Production and Perception of Prosodic Cues in Narrow & Corrective Focus in Urdu/Hindi

Farhat Jabeen, Bettina Braun

University of Konstanz

`firstname.lastname@uni-konstanz.de`

Abstract

This study investigates the production and perception of prosodic cues to realize narrow and corrective focus in Urdu/Hindi. We recorded SOV sentences with the target constituents at the preverbal position. Our results show that correctively focused nouns have longer syllable duration, wider F0 range, early alignment of F0 peaks, variant production of downtrend, and less steep post-focal compression as compared to narrowly focused nouns. We further set up a perception experiment to investigate if the difference in syllable duration of narrowly and correctively focused constituents is perceptible to Urdu/Hindi speakers. We manipulated syllable duration of the target constituents and presented them in contexts via a web-based interface. Twenty-nine respondents rated the naturalness of manipulated sentences in the given contexts. The analysis of respondents' ratings indicated that while they accepted both long and short durations in narrow focus, they rated long duration significantly better in corrective focus. Our results lend support to earlier claims about the prosodic cues of corrective focus in Urdu/Hindi [1, 2] and bring new evidence regarding the perceptual relevance of duration to cue corrective focus.

Index Terms: narrow focus, contrastive focus, perception, duration, F0 range, F0 alignment, downtrend, Urdu, Hindi

1. Introduction

Urdu/Hindi[1] is an Indo-Aryan language that offers an interesting interplay of prosody, word order, and i(nformation) -structure). In Urdu/Hindi, all the major constituents in a sentence can scramble and this influences its i-structure [3]. The sentence initial constituents are topicalized, the immediately preverbal position is reserved for focus, and the sentence final position is used to convey background information and givenness. This is also the position for heavy NP shift [4, 5, 6]. While this notion of i-structure is open to discussion, it shows that an investigation of prosody in Urdu/Hindi should take word order and i-structure into account. However, most of the existing researchers investigating the prosody of i-structure in Urdu/Hindi do not consider the influence of structural positions. This has resulted in contradictory claims about the interplay of prosody, word order, and i-structure. The present research examines this interplay while controlling for variation in i-structure and word order by placing the target constituents only at the preverbal position in SOV sentences. It also investigates the production and perception of the prosody of narrow and corrective focus.

---

[1]Apart from differences in script and loan words due to historical reasons, Hindi and Urdu are structurally very similar. The linguistic claims about one languages are generally accepted to be true for the other too. We here will refer to the pair as Urdu/Hindi.

2. Focus and Prosody in Urdu/Hindi

There are few systematic investigations in the prosody of Urdu/Hindi. Earlier researchers such as [7] and [8] studied the prosody of declaratives and questions and found that the non-final constituents in Urdu/Hindi have a default LH F0 contour. The association of L with lexical accent and the interpretation of H as either a phonological phrase boundary or the trailing end of a bitonal unit are open to discussion. [8] placed the focused constituent at the sentence initial as well as preverbal position. His data showed that focus at the preverbal position is marked by wider F0 span and compression of F0 register on the post-focal constituents. He claimed that this compression of register is different from deaccentuation as the LH pattern is still realized on the post-focal constituents, albeit the F0 range is compressed. He found the same results for the focused constituents at the sentence initial position. But the prosody of sentence initial focus is tricky as one needs to distinguish between the prosodic effects of focus and topicalization. Moreover, his stimuli were presented out of context and it's difficult to determine the focus type his respondents had produced.

[9] studied focus types in a controlled experiment and investigated the interplay of prosody with focus as well as word order in Urdu/Hindi. They studied the prosody of subject and object focus in comparison with broad focus in SOV and OSV sentences. They could not find any systematic difference in the F0 and duration of the narrowly focused constituents and their broadly focused counterparts. They concluded that focus in Urdu/Hindi is marked by post-focal deaccentuation than by manipulating the F0 or duration of the focused constituent. The problem with [9]'s analysis is that they also do not distinguish between the i-structure of sentence initial focus (also topicalized) and the preverbal focus (also structurally focused). [10] attempts to explain [8] and [9]'s findings in terms of i-structure positions. She claims that prosodic focus marking is optional when focus is marked at the structural position i.e. preverbal. But when focus is marked prosodically, it can use a few or all of the following strategies: F0 raising, wider F0 span, alignment of F0 peak with the end of noun as opposed to the following case marker in a constituent, and post-focal compression. However, [10]'s claims are relevant for new information focus only. She also does not clarify what happens when focus is placed at a non-preverbal position in the sentence leading to a clash between structural and prosodic focus.

An interesting contribution to understanding the prosody of different focus types in Urdu/Hindi is made by [1]. They studied the prosody of contrastively focused adjectives in comparison with broad focus and found that contrastively focused adjectives were realized with the same F0 contour as the ones in broad focus. However, in contrastive context, adjectives had significantly longer syllable duration and wider F0 span. The contrast in the findings of [1] and [9] confirms that it is per-

tinent to distinguish between different focus types to reach a better understanding of prosodic focus in Urdu/Hindi.

A detailed investigation in this regard is provided by [2] who compared the prosodic realization of new information (answer to a wh-question), selective (selection between two items), and corrective focus (correction of a previously stated constituent) in SOV and OSV sentences in Urdu/Hindi. Similar to [9], they found that word order does not affect prosody and that F0 is not used to cue different focus types. However, they showed that the relative duration of the constituent played a significant role in distinguishing between corrective and other focus types. Thus, correctively focused constituents were the longest, whereas selective and new information focus had the same relative constituent duration. While these results offer an interesting interplay of prosodic cues, a recent research by [11] has further added to the mass of contradictory findings.

[11] also studied the prosodic cues related to marking new information, selective, and corrective focus in Urdu/Hindi. They found that focused constituents had higher intensity (the only significant cue) compared to their given counterparts, but their analysis did not reveal any significant difference in the F0 or duration of different types of focus. This overview shows that the issue of prosodic realization of focus types in Urdu/Hindi is far from resolved. It is noticeable that the studies cited above do not take the structural focus position into account and the fact that the sentence initial focus is placed at a position otherwise marked for topicalization remains unnoticed and undiscussed. Therefore we set up this experiment to investigate the production of different types of focus only at the preverbal position. In order to tease apart the relevant issues individually, we investigated the prosody of new information narrow focus in comparison with corrective focus. We here report the results of a production experiment. We also present the findings of a perception study regarding the association of syllable duration with narrow and corrective focus.

3. Production

3.1. Materials

The dataset consisted of twelve SOV sentences each presented in narrow and corrective contexts. All the target nouns were disyllabic (CV.CV) with stress on the first syllable. The nouns were followed by a case marker (*ko*) and placed at the preverbal position. An example sentence in both narrow and corrective focus contexts is given below:

Narrow focus: The gardener was working in the garden when someone asked him to call a resident of the house. Your mother asks whom the gardener had called. You reply:

Corrective focus: The gardener was working in the garden when someone asked him to call a resident of the house. Your mother thinks that the gardener had called Ali. You correct her and say that, in fact:

mali=ne   [zara=ko]$_F$ bʊla   ḍi.ja   tʰa
gardener.M.S=Erg Zara.F.S=Acc call-Perf give-Perf be.Past.M.S
'The gardener had called Zara.'

3.2. Participants

Four speakers of Urdu (3 females) were recorded for this experiment. All the participants were Pakistanis living in Germany for at least three years. They were multilingual and spoke Urdu, English, and at least one regional language from Pakistan. We understand that their multilingual background affects their language usage. But Pakistan is a multilingual country and it's hard

to find a monolingual educated speaker of Urdu within the country either. Urdu is a lingua franca that connects communities of different linguistic and ethnic backgrounds in Pakistan. All the participants spoke Urdu with family members and friends. The participants were unaware of the purpose of the experiment.

3.3. Data collection

The data was recorded with a head-mounted Schure microphone and ZoomH6 at the sampling frequency of 44,100Hz in a quiet room. The stimuli were presented in Urdu script via Ms PowerPoint. The target sentences were interspersed with twenty-nine fillers (declaratives and mathematical equations). The participants read the context loudly and said the target sentence as a response in the given context. They were directed to speak as naturally as possible and repeat in case of coughing, stuttering or laughter. Participants controlled the pace of the experiment. The average duration of the experiment was 11 minutes.

3.4. Data analysis

The target sentences were labeled manually to measure syllable duration in pre-focal and focused constituents. To analyze the F0 contour, the F0 valleys and peaks were labeled as L and H respectively. F0 range in pre-focal and focused constituents was calculated in semitones. Moreover, the alignment of the H tone was investigated. We also measured the downtrend in semitones between the F0 peaks in the first and the second constituents. When the first peak was at least 1 semitone higher than the second, the sentence was coded as showing downtrend. Moreover, in order to determine post-focal compression, the difference in F0 of the last syllable of the focused constituent and the first syllable of the following verb was calculated in semitones. Six sentences were removed because the participants did not pause between the context and the target sentence or they produced the bisyllabic subject noun as trisyllabic. For each constituent, we fitted a series of linear mixed effects regression models (lmers) for the dependent variables i.e. duration, F0 range, and post-focal compression. Focus was the fixed factor and participants and items were added as random factors [12]. Duration was modelled separately for the first and second syllables of the focused and the pre-focal constituents and the case markers. Downtrend and alignment of F0 peak were modelled with a logistic mixed-effects regression model (glmer). When it was not possible (empty cells), we ran a χ^2 test.

3.5. Results

3.5.1. Syllable duration

The analysis of syllable duration indicated that the first (stressed) syllable was significantly shorter in narrow focus than in corrective focus context (β: -0.02, $p < 0.0001$).The average syllable duration is given in Table 1. Similarly, the second syllable was significantly short in narrow focus context (β: -0.008, $p = 0.01$). Interestingly though, the case marker following the focused noun was significantly long in narrow focus (β: 0.01, $p = 0.04$). No significant difference was found in the syllable duration of the pre-focal constituents.

Table 1: *Average duration(ms) of syllables in narrow and corrective contexts.*

Context	1st syllable	2nd syllable	Case Marker
Narrow	.195	.132	.136
Corrective	.238	.158	.132

3.5.2. F0 (contour & alignment)

As observed by [2], [9], and [11], our data also showed that each non-final constituent has an L tone on the stressed syllable followed by an H tone at the end of the constituent (Figure 1).
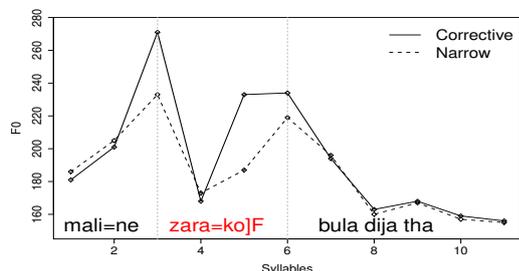

Figure 1: *F0 (measured in the middle of the vowel) contour of a correctively and narrowly focused sentence.*

The H tone either aligned with the case marker immediately following the focused noun (late as shown in the narrowly focused "zara=ko" in Figure 1) or with the last syllable of the focused noun (early as shown in the correctively focused "zara=ko" in Figure 1). This variation was associated with focus condition (Table 2). When the target noun was correctively focused, H always had early alignment. However, when the target noun had narrow focus, the alignment was either early or late.

Table 2: *H alignment in focused and pre-focal constituents.*

	Focused		Pre-focal	
	Corrective	Narrow	Corrective	Narrow
Early	100	48	0	3
Late	0	52	100	97

A χ^2 test showed that the difference in the distribution of H alignment on the focused constituent with relation to focus type was significant ($p < 0.0001$). The distribution of alignment in the focused constituent becomes more interesting when compared with almost universal late alignment in the pre-focal constituent as shown in Table 1 above.

3.5.3. F0 range

The F0 range in semitones was measured for the pre-focal and the focused constituents. The results showed that the F0 range between L and H in the focused constituent was significantly narrower in narrow focus than in corrective focus (β: -0.44, p = 0.02). Similarly, the F0 range in the pre-focal constituent was narrower in narrow focus than in corrective focus (β: -0.64, p = 0.03). Thus the sentences with a correctively focused constituent overall had a wider F0 range on the pre-focal and focused constituent. This finding about the wider F0 range in corrective focus confirmed the findings of [1].

3.5.4. Post-focal compression

We measured the difference in semitones in F0 between the last peak on the focused constituent and the F0 on the first syllable of the following verb. We found that the degree of fall on the verb following the focused constituent was greater in narrow focus than in corrective focus (β: 0.26, p < 0.0001). Thus post-focal compression was greater after the narrowly focused constituents than after their correctively focused counterparts. This is in line with the observation in *3.5.3* that the overall range of F0 is higher in corrective as compared to narrow focus.

3.5.5. Downtrend

[13] explained that the downtrend between successive constituents in a sentence in South Asian languages differed from downstep in its predictability. [8] and [14] claimed that Urdu/Hindi depicted a universal pattern of downtrend. Contrarily, we observed sentences where the second (focused) constituent had a higher F0 peak than the first or the second peak was scaled at the same height as the first peak. Our analysis showed that narrowly focused sentences had significantly more occurrences of downtrend than the correctively focused sentences (β: 1.18, p = 0.01).

This lack of downtrend was also observed by [15] in her investigation of broad and corrective focus in Hindi. Considering [14]'s claims about the lack of recursive prosodic phrasing due to downtrend, our finding has implications for the prosodic structure in Urdu/Hindi. But a discussion of prosodic structure is beyond our scope here.

3.6. Discussion

The results shown above indicate that the correctively focused nouns have longer syllable duration and wider F0 span but a lower degree of post-focal compression as well as variable occurrence of downtrend. Moreover, correctively focused nouns have early H alignment whereas in narrow focus, the H tone alignment is variable. We explain this variation in prosodic marking in terms of the structural focus position for narrow focus (cf. [10]). As narrow focus can be rendered by position, the prosodic marking of narrow focus becomes optional. But there is no structural cue for corrective focus as it can be marked anywhere in the sentence. So the prosodic cueing is essential to convey corrective focus. This phenomenon has also been observed in other typologically different languages that use a structural position to convey narrow focus (see [16] for Hungarian, [17] for French). The optional marking of narrow focus may also be associated with the elastic word order in Urdu/Hindi. [18] claim that while languages with relatively free word order do mark contrastive focus prosodically, it appears that their prosodic marking of narrow focus is optional. Our data bears evidence to this claim for Urdu/Hindi as well.

4. Perception

The results of our production experiment discussed above indicate that corrective focus is associated with certain prosodic markings. However, the prosodic cues observed in speech production may not be perceptually relevant to the speakers of a language [19]. Therefore, we investigated if the prosodic cues mentioned in section 3 above are perceptible to speakers of Urdu/Hindi to mark corrective focus. We here report the results of the first experiment from a series of perception experiments planned for the future. In this experiment, we manipulated the duration of syllables in the target constituents and presented them in narrow and corrective focus.

4.1. Stimuli & manipulation

A female speaker from the production experiment recorded the stimulus sentences. We used the same set of sentences and contexts as the ones used for the production experiment. These contexts will henceforth be referred to as the 'recording context'. We measured the syllable duration of the focused constituent produced by this speaker. For each sentence pair, we calculated the ratio of difference in duration between correctively and nar-

rowly focused syllables (Table 3) and used it for duration manipulation. So the originally long syllables were shortened and the originally short syllables were elongated. Recall that the case markers showed reverse duration patterns in the production experiment. Accordingly the case markers recorded in narrow focus were shortened but elongated in the correctively focused constituents.

Table 3: *Ratio for manipulation of syllable duration.*

Context	1st syllable	2nd syllable	Case Marker
Narrow	0.81	0.83	1.02
Corrective	1.22	1.19	0.97

The syllable duration was manipulated using PRAAT (v. 6.0.28) [20]. The duration ratios were used to add duration points to the beginning and end of syllables in the target constituents.Then a resynthesis (overlapp-add) of the manipulated sentence was published. This sentence was used as stimulus for the perception experiment. We are aware that our manipulation of syllable duration affected the F0 contour and stretched or squeezed it when syllables were elongated or shortened respectively. Apart from this unavoidable artefact of the experimental manipulation, F0 scaling and alignment remained unchanged.

4.2. Apparatus

The recordings with original and modified durations were presented in narrow and corrective focus (presented context), resulting in eight conditions (12 sentences x 2 recording contexts x 2 presented contexts x 2 duration conditions). All factors were manipulated within-subjects but between-items. We constructed two lists, such that each participant was presented with all conditions, but for different items. Each list contained 48 experimental and 14 filler items. In order to avoid comparing manipulated speech with natural speech, the overall F0 of the sentences with original duration was raised by 10Hz.

4.3. Procedure

The experiment was carried out via a web-based interface. The participants read the context (presented in Urdu script), listened to the utterance, and rated its naturalness relative to the context on a scale from 1 (most unnatural) to 5 (most natural). Each sentence had to be played at least once but no more than three times. Each participant responded to only one list. The average duration of the experiment was 10 minutes and the average time spent per item was 13 seconds.

4.4. Participants

Twenty-nine respondents (24 male) participated in the experiment. All the participants were 21-30 years old Pakistanis living in Germany. They were multilingual and used Urdu regularly in their daily life. They spoke English, German, and at least one regional language from Pakistan.

4.5. Data analysis

For our analysis, we used the respondents' ratings of target sentences in the presented contexts. We also measured the response times (relative to the onscreen display of the context) for each rated item. The average duration of the stimuli sentences was 2 seconds. Ratings with response times shorter than 5 seconds were excluded. For statistic analysis, we first fitted an lmer model with duration, recorded and presented contexts, two lists, and their interactions as fixed factors and participants and items

as random factors. The model was simplified by eliminating interactions and main effects that were not significant at p = 0.1.

4.6. Results

The results showed a significant interaction between duration condition and presented context (p = .004). As shown in Figure 2, both long and short syllable durations were acceptable in narrow focus context. But in corrective focus context, long durations were rated significantly better than short durations (β: 0.24, p = 0.001). Recording condition had a main effect, with lower ratings for narrow focus than for corrective focus (β: -0.21, p = 0.017). No significant effect of list was found.
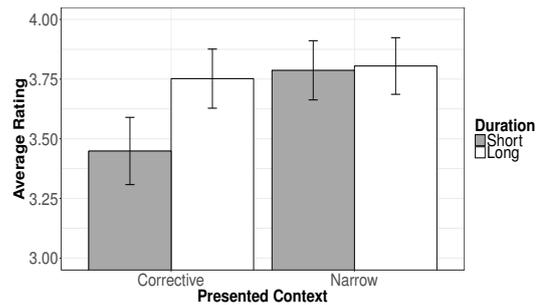

Figure 2: *Average ratings for duration manipulation in narrow and corrective contexts. The whiskers indicate CI (95%).*

4.7. Discussion

Our data lend further support to the results of our production experiment as well as the findings of [1] and [2], who reported longer syllable durations in corrective than broad and narrow focus. The lacking interaction between presented and recorded context and the interaction between presented context and duration condition shows that duration is a vital cue. As stated in *3.6*, the observed asymmetry in the interpretation of duration could be a consequence of the structural focus position. Here no durational marking is necessary and is not interpreted. Corrective focus needs to be marked by duration; durations that are too short are less acceptable. An alternative interpretation is that listeners are more sensitive to the correct prosodic realization of more marked forms (corrective focus) but accept overly marked prosodic forms in less marked pragmatic conditions (narrow focus). (cf. [21] for contrastive vs. non-contrastive topics.)

5. Conclusion & future work

Our investigation of the prosody of narrow and corrective focus in Urdu/Hindi brings in new evidence regarding the variable alignment of H tone in the narrowly and correctively focused constituent as well as the perceptual relevance of duration to mark corrective focus. It also confirms the findings of previous researches and raises interesting questions for future research. We are planning further production experiments to investigate the prosody of focus types in OSV and other non-canonical word orders in Urdu/Hindi. We are also planning more perception experiments to investigate if Urdu/Hindi speakers are sensitive to variation in the alignment of H tone and F0 range for narrow and corrective focus.

6. Acknowledgment

7. References

[1] S. Genzel and F. Kügler, "The prosodic expression of contrast in Hindi," in *The Proceedings of Speech Prosody 2010*, Chicago, 2010.

[2] A. Choudhury and E. Kaiser, "Comparing languages and comparing focus types: Insights from an interactive task," 2016, talk presented at Prosody and Information Structure workshop, University of Stuttgart.

[3] V. Gambhir, "Syntactic restrictions and discourse functions of word order in standard Hindi," Ph.D. dissertation, University of Pennsylvania, Philadelphia, 1981.

[4] A. Kidwai, *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press, 2000.

[5] M. Butt and T. H. King, "Structural topic and focus without movement," in *Proceedings of the First LFG Conference*, M. Butt and T. H. King, Eds. Stanford: CSLI Publications, 1996.

[6] ——, "Null elements in discourse structure," 1997, written to be part of a volume that never materialized, http://ling.uni-konstanz.de/pages/home/butt/main/papers/nulls97.pdf.

[7] R. R. Moore, "A study of Hindi intonation," Ph.D. dissertation, University of Michigan, 1965.

[8] J. D. Harnsberger, "Towards an intonational phonology of Hindi," Master's thesis, Unversity of Florida, 1994.

[9] U. Patil, G. Kentner, A. Gollrad, F. Kügler, C. Féry, and S. Vasishth, "Focus, word order and intonation in Hindi," *Journal of South Asian Linguistics*, vol. 1, no. 1, pp. 55–72, 2008.

[10] F. Jabeen, "Position vs. prosody: Focus realization in Urdu and Hindi," in *Phonetics and Phonology in Europe*, Cologne, Germany, 2017.

[11] C. Féry, P. Pandey, and G. Kentner, "The prosody of focus and givenness in Hindi and Indian English," *Studies in Language*, vol. 40, no. 2, pp. 302–339, 2016.

[12] H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, 2008.

[13] S. ud Dowla Khan, *The intonational phonology of Bengladeshi Standard Bengali*. Oxford University Press, 2014, ch. Prosodic Typology II, pp. 81–117.

[14] C. Féry and F. Schubö, "Hierarchical prosodic structure in the intonation of center-embedded relative clauses," *The Linguistic Review*, vol. 27, pp. 293–317, 2010.

[15] V. Puri, "Intonation in Indian English and Hindi late and simultaneous bilinguals," Ph.D. dissertation, University of Illinois, Urbana Champaign, 2013.

[16] K. Mády, "Prosodic (non-)realization of broad, narrow and contrastive focus in Hungarian: a production and a perception study," in *InterSpeech 2015*, Dresden, Germany, 2015.

[17] J. V. Klok, H. Goad, and M. Wagner, "Prosodic focus in English vs. French: A scope account," *Glossa*, 2017.

[18] A. M. Devine and L. D. Stephens, Eds., *The prosody of Greek speech*. USA: Oxford University Press, 1994.

[19] T. Bögel, "The syntax-prosody interface in lexical functional grammar," Ph.D. dissertation, University of Konstanz, 2015.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program, Version 6.0.05]," 2015, available at http://www.praat.org/ [retrieved 08.11.2015].

[21] B. Braun, "Answers to the perception of thematic contrast and questions regarding the perception of thematic "non-contrast"," in *Proceedings of the 2nd International Conferences on Speech Prosody*, Nara, Japan, 2004, pp. 685–688.