A phonetic description of Pashto

Ellen O'Connor

1. INTRODUCTION

Pashto is an Indo-Iranian language spoken in southern Afghanistan and parts of Pakistan by approximately 40 to 50 million people. Pashto is closely related to neighboring languages Iranian Persian/Farsi, Dari, and Tajik. It is one of two official languages of Afghanistan, and can be divided broadly into two dialect groups: Kandahari or 'soft' Pashto; and Yusufzai or 'hard' Pashto, the literary dialect. Dialect groups are defined mostly by phonological differences, primarily in variation among obstruents [ʒ, d͡ʒ, g], [ç, ʂ, x, χ] and vowel inventories.

In the present paper we focus primarily on the dialect of a speaker born in the Nangarhar region in the Northeast and raised after age twelve in the capital, Kabul (referred to as S1). He is a bilingual speaker of both Pashto and Dari. We will also compare pronunciation by another speaker, S2, of a similar dialect, using data from an online resource (http://famdliflc.lingnet.org/index.aspx). This is primarily due to limitations in the data we received from S1 (words were occasionally skipped or misread, rendering some data unusable).

2. CONSONANTS

2.1 Obstruents

A rough description of the phonemic contrasts among obstruents is given below. Examples of narrowly transcribed minimal pairs are provided in the following sections along with discussion of the relevant contrasts.

	BILABIAL	ALV./DENTAL	POST-ALVEOLAR		VELAR	UVULAR
		LAMINAL	LAMINAL	APICAL		
STOP	p b	t̪ d̪		ʈ ɖ	k g	
FRICATIVE		s z	ʃ		x ɣ	χ
AFFRICATE			t͡ʃ d͡ʒ			

Contrasts in constriction location & orientation

Obstruents contrast roughly[1] five constriction locations: bilabial, alveolar-dental, post-alveolar, velar, and uvular. Foreign loanwords (in particular those borrowed from Farsi, Arabic or Dari) sometimes include the uvular stop [q] and labio-dental fricative [f]; however these sounds are mostly borrowed into Pashto as [k, p] except in "elegant" pronunciation.

Some grammars include alveolar/dental affricates [ts, dz], which are distinguished

[1] See discussion below on whether broadly alveolar consonants /s, z, n, l/ are alveolar or dental. To simplify the contrastive places of orientation I will consider them to be dental throughout.

orthographically from [s, z]. The spectrogram and waveforms below for *wradz* ‘day’ and *tse* ‘what’ suggest that these affricates have merged with fricatives. The final affricate has also been devoiced, and at least for one speaker – S2 – the cluster *wr* simplified. (See folder *Affricates* for audio files.)

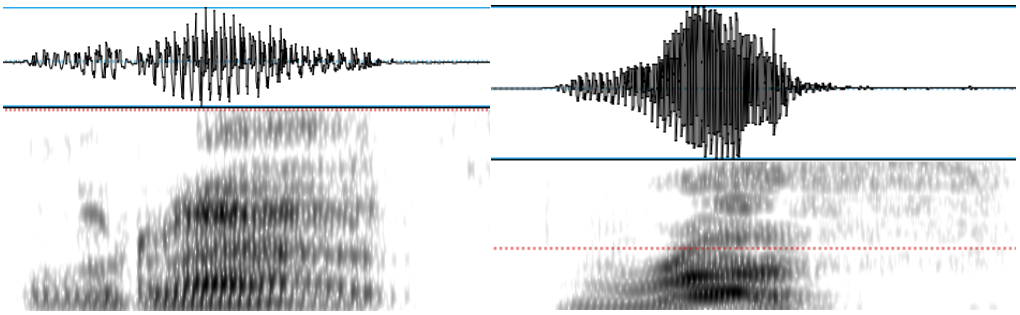


Figure 1. Pronunciation of *wradz* 'day' as [wrɑs̪]/[wrɑθ] (left, S1) and as [ɹas] (right, S2).

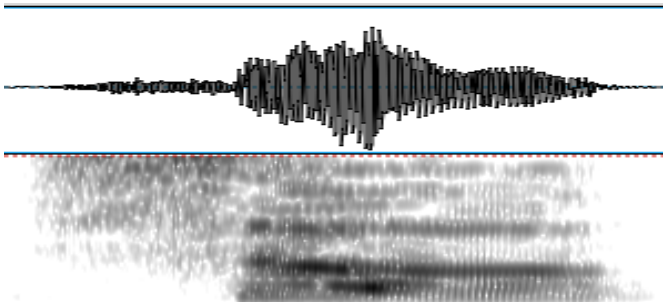
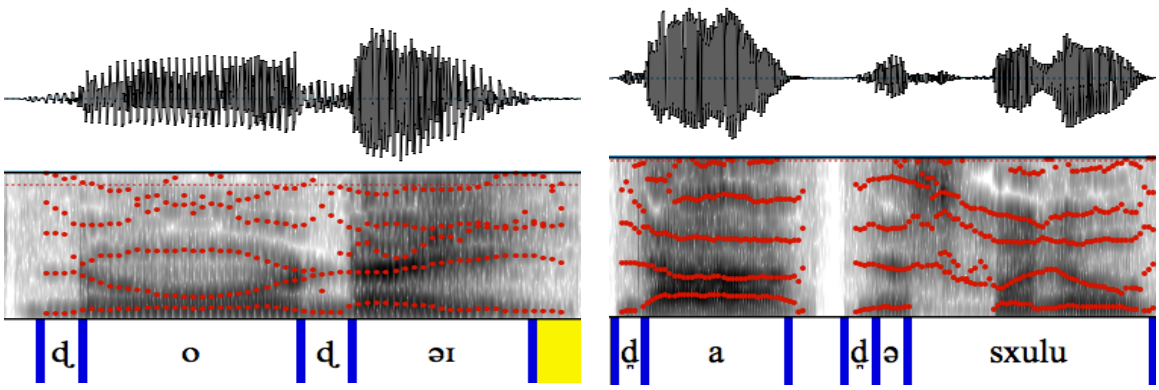


Figure 2. Pronunciation of *tse* 'what' by S2 as [sə]

Retroflex stops contrast with non-retroflex stops in both constriction location and orientation. /t̪, d̪/ are generally described as dental and sound most different from alveolar /t, d/ at the ends of words. /ʈ, ɖ/ are traditionally considered retroflex stops, but are more clearly articulated as such by S2 than S1. The file *Retroflex* exemplifies these contrasts in similar environments, using data from S2.

	Apical post-alveolar	Laminal dental
Voiceless	ʈɑŋk ‘tank’	t̪ɑn ‘hesitation’
Voiced	ɖoɖəɪ ‘food’	d̪ɔɪ d̪i 'they are'

It is clear that these sounds are retroflex by the abrupt lowering of F3 around [ɖ] below:



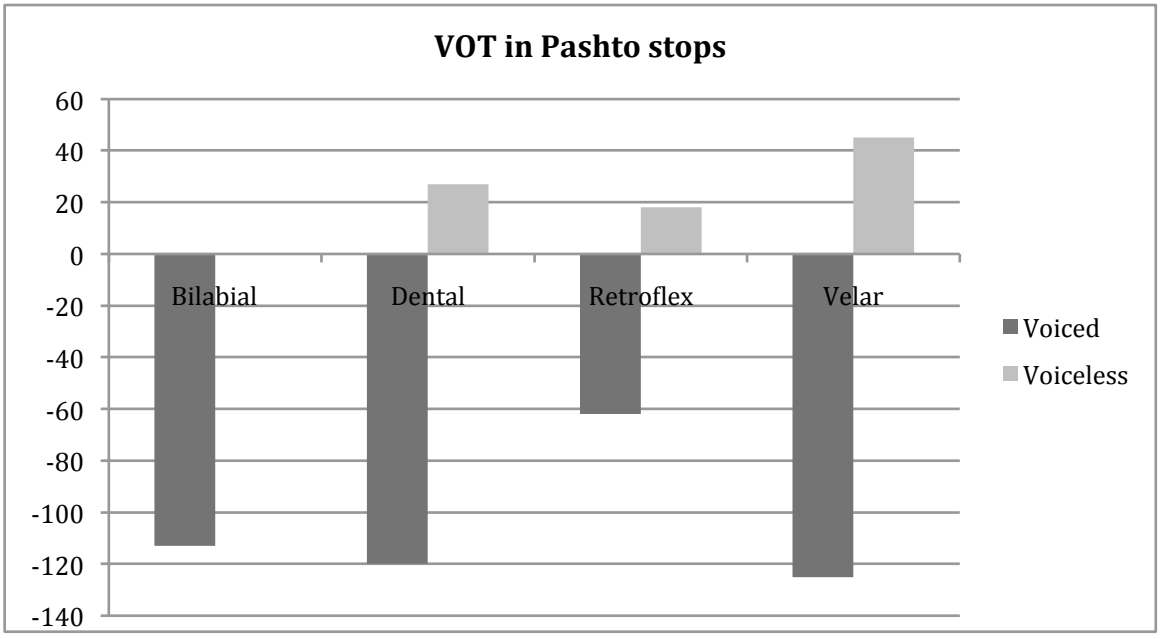
It is not obvious whether other alveolar consonants (e.g., /s, z, n/) are also primarily dental, since again it is difficult to hear a contrast between Pashto /s, z/ and English alveolar fricatives /s, z/ in word-initial position. Word-finally, however, the /z/ and /s/ fricative sounds close to /θ, ð/; therefore I will consider them to be dental fricatives throughout the paper, which also simplifies the number of contrastive constriction locations.

Contrasts in voicing

Stops and affricates are typically thought to contrast in voicing, and for S2 all of the voiced stops are indeed voiced throughout. S1 seems to show little voicing in the waveform, but it is not clear whether the lack of voicing is due to the recording quality, since the spectrogram shows a voice bar for ~130ms before release (and on some occasions the stops also look fully voiced in the waveform). VOT typically does not exceed 40 ms, but may be up to 80 ms for initial stressed voiceless velar stops. We will consider most Pashto voiceless stops to be underlyingly unaspirated (laryngeal gesture: *small opening*) but provide aspiration in narrow transcription as needed below. VOTs for S2 are given in the table below, averaged over ~8 tokens per phoneme.

Table 1. Voicing contrasts in stops for S2.

	Bilabial		Dental/Alveolar		Retroflex		Velar	
Stop	p 0 ms	b -113 ms	t̪ 27 ms	d̪ -120 ms	ʈ 18 ms	ɖ -62 ms	k 45 ms	g -125 ms
Affricate					tʃ 63 ms	dʒ -150 ms		

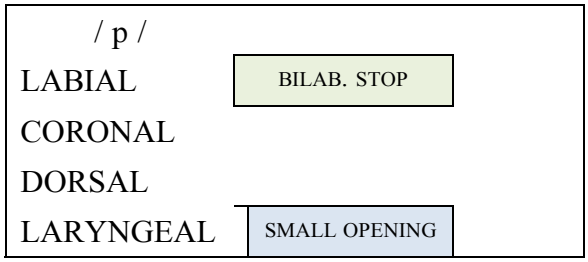


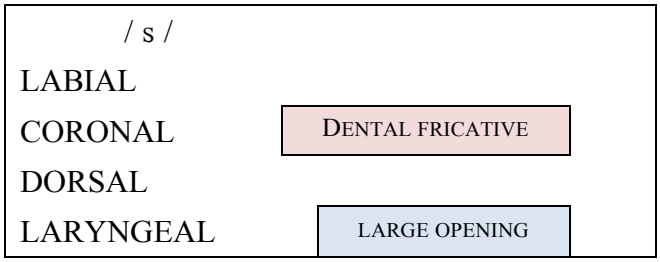
Full obstruent inventory

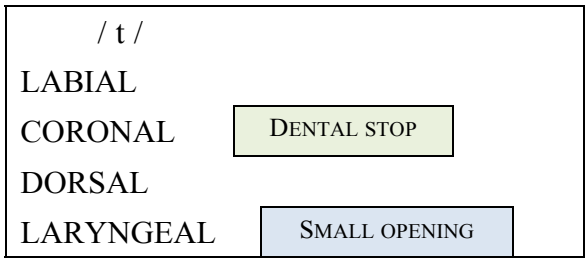
See file *Obstruents.mp3* for pronunciation by S1.

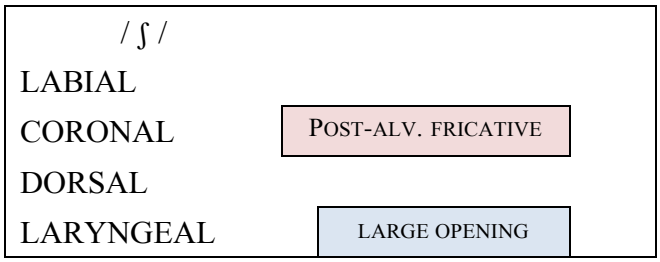
	Example	Traditional	Gestural
b	[baˌɾ] ‘up’	Pulmonic egressive voiced bilabial central oral stop	Lips: labial, stop
p	[paˌɾ] ‘on’	Pulmonic egressive voiceless bilabial central oral stop	Lips: labial, stop Laryngeal: small opening
d̪	[d̪aˌɾ] ‘door’	Pulmonic egressive voiceless dental laminal central oral stop	Tongue tip: dental, stop, laminal
t̪	[t̪əɾ] ‘from’	Pulmonic egressive voiceless dental laminal central oral stop	Tongue tip: dental, stop, laminal Laryngeal: small opening
ɖ	[ɖɑːɾ] ‘threat’	Pulmonic egressive voiceless apical post-alveolar central oral stop	Tongue tip: post-alveolar, stop, apical
ʈ	[ʈʰaʈʰaɾ] ‘breast’	Pulmonic egressive voiceless apical post-alveolar central oral stop	Tongue tip: post-alveolar, stop, apical Laryngeal: small opening
g	[gaˌɾ] ‘maker’	Pulmonic egressive voiceless velar central oral stop	Tongue dorsum: velar, stop
k	[kʰaˌɾ] ‘farming’	Pulmonic egressive voiceless velar central oral stop	Tongue dorsum: velar, stop Laryngeal: small opening
dʒ	[dʒaˌɾ] ‘drain’	Pulmonic egressive voiceless laminal post-alveolar central oral affricate	Tongue tip: dental, stop, laminal Tongue tip: post-alveolar, fricative, laminal
t͡ʃ	[tʃaˌɾ] ‘chirping’	Pulmonic egressive voiceless laminal post-alveolar central oral affricate	Tongue tip: dental, stop, laminal Tongue tip: post-alveolar, fricative, laminal Laryngeal: large opening
z	[zaˌɾ] ‘jewelry’	Pulmonic egressive voiced apical dental central oral fricative	Tongue tip: dental, fricative, apical
s	[saɾ] ‘point’	Pulmonic egressive voiceless apical dental central oral fricative	Tongue tip: dental, fricative, apical Laryngeal: large opening
ɣ	[ɣaˌɾ] ‘mountain’	Pulmonic egressive voiced velar central oral fricative	Tongue dorsum: velar, fricative
x	[xaˌɾ] ‘donkey’ [xoˌɾ] ‘sister’	Pulmonic egressive voiceless velar central oral fricative	Tongue dorsum: velar, fricative Laryngeal: large opening
ʃ	[ʃoˌɾ] ‘noise’	Pulmonic egressive voiceless laminal post-alveolar central oral fricative	Tongue tip: post-alveolar, fricative, laminal Laryngeal: large opening
χ	[χoɾ] ‘spread’	Pulmonic egressive voiced uvular central oral fricative	Tongue dorsum: uvular, fricative Laryngeal: large opening

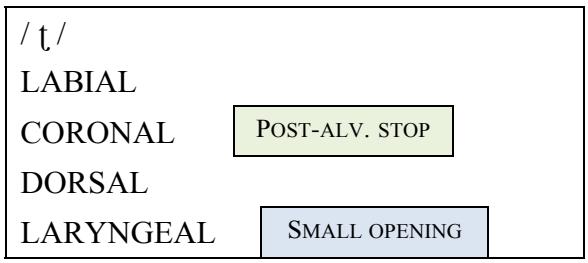
Timing of laryngeal & supralaryngeal gestures: To account for the slight aspiration in stops and closure that begins before voicing ceases, I consider the two gestures slightly to be offset.

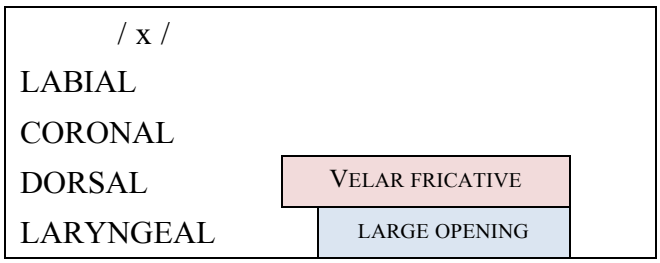


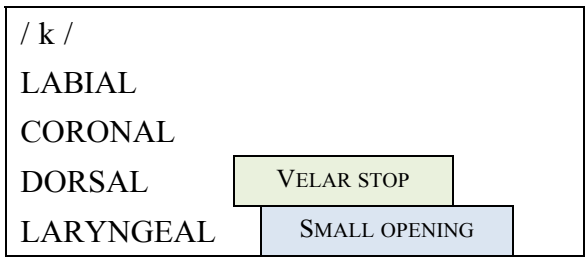


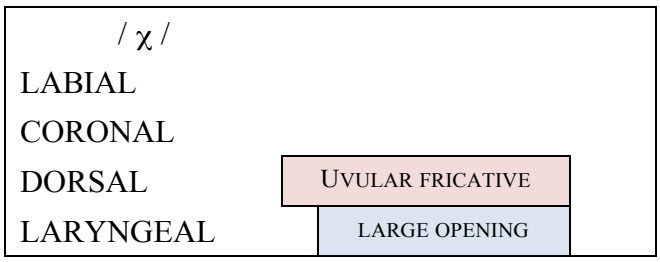




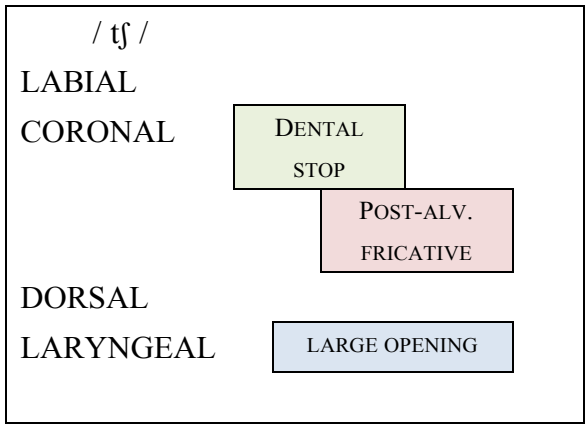


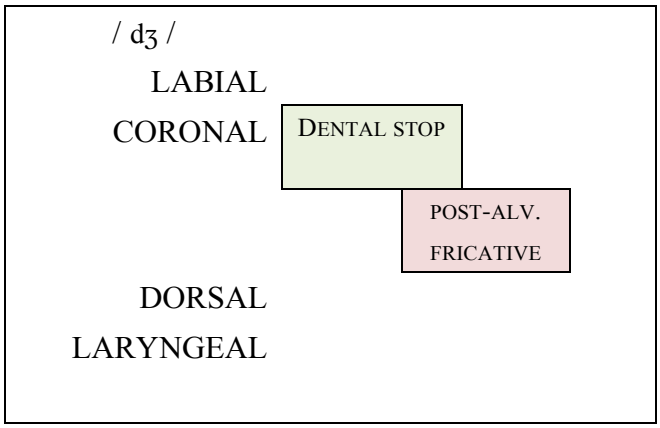






Timing of multiple supralaryngeal gestures





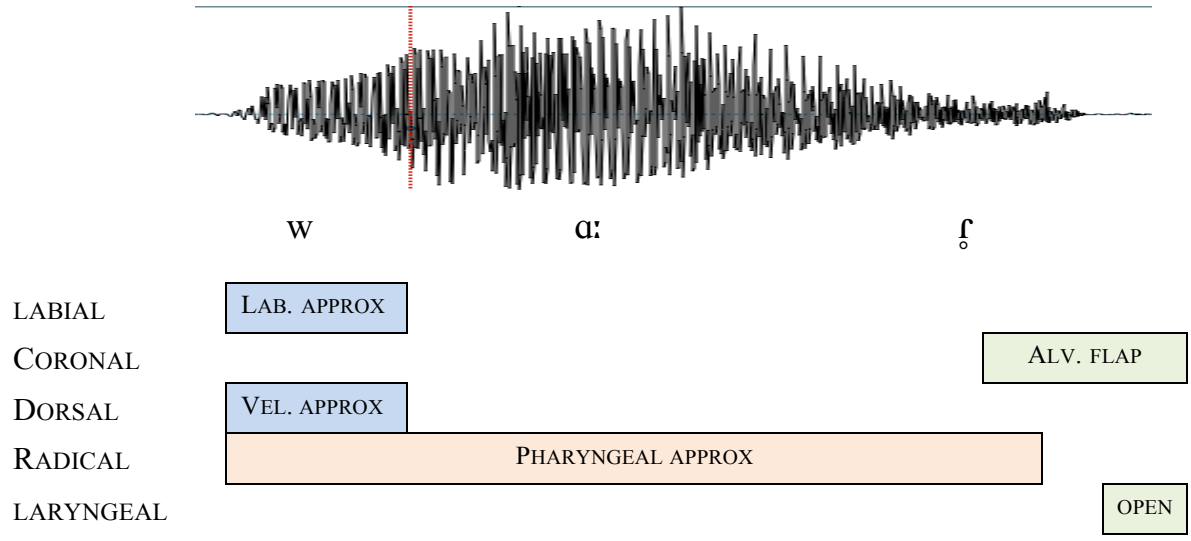
2.2 Approximants

Semivowels

Pashto contrasts two semivowels: palatal [j] and labial-velar [w], as shown below.

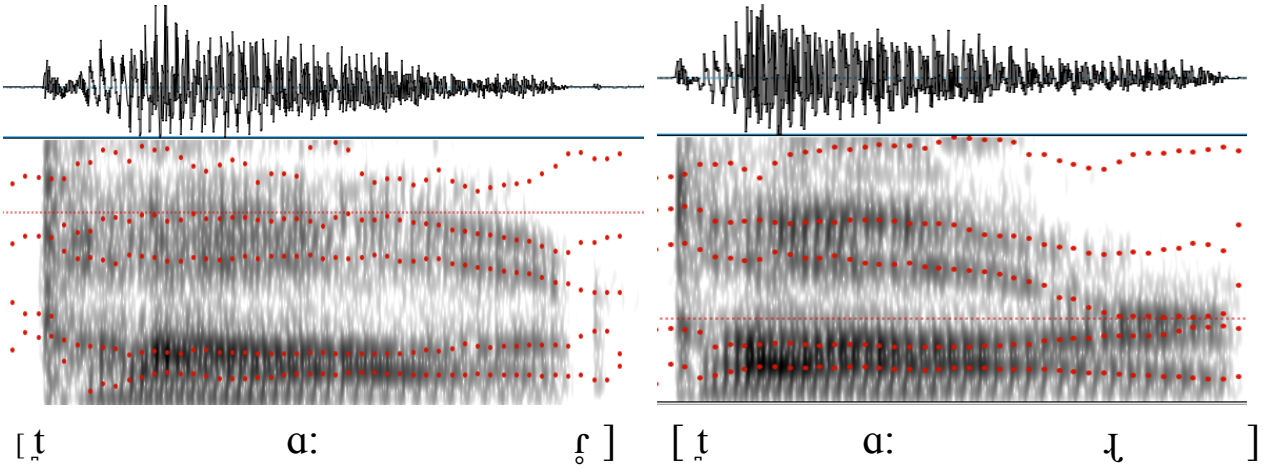
	Example	Traditional description	Gestural description
w	[wɑːɾ̥] ‘time’	Pulmonic, egressive, voiced, N/A, labial/velar, central, oral, approx.	Tongue dorsum: velar, approximant Lips: labial, approximant
j	[jɑːɾ̥] ‘friend’	Pulmonic, egressive, voiced, N/A, palatal, central, oral, approximant	Tongue dorsum: palatal, approximant

An approximate gestural score for [wɑːɾ̥] ‘friend’ is given below:

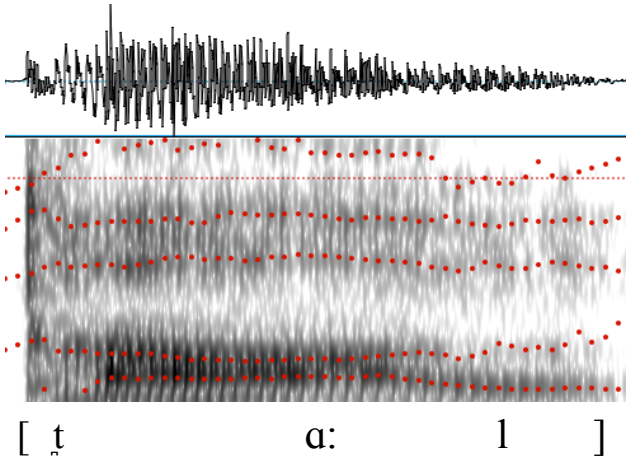


Liquids

For speaker S1, there are three phonemic liquids which are broadly transcribed as /l, r, ɻ/. The latter two may alternate with partially-devoiced taps. A spectrogram comparing these two sounds is given below. The difference between them is evident in the frequency of F3 and the voicing. While /ɻ/ is voiced throughout closure, /r/ is largely devoiced toward the end. The retroflex nature of /ɻ/ is again clearly apparent from the abrupt lowering of F3.



The lateral approximant /l/ is voiced throughout. F2 seems to be slightly lower than the rhotics, and F3 slightly higher, possibly due to retraction of the tongue dorsum.



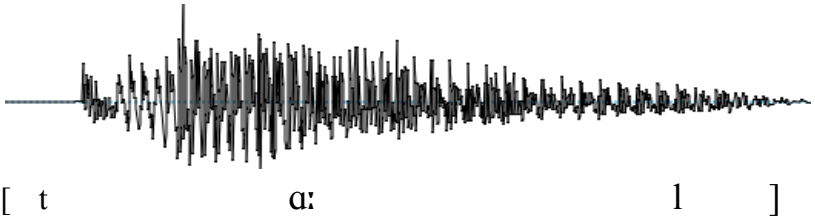
Average formant values[2] for the liquids are given below along with the labio-velar approximant [w]. With respect to the contrast between /r, ɻ/ we can see again that F3 is much lower in the retroflex approximant. /l/ and /w/ have a lower F2 and higher F3 than /r/, presumably due to the dorsal retraction. Additionally, /w/ has a lower F2 than /l/, possible because of the lip-rounding.

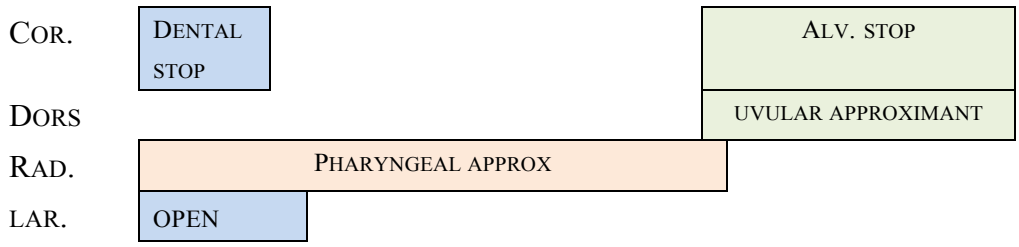
	/r/	/ɻ/	/l/	/w/
F1	530	478	399	324
F2	1472	1325	1156	911
F3	2393	1706	2666	2518

[2] We used the average of four tokens with /r/, four with /ɻ/, six with /w/ and one value for /l/ (based on available data from S1).

A gestural description of the liquids is provided below. Again it is unclear whether [l] is alveolar or dental, so to simplify the phoneme inventory we assume it is laminal dental.

	Example	Traditional	Gestural
l	[t̪ɑːl] ‘beating time in music’	Pulmonic egressive voiced dental lateral oral approximant	Tongue tip: dental laminal stop Tongue dorsum: uvular approximant
r	[t̪ɑːr̥] ‘thread’	Pulmonic egressive voiceless alveolar central oral trill	Tongue tip: alveolar trill/tap
ɻ	[t̪ɑːɻ] ‘gang of robbers’	Pulmonic egressive voiceless dental laminal central oral stop	Tongue tip: post-alveolar apical approximant





Nasals

There are four nasals: /m, n, ɳ, ŋ/. The latter three are found mostly word-finally. The retroflex nasal is especially rare, and never occurs word-initially; it is also found as an allophone of /n/ before another retroflex consonant (e.g. [puɳɖ] ‘fat’). [The recording for [mãˈɳã] comes from an outside resource]

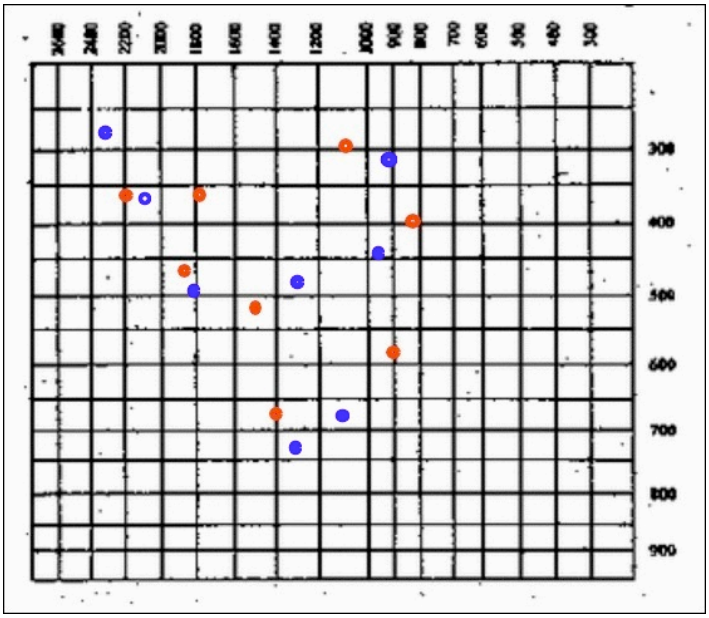
	Example	Traditional	Gestural
m	t̪ɑ̃m ‘complete’	Pulmonic egressive voiced bilabial central nasal stop	Lips: labial, stop Lowered velum
n	t̪ɑ̃ːn ‘hesitation’ ˈmana ‘forbidden’	Pulmonic egressive voiced laminal dental central nasal stop	Tongue tip: dental, laminal, stop Lowered velum
ɳ	mãˈɳã ‘apple’	Pulmonic egressive voiced apical	Tongue tip: post-alveolar, apical, stop

		post-alveolar central nasal stop	Lowered velum
ŋ	t̪ãŋ ‘narrow’	Pulmonic egressive voiced velar central nasal stop	Tongue dorsum: velar, stop Lowered velum

3. VOWELS

We were able to contrast eight vowels: five front and three back. Average formant values for both speakers are given below, and are plotted on the following formant chart (red: S1, blue: S2). For both speakers there were regular durational differences between /a/ and /ɑ/, the latter about the length of a diphthong, but not double the duration of /a/; /e, o/ also seem to be slightly longer than /ɛ, ə/. On the whole, however, we did not find evidence for treating length as phonemic in Pashto, contra claims in early grammars (which are most likely describing length in the orthography system borrowed from Arabic). It seems more likely that durational differences are caused by tense/lax properties of the vowels. /i/ is reduced to /ɪ/ in e.g. closed syllables, and /o, u/ occasionally to [ʊ] in fast speech, but no minimal pairs were found to justify treating them as separate phonemes, also contra claims by other authors. We demonstrate the two vowels by plotting [ɪ] for S1 and [i] for S2.

FRONT VOWELS				BACK VOWELS			
	F1	F2	F2-F1		F1	F2	F2-F1
ɪ	369	1792	1423	u	292	1084	792
i	287	2278	1990		313	913	600
e	365	2203	1837	o	396	812	415
	369	2101	1731		461	959	498
ɛ	459	1833	1373				
	507	1799	1291				
ə	522	1499	977	ɑ:	573	935	362
	490	1342	852		691	1172	480
a	677	1444	766				
	745	1361	615				



The following pairs demonstrate the vowel contrasts. Gestural descriptions are provided; I consider schwa to lack a constriction gesture, and differentiate /a/ and /ɑ/ via constriction degree.

	Example	Continuous vowel space	Gestural description
i	[d̪i] ‘(they) are’	Front, closed, unrounded	Tongue dorsum: palatal, approximant, (tense or lax)
e	[d̪e] ‘village’	Front, close-mid, unrounded	Tongue dorsum: palatal, mid, tense
ɛ	[d̪ɛ] ‘this (oblique)’	Front, open-mid, unrounded	Tongue dorsum: palatal, mid, lax
a	[d̪a] ‘(she) is’	Front, open, unrounded	Tongue dorsum: pharyngeal, mid, tense
ə	[d̪ə] ‘of’	Central, mid, unrounded	(none?)
u	[suˠ] ‘red’	Back, closed, rounded	Tongue dorsum: velar, approximant, (tense or lax) Lips: bilabial approximant
o	[soːʁ̞] ‘raised’ d̪o ‘game’	Back, closed-mid, rounded	Tongue dorsum: uvular, approximant, low jaw/mid, tense Lips: bilabial approximant
ɑ	[sɑːʁ̞] ‘good news’ [d̪ɑ] ‘this’ (nominative)’	Back, open, unrounded	Radical: pharyngeal, approximant, tense

4. CLUSTERS

Pashto is known to allow complex onset clusters that violate the Sonority Sequencing Principle, and may include any of the following:

1. Fricative + Stop combination, e.g: /xpəl/ ‘own’, /zda/ ‘study’, /zbəxi/ ‘sucks’
2. Nasal + Obstruent combinations, e.g.: /nɣararɪ/ ‘burner’, /ngor/ ‘sister-in-law’
3. Affricate + Stop combinations, e.g.: /tskɑk/ ‘drinking’
4. Liquid + Nasal combinations, e.g.: /lmar/ ‘sun’
5. Glide + Liquid combinations, e.g.: /wruna/ ‘doors’, /wluna/ ‘curls’

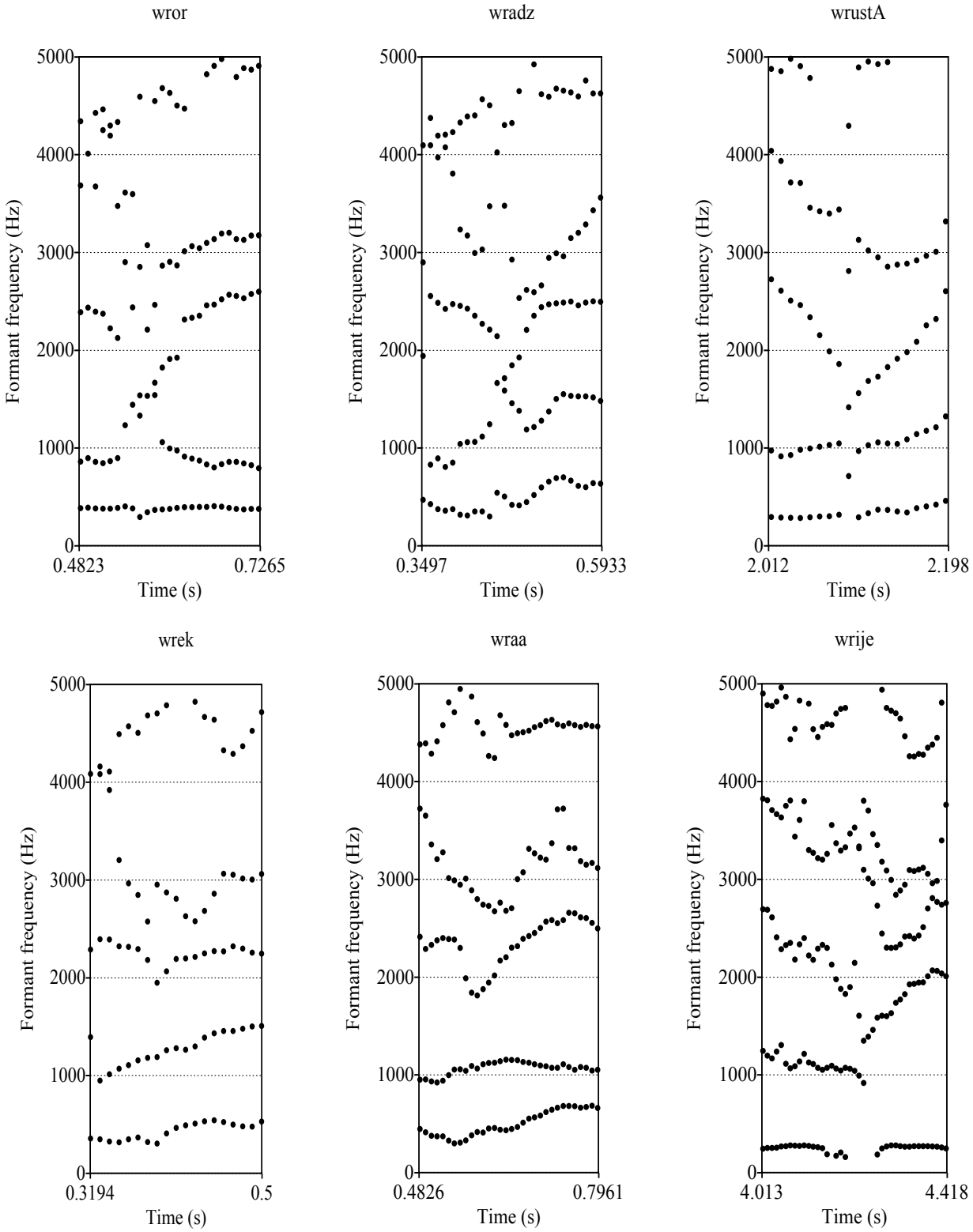
This is especially surprising given that a number of examples can be found of clusters that do *not* violate sonority sequencing principles, but are nevertheless simplified:

/zmɑ:/ → [zəmɑ:] ‘my’ /dwa/ → [dəwa] ‘two’
/mrasta/ → [mərasta] ‘help’ /xrɑ:ba/ → [xərɑ:ba]
/tsə/ → [sə] ‘what’

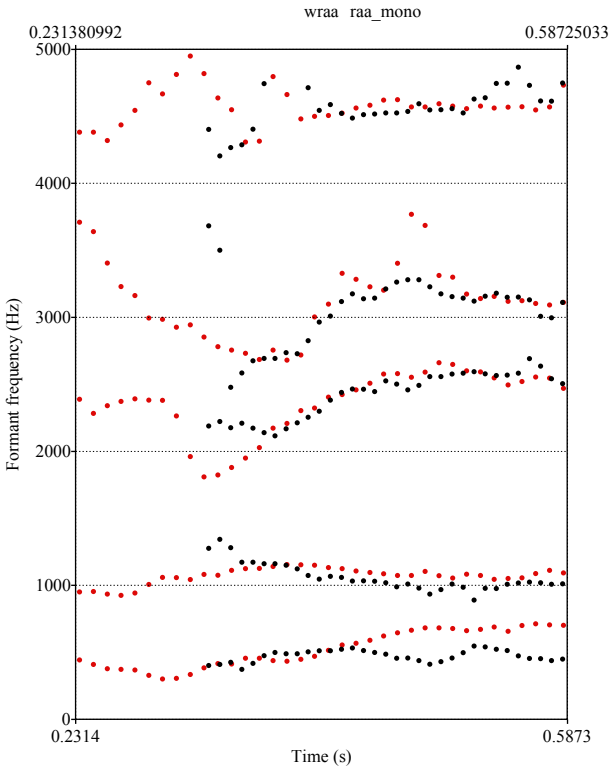
Of the cluster types outlined above, fricative + stop combinations were the most common; affricate + stop also reduces into fricative + stop. Clusters including the glide [w], which have previously been analyzed by Bell & Saka (1982) as true consonant clusters, exhibited a substantial amount of gestural overlap. To demonstrate the latter, the formants and formant transitions for single liquids and glides were compared with the cluster /wr/. We repeat below the average formants for various Pashto approximants:

	/r/	/ɻ/	/l/	/w/
F1	530	478	399	324
F2	1472	1325	1156	911
F3	2393	1706	2666	2518

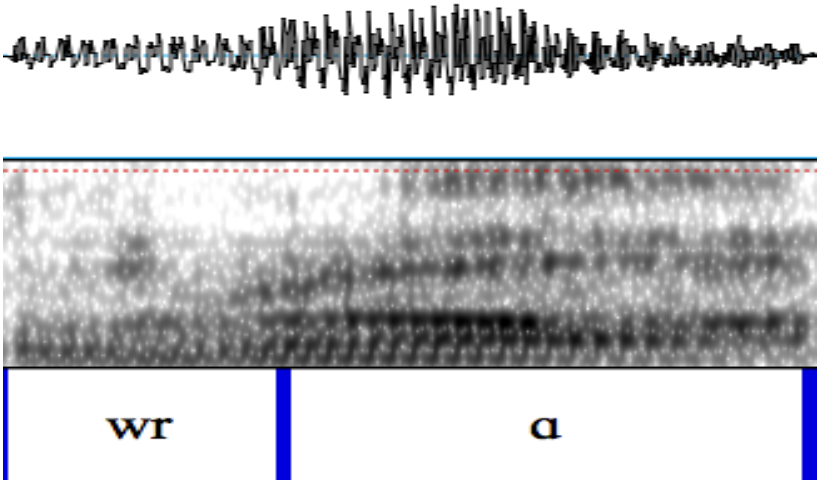
The graphs below show the formant contours for six onset + vowel combinations: [wro], [wra], [wru], [wrə], [wri] and [wrɑ:]. At the onset of the word there are resonances at approximately 400 Hz, 900 Hz, and 2400 Hz; F2 then rises while F3 lowers dramatically.

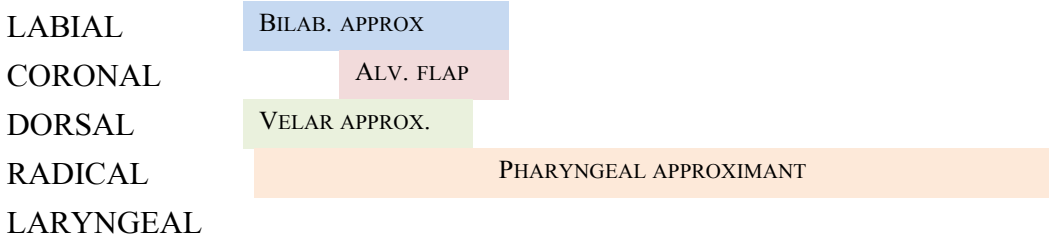


The onset of the word has roughly similar formant values to single [w]; however, when we overlay the contour for /rɑ/ onto /wrɑ/, we find that in the complex cluster, F2 and F3 are lower at the point of release [red: /wrɑ:/, black: /rɑ:/]. We interpret this as evidence for gestural overlap, such that the release of [r] is also accompanied by lip rounding, lowering the F2 & F3 values.



Gestural overlap will also explain why the duration of [wr] is not much longer than a singleton onset. A gestural score is given below for /wrɑ:/:





5. CONCLUSION

This paper provides a basic description of the phonetic system in Pashto, an Indo-Iranian language spoken in Afghanistan. Although there are descriptive resources on Pashto, very few are based on acoustic analysis, and conclusions are often complicated by issues pertaining to dialect and orthography. Contra previous work, it does not look like vowel length is phonemic, and in particular the length implied in the orthography is mostly orthogonal from vowel duration, a fact that may seem obvious but has been a point of confusion for applied and computational work on Pashto. We also briefly examined initial clusters and found evidence for a large amount of overlap between the consonants, such that lip rounding is still present at the release of the liquid.

There were some differences between pronunciation by S1 and S2, despite speaking similar dialects. For example, S2's /ɑ/ is distinguished from /a/ more by duration than quality; while S1 shows the opposite pattern. One possible source of these differences lies in the fact that S1 is a bilingual Dari/Pashto speaker, who has most likely had more exposure to Dari than Pashto (because of his upbringing in Kabul).