



Pashto intonation patterns

Luca Rognoni, Judith Bishop, Miriam Corris

Appen Limited, Australia

lrognoni@appen.com, jrbishop@appen.com, mcorris@appen.com

Abstract

A hand-labelled Pashto speech data set containing spontaneous conversations is analysed in order to propose an intonational inventory of Pashto. Basic intonation patterns observed in the language are summarised. The relationship between pitch accent and part of speech (PoS), which was also annotated for each word in the data set, is briefly addressed.

The results are compared with the intonational literature on Persian, a better-described and closely-related language. The results show that Pashto intonation patterns are similar to Persian, as well as reflecting common intonation patterns such as falling tone for statements and WH-questions, and yes/no questions ending in a rising tone. The data also show that the most frequently used intonation pattern in Pashto is the so-called hat pattern. The distribution of pitch accent is quite free both in Persian and Pashto, but there is a stronger association of pitch accent with content than with function words, as is typical of stress-accent languages.

The phonetic realisation of focus appears to be conveyed with the same acoustic cues as in Persian, with a higher pitch excursion and longer duration of the stressed syllable of the word in focus. The data also suggest that post-focus compression (PFC) is present in Pashto.

Index Terms: Pashto, intonation patterns, Persian, prosody, pitch contour, RaP, ToBI, PFC, hat pattern, part of speech

1. Introduction

To the best of the authors' knowledge, there are no significant acoustic-phonetic descriptions of Pashto intonation in the research literature published to date (either in English or French). Pashto is an Indo-Iranian language closely related to Persian. Given this relationship, we hypothesised that the two languages might have a similar inventory of intonation patterns. The intonation of Persian has been recently described in a series of studies based on the annotation of scripted speech, using ToBI-like systems of annotation and description ([6], [7], [8]).

Some information regarding Pashto prosody can be found in general descriptive grammars of Pashto (cf. [2] and [10]), although the suprasegmental aspects are treated marginally as compared to syntax and morphology. In particular, [10] reports that word stress is not predictable in Pashto: generally, it is located on the last or on the penultimate syllable of a word, but (as in Persian), the presence of affixes and other particles often reconfigures stress placement. As for pitch patterns, [2] and [10] confirm that the distinction between statements and Yes/No Questions (YNQ) in speech is conveyed through intonation. As for WH-questions, [2]

reports that in default conditions, the question word has the major prominence, while the intonation pattern is falling, similar to statements. Preliminary observations of the data and discussion with Pashto native linguists indicates that any word of a sentence can be emphasised in Pashto, depending on the communicative context.

The data presented in this study are novel in that they contain spontaneous conversational speech (collected via telephony) rather than read or scripted speech, as used in the studies of Persian intonation ([6], [7], [8]). Spontaneous speech can be expected to be better representative of intonation patterns as used in ordinary speech. The patterns can also be expected to differ from those found in more careful speech; conversational speech may have a faster tempo, it often contains grammatically incomplete phrases, and the dialogic context influences the use of focus and other modes of emphasis.

Since this is one of the first attempts to describe Pashto intonation, the authors propose a basic inventory of intonation patterns, based on the model of Persian adopted in [7], and focusing on three main types of sentence: statements, Yes/No questions (YNQ) and WH-questions (WHQ).

In line with what was done for Persian by [7], this study also provides some preliminary information regarding the prosodic marking of focus in Pashto.

Finally, we examine the relationship between pitch accents and parts of speech (PoS) in the data.

2. Methodology

2.1. Data set

The data set used in this study consists of a collection of spontaneous telephone conversations, sampled at 8000 Hz. The recordings include 28 speakers (14 male and 14 female) speaking for a total of ~100 minutes (excluding silences) and 3111 intonation phrases. The speakers were native speakers of the main four dialects of Pashto [2]: Northwest (6 speakers), Northeast (5 speakers), Southwest (5 speakers) and Southeast (12 speakers).

2.2. Annotation system and processes

Rhythm and pitch were annotated using a simplified version of the Rhythm and Pitch (RaP) labelling system, which was chosen for its capacity to capture both rhythmic and intonational aspects of speech in parallel [4]. In the RaP system, pitch annotation departs from ToBI conventions in a number of ways. The main differences include a specific notation for phrase-initial tones and the introduction of E tones (where E stands for "equal"). Tones at the onset of a new phrase are marked with a colon preceding the tone (e.g. :L+),

:H*). As for E tones, E* “marks a metrically prominent syllable which has a pitch that is about the same as the immediately preceding syllable or syllables” [4, p.7]. E+ and +E tones are used to mark leading and trailing pitch movements (or boundary tones) at the same pitch height as a preceding tone marked high (H) or low (L). In this study, E markers were generally assigned to tones whose difference in F0 compared to the preceding assigned tone was smaller than 25 Hz.

We made a few notable departures from the original RaP labelling system. First, no use was made of uncertainty labels. Annotators were trained how to make decisions in cases of uncertainty. Second, leading tones (L+ and H+) were reserved for tone targets that are closely associated with a following starred tone (within 1-2 syllables of the starred tone).

Our simplified RaP format was implemented in Praat [1], as suggested by the original labelling guide [4]. Praat was also used to write and run scripts that were used to check the consistency of the annotations with the labelling guidelines and to get acoustic measurements using the standard acoustic algorithms offered by the program [1]. We did not normalise for differences between male and female speakers, as we were not comparing acoustic values among speakers.

The data set was annotated in two passes. First, two trained Pashto native speakers with a solid background in acoustic phonetics annotated rhythm. Then, two linguists trained in pitch annotation annotated the pitch contour of the utterances. IAC (Inter-Annotator Consistency) levels were calculated at regular intervals using Cohen’s kappa and benchmarked against the original RaP levels of consistency for both rhythm and pitch annotation [3]. The IAC levels resulted in $k=.88$ for rhythm presence and $k=.87$ for rhythm strength, $k=.75$ for pitch accent presence and $k=.72$ for pitch accent type. Based on the ranges specified in the original study, these results indicate good and substantial inter-annotator consistency for rhythm and pitch annotation respectively.

3. Analysis

3.1. Pitch accents and boundary tones

The data set used in this study contained a total of 9504 tone markers, not including boundary tones. The distribution of the single tones is summarised in Table 1.

Table 1: *Distribution of tone markers in the data set*

Marker	Count	% of total
H*	2427	25.54%
E*	1630	17.15%
:L+	997	10.49%
:H*	965	10.15%
L*	652	6.86%
:L*	630	6.63%
+L	422	4.44%
L+	403	4.24%
:L	282	2.97%
+H	276	2.90%
!H*	264	2.78%
H+	197	2.07%
:H+	176	1.85%
+E	80	0.84%

:H	61	0.64%
E+	42	0.44%
TOTAL	9504	100%

The main types of Pashto pitch accent (PA) found in the data are summarised in Table 2.

Table 2: *Pitch accent types and definitions*

PA	Description
H*	Marks a rise in the pitch trace corresponding to a major or minor prominence.
L*	Marks an audible step down in pitch corresponding to a major or minor prominence; appears as a dip in the pitch trace in the lower part of the speaker’s pitch range.
E*	Marks a tone that is perceptually equivalent to the pitch of the preceding tone label; appears at about the same pitch height as the previous tone.
!H*	Marks an audible step down in pitch, where the accent shape and pitch height (typically mid in the speaker’s pitch range) remain congruent with a high pitch accent (i.e., it is not a dip).
H* preceded by L+	Marks a bitonal sequence that is often used for contrastive focus, that is, when a word is emphasised or contrasted with other information. L+ is marked on the syllable preceding the H*.

The data set included the following boundary tones (BT), corresponding to the end of each intonation phrase (intermediate phrase tones were not marked). The distribution of the boundary tones is summarised in Table 3 and their functions are described in Table 4.

Table 3: *Distribution of boundary tones in the data set*

Marker	Count	% of total
+L (BT)	2868	92.21%
+E (BT)	161	5.18%
+H (BT)	82	2.64%
TOTAL	3111	100%

Table 4: *Boundary tones types and functions*

BT	Description
+L	Marks a falling intonation at the end of a phrase.
+H	Marks a rising intonation at the end of a phrase.
+E	Marks a suspended intonation at the end of a phrase, often signalling the continuation of a speaker’s turn or an interruption.

3.2. Intonation patterns

Based on an analysis of the distribution of tone sequences and visual inspection of the corresponding utterances, it was possible to identify the typical intonation patterns for statements, WH-questions (WHQ), Yes/No questions (YNQ) and contrastive focus.

3.2.1. Statements and WH-questions

As reported in [10], Pashto statements end with a falling intonation, which is captured in RaP with the final +L boundary tone. The statements analysed in this study are in line with this general claim.

Similar to statements, WHQ end in a falling tone (see Figs. 1 and 2). As in Persian, the major prominence is on the question word [7].

The so-called hat pattern is a frequent configuration of the pitch contour in both statements and WHQ in Pashto. The hat pattern consists of a rise and fall in pitch, concatenated by a high plateau [9]. Hat patterns are represented in RaP by the tone sequence :L+/:L* H* E* (E*) +L. This intonation pattern occurs in 12.92% of all intonation phrases with more than two tones. The hypothesis that the low-high tone sequence (L+)H* could be the basic intonation building block, as theorised for Persian in [7], does not seem confirmed for Pashto.

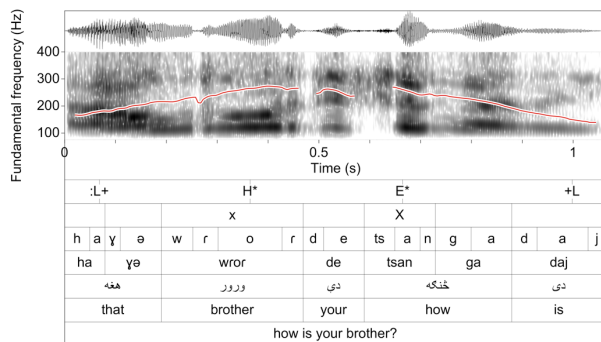


Figure 1: Hat Pattern in a WHQ question (1)

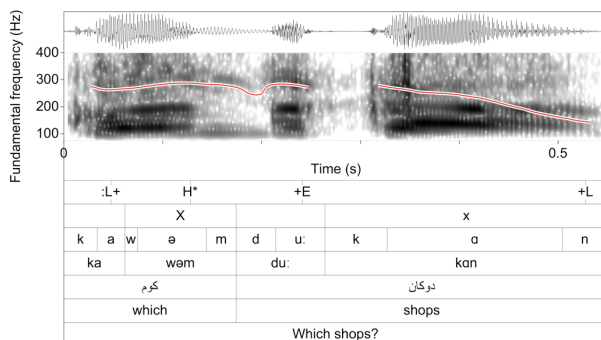


Figure 2: Hat Pattern in a WHQ question (2)

3.2.2. Yes/no questions

YNQ questions end in a rising or high level intonation (+H or +E following an H* accent, see Figs. 3 and 4 respectively).

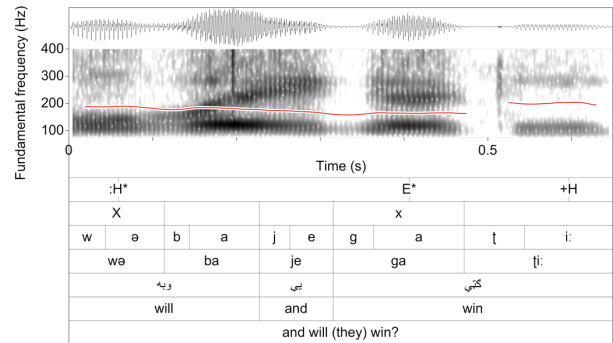


Figure 3: YNQ ending with a rising tone

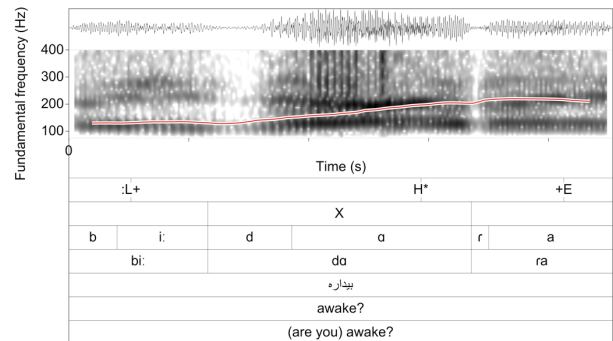


Figure 4: YNQ ending with a high level intonation

The same pattern (in this case, a rising intonation) applies to question tags attached to statements (Fig. 5).

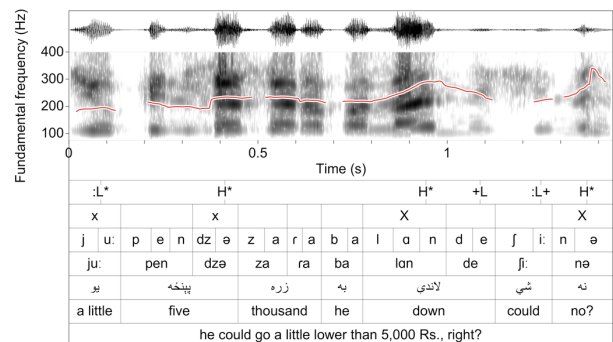


Figure 5: Statement followed by a question tag

3.2.3. Contrastive focus

In Pashto, contrastive focus is typically realised with a higher pitch excursion, longer duration and an L+H* sequence, similar to Persian [7] (and English, among other languages). When a word is contrasted with another word, for example in a correction, it normally carries an L+H* pitch accent, followed by a sharp fall in pitch, represented by a +L tone label (see Fig. 6).

The data also show that words after the focused word are produced with a generally lower and compressed pitch range, as also shown by the example in Fig. 6. This can be seen as evidence in support of the claim that Pashto relies on post-focus compression (PFC, [11]) as a strategy to highlight focused information (in RaP pitch annotation, PFC is captured by the presence of sparse sequences of L and E tones following the focused constituent).

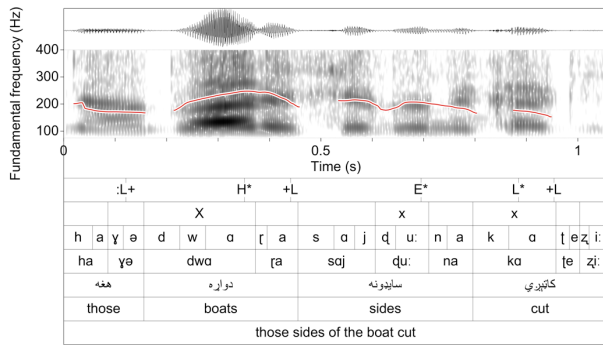


Figure 6: Statement with a contrastive focus

3.3. Part of speech and pitch accent

As mentioned earlier, each word in the database was annotated with a part of speech (PoS) tag. The set of PoS tags, their expansions and their frequency in the data set are given in Table 5.

Table 5: Word counts by PoS types

PoS	Count	Tag expansion	PoS	Count	Tag expansion
ADJ	1001	Adjective	DET	400	Determiner
ADV	1596	Adverb	NUM	266	Numeral
NOUN	2773	Noun	PART	330	Particle
IJ	1233	Interjection	PRON	2267	Pronoun
ADP	1815	Adposition	NON	5	Nonsense word
CONJ	1182	Conjunction	FRAG	21	Fragment
VERB	3560	Verb	FOR	151	Foreign word
COV	423	Coverb	PUNC	1705	Punctuation marker
LVERB	428	Light Verb	REP	10	Repeated word

The following categories were excluded from analysis: nonsense words, fragments, foreign words, punctuation markers, repeated words. We also excluded one-word utterances from analysis, as these must bear an accent by convention; they are treated as full intonation phrases, which therefore require an accent. We used the chi-square measure in order to test the association between frequency of accent status (accented/unaccented) and part of speech. Our hypothesis was that the more information-rich content word categories, with the exception of verbs, would show a positive association with accent (a relatively stronger propensity to be accented). The verb category was not expected to be accented at a higher rate, because Pashto is an SOV language [10] in which the object preceding the verb is a likelier target for accent [5], leading (we hypothesised) to a reduced frequency of accent on the verb.

The data allow us to reject the null hypothesis, showing a lack of any association with accent status for most function word categories (with the exception of adpositions, conjunctions and interjections), and a higher than expected rate of accentuation only for nouns, adjectives and adverbs among content word categories. Verbs are indeed less often accented than would be expected based on the overall proportion of accented to unaccented words in the data set (see Table 6), suggesting there is a positive tendency for this category to reject, rather than attract, pitch accent. There is no significant association between rate of accentuation and PoS

for coverbs, determiners, numerals, particles and pronouns, which are therefore excluded from Table 6.

Table 6: PoS types and frequency of accentuation

PoS	Level of significance		Frequency of accentuation (observed relative to expected)
	p=0.05	p=0.01	
ADJ	*		higher
ADV	*	*	higher
NOUN	*	*	higher
IJ	*	*	higher
ADP	*	*	lower
CONJ	*	*	lower
VERB	*	*	lower
LVERB	*		lower

4. Discussion and Conclusions

Most sentences in our data are declarative or WH-questions, ending in a low tone (+L). The data show that the so-called "hat-pattern" intonation contour is the typical intonation pattern both for declarative sentences and WH-questions with more than two tones (not including the mandatory final boundary tone). Yes/no questions exhibit a final rising or high level contour in most cases, while contrastive focus is typically realised with a rising accent (L+ followed by H*).

Our preliminary examination of the association between frequency of accentuation and PoS allows us to reject the null hypothesis for the PoS categories that may be considered most information-rich, namely nouns, adjectives and adverbs. It is interesting that the category of coverb groups together with most function word categories other than adpositions and conjunctions in showing no significantly stronger (or weaker) affiliation for accent than expected, while light verbs show a lower rate of accent than expected.

One notable observation is that overall, we observed a somewhat higher proportion of unaccented to accented words than anticipated. The ratio of unaccented to accented words in the data set analysed (excluding one-word utterances) was approximately 2:1. It may be that the RaP annotation system, with its use of E*, results in somewhat lower accent counts than would have been the case with ToBI. RaP disallows more than two E* accents in a row in monotonic stretches of speech, where ToBI would allow multiple consecutive H* or L* labels.

The data set presented in this paper represents a first acoustic-phonetic analysis of Pashto intonation patterns, using spontaneous speech data. It provides confirmation of the presence of typical intonation patterns observed in better-known languages, and a basis for future research into the phonetic realisation of accent and the association between accent and part of speech in Pashto.

5. References

- [1] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.0.27, retrieved 17 March 2017 from <http://www.praat.org/>, 2017.
- [2] A. Boyle David, *Descriptive Grammar of Pashto and its Dialects*. Göttingen: De Gruyter Mouton, 2014.
- [3] L. Dilley, M. Breen, M. Bolivar, J. Kraemer, E. Gibson, E. "A comparison of inter-transcriber reliability for two systems of prosodic annotation: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices)", in *INTERSPEECH 2006 – 9th Annual Conference of the International Speech Communication*

Association, September 17-21, Pittsburgh, PA, Proceedings, pp. 317-320, 2006.

- [4] L. Dilley and M. Brown, *The RaP (Rhythm and Pitch) labeling system*, retrieved 21 March 2017 from https://www.academia.edu/2730467/The_RaP_Rhythm_and_Pitch_Labeling_System, 2005.
- [5] E. Koktova, *Word-order based grammar*. Berlin: De Gruyter Mouton, 1999.
- [6] B. Mahjani, "An instrumental study of prosodic features and intonation in modern Farsi (Persian)", MSc thesis, University of Edinburgh, 2003.
- [7] N. Sadar-Tehrani, "The intonational grammar of Persian", PhD thesis, University of Manitoba, 2007.
- [8] R. Scarborough, "The intonation of focus in Farsi", *UCLA Working Papers in Phonetics*, 105, 19-34, 2007.
- [9] J. t'Hart, R. Collier, A. Cohen, *A perceptual study of intonation*. Cambridge: Cambridge University Press, 1990.
- [10] H. Tegey and B. Robson, *A reference grammar of Pashto*. Washington, DC: Center for Applied Linguistics, 1996.
- [11] Y. Xu, "Post-focus compression: cross-linguistic distribution and historical origin", *Proc. of 17th International Conference of Phonetic Sciences (ICPhS), August 17-21, Hong Kong, PRC*, 152-155, 2011.