Interspeech 2024
1-5 September 2024, Kos, Greece

Urdu Alternative Questions: A Hat Pattern

Benazir Mumtaz, Miriam Butt

University of Konstanz, Germany
firstname.lastname@uni-konstanz.de

Abstract

This study explores the prosody of alternative questions (AltQs) in Urdu. We found that AltQs are characterized by a hat pattern that is linked to double contrast sentences, featuring two contrastive foci, similar to German. However, unlike German, the selection of the phrasal accent (PA) is fixed. In Urdu the first PA features an L*(+H)Ha while the second accent has an L*(+H). The hat pattern's full realization depends on the spacing between PAs: greater spacing leads to a prolonged high peak, while reduced spacing results in a compressed hat pattern. Moreover, both the accentual phrase (AP) and the hat pattern carry an L*H contour, with differences in H tone realization: with the hat pattern, the H is significantly higher than that of the AP. We explain the presence of this shared phonological L*H contour by noting that the basic intonational phonology of Urdu is defined by a series of L*H contours, that can then be realized in different ways to show, for example, focus or AltQs.
Index Terms: Alternative questions, Urdu TTS, Hat patterns

1. Introduction

As part of a cooperative project, we have been investigating the prosody of Urdu to aid the development of a Text-to-Speech (TTS) system for Urdu. Urdu is the national language of Pakistan, whose literacy rate hovers around 50%, so a natural sounding TTS is crucial to help ameliorate literacy barriers and to provide access to written on-line Urdu content. We here focus on the prosody of alternative questions (AltQs).

Next to no previous work exists on the prosody of AltQs in Urdu, except for claims in [1] that Urdu/Hindi AltQs use similar prosodic cues to English, including a pitch accent on each disjunct, a prosodic break between the disjuncts, and a final fall. However, results gathered as part of previous research showed that in addition to these three crosslinguistically signature properties of AltQs, Urdu seems to use a fourth cue: a hat pattern.

In Urdu, L*Ha serves as the default phrasal accent and neither focus nor the syntactic constituency types can disrupt this phrasal accent [2, 3]. Given this fixed intonational phonology, it becomes intriguing to investigate the influence of the hat pattern on the choice of phrasal accent in Urdu. Additionally, if the hat pattern also bears a L*Ha contour, the question is what distinguishes the L*Ha of the hat pattern and the typical L*Ha of a phrasal accent. We report results from two production experiments to resolve these questions.

2. Background

Urdu/Hindi intonation is defined by a series of rising contours (L*Ha) recognized as repeated rising contours (RRCs). The prosodic unit corresponding to this rise is called an Accentual Phrase (AP). There is a general censensus in the literature that each constituent in declaratives, with the exception of the clause-final constituent (which undergoes a fall), is associated with a rising contour (as in Fig1(a)) [4, 2, 5]. Even the focused constituent carries the basic L*H contour but with increased F0 range, longer syllable duration, greater intensity, and post-focal compression [2, 3, 4, 6, 5].

Moreover, while the literature does not document the occurrence of two constituents carrying a single rising contour, our ongoing investigations to understand the prosody of AltQs reveal an interesting observation. In AltQs, two constituents (NP_1 and V along with a cliticized conjunction (Conj) as shown in Fig1(b)) can constitute a single AP. This prosodic incorporation of two constituents and a Conj into one AP creates a sustained high pitch between two accents forming a hat pattern.

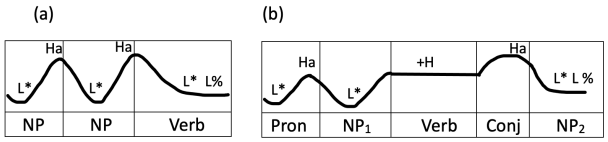

Figure 1: *Fig1(a) depicts the default tonal structure of Urdu declarative: each constituent forms a single AP. Fig1(b) shows the tonal structure of Urdu AltQs, where NP_1 V Conj form a single AP.*

The hat pattern, also known as a bridge accent ([7]), is defined as instances in which the F0 does not drop down to a low-level target between two accents [8]. In German, the first accent in the hat pattern is phonologically described as H* or L*+H, while the second accent is described as (H+)L* or H*L. [9] differentiates between two hat contours: (a) H* H*L (a complete linking of two pitch accents) and (b) L*H H*L (two fully realized accents). In contrast, [7] described the hat pattern as H*HL*. According to [7], H* in the hat pattern is a thematic accent whereas the fall (HL*) is the rhematic accent. [10] describes the hat pattern as H*>H*+L, indicating a deleted trailing tone of the first accent. Based on this literature, pitch accents in a hat pattern can carry any combination of tones; the only reliable cue of the hat pattern seems to be the absence of a low tonal target between the two accents.

Moreover, different languages attribute distinct meanings to hat patterns. In German, hat patterns are linked to contrast [8], [7], while in English, they signify late focus [11]. In Dutch, a neutral intonation contour is associated with hat patterns [12].

3. Production Experiment 1

Experiment 1 (Exp1) investigates the basic structure of the hat pattern in Urdu AltQs via a production experiment.

10.21437/Interspeech.2024-2172

3.1. Material

The material for Exp1 consisted of seven AltQ sentences we constructed for the experiment, see Table 1. The content within the AltQs was carefully controlled. Some AltQs are in principle ambiguous between a polar question (PolQ) and AltQ reading. We avoid the possibility of such ambiguity by only using sentences of the following form: Pron NP_1 V Conj NP_2, which unambiguously expresses an AltQ. Additionally, each target sentence begins with the pronoun *t̪ʊm* 'you', and all sentences end on an /i/. The NPs are consistently bisyllabic with stress on the first syllable, featuring a CV.CV syllable structure. Both the NPs preceding and following the Conj *ja* 'or' also end on an /i/. Additionally, all verbs end on the future marker *gi*. All target sentences were embedded in contexts, as illustrated in (1).

Table 1: *Target sentences for AltQs*

AltQ	Translation
t̪ʊm məri dʒɑo-gi jɑ bɑli?	Will you go Murree or Bali?
t̪ʊm pɑni pɪjo-gi jɑ kofi?	Will you drink water or coffee?
t̪ʊm muli kʰɑo-gi jɑ gobʰi?	Will you eat radish or cauliflower?
t̪ʊm lɑri betʃo-gi ja gɑɽi?	Will you sell a lorry or a car?
t̪ʊm gʰɔri mɑ̃go-gi jɑ kʰoɽ̃i?	Will you ask for a mare or a donkey?
t̪ʊm roɽi kʰɑo-gi jɑ boɽi?	Will you eat bread or meat?
t̪ʊm bɑli dekʰo-gi jɑ ʧuɽi?	Will you see an earring or a bangle?

(1) **AltQ Context**: *Ask your friend where she plans to go, considering her classmates are urging her to join them on a trip to Murree, while her family prefers spending the vacation together in Bali.*
t̪ʊm məri dʒɑo-gi jɑ bɑli?
you Murree go-Fut or Bali
'Will you go Murree or$_{ALT}$ Bali?'

3.2. Procedure

Participants were given randomized target sentences accompanied by contexts, interspersed with fillers such as mathematical equations, counting, and dates. The sentences were written in the Urdu script and presented on slides. Each slide contained a single sentence and its corresponding context. To ensure participants were familiar with the material, they received the content in advance before the start of the experiment. All recordings took place in a soundproof room in Lahore, Pakistan. Participants were instructed to read the context silently and to articulate the target sentence aloud during the recording.

3.3. Participants

Sixteen participants (11 F, avg. age = 22.3 years) took part. All participants were born and raised in Lahore, Pakistan, are fluent in Urdu and Punjabi and know some English.[1]

3.4. Data treatment

A total of 224 productions were collected (1 context x 7 sentences x 16 speakers; all participants recorded the targets twice). The data was annotated manually at word, tone, and utterance levels using Praat [13]. Word and utterance tiers were marked using the guidelines in [14] by two trained linguists, each of whom independently marked different tiers. Tone labels were marked based on the UTOBI model [5]. The F0 values (10 per word) were extracted using Prosody Pro [15].

[1] It is virtually impossible to find monolingual speakers in Pakistan.

3.5. Results

3.5.1. Hat Pattern

A hat pattern is described as lack of a low tonal target between the two accents [8]. Overall 190 instances (84.8%) of hat patterns were identified in the data (T=224). To analyze the hat pattern in more detail, we selected all the sentences without pauses (AltQ:151) and extracted ten evenly distributed F0 measurements per word. The resulting ten data points were converted into semi-tones (ST) and the mean values for each point were calculated across all speakers and sentences. The black line in Figure 2 shows the aggregated pitch in ST. The corresponding red line has been drawn manually and avoids disruptions in pitch caused by consonants.

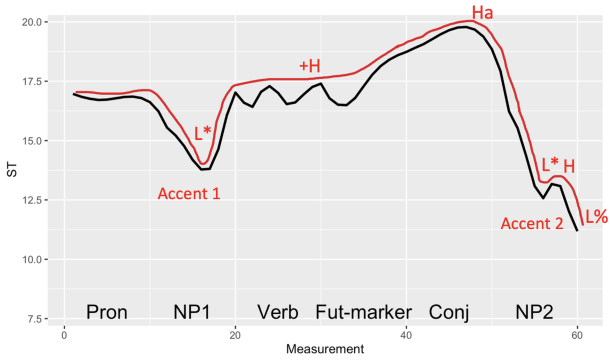

Figure 2: *Aggregated pitch contours of AltQs; red line smoothes perturbations from consonants*

In 84.8% of data, two accents are present, with the first aligning with NP_1 and the second aligning with NP_2. Notably, there is an absence of a low-level target between these two accents. This lack of low tone between NP_1 and NP_2 in AltQs creates a sustained high pitch forming a hat pattern. In the remaining 16% of the data where the hat pattern is absent, the Verb+Conj carry a narrow LH contour, but the verb does not sound as prominent as the NPs.

3.5.2. Pauses in the Hat Pattern

Pauses can affect the prosodic boundaries [16, 17]. To understand the effect of pauses on the hat pattern we focused on the 84.8% hat pattern of data, dividing it into two categories: with pauses (TS=39) and without pauses (TS=151). Pauses were marked using the guidelines developed by [14]. We found that when there is no pause (Fig2), the peak of the hat pattern is realized on the Conj *ja* 'or'. When there is a pause (Fig3) the peak of the hat pattern is realized on the final syllable of the verb (the future marker *gi*) indicating that the pause changes the placement of the peak, but does not affect the hat pattern.

3.5.3. Phrasal Accents in the Hat Pattern

We found that in without pauses the first phrasal accent is consistently placed on NP_1+V+Conj and is always L*+HHa, while the second phrasal accent is on the NP_2 and consists of L*+HL%. In the data with pauses, the first phrasal accent is placed on NP_1+V and is also L*+HHa, while the second phrasal accent is on the NP_2 and carries L*H%.

3.6. Discussion

Each constituent in Urdu/Hindi carries an L*H contour except the final one [4, 2, 5]. In Exp1, the verb appears in a non-final

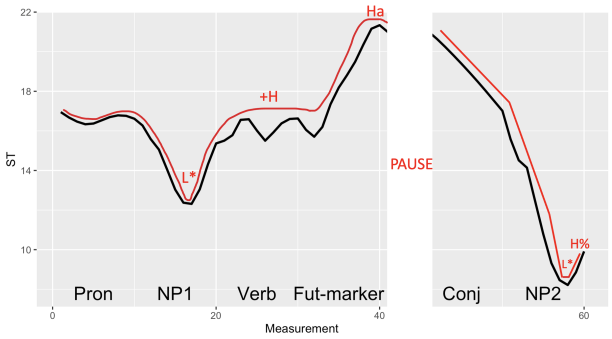

Figure 3: *Hat pattern and pauses in AltQs*

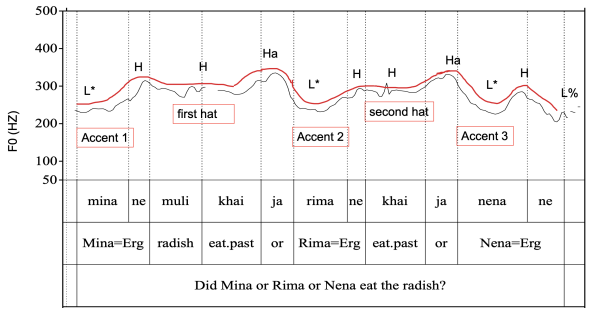

Figure 4: *Hat Pattern in the sentence* minɑ=ne muli k^hɑi jɑ rimɑ=ne k^hɑi jɑ nenɑ=ne?/*'Did Mina or Rima or Nena eat the radish?' with two occurrences of the Conj* jɑ.

position and is thus expected to bear an L*H contour. However, the verb instead is unaccented, carries a floating H tone, and forms an AP with NP_1 suggesting that the L*+HHa is not just a version of a basic L*H. As the contour fits the definition of a hat pattern, we propose to analyze it as one.

One might claim that the verb is actually in a clause-final position in our examples even though it is not linearly final (given that the Conj signals the beginning of the next subclause) and, therefore, cannot bear an L*H contour. However, the data in the existing literature does not indicate an obligatory fall anywhere except for in the utterance final position in declaratives and constituent questions, with the basic L*H contour in evidence elsewhere (though in variations to accommodate focus marking, etc.). One potential explanation for the presence of the H on the Conj might be that the Conj is functioning as a clitic (cf. Fig1(b)). This analysis is supported by what is known about the items that have independently been established as clitics, for example Urdu/Hindi case markers [18]. The prosodic properties of the Conj *ja* are similar to that established for case clitics [19], so we assume a clitic analysis for the Conj *ja*.

We also propose that in Urdu the hat pattern cue is used to highlight the double contrast in an utterance. The first AP (L*+HHa) with the hat pattern signifies the first contrast (NP_1 V Conj/NP_1 V), whereas the second AP (L*(H)) signifies the second contrast (NP_2). The occurrence of the hat pattern serves as a prosodic marker of an item being set into a disjunction with another item. The results also show that the selection of the phrasal accent in the hat pattern is fairly fixed in Urdu. This contrasts with the more flexible positioning of pitch accents in a hat pattern in German [7, 9, 10].

To check if the hat pattern remains evident in sentences with multiple conjunctions, we conducted an exploratory analysis of an utterance containing two occurrences of the Conj *ja*, as depicted in Fig4 (more detailed follow-up work will need to be conducted in future work). We observe two Conjs and two corresponding hat patterns. The initial hat pattern is realized on the prosodic phrase *minɑ=ne muli k^hɑi jɑ*, characterized by L*HHa. The second hat is realized on *rimɑ=ne k^hɑi jɑ*, also featuring the L*HHa accent. This indicates that the hat pattern is indeed used as a prosodic cue for AltQs in Urdu.

4. Production Experiment 2

In Exp2 we set out to further investigate the nature of the hat pattern. One method to assess this is by comparing the hat pattern with the well-known wide-pitch excursion associated with focus and that is observed in polar questions (PolQs). PolQs with a final verb show a final rise as the verb is in focus by default. Thus, for Exp2 we constructed sentences that are ambiguous between AltQ and PolQ readings and have a final verb. Additionally, we constructed the sentences so that we could compare (potentially) different intonations across segmentally otherwise identical stimuli.

4.1. Materials

We used the same set of seven sentences created for Exp1, but with a different word order: Pron NP_1 Conj NP_2 V. All target sentences were embedded in appropriate and disambiguating contexts as shown in Table 2.

Table 2: *Contexts for AltQ vs. PolQ*

Context for AltQ	**Context for PolQ**
Your friend has bought tickets to Murree and Bali but she can only go to one place. Ask your friend where she will go?	You are going on a trip but you do not know whether your friend will go on this trip with you or not. Ask your friend:
t̪ʊm mərɪ jɑ bɑli dʒɑo-gi?	
you Murree or Bali go-Fut	
AltQ: Will you go Murree or_{ALT} Bali?	
PolQ: Will you go to either Murree or Bali?	

4.2. Procedure & Participants

The procedure and the participants were the same as in Exp1.

4.3. Data treatment

A total of 476 productions were recorded (2 contexts x 7 sentences x 16 speakers: 14 participants recorded the target sentences twice and 2 participants recorded the target sentences thrice). Out of the total, 16 AltQs (6.7%) and 14 PolQs (5.8%) were excluded due to perceptual testing failures (when the PolQ or AltQ could not be perceived as one). The remaining dataset consisted of 446 questions (AltQs: 222; PolQs:224). Annotations were performed at the word, tone, and utterance levels, using the same methods as in Exp1.

For word duration we calculated an lmer model with Qtypes (2 levels: AltQs and PolQs), and word class (5 levels: Pron NP_1, Conj, NP_2, V) as fixed effects. Participants and items were added as crossed random effects. The computation used the R packages lme4 and lmerTest [20, 21]. Another cue analyzed was the F0 range, which was converted into semitones (ST) before the analysis. For the F0 range, we concentrated on the NP_1 and NP_2 in PolQs and AltQs. An lmer model was calculated

with 'range difference in nouns' as a dependent variable and Qtype (2 levels: AltQs and PolQs) and Nouns (2 levels: NP1 and NP2) as independent variables. Participants and items were added as crossed random effects. For pauses we calculated a logistic mixed effects (glmer) model with Qtypes as fixed effect and participants and items as crossed random effects. For the H peak analysis on the Conj, an lmer model with 'H tone value in ST' as a dependent variable and Qtype (2 levels: AltQs and PolQs) as independent variables was calculated.

4.4. Results

4.4.1. Word Duration

The lmer model showed a significant interaction between Qtype and word class (x2=227, df=4, p<0.001). Posthoc pair-wise analyses showed that AltQ pronouns do not differ from PolQ pronouns (p=0.8). NP1 (β=0.06, SE=0.004, p<0.0001) and NP2 (β=0.01, SE=0.004, p<0.005) in AltQs are significantly longer than in PolQs. The Conj *ja* is also significantly longer in AltQs than in PolQs (β=0.02, SE=0.004, p<0.001). However, verbs are significantly longer in PolQs than AltQs (β=-0.01, SE=0.004, p<0.0001).

4.4.2. F0 Excursion of the NPs

The results of the lmer model showed a significant interaction between Qtype and Nouns (x2=6.8, df=1, p<0.008). NP_1s in AltQs have a significantly wider range than NP_1s in PolQs (β=1.34, SE=0.24, p<0.0001) as depicted in Fig5. No significant difference was found for the NP_2 in AltQ vs. PolQ (p=0.1).

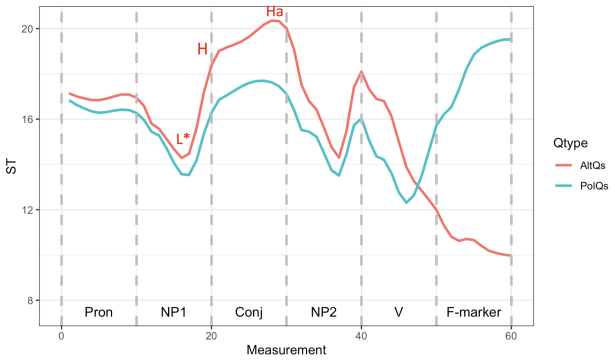

Figure 5: *F0 Contour of string identical AltQs vs. PolQs*

4.4.3. F0 Peak on the Conj

The lmer test results showed that the H on the Conj in AltQs is significantly wider than the H on the Conj in PolQs (x2=12, df=1, p<0.0001) as illustrated in Fig5.

4.4.4. Break/Pauses

A glmer model showed a significant effect of Qtype on pauses (x2=60.7, df=1, p<0.0001). In AltQs we found 93 pauses after the NP_1, see Fig6. As in Exp1, Exp2 showed that the pauses induce a shift in the highest peak. In this case, the peak shifts to the NP1, but it does not eliminate the high peak entirely.

4.5. Discussion

Our F0 results show that NP_1+Conj has a wider range in AltQs compared to PolQs, suggesting a focus on NP_1+Conj in AltQs. This claim is also supported by our duration analysis, showing

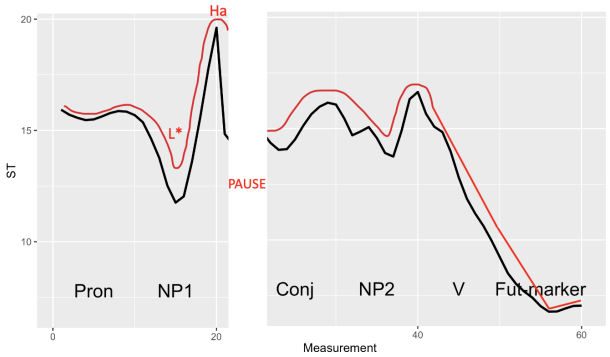

Figure 6: *Alignment of the highest F0 peak with NP_1+pause*

longer durations for both NPs and the Conj in AltQs compared to PolQs. Moreover, we observed a wider range for V, see Fig5, with the absence of an accent on V in AltQs. The final rise on the V in PolQs is consistent with the established pitch excursion on foci in Urdu. We propose that the highest peak realized on the Conj represents the hat pattern in AltQs, supported by F0 peak analysis, where the H tone on the Conj in AltQs significantly differs from that in PolQs. The NP_1+Conj in PolQ has an L*H that is consistent with the usual LH on APs, the intonation of NP_1+Conj in AltQs differs substantially from this.

We also noted phonetic distinctions in the realization of the hat patterns across the experiments. In Exp1, we observed a sustained H pitch on the V+Conj whereas in Exp2 we found a compressed hat pattern. We posit that since in Exp1 a V+Conj appear between two accented NPs, this creates enough space for the complete realization of a hat pattern. In contrast, only a cliticized Conj was placed between two accents in Exp2 and this offers a very reduced space for the full realization of a hat pattern.

Finally, we note that that in both AltQs and PolQs, the F0 contours bear an L*Ha phrasal accent on NP_1+Conj (Fig(5)). We explain the presence of this shared phonological L*H contour by noting that clauses in Urdu are realized via a series of LH contours and by proposing that these basic LH contours can be realized in different ways to signal information prosodically. The narrower L*H contour in PolQs on NP_1+Conj results from Urdu's default L*H prosody, while the wider pitch range on NP_1+Conj in AltQs is a result of an extension and excursion of the basic LH contour to effect a hat pattern that signals a disjunction between alternatives.

5. Conclusion

We set out to investigate the basic structure of the hat pattern in Urdu AltQs and conducted two production experiments. We found that the hat pattern in Urdu is robust across different types of AltQs, but appears with distinct phonetic realizations. In Exp1, we found a hat pattern characterized by sustained high-pitch between two accents, while in Exp2 we noticed a compressed hat pattern. We also noticed the differences between the default L*H accent in Urdu from the L*H found in the hat pattern and we linked these differences to focus. Our results furthermore show that in contrast to German, the choice of phrasal accents in the hat pattern is fairly fixed in Urdu where the hat pattern is a prosodic marker to signal the double contrast and therefore the disjunction between alternatives in an AltQ.

6. Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding within project BU 1806/9-2 "Information Structure and Questions in Urdu/Hindi" of the FOR 2111 *Questions at the Interfaces*. We also thank Tina Bögel and Bettina Braun for their comments and valuable discussions.

7. References

[1] R. Bhatt and V. Dayal, "Polar question particles: Hindi-Urdu *kya:*," *Natural Language and Linguistic Theory*, vol. 38, pp. 1115–1144, 2020.

[2] 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, 2008.

[3] C. Féry, "The intonation of Indian languages: An areal phenomenon," in *Festschrift for Ramakant Agnihotri*, I. Hasnain and S. Chaudhury, Eds. Akar Publishers, 2010.

[4] J. D. Harnsberger, "Towards an intonational phonology of Hindi," *Ms., University of Florida*, 1994.

[5] S. Urooj, B. Mumtaz, and S. Hussain, "Urdu intonation," *Journal of South Asian Linguistics*, vol. 10, pp. 3–22, 2020.

[6] S. Khan, "The intonation of South Asian languages," *Proceedings of FASAL*, vol. 6, pp. 23–36, 2016.

[7] D. Wunderlich, "Intonation and contrast," *Journal of Semantics*, vol. 8, pp. 239–251, 1991.

[8] B. Braun, "Phonetics and phonology of thematic contrast in German," *Language and Speech*, vol. 49, no. 4, pp. 451–493, 2006.

[9] C. Féry, *German Intonational Patterns*, ser. Linguistische Arbeiten. Niemeyer, Tübingen, 1993, vol. 285.

[10] E. Grabe, "Comparative intonational phonology: English and German," Ph.D. dissertation, Katholieke Universiteit Nijmegen, The Netherlands, 1998.

[11] S. Jannedy, "Hat patterns and double peaks: The phonetics and psycholinguistics of broad versus late narrow versus double focus intonations," Ph.D. dissertation, The Ohio State University, 2002.

[12] A. Cohen and J. Hart, "On the anatomy of intonation," *Lingua*, vol. 19, no. 1-2, pp. 177–192, 1968.

[13] W. Styler, "Using Praat for linguistic research," *University of Colorado at Boulder Phonetics Lab*, 2013.

[14] B. Mumtaz, A. Hussain, S. Hussain, A. Mahmood, R. Bhatti, M. Farooq, and S. Rauf, "Multitier annotation of Urdu speech corpus," in *Conference on Language and Technology (CLT14), Karachi, Pakistan*, 2014.

[15] Y. Xu, "ProsodyPro — a tool for large-scale systematic prosody analysis," in *Tools and Resources for the Analysis of Speech Prosody*, Aix-en-Provence, France, 2013.

[16] R. Carlson, J. Hirschberg, and M. Swerts, "Cues to upcoming swedish prosodic boundaries: Subjective judgment studies and acoustic correlates," *Speech Communication*, vol. 46, pp. 326–333, 2005.

[17] C. Männel, C. S. Schipke, and A. D. Friederici, "The role of pause as a prosodic boundary marker: Language ERP studies in German 3-and 6-year-olds," *Developmental Cognitive Neuroscience*, vol. 5, pp. 86–94, 2013.

[18] M. Butt and T. H. King, "The status of case," in *Clause Structure in South Asian Languages*, V. Dayal and A. Mahajan, Eds. Berlin: Springer Verlag, 2005.

[19] B. Mumtaz, M. Canzi, and M. Butt, "Prosody of case markers in Urdu," in *INTERSPEECH*, 2021, pp. 2661–2665.

[20] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.

[21] A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen, "lmertest package: tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, pp. 1–26, 2017.