). Kalasha (Pakistan) – Language Snapshot. *Language Documentation and Description, 17,* 66–75.

Hussain, Q., & Mielke, J. (2021). An acoustic and articulatory study of rhotic and rhotic-nasal vowels of Kalasha. *Journal of Phonetics, 87,* 1–45. DOI: https://doi.org/10.1016/j.wocn.2020.101028

Hussain, Q., & Mielke, J. (2022). The emergence of bunched vowels from retroflex approximants in endangered Dardic languages. *Linguistics Vanguard, 8*(s5), 597–610. DOI: https://doi.org/10.1515/lingvan-2021-0022

Jackson, M. T.-T., & McGowan, R. S. (2012). A study of high front vowels with articulatory data and acoustic simulations. *The Journal of the Acoustical Society of America, 131*(4), 3017–3035. DOI: https://doi.org/10.1121/1.3692246

Jang, H. (2022). A tutorial on articulatory muscles and ArtiSynth: Tongue and suprahyoid muscles, and 3D tongue model. *Language and Linguistics Compass, 16*(3), e12447. DOI: https://doi.org/10.1111/lnc3.12447

Jiang, S., Chang, Y., & Hsieh, F. (2019). An EMA study of Er-suffixation in Northeastern Mandarin monophthongs. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences,* (pp. 2149–2153). Canberra: Australasian Speech Science and Technology Association Inc.

Kochetov, A., Arsenault, P., Petersen, J. H., Kalas, S., & Kalash, T. K. (2021). Kalasha (Bumburet variety). *Journal of the International Phonetic Association, 51*(3), 468–489. DOI: https://doi.org/10.1017/S0025100319000367

Krakow, R., Beddor, P., Goldstein, L., & Fowler, C. (1988). Coarticulatory influences on the perceived height of nasal vowels. *The Journal of the Acoustical Society of America, 83*, 1146–1158. DOI: https://doi.org/10.1121/1.396059

Krámskỳ, J. (1939). A study in the phonology of Modern Persian. *Archiv Orientální, 11*(1), 66.

Laver, J. (1980). *The phonetic description of voice quality*. London: Cambridge Studies in Linguistics.

Lehiste, I. (1962). *Acoustical characteristics of selected English consonants*. Ann Arbor: The University of Michigan Communication Sciences Laboratory.

Leidner, D. R. (1976). The articulation of American English /l/: A study of gestural synergy and antagonism. *Journal of Phonetics, 4*(4), 327–335. DOI: https://doi.org/10.1016/S0095-4470(19)31259-8

Leitner, G. W. (1880). A sketch of the Bashgali Kafirs and of their language. *Journal of the United Service Institution of India, IX*(43), 143–190.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America, 35*(11), 1773–1781. DOI: https://doi.org/10.1121/1.1918816

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). New York: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-009-2037-8_16

Liu, C., & Kewley-Port, D. (2004). Vowel formant discrimination for high-fidelity speech. *The Journal of the Acoustical Society of America, 116*(2), 1224–1233. DOI: https://doi.org/10.1121/1.1768958

Lloyd, J. E., Stavness, I., & Fels, S. (2012). Artisynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In Y. Payan (Ed.), *Soft tissue biomechanical modeling for computer assisted surgery* (pp. 355–394). Berlin: Springer. DOI: https://doi.org/10.1007/8415_2012_126

Lotto, A. J., Holt, L. L., & Kluender, K. R. (1997). Effect of voice quality on perceived height of English vowels. *Phonetica, 54*(2), 76–93. DOI: https://doi.org/10.1159/000262212

MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *The Journal of the Acoustical Society of America, 45*(5), 1217–1233. DOI: https://doi.org/10.1121/1.1911593

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511753459

MATLAB. (2019). *version 9.7.0.1261785 (r2019b)*. Natick, Massachusetts: The Math-Works Inc. Retrieved from https://www.mathworks.com/products/new_products/release2019b.html

Mielke, J. (2013). Ultrasound and corpus study of a change from below: Vowel rhoticity in Canadian French. In *Penn Working Papers in Linguistics 19.2: Papers from NWAV 41* (pp. 141–150).

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

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language, 92*(1), 101–140. DOI: https://doi.org/10.1353/lan.2016.0019

Moisik, S. R. (2013). *The epilarynx in speech* (Unpublished doctoral dissertation). University of Victoria, Canada.

Moran, S., McCloy, D., & Wright, R. (2014). *PHOIBLE Online.* (Leipzig: Max Planck Institute for Evolutionary Anthropology). http://phoible.org (accessed on 2023-01-30)

Morgenstierne, G. (1954). The Waigali language. *Norsk Tidsskrift for Sprogvidenskap, 17*, 146–219.

Morgenstierne, G. (1973). *The Kalasha language (Indo-Iranian frontier languages, vol. 4)*. Oslo: Universitetsforlaget.

Nazari, M. A., Perrier, P., Chabanas, M., & Payan, Y. (2010). Simulation of dynamic orofacial movements using a constitutive law varying with muscle activation. *Computer Methods in Biomechanics and Biomedical Engineering, 13*(4), 469–482. DOI: https://doi.org/10.1080/10255840903505147

Nye, G. E. (1955). *The phonemes and morphemes of modern Persian: A descriptive study*. University of Michigan, USA.

Ong, D., & Stone, M. (1998). Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope, 1*(1), 1–13.

Padgett, J., & Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica, 62*(1), 14–54. DOI: https://doi.org/10.1159/000087223

Perder, E. (2013). *A grammatical description of Dameli* (Unpublished doctoral dissertation). Stockholm University, Stockholm, Sweden.

Sanders, I., & Mu, L. (2013). A three-dimensional atlas of human tongue muscles. *The Anatomical Record, 296*(7), 1102–1114. DOI: https://doi.org/10.1002/ar.22711

Scobbie, J. M., Wrench, A. A., & van der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the 8th International Seminar on Speech Production* (pp. 373–376). Strasbourg, France.

Serrurier, A., & Badin, P. (2008). A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. *The Journal of the Acoustical Society of America, 123*(4), 2335–2355. DOI: https://doi.org/10.1121/1.2875111

Smith, K. K., & Kier, W. M. (1989). Trunks, tongues, and tentacles: Moving with skeletons of muscle. *American Scientist, 77*, 28–35.

Stavness, I., Gick, B., Derrick, D., & Fels, S. (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America – Express Letters, 131*(5), EL355–EL360. DOI: https://doi.org/10.1121/1.3695407

Stavness, I., Lloyd, J. E., & Fels, S. (2012). Automatic prediction of tongue muscle activations using a finite element model. *Journal of Biomechanics*, *45*(16), 2841–2848. DOI: https://doi.org/10.1016/j.jbiomech.2012.08.031

Stavness, I., Lloyd, J. E., Payan, Y., & Fels, S. (2011). Coupled hard–soft tissue simulation with contact and constraints applied to jaw–tongue–hyoid dynamics. *International Journal for Numerical Methods in Biomedical Engineering*, *27*(3), 367–390. DOI: https://doi.org/10.1002/cnm.1423

Strand, R. F. (2011). *The sound system of Nišei-alâ.* (https://nuristan.info/lngFrameL.html).

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, *16*(3), 397–420. DOI: https://doi.org/10.1044/jshr.1603.397

Takano, S., & Honda, K. (2007). An MRI analysis of the extrinsic tongue muscles during vowel production. *Speech Communication*, *49*(1), 49–58. DOI: https://doi.org/10.1016/j.specom.2006.09.004

Thomas, E. R. (2001). *An acoustic analysis of vowel variation in New World English*. Durham, N.C.: Duke University Press. (Publication of the American Dialect Society 85).

Toosarvandani, M. D. (2004). Vowel length in modern Farsi. *Journal of the Royal Asiatic Society*, *14*(3), 241–251. DOI: https://doi.org/10.1017/S1356186304004079

Trail, R., & Cooper, G. R. (1985). *Kalasha phonemic summary.* (ms., Summer Institute of Linguistics)

Walker, R., & Proctor, M. (2019). The organisation and structure of rhotics in American English rhymes. *Phonology*, *36*(3), 457–495. DOI: https://doi.org/10.1017/S0952675719000228

Wells, J. C. (1982). *Accents of English: Volume 1*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511611759

Winges, S. A., Furuya, S., Faber, N. J., & Flanders, M. (2013). Patterns of muscle activity for digital coarticulation. *Journal of Neurophysiology*, *110*(1), 230–242. DOI: https://doi.org/10.1152/jn.00973.2012

Wood, S. (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, *7*, 25–43. DOI: https://doi.org/10.1016/S0095-4470(19)31031-9

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4th ed.). Boston, MA: Allyn and Bacon.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *The Journal of the Acoustical Society of America*, *123*(6), 4466–4481. DOI: https://doi.org/10.1121/1.2902168