Interspeech 2022
18-22 September 2022, Incheon, Korea

Hesitations in Urdu/Hindi: Distribution and Properties of Fillers and Silences

Farhat Jabeen[1], *Simon Betz*[1]

[1] Phonetics Workgroup, Center for Cognitive Interactive Technology (CITEC)
Faculty of Linguistics and Literary Studies, Bielefeld University, Germany
farhat.jabeen@uni-bielefeld.de, simon.betz@uni-bielefeld.de

Abstract

This research presents an analysis of hesitations in Urdu/Hindi semi-spontaneous dialogues. We annotated and analyzed twenty-five minutes of speech to investigate the frequency of hesitations and the properties of fillers as well as the formants in fillers' vocalic intervals to determine their vowel quality. We found that our participants used fillers, silences, and prolongations with varying frequency. Moreover, Urdu/Hindi speakers used the fillers with only vocalic intervals (*uh*) more frequently than the ones with vocalic intervals followed by nasals (*um*). The regression analysis showed that the *um*_type fillers were significantly longer and followed by longer silences as compared with the *uh*_type fillers. Furthermore, the *um*_types were placed more frequently at the turn medial position, whereas the *uh*_type fillers occurred at turn initial or medial position with similar frequency. The analysis of their formants showed that the vocalic intervals used in the fillers differed from other vowels in the inventory of Urdu/Hindi. Our data confirms the existing claim that *uh* and *um* are two distinct types of fillers. Our results are relevant for developing speech synthesis systems for Urdu/Hindi as well as improving the existing models seeking to incorporate hesitations and fillers in a realistic manner.

Index Terms: hesitation, fillers, silences, vowel quality, formants, turn units, dialogue, discourse, Urdu, Hindi

1. Introduction

In this study, we report the frequency of different types of hesitations in Urdu/Hindi[1] with special focus on the structure of fillers, their relationship with silences, and their position in turn units. [6] describes hesitations as "the temporary suspension of flowing speech" (p. 456), categorizing them as silent pauses, filled pauses (*uh* and *um*), elongation, and repetitions. Fillers have traditionally been seen as detrimental for communication [7]. However, recent research has highlighted the usefulness of hesitations as they have been found to allow both speakers and listeners extra time for micro-managing communication (cf. [8, 9, 10, 11, 12, 13]). Hesitations have also been proven to be useful in human-machine interaction, thus making it desirable to gain insights into their distribution in Urdu/Hindi, an under-researched language in this regard.

Various types of hesitations differ in terms of their frequency and distribution [14, 15]. Furthermore, [6] reports that the use of hesitations is speaker and context specific, advocating for the analysis of hesitations at the discourse level, arguing that conversation planning by speakers involves "timing their speaking turns to ensure optimal conditions for communicating their message" (p. 450). This alludes to the variable use of hesitations in discourse as they may be placed at different positions in turn units and utterances. The occurrence of different types of hesitations and their combinations thereof has also been found to vary between languages [11, 16, 17]. Recent analyses have shown cross-linguistic differences in the duration of silences preceding and following fillers in an utterance. [18, 17] found that there is an interplay between fillers and associated silences in English and German as silences following fillers are significantly longer compared with the preceding ones. In French, on the other hand, the silences preceding fillers have been found to be significantly longer [17].

As for the form of fillers, Clark and Fox Tree [11] have advocated for fillers to be given the status of words at par with functional words such as articles in English. They argue that *uh* and *um* follow the lexical, syntactic, phonological, and prosodic rules of the language. Moreover, fillers are used to convey meta-information and therefore carry meaning in their own right. Clark and Fox Tree further claim that *uh* and *um* are distinct and not just the variants of one filler type. This argument is supported by their finding that *uh* and *um* signal short versus long upcoming delays respectively. However, subsequent studies could not replicate these findings consistently. For example, [19, 20]'s results contradict this claim, whereas the findings of [18] confirm the pattern reported by Clark and Fox Tree. So far, the source of this variation is unclear. The current study contributes to this ongoing discussion bringing in evidence from a language belonging to a different family.

The inconsistent findings notwithstanding, the analysis of their phonetic properties may be used to verify if *uh* and *um* are two distinct filler types. [6] argues that transcriptions of these filler types are overgeneralized as they are mostly annotated with the centralized vowel in English. He provides examples from Standard British and Scottish English to show that the vowels produced in fillers are distinct in terms of their duration and stress as compared with the unstressed centralized vowel. However, [21] reports that the filler *um* in English continues to be transcribed as a schwa followed by a bilabial nasal consonant, whereas *uh* is annotated as [ɔ]. Bringing in evidence from the synthetic speech perspective, [21] shows that the vowels in fillers cannot be modeled after schwa and need their own specific F0 and duration properties for successful synthetic modeling. This finding is relevant for other languages such as German where [22] has shown that the vowel quality of fillers does not correspond to that of the other vowels found in this language's phoneme inventory.

[23] is the only existing study that refers to the presence of hesitations in Urdu in their discussion of the annotation scheme

[1] Urdu is an Indo-Aryan language spoken in Pakistan and India and associated with the Muslim religious identity. Hindi is a close relative mainly associated with Hindus in India and the diaspora. Both languages are mutually intelligible. [1] explains that these languages mainly differ in their script and formal vocabulary. The linguistic analyses of one language are generally claimed to be applicable to the other (cf. [2, 3, 4, 5] among others), hence our use of Urdu/Hindi in the current study. We use the term 'Urdu' only when the data has specifically been collected in Pakistan.

10.21437/Interspeech.2022-805

for a multi-genre corpus of Urdu speech. While they do not explain which hesitation phenomena they found in this corpus, the tag they use to represent hesitations is revealing in itself. The annotation used by [23] may be transliterated as [m:], an elongated bilabial nasal consonant. The use of this tag leads us to believe that it refers to the occurrence of fillers in this corpus. Although [23] do not present any data regarding the frequency and distribution of hesitations in Urdu, they do report that both human annotators and the automatic speech recognition system misidentified and mislabeled hesitations. This confirms the imminent need for a detailed analysis of the form and distribution of hesitations in Urdu in general and fillers in particular.

The studies presented above show that the form and distribution of hesitations and fillers differ cross-linguistically. Furthermore, their use is shown to be affected by speaker, context, and discourse based differences. Therefore, it is pertinent to analyze the use of hesitations and fillers in Urdu/Hindi, a language under-researched with regard to these phenomena.

1.1. Research questions and hypotheses

In the current study, we aim to analyze the occurrence of hesitations in Urdu/Hindi while focusing on the distribution and form of fillers. The following research questions are addressed here:

1. What kind of hesitations are found in Urdu/Hindi semi-spontaneous dialogues? (cf. 3.1)
2. If found in this type of speech:
   (a) What type of fillers are used and what is the frequency of each filler type? (cf. 3.2)
   (b) Do filler types vary in their duration? (cf. 3.2)
   (c) Do different types of fillers occur at specific positions in turn units i.e. initial, medial, final? (cf. 3.3)
   (d) What is the nature of the association between fillers and silences within an utterance? (cf. 3.4)
   (e) Do the vowels produced in fillers differ from or match with the other vowels in the phoneme inventory of Urdu/Hindi? (cf. 3.5)

We hypothesize that, as per [11]'s claim based on English, *uh* and *um* are two distinct types of fillers in Urdu/Hindi as well. Moreover, we expect an interplay between fillers and silences. It is interesting to investigate if Urdu/Hindi, an Indo-Aryan language, may be grouped with West-Germanic languages such as English and German (longer silences following fillers) or if it uses longer preceding silences as exhibited by French, a Romance language. In what follows, we present our analysis of a corpus of Urdu semi-spontaneous speech to address the questions and hypotheses outlined here.

2. Methods

2.1. Speech corpus and participants

This analysis is carried out using a corpus of interview-style semi-spontaneous speech in Urdu/Hindi. The data was collected as part of a larger research project not reported here. As the project was carried out during the Pandemic, it was not possible to meet the respondents in-person and the recordings took place using the Zoom platform.

Thirteen female Urdu/Hindi speakers were recorded for this study. Each of them was shown a form on screen inquiring about their name, age, gender, mother tongue, their age when they started learning Urdu/Hindi as well as their use of Urdu/Hindi in everyday life for personal and official communication. At the end of the questionnaire, the participants were asked to confirm that their data may be used for academic research. The questionnaire on the screen was used only to avoid any potential confusion for the participants. The questions in the form were asked by a female Urdu/Hindi speaker who also provided further explanations for those questions when asked by the participants. This communication between the interviewer and the participants resulted in semi-structured conversations based on the items in the questionnaire. However, as participants asked for explanations and clarifications, these may also be viewed as instances of spontaneous conversation between the interviewer and the interviewees. In the analysis reported below, we have used the recordings from the interviewer as well as the interviewees, resulting in a data-set of fourteen Urdu/Hindi speakers. The total duration of the interviews is twenty-five minutes.

2.2. Data analysis

The data was analyzed by the first author using Praat [24]. She annotated all instances of hesitations as well as types of fillers for each interview. In order to analyze the annotations' reliability, the second author, a non-Urdu/Hindi speaker, annotated fillers in the entire data set. Only the instances where both authors agreed in their annotation were included in the analysis.

To determine the quality of vocalic intervals produced in fillers, the first two formants were extracted using a Praat script written by Katherine Crosswhite (crosswhite at ling.rochester.edu). As all our participants were female, the maximum frequency was set to 5500Hz. Both formants were measured in the middle of each vocalic interval produced in a filler. The formant values were Lobanov-normalized to account for variation [25] based on different speakers and filler types. The resulting data was used to analyze the quality of vocalic intervals in fillers and to compare them with the vowels reported in the phonemic inventory of Urdu.

The position of fillers within turn units (beginning, middle, or end of a speaker's turn) was marked manually. The position of silences preceding and following the fillers within an utterance was identified in terms of the number of words placed between a filler and the associated silences. The silence immediately after a filler was annotated as 0 and the silence immediately preceding it was marked as -1. Thus the positive integers indicate the number of words placed to the right of the filler and the following silences, whereas the negative integers refer to the number of words to the left of the filler and the preceding silences in an utterance.

2.3. Statistical analysis

Using R [26], Linear Mixed Effects Regression models [27] were run for the log duration of different types of fillers as well as for the log duration of silences associated with those fillers. Logistic Mixed Effects Regression analysis was carried out to analyze the occurrence of fillers at different positions in turn units. Participants were added as random effects to all the models. Random slopes were retained at the α-level of 0.05.

3. Results

3.1. Frequency of hesitations

The analysis revealed the occurrence of different types of hesitations in our data. Table 1 presents the frequency of each

type along with percentages. It shows that silence was the most frequent type, followed by fillers, and lengthening of segments (consonants and vowels). In what follows, we offer a detailed analysis of the form and distribution of fillers only.

Table 1: *Frequency of different types of hesitations.*

Hesitation type	Frequency	Percentage
Silences	377	63
Fillers (*uh*, *um* etc.)	133	22
Prolongations	93	15
Total	603	

3.2. Types and duration of fillers

The fillers and their frequency is given in Table 2. It illustrates that *uh* and *um* are the most frequently found fillers in semi-spontaneous dialogues in Urdu/Hindi. Further analysis (not shown in Table 2) showed that only 37% of these fillers occurred in clusters, whereas 63% of them were produced in isolation. Moreover, when found in clusters, they frequently co-occurred with silences (79%).

Table 2: *Frequency and duration of different fillers.*

Fillers	Frequency	Percentage	Avg. duration(ms)
uh	92	69%	329
um	33	25%	582
un	1	1%	390
m:	4	3%	620
eə	3	2%	532

In existing analyses (cf. [21]), fillers are categorised into two types: vowel-only and vowel + nasal. We here refer to the former as the *uh*_type (*uh* and *eə*) and the latter as the *um*_type fillers to collectively group *um*, *un*, and *m:* together. The LMER analysis of log duration of filler types showed that the *uh*_type fillers were shorter (Est. 5.78, SE = 0.06) than the *um*_type fillers (β: 0.56, SE = 0.07, t = 7.3, $p < 0.0001$).

3.3. Position of fillers in turn units

The position of fillers in turn units is presented in Figure 1.

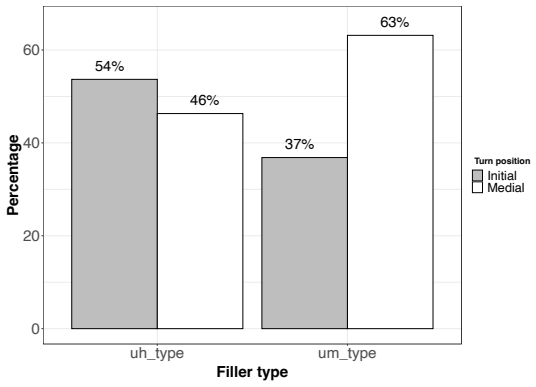

Figure 1: *Fillers' position in turn units.*

It shows that fillers were found only at the beginning and the middle of turn units and never at turn-endings in our data. Moreover, their position was affected by the type of fillers as the *uh*_types occurred with almost similar frequency at the turn initial and medial positions, whereas the *um*_type fillers were placed mostly at the turn medial position. The logistic regression analysis showed that this difference in the distribution of fillers was marginally significant (β: 0.87, SE = 0.4, z = 2.0, p = 0.04). Their position at the beginning or middle of turn units had no statistically significant effect on the duration of fillers.

3.4. Position & duration of silences

Figure 2 illustrates the distance between silences and fillers in terms of the number of words placed between them within an utterance. It shows that overall, silences occurred more often on the right side of fillers as compared to the left. Having said that, silences occurred most frequently either on the immediate right (position 0) or left (position -1) of fillers in this data.

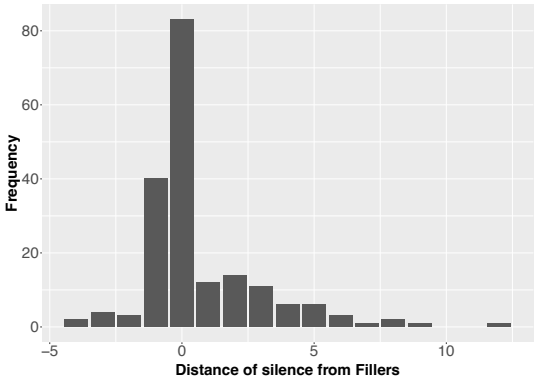

Figure 2: *Position of silences (n = 189) preceding and following fillers within an utterance.*

Figure 3 indicates that the duration of silences differs on the basis of their position with reference to fillers.

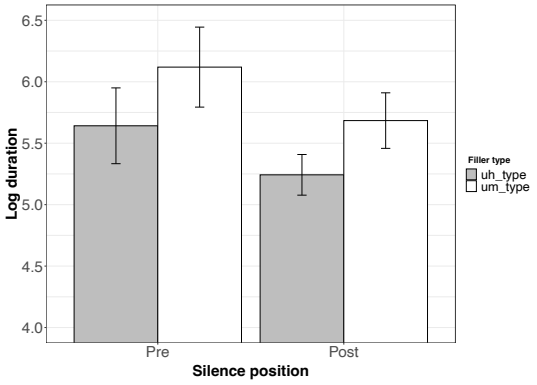

Figure 3: *Duration of silences preceding and following different filler types. The whiskers represent 95% Confidence Interval.*

The analysis of log duration showed an effect of position as the preceding silences were significantly longer (Est. 5.97, SE

= 0.19) than the silences following fillers (β: -0.45, SE = 0.13, t = -3.4, p = 0.0006). Moreover, filler type played a role in this regard as the silences associated with *uh*_type fillers were significantly shorter (Est. 5.4, SE = 0.15) than their counterparts produced in conjunction with *um*_type fillers (β: 0.40, SE = 0.11, t = 3.5, p = 0.0005).

3.5. Vowel quality of fillers

In her analysis of the vowel inventory of Urdu, [25] has offered evidence for ten distinct vowels in this language. Figure 4 illustrates the Lobanov-normalized mean F1 and F2 for these vowels as reported by [25]. It shows that Urdu has four front vowels (iː, ɪ, eː, ɛ/æ), four back vowels (ɑ, o/ɔː, ʊ, u), and two central vowels (ə, and ʌ). Figure 4 further shows that the vocalic intervals produced in fillers in our data differ both in their F1 and F2 from the other vowels in Urdu[2]. The regression analysis showed no significant effect of filler type on either the F1 or F2 of vocalic intervals, thus indicating that both the *um*_type and the *uh*_type fillers carry the same vowel.

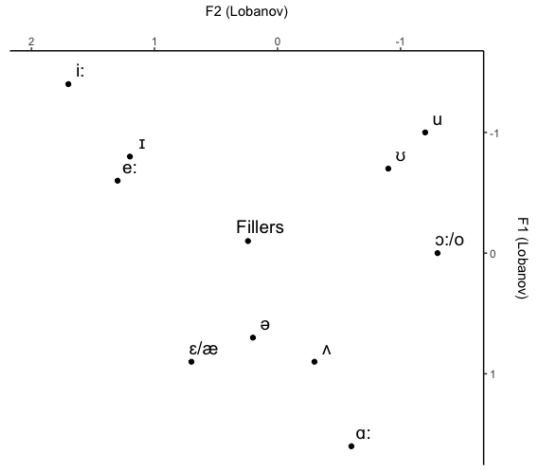

Figure 4: *Lobanov-normalized mean F1 and F2 values for the vowels produced in Urdu as reported by [25]. The vowels produced in fillers in our data are labeled as such.*

4. Discussion

In this study, we have presented the results of our preliminary research on hesitations in Urdu/Hindi. The results reported in the last section offer an interesting insight into the distribution of hesitations as well as the structure and properties of fillers in Urdu/Hindi, their association with silences, and their position in turn units. The overall frequency of hesitations in our data is in line with the findings of existing research on other languages. Silence is usually found to be the most frequent hesitation phenomenon, followed by fillers and lengthening (cf. [14, p. 215], [15]). It is pertinent to remember, however, that the frequency and distribution of hesitations vary greatly across different languages and dialects of the same language [11] and is also influenced by speaker and context dependent factors [6, p. 458]. Therefore, further research is needed to understand these aspects of different hesitation phenomena in Urdu/Hindi.

[11] had reported that the *uh*_type fillers were more frequent than the *um*_type in English. Accordingly, in our analysis of Urdu/Hindi, we found that the frequency of vowel-only *uh*_type fillers exceeded that of the vowel + nasal type *um*. We have also shown that the *um*_ and *uh*_type fillers differ in terms of their form and distribution. Predictably, the *um*_types were longer than the *uh*_type fillers. Furthermore, we found that fillers frequently occur in isolation. This is also in line with previous research as fillers have been reported to occur more frequently on their own and do not tend to form clusters in other languages [15]. Nevertheless, when they do appear in clusters, fillers are mostly grouped with silences in Urdu/Hindi. As for the interplay between these two hesitation phenomena, our data illustrated that silences that co-occur with fillers in the same utterance are most likely to be placed immediately before or after a filler. Co-occurrences with one or more words in between are rather infrequent. In terms of filler type and its influence on silence duration, we showed that Urdu/Hindi conforms to [11]'s findings on English as the *um*_types lead to longer silences than the *uh*_type fillers. This lends support to Clark and Fox Tree's claim that these are two distinct filler types. On the other hand, Urdu/Hindi differs from English (and German) with reference to the duration and position of silences relative to fillers. The longer duration of silences preceding fillers in Urdu/Hindi puts it in league with French [17].

The fact that the fillers in our data occur only at turn initial and medial positions and never at the final position may be explained as an artefact of the interview style of speech analyzed here. In non-interview styles, turn shifts may occur when a speaker produces a filler but fails to continue before the other speaker barges in. In interviews, the interviewer is likely to be more patient to wait for interviewees to continue their turn, which might explain the lack of turn-final fillers. Future research on larger data sets comprising different speech genres can shed further light on this.

The results regarding vowel quality reported here provide evidence against the prevailing annotation of fillers. The analysis of F1 and F2 illustrates that the two filler types carry the same close-mid central vowel, thus raising questions against the use of [ɔ] and [ə] for the *uh* and the *um*_type fillers respectively. It could be argued that in Urdu/Hindi, the vowel used in fillers is a distinct phoneme. However, detailed research is needed to confirm if this pattern holds true in a larger speech corpus.

5. Conclusions

The current study is the first investigation into the frequency of hesitations as well as the distribution and properties of fillers in Urdu/Hindi. Our results confirm the existing understanding of fillers and show that the *um*_type fillers differ in their properties from the *uh*_type fillers in Urdu/Hindi. There are very few analyses, to date, investigating the vowel quality of fillers and our findings make an important contribution to understanding this phenomenon. As evident by the misidentification of hesitation in the Urdu speech corpus developed by [23], our analysis of hesitations and fillers in Urdu/Hindi can help pave the way to improved speech recognition systems and acoustic modeling for this under-researched language spoken by millions of people.

[2] In this comparison of the vowel quality of fillers and Urdu vowels, we have excluded the filler with only a nasal consonant [m:] (n = 4).

6. References

[1] M. Ohala, *Aspects of Hindi Phonology*. Dordrecht: Indological publishers and booksellers, 1983.

[2] M. Butt, "Questions and information structure in Urdu/Hindi," in *Proceedings of the LFG14 Conference*, M. Butt and T. H. King, Eds. CSLI Publictions, 2014.

[3] R. Bhatt and V. Dayal, "Polar Questions and Disjunction: clues from Hindi-Urdu polar *kyaa*," 2015, talk given at the University of Texas, Arlington.

[4] M. Butt, F. Jabeen, and T. Bögel, "Verb cluster internal wh-Phrases in Urdu: Prosody, syntax and semantics/pragmatics," *Linguistic Analysis*, vol. 40, no. 3–4, pp. 445–487, 2016.

[5] M. Butt, F. Jabeen, and T. Bögel, "Ambiguity resolution via the syntax-prosody interface: The case of kya 'what' in Urdu/Hindi," in *Prosody in Syntactic Encoding*, G. Kentner and J. Kremer, Eds. De Gruyter, 2020, pp. 85–117.

[6] R. J. Lickley, "Fluency and disfluency," in *The Handbook of Speech Production*, M. A. Redford, Ed. John Wiley & Sons, Inc, 2015, pp. 445–474.

[7] K. Fischer, O. Niebuhr, E. Novák-Tót, and L. C. Jensen, "Strahlt die negative reputation von häsitationsmarkern auf ihre sprecher aus?" in *Proc. 43rd Annual Meeting of the German Acoustical Society (DAGA), Kiel, Germany*, 2017, pp. 1450–1453.

[8] W. Chafe, "Some reasons for hesitating," *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, pp. 169–180, 1980.

[9] J. Allwood, J. Nivre, and E. Ahlsén, "Speech management—on the non-written life of speech," *Nordic Journal of Linguistics*, vol. 13, no. 1, p. 3–48, 1990.

[10] J. E. Fox Tree, "The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech," *Journal of Memory and Language*, vol. 34, no. 6, pp. 709–738, 1995.

[11] H. H. Clark and J. E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, 2002.

[12] J. E. Arnold, C. L. H. Kam, and M. K. Tanenhaus, "If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 5, p. 914, 2007.

[13] S. Götz, *Fluency in native and nonnative English speech*. John Benjamins Publishing, 2013, vol. 53.

[14] R. Eklund, "Disfluency in Swedish human–human and human–machine travel booking dialogues," Ph.D. dissertation, Linköping University Electronic Press, 2004.

[15] S. Betz, P. Wagner, and D. Schlangen, "Micro-structure of disfluencies: Basics for conversational speech synthesis," in *Proceedings of the 16th Annual Conference of the International Speech Communication Association (Interspeech 2015, Dresden)*, 2015, pp. 2222–2226.

[16] L. Schettino, S. Betz, and P. Wagner, "Hesitations distribution in Italian discourse," in *DiSS 2021*, 2021.

[17] S. Betz, N. Bryhadyr, L. Kosmala, and L. Schettino, "A crosslinguistic study on the interplay of fillers and silences," *Proceedings of DiSS 2021*, 2021.

[18] S. Betz and L. Kosmala, "Fill the silence! Basics for modeling hesitation," in *Proceedings of Disfluency in Spontaneous Speech 2019*, 2019, pp. 12–14.

[19] D. C. O'Connell and S. Kowal, "Uh and um revisited: Are they interjections for signaling delay?" *Journal of Psycholinguistic Research*, vol. 34, no. 6, pp. 555–576, 2005.

[20] M. Belz and M. Klapi, "Pauses following fillers in L1 and L2 German map task dialogues," in *Sixth Workshop on Disfluency in Spontaneous Speech*, 2013.

[21] R. Dall, M. Tomalin, and M. Wester, "Synthesising filled pauses: Representation and datamixing," in *Proceedings of SSW9*, Cupertino, CA, USA, 2016, pp. 7–13.

[22] M. Belz, *Die Phonetik von äh und ähm: Akustische Variation von Füllpartikeln im Deutschen*. Springer Nature, 2021.

[23] E. Khan, S. Rauf, F. Adeeba, and S. Hussain, "A multi-genre Urdu broadcast speech recognition system," in *24th Conference of the Oriental COCOSDA International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, 2021, pp. 25–30.

[24] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer [computer program, [v. 6.0.56]," 2013, available at http://www.praat.org/ [retrieved 20.11.2017].

[25] I. Rehman, "Urdu Vowel System and Perception of English Vowels by Punjabi-Urdu Speakers," Ph.D. dissertation, University of Kent, 2019.

[26] R Core Team, *R: A language and environment for statistical computing[v. 3.6.1]*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[27] 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.