

# Urdu Pitch Accents and Intonation Patterns in Spontaneous Conversational Speech

Luca Rognoni, Judith Bishop, Miriam Corris, Jessica Fernando, Rosanna Smith

Appen Limited, Australia

Level 6, 9 Help Street Chatswood NSW 2067, Australia

{lrognoni, jbishop, mcorris, jfernando, rsmith}@appen.com

## Abstract

An intonational inventory of Urdu for spontaneous conversational speech is determined based on the analysis of a hand-labelled data set of telephone conversations. An inventory of Urdu pitch accents and the basic Urdu intonation patterns observed in the data are summarised and presented using a simplified version of the Rhythm and Pitch (RaP) labelling system. The relation between pitch accents and parts of speech (PoS) is also explored.

The data confirm the important role played by low pitch accents in Urdu spontaneous speech, in line with previous studies on Urdu/Hindi scripted speech. Typical pitch contours such as falling tone in statements and WH-questions, and rising tone for yes/no questions are also exhibited. Pitch accent distribution is quite free in Urdu, but the data indicate a stronger association of pitch accent with some PoS categories of content word (e.g. Nouns) when compared with function words and semantically lighter PoS categories (such as Light Verbs).

Contrastive focus is realised by an L\*+H accent with a relatively large pitch excursion for the +H tone, and longer duration of the stressed syllable. The data suggest that post-focus compression (PFC) is used in Urdu as a focus-marking strategy.

**Keywords:** Urdu, intonation, prosody annotation, pitch contour, RaP, ToBI, PFC, part of speech

## 1. Introduction

Recent studies of Urdu intonation describe scripted and laboratory speech, using ToBI-like annotations to characterise Urdu pitch accents and intonation patterns (Nayyer and Madni, 2002; Jabeen, 2010). The novelty of the data presented here is that they contain spontaneous conversational speech rather than read or scripted material. Speech recorded in this manner (unscripted telephone conversations between friends or relatives) is likely to represent better the intonation patterns of ordinary speech.

It has been argued that, based on the structural similarity between Urdu and Hindi, intonation patterns and prosodic marking strategies observed in Hindi can also be expected in Urdu (Masica, 1991; Dyrud, 2001; Nyar, 2002; Hock and Bashir, 2016). Sengar and Mannel (2012) have proposed L\*+H as the “default pitch accent” for Hindi, in agreement with Patil et al. (2008), who suggest the p-phrase LH as the main prosodic building block of Hindi. The data presented by Nayyer and Madni (2002) and Jabeen (2010) suggest that this could also apply to Urdu. The latter two studies specifically analyse Urdu intonation patterns in scripted speech. They confirm that statements and WH-questions (WHQ) are realised with a falling intonation, while yes/no Questions (YNQ) are characterised by a rising contour. Studies of the acoustic-phonetic realisation of word stress in Urdu show that this is often aligned with a low tone (L\*) (Hussain (1997); Dyrud, 2001). In a study of Hindi intonation, Patil et al. (2008) argue that downstepped intonation contours appear to be compulsory, independent of the focus condition, and that post-focus compression (PFC) is present. Jabeen and Braun (2018) show that duration appears to be a strong acoustic cue for perceiving focus in Urdu.

We propose a basic inventory of the intonation patterns found in our spontaneous speech data set, presenting three main types of sentence: statements, WH-questions and yes/no questions (YNQ). Some of the claims proposed for

Urdu, based on the prosodic marking of focus in Hindi (Patil et al., 2008; Sengar and Mannel, 2012), are assessed using our spontaneous data.

Finally, this study will investigate the connection between frequency of pitch accent and part of speech (PoS).

## 2. Methodology

The methodology adopted in this study matches that applied by the authors to a similar data set in Pashto in Rognoni et al. (2017).

### 2.1. Data Set

This study uses a data set consisting of a collection of audio recordings of spontaneous telephone conversations between friends or relatives, sampled at 8000 Hz. The set includes a total of ~50 minutes of speech (excluding silences) and 2188 intonation phrases. The total number of speakers is 28 (14 female and 14 male). The speakers are native speakers of several Urdu dialects.

### 2.2. Annotation System and Processes

The Rhythm and Pitch (RaP) labelling system was adopted for its capacity to account for both the rhythmic and intonational levels of speech (Dilley and Brown, 2005). The main differences between this annotation model and ToBI have been described in depth in Rognoni et al. (2017) and Dilley and Brown (2005). Two of the main formal differences between the two systems are important in order to understand the results presented in this study. First, the RaP system has a specific notation for phrase-initial tones, which are marked with a colon preceding the appropriate marker (e.g. :H+, :L\*). Second, RaP includes E (“equal”) tones: E\* is used to mark a pitch accent whose pitch is perceptually equivalent to the one of the immediately preceding tone (Dilley and Brown, 2005). Leading and trailing pitch segments that appear at the same (or at a very similar) pitch height of a preceding tone are marked with E+ and +E respectively. In this study, we limited the use of

E markers to the tones with a difference in fundamental frequency smaller than 25 Hz as compared to the preceding assigned tone marker.

This study adopted the simplified version of the RaP labelling system described in detail in Rognoni et al. (2017). This annotation format was implemented in Praat (Boersma and Weenink, 2019), as recommended in the RaP labelling guide (Dilley and Brown, 2005). Scripts were also written and run within the program in order to check the consistency of the annotations and to obtain acoustic measurements. Differences between male and female speakers were not normalised because acoustic values were not compared among speakers.

Two passes of annotations were performed. First, rhythm was annotated by two trained Urdu native speakers with a strong background in acoustic phonetics. Then, two different linguists trained in pitch annotation labelled the pitch contour of the same utterances. IAC (Inter-Annotator Consistency) checks were performed at regular intervals using Cohen’s kappa. The results were compared to levels of consistency of the original RaP, both for rhythm and for pitch annotation (Dilley et al., 2006). The final IAC results were  $k=.88$  for rhythm presence and  $k=.87$  for rhythm strength;  $k=.72$  for pitch accent presence and  $k=.62$  for pitch accent type. These results indicate strong inter-annotator consistency for rhythm and substantial inter-annotator consistency for pitch annotation.

### 3. Analysis

#### 3.1. Pitch Accents and Boundary Tones

The rhythm and pitch analysis adopted in this study allows us to create a data-based inventory of the nuclear pitch accents (NPA) and boundary tones (BT) in Urdu spontaneous speech.

The distribution of the main tone types is shown in Table 1, while Table 2 presents their description. The tone sequences that account for less than 1% of the total number of nuclear pitch accents were discarded as not significantly representative. The remaining set of pitch accents covers the 93.33% of the nuclear accents in the data set.

NPA	Count	% of total
E*	471	21.53%
:L*	425	19.42%
:H*	190	8.68%
L*	186	8.50%
H*	173	7.91%
:L+E*	127	5.80%
:L*+H	85	3.88%
:L+H*	81	3.70%
:H+E*	77	3.52%
E*+H	51	2.33%
!H*	47	2.15%
L*+H	45	2.06%
:H+L*	37	1.69%

H+L*	24	1.10%
H+E*	23	1.05%
<b>TOTAL</b>	<b>2042</b>	<b>93.33%</b>

Table 1: Distribution of nuclear pitch accents in the data set

Tone	Description
:L*	Initial tone realised in the lower region of the speaker's pitch range.
:H*	Initial tone realised in the higher region of the speaker's pitch range.
L*	A perceptible step down in pitch; appears as a dip in pitch in the lower region of the speaker's pitch range.
H*	A rise in the pitch trace.
E*	A tone that has a very similar pitch height to the preceding tone label.
!H*	Pitch downstep following H*. The pitch height is typically mid in the speaker's pitch range, and its shape is similar to H*.
+L	Low trailing tone following a L* or E* tone
+H	High trailing tone following a L* or E* tone
+E	Equal trailing tone following a L* or E* tone
L+	Low leading tone preceding a L* or E* tone
H+	High leading tone preceding a L* or E* tone
E+	Equal leading tone preceding a L* or E* tone

Table 2: Nuclear pitch accent types and definitions

If one excludes single tones and tone sequences including E\*, whose peculiarities will be discussed in Section 4, the single-tone pitch accents are distributed as follows: :L\* and L\* correspond to 27.92% of NPAs, while :H\*/H\*/!H\* cover 18.74%. As for two-tone sequences, :L\*+H and L\*+H sequences represent 5.94% of the total number of NPAs. The other main two-tone sequence found in the data, captured by the :H\*+L and H\*+L notations, covers 2.79% of the NPAs.

The end of each intonation phrase in the data set was also marked with the appropriate boundary tone (BT), while intermediate phrase boundary tones were not labelled. Table 3 presents the distribution of the boundary tones. The higher number of +L BTs reflects the majority of intonation phrases characterised by a falling intonation in the data set, while the +H BT accounts for rising final intonation, found in polar questions (see 3.2.2) and for phrases realised with suspended intonation (as in lists or in incomplete or interrupted utterances).

BT	Description	Count	% of total
+L	Falling intonation	1227	56.08%
+H	Rising intonation	961	43.92%

Table 3: Boundary tones types at phrase end: description and distribution in the data set

### 3.2. Intonation Patterns

Intonation patterns were identified based on analysis of the distribution of pitch accent sequences and boundary tones, paired with visual inspection of the corresponding utterances. We focused on declarative sentences, polar and non-polar questions and on the realisation of contrastive focus.

#### 3.2.1. Statements and WH-questions

Nayyer and Madni (2002) observe that Urdu statements end with a falling intonation. As mentioned in 3.1, this pattern is captured in RaP with a final +L boundary tone. The pitch contours found in the declarative sentences in this study agree with this claim (see Fig. 1).

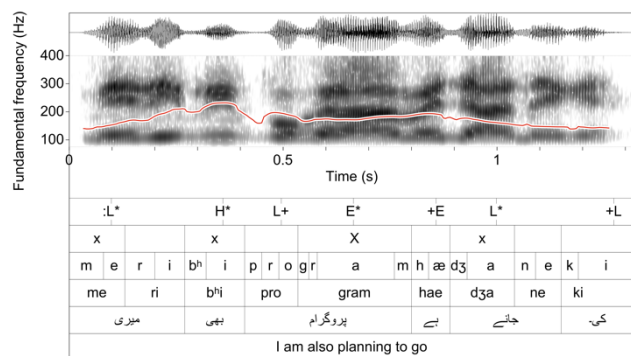


Figure 1: Falling tone in a statement

Our data show that the so-called hat pattern is a recurring configuration of the pitch contour in Urdu statements (see Fig. 2). The hat pattern is represented by a rise and fall in pitch, connected with a high plateau (t'Hart et al., 1990). This intonation pattern is captured by the tone sequence :L+/:L\* H\* E\* (E\*) +L (Rognoni et al., 2017).

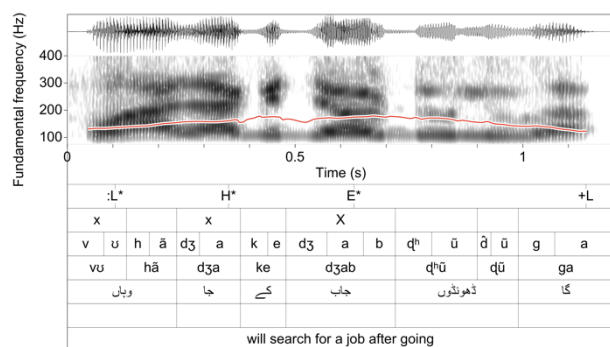


Figure 2: Hat Pattern in a statement

Similar to statements, a falling tone is used to end a WH-question (see Fig. 3). The major prominence is usually on the question word.

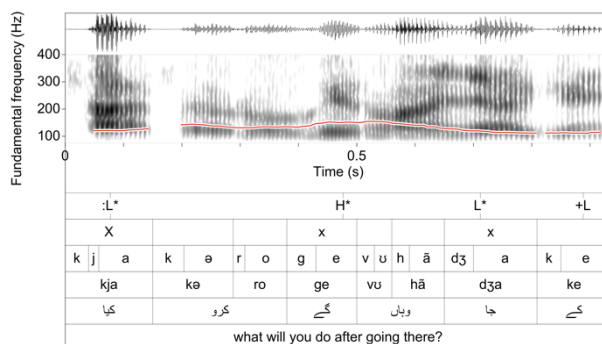


Figure 3: Falling pattern in a WHQ

#### 3.2.2. Yes/no Questions

Polar questions are characterised by a final rising or high level intonation (+H, see Fig. 4).

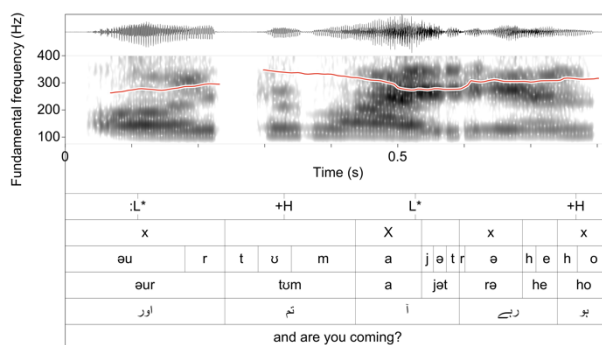


Figure 4: YNQ ending with a rising tone

Question tags in statement-final position also show a rising pattern (Fig. 5).

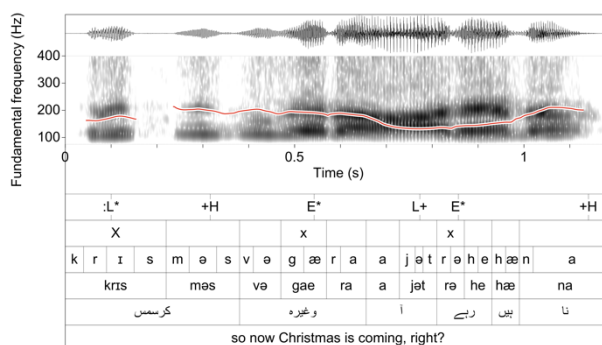


Figure 5: Statement followed by a question tag

#### 3.2.3. Contrastive Focus

Larger pitch excursion and longer duration are the acoustic features that are typically used to realise contrastive focus in Urdu (Jabeen and Braun, 2018), similar to other languages (cf. Rognoni et al., 2017). An emphasised word normally carries an L\*+H pitch accent, characterised by a sharp rise in pitch after the starred tone (see Fig. 6).

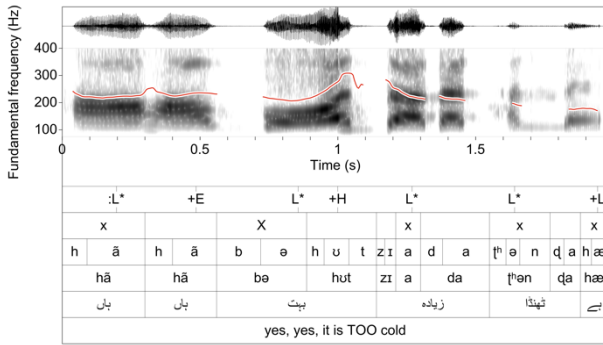


Figure 6: Statement with a contrastive focus

The examples in the data also show that, following a focused word, words are produced with a generally lower pitch and compressed pitch range, as seen in Fig. 6 and in Fig. 7.

Such utterances support the claim that Urdu relies on post-focus compression (PFC, Yu, 2011) and downstep as strategies to highlight the preceding focused information (cf. Patil et al., 2008; Jabeen and Braun, 2018).

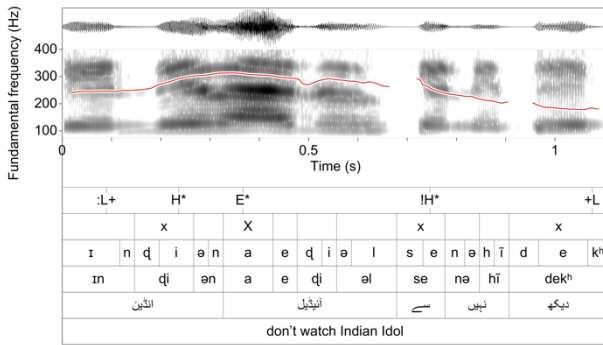


Figure 7: Statement with focus followed by downstep and compressed pitch range

### 3.3. Part of Speech and Pitch Accent

Every word of the data set was annotated with a part of speech (PoS) tag. The complete set of PoS tags, along with their expansions and their frequency, are presented in Table 4.

PoS	Category	#	PoS	Category	#
ADJ	Adjective	658	LVERB	Light Verb	637
ADP	Adposition	1107	NOUN	Noun	2134
ADV	Adverb	1181	NUM	Numeral	197
CONJ	Conjunction	422	PART	Particle	937
COV	Co-verb	599	PRON	Pronoun	1649
DET	Determiner	393	VERB	Verb	3021
IJ	Interjection	589			

Table 4: Word counts by PoS types

The *chi-square* statistic was used to test for associations between frequency of accent status (accented vs unaccented) and part of speech. We hypothesised that the more information-rich PoS categories, except verbs, would show a positive association with pitch accent (i.e. a stronger

than expected tendency to be accented). Verbs were not expected to carry accents at a higher rate: the structure of Urdu sentences is typically SOV (Masica, 1991), and the object that precedes the verb is more likely to attract accent (Patil et al., 2008; Jabeen et al., 2016), resulting in a reduced frequency of accent on the verbs that follow.

The data show that most function word categories (with the exception of interjections) lack a positive association with accent, and, among content word categories, only nouns and co-verbs showed a higher than expected rate of accentuation. Co-verbs are meaningful elements combining with a light (semantically bleached) verb to form a complex predicate. In particular, Urdu co-verbs are generally drawn from the Noun and Adjective classes, although they can also be other verbs. In contrast, based on the overall proportion of accented to unaccented words, verbs are confirmed to be accented less frequently than expected (see Table 5). This suggests a positive tendency for verbs to reject, rather than attract, pitch accent. No significant association is found between frequency of accent status and PoS for adjectives, adverbs, conjunctions, determiners, light verbs, numerals and pronouns. For this reason, these categories were excluded from the results summarised in Table 5.

PoS	Level of significance		Frequency of accent (observed relative to expected)
	p=0.05	p=0.01	
ADP	*	*	Lower
COV	*		Higher
IJ	*	*	Higher
NOUN	*	*	Higher
PART	*	*	Lower
VERB	*		Lower

Table 5: PoS types and frequency of accent

## 4. Discussion and Conclusions

The majority of the sentences in our data are statements or wh-questions. Both these types of sentences end in a low tone (+L). Yes/no questions show a final rising contour, while contrastive focus is realised with a low pitch accent followed by a sharp rise (L\*+H).

Urdu intonation is commonly claimed to comprise sequences of low-high prosodic phrases (L\*+H) (Nayyer and Madni, 2002; Jabeen, 2010). This is only partially confirmed by our data, where L\*+H/:L\*+H were indeed the most frequent bitonal sequences, but single-tone nuclear pitch accents L\* and H\* occurred with a considerably higher frequency. These results show that the high number of L\* tones found in our data confirms the important role played by low pitch accents in Urdu in spontaneous speech. It may be that L\*+H phrases are more often realised in careful speech than in spontaneous conversation. Alternatively, the extensive use of the E\* labels allowed by the RaP system might conceal what could otherwise be interpreted as L\*(+H) accents, albeit with compressed pitch range (see 2.2). Indeed, many instances of L+E sequences annotated using the RaP system could be L+H structures in a binary L/H annotation system such as ToBI (Dilley and Brown, 2005). Finally, the frequency of E\* labels also suggests that Urdu spontaneous speech could be

characterised by relatively flatter pitch contours as compared to careful or scripted speech.

To the best of our knowledge, there have been no previous studies touching on the association of PoS with pitch accent in Urdu/Hindi. Our preliminary analysis of this association indicates that nouns, in particular, tend to attract pitch accent. The category of co-verb also shows a significantly stronger affiliation for accent than expected, while verbs are characterised by a lower rate of accent.

In general, we found a higher number of unaccented words as compared to accented words. Excluding one-word utterances, the ratio of words without accent to accented words was higher than 2:1. It may be that the use of E\* labels in RaP results in relatively low accent counts. As noted in Rognoni et al. (2017), in stretches of speech where differences in pitch are subtle, the RaP system does not allow more than two equal accents in a row, where ToBI annotations would still allow multiple consecutive L\* or H\* labels. This may also be another reason why the L\*+H pattern found in Jabeen (2010) and in Nayyer and Madni (2003) was not consistently present in our data.

This paper presented an acoustic-phonetic analysis of Urdu intonation patterns in spontaneous conversational speech. Claims about the phonetic realisation of Urdu pitch accents and contrastive focus based on Urdu/Hindi scripted speech were confirmed by the data presented in this study. This study also provided an initial analysis of the strength of the relation between pitch accent and PoS in Urdu.

## 5. Bibliographical References

- Boersma, P. and Weenink, D. (2019). "Praat: doing phonetics by computer [Computer program]. Version 6.1.07," 2019. [Online]. Available: <http://www.praat.org/>. [Accessed 2 December 2019].
- Dilley, L., Breen, M., Bolivar, M., Kraemer, J. and Gibson, E. (2006). A comparison of inter-transcriber reliability for two systems of prosodic annotation: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices). Proceedings of Interspeech 2006 – 9th Annual Conference of the International Speech Communication Association, September 17-21, Pittsburgh, USA, pp. 317-320.
- Dilley, L. and Brown, M. (2005). The RaP (Rhythm and Pitch) labeling system. [Online]. Available: [https://www.academia.edu/2730467/The\\_RaP\\_Rhythm\\_and\\_Pitch\\_Labeling\\_System](https://www.academia.edu/2730467/The_RaP_Rhythm_and_Pitch_Labeling_System). [Accessed 2 December 2019].
- Dyrud, L. (2001). Hindi-Urdu: Stress Accent or Non-Stress Accent?. University of North Dakota, MSc Dissertation.
- Hock, H. and Bashir, E. (2016). The Languages and Linguistics of South Asia: A Comprehensive Guide, Berlin: de Gruyter Mouton, 2016.
- Hussain, S. (1997), Phonetic correlates of lexical stress in Urdu. PhD dissertation, Northwestern University, Evanston, USA.
- Jabeen, F. (2010). An acoustic study of the influence of Urdu on the intonation patterns of English in Pakistan. MSc dissertation, GC University, Faisalabad, Pakistan.
- Jabeen, F., Bögel T. and Butt, M. (2016). Variable Prosodic Realization of Verb Focus in Urdu. *Speech Prosody 2016*, Boston, USA.
- Jabeen, F. and Braun, B. (2018). "Production and Perception of Prosodic Cues in Narrow & Corrective Focus in Urdu/Hindi," in *Speech Prosody 2018*, Poznań, Poland.
- Masica, C. (1991). The Indo-Aryan Languages, Cambridge: Cambridge University Press.
- Nayyer, R. and Madni, F. (2003). Analysis of Intonation Patterns in Urdu. [Online]. Available: [http://crulp.org/Publication/Crulp\\_report/CR03\\_05E.pdf](http://crulp.org/Publication/Crulp_report/CR03_05E.pdf). [Accessed 22 March 2018].
- Nyyar, S. (2002). Syllable stress in Urdu. [Online]. Available: [www.cle.org.pk/Publication/Crulp\\_report/CR02\\_24E.pdf](http://www.cle.org.pk/Publication/Crulp_report/CR02_24E.pdf). [Accessed 22 March 2018].
- Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C. and Vasishth, S. (2008). Focus, Word Order and Intonation in Hindi. *Journal of South Asian Linguistics*, vol. 1, no. 1, pp. 55-72.
- Rognoni, L., Bishop, J. and Corris, M. (2017). Pashto Intonation Patterns. Proceedings of Interspeech 2017 – 18th Annual Conference of the International Speech Communication Association, August 20-24, Stockholm, Sweden, pp. 1228-1232.
- Sengar, J.A. and Mannel R. (2012). A Preliminary Study of Hindi Intonation. Proceedings of the 14th Australasian International Conference on Speech Science and Technology, December 3-6, Macquarie University, Sydney, Australia, pp. 149-152.
- t'Hart, J., Collier, R. and Cohen, A. (1990). A perceptual study of intonation, Cambridge: Cambridge University Press,.
- Xu, Y. (2011). Post-focus compression: cross-linguistic distribution and historical origin. Proceedings of the 17th International Conference of Phonetic Sciences (ICPhS), August 17-21, Hong Kong, PRC, pp. 152-155.