ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL

Vol.8. No.4.2025

ACOUSTIC AND DISTRIBUTIONAL PATTERNS OF FILLED PAUSES IN URDU-ENGLISH BILINGUALS' SPEECH

Saira Zahid[1,2], Muhammad Asim Mahmood[1]

[1]Department of Applied Linguistics, Government College University Faisalabad, Pakistan.

[2]Basic Sciences and Humanities, University of Engineering and Technology Lahore, Faisalabad Campus, Pakistan.

Abstract

This study investigates the production of filled pauses (FPs) in sequential bilinguals to assess whether hesitation behavior is primarily shaped by language-specific norms or stable speaker-specific habits. Eighteen female speakers of Urdu (L1) and English (L2) were recorded while performing identical spontaneous speech tasks in both languages, generating 216 minutes of data. FPs were manually annotated and analyzed according to phonetic type (vocalic-only, vocalic-nasal, nasal-only), segmental context (e.g., silence–word, word–word), utterance position (initial, medial, final, single), and acoustic properties (duration, pitch, formants, and intensity). Results revealed a higher frequency of FPs in English than in Urdu, reflecting increased planning demands in L2 speech. Both languages showed a strong preference for vocalic-only fillers, with Urdu FPs being longer and English FPs displaying lower pitch and more fronted vowel quality. Contextually, English exhibited more isolated (silence-bound) FPs, while Urdu favored embedded placements. Positional patterns showed that medial fillers dominated across both languages, though English had a higher proportion of initial and stand-alone tokens than Urdu. These findings suggest that sequential bilinguals adapt phonetic aspects of FPs to L2 norms while maintaining L1-derived structural preferences, highlighting the dual influence of language and speaker identity on disfluency production. The study contributes to bilingual fluency research and supports the development of language-specific models of disfluency.

1. Introduction:

FPs—such as *uh*, *um*, and their language-specific variants—are among the most frequent types of disfluency in spontaneous speech. They have been studied extensively across languages for their potential roles in signaling planning difficulty (Clark & Fox Tree, 2002), discourse structuring (Swerts, 1998), and turn-holding (Maclay & Osgood, 1959). While individual speakers show consistent patterns in FP use (Shriberg, 2001), a growing body of cross-linguistic research demonstrates that these pauses also reflect language-specific conventions in their frequency, form, placement, and phonetic realization (De Leeuw, 2007; Belz, 2021; Torreira et al., 2010).

Urdu, despite being spoken by over 70 million people worldwide, remains underrepresented in the study of disfluency. Prior work on Urdu FPs (Jabeen & Betz, 2022) has shown a preference for vocalic-only forms, mirroring patterns observed in Dutch (De Leeuw, 2007) and French (Torreira et al., 2010), but diverging from English and German, where vocalic-nasal forms such as *um* are more frequent (Clark & Fox Tree, 2002; Fischer et al., 2017). These language-specific tendencies suggest that FPs are not merely idiosyncratic hesitations but are shaped by the phonological and prosodic constraints of each language.

The question of how such patterns are acquired becomes especially salient in bilinguals. For sequential bilinguals—speakers who acquire an L2 after the establishment of their L1—L1 influence or transfer has been widely documented in phonetics, syntax, and prosody (Flege, 1995; Grosjean, 2001). However, relatively little is known about whether FPs are susceptible to such transfer. Do bilinguals produce L2 FPs according to the norms of that language, or do they carry over L1 patterns into L2 speech? Recent work by De Jong (2016) and Watanabe et al. (2008) suggests that hesitation behavior may be more automatic than strategic, raising the

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

possibility that L1-based FP habits could persist in L2.

To address this gap, the present study investigates the use of FPs in Urdu and English spontaneous speech by sequential bilinguals. It examines whether FPs are produced in a language-specific manner in both L1 and L2, or whether speaker-specific tendencies dominate. By analyzing both distributional and acoustic properties of vocalic-only, vocalic-nasal, and nasal-only FPs, this study provides new insight into the interaction between speaker identity, language structure, and bilingual processing.

2. Literature Review and hypothesis development:

2.1.Language-specific variation of FPs:

Research consistently highlights the language-specific nature of filled pause (FP) distribution. FPs vary across languages in both their frequency patterns and spectral characteristics (e.g., Swerts, 1998; Clark and Fox Tree, 2002; Candea et al., 2005). Most languages exhibit a preference for at least two core types of FPs: a purely vocalic form (e.g., *uh*) and a vocalic-nasal form (e.g., *um*). Within a single language, speakers often display a consistent inclination toward one form, and these preferences differ across languages. For example, Dutch speakers consistently favor *uh* over *um* across various speech styles (de Leeuw, 2007; Swerts, 1998). French data show a similar trend, with speakers producing *uh* substantially more often than *um* (Torreira et al., 2010). Urdu speakers likewise show a marked preference for *uh* (Zahid, in press). By contrast, German speakers tend to favor the vocalic-nasal form *um* over *uh* (de Leeuw, 2007; Wieling et al., 2016). A similar tendency is reported for British English speakers, who generally prefer *um* (de Leeuw, 2007). American English speakers tend to use *um* less frequently than British English speakers (Shriberg, 1994); however, in both English varieties, *um* appears more often than Dutch (Wieling et al., 2016).

Beyond their distributional patterns, FPs also exhibit language-specific variation in vowel quality. Several studies have shown that the phonetic realization of FPs differs systematically across languages and dialects (Candea et al., 2005). In English, FPs typically involve a schwa-like vowel [ə] (Maclay & Osgood, 1959; Shriberg, 1994, 2001), though regional variation is notable. American English often features a more backed and open vowel, such as [ʌ] (Vasilescu & Adda-Decker, 2007). Italian also displays regional differences, with southern speakers tending toward central vowels and central Italian speakers using more fronted ones (Giannini, 2003). In contrast, French FPs often involve a rounded vowel such as [ø] or [œ] (Candea et al., 2005). Hence FPs type and vowel quality appear to be constrained by language-specific patterns.

2.2.Speaker Specific variation of FPs

In addition to language-specific patterns, FPs are widely recognized for their speaker-specific characteristics. Within a single language or dialect, speakers display substantial individual variation in hesitation behavior, including the frequency of occurrence, the relative distribution of different hesitation types, and the phonetic quality of the sounds produced (Braun & Rosin, 2015; de Leeuw, 2007; Hughes et al., 2016; Shriberg, 1994).

Several factors contribute to the relatively low variability of FPs within individual speakers. Because FPs are often surrounded by silence (O'Connell & Kowal, 2005; Swerts, 1998), they are less affected by coarticulatory influences from adjacent segments. Another reason could be the automatic and unconscious nature of FP production (Hughes et al., 2016; Jessen, 2008), which limits speakers' conscious control and reduces potential variability—even under conditions involving voice disguise. This led to several linguists predicting that they may be transferred from the L1 and thus be language independent (e.g., Clark and Fox Tree, 2002; de Leeuw, 2007). Nonetheless, the extent of volitional control over FP production remains debated. Given their high inter-speaker variability and low intra-speaker variability, FPs hold

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

considerable value in forensic speaker comparison across languages. Their acoustic properties have demonstrated strong discriminatory potential (Cicres, 2014; Hughes et al., 2016). These findings reinforce the idea that patterns of FP use can serve as reliable markers of speaker identity.

2.3.Filled Pauses in L2 speakers

For bilingual speakers, the intersection of language- and speaker-specific characteristics raises important questions about the production of FPs across languages. Just as speakers adapt to distinct phonological and syntactic systems, they may also acquire separate, language-specific patterns of hesitation when the use and form of FPs diverge between languages. At the same time, individual speaker habits may exert a stabilizing effect, resulting in consistent hesitation behavior across both languages. Understanding how these two forces—language-specific acquisition and speaker-driven consistency—interact is the primary aim of this study.

Clark and Fox Tree (2002) suggest that within-speaker consistency extends across languages; based on observations of L2 English speakers, the authors argue that such individuals can often be recognized by their transfer of FPs from their first language (see also de Leeuw, 2007). However, empirical acoustic support for this claim remains limited due to the scarcity of cross-linguistic research on FPs. One relevant study by Lo (2020), involving 15 simultaneous German-French bilinguals, found that these speakers developed distinct FPs for each language. Another study on Dutch (L1)–English (L2) sequential bilinguals by de Boer and Heeren (2020) revealed that speakers adapted their vowel formants (F1, F2) to the language being spoken, rather than transferring them directly from L1.

Empirical studies investigating the characteristics of FPs in L2 speech remain limited, often involving small and demographically varied samples. In a study, García-Amaya and Lang (2020) examined 14 bilingual speakers of Afrikaans (L1) and Spanish (L2) living in Spanish-speaking environment. These speakers produced intermediate vowel realizations compared to L1 controls and demonstrated distinct F1 and F2 values across their two languages. They also used language-appropriate FP types, producing nasal-only fillers exclusively in Spanish. Wong and Papp (2018) analyzed the use of *uh* and *um* among 21 individuals speaking English and te reo Māori, where English was typically their L1. Speakers used *um* more frequently in English than in Māori, aligning with language-specific patterns. Their F1 and F2 values were slightly higher in English, indicating cross-linguistic shifts. Rose (2017) reported that Japanese learners of English with low proficiency did not adjust F1 and F2 in their English FPs, whereas high-proficiency speakers produced FPs closer to native norms. These findings suggest that advanced L2 speakers tend to adapt FP usage toward the target language. However, broader studies with more homogeneous participant samples in terms of age and gender are still lacking. Speakers vary preference for using *uh* or *um* (Clark & Fox Tree, 2002), and such variation in L2 input can delay the adaptation of FP realizations by second language learners. Moreover, interjections like FPs—and particularly their phonetic forms—are rarely taught explicitly in language classrooms (cf. Chen, 2009), which can further hinder acquisition. Even in the absence of formal instruction, FPs in L2 speech may differ from those in L1 due to reduced fluency. L2 speech is often less automatic (Guz, 2015) and places greater cognitive demands on the speaker (Fehringer & Fry, 2007), resulting in increased planning difficulties. As a consequence, FPs in L2 may differ in frequency, duration, and positional distribution.

Empirical evidence supports this. Guz (2015) found that L1 Polish speakers, even at high proficiency levels in English, produced more and longer FPs in L2. Similarly, De Jong et al. (2015) observed that intermediate-to-advanced Dutch L2 learners (L1: Turkish or English) showed a higher incidence of FPs in L2. However, they also found strong correlations between L1 and L2 FP rates, suggesting that hesitation behavior partly reflects individual speaking style

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

(cf. Fehringer & Fry, 2007). De Jong (2016) further noted that L2 Dutch speakers used more FPs within utterances, but not between them, which may be linked to lexical retrieval difficulties. Consistent ratios across hesitation types—such as repetitions, lexical fillers, and FPs—have also been observed in highly proficient French–German bilinguals (Fehringer & Fry, 2007), indicating possible L1 influence on L2 hesitation profiles (cf. Wiese, 1984).

One of the central aims of research on second language (L2) fluency is to gain a clearer understanding of L2 speech production processes. Increasing evidence suggests that the location of pauses—rather than their overall frequency or duration—offers more diagnostic value when assessing fluency differences across proficiency levels or between L1 and L2 speech (Davies, 2003; De Jong, 2016; Kahng, 2014, 2018; Pawley & Syder, 2000; Skehan et al., 2016). Comparative studies involving L2 learners and native speakers show that although both groups tend to pause at clause or message boundaries, L2 speakers are more likely to insert pauses within clauses or messages—and these pauses are often longer (De Jong, 2016; Kahng, 2014; Tavakoli, 2011). Such within-clause pauses are typically linked to challenges in formulation, including grammatical and lexical encoding.

Overall, although FPs may be influenced by L1 transfer, accumulating evidence suggests they can also shift in the L2 due to fluency constraints or language-specific adaptation. The current study aims to examine whether FP in L2 English by L1 Urdu speakers differ systematically from those in their L1 or whether they remain stable across languages, supporting the L1 transfer hypothesis.

Drawing on prior research highlighting both language- and speaker-specific patterns in filled pause production, this study examines how these factors intersect in bilingual speakers and what this reveals about the acquisition of such elements. Focusing on sequential bilinguals—individuals who acquire their first language (L1) from birth and begin learning a second language (L2) later, typically in academic settings—this study investigates whether and how these speakers differentiate the distribution and phonetic realization of FPs across their two languages.

2.4.The present study:

To investigate how sequential bilinguals make use of FPs in different languages, the present study analyzes speakers of Urdu and English, two languages that exhibit marked differences in FPs. Regarding vowel quality the British English FP uh has been described as a mid-central (McDougall and Duckworth, 2017) or schwa-like vowel, but longer in duration (Hughes et al., 2016; Shriberg, 2001). Hughes et al. (2016) reported that FPs of 60 Standard Southern British English (SSBE) male speakers had mean F1 values in the range of 450–700Hz, and mean F2 values of 1250–1550Hz, and were relatively stable throughout the vowel. In American English, FPs have been described as a mid-open vowel between /ʌ/ and /æ/ (Candea et al., 2005): this would mean that especially their F1 is higher than in British English. In Urdu language, FPs have been described as close-mid central vowel (Jabeen & Betz, 2023): this would mean that their F1 is lower than these two varieties.

Regarding the position, i.e., occurrence within/between phrases, languages show different patterns. In general, silent pauses are considered more typical within phrases than FPs, which tend to occur at phrase boundaries (Maclay and Osgood, 1959). de Leeuw (2007) found that the majority of FPs in British English occurred in combination with at least one silent pause, whereas FPs without any silent pauses— typical in mid-phrase position—occurred in only 15% of the cases. When de Leeuw (2007) considered the type of FPs in regard to context (surrounded with silence or words), she found that FPs surrounded by silent pauses are likely to be um.

The difference in filled pause type, placement and also their phonetic realization makes this an

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

interesting case to study whether Urdu L1 speakers shift to L2 filled pause patterns or they exert their L1 influence on L2 comprehension.

3. Corpus and Methods:

The present study examines the cross-linguistic realization of FPs (FPs) by analyzing spontaneous speech from native speakers of Urdu and their corresponding second-language English speech. It builds on the dataset reported in Zahid (2025, in press), which included L1 Urdu speech from 18 female undergraduate students at Government College University Faisalabad, aged between 18 and 25. In this study, the same participants performed an equivalent spontaneous speech task in English, enabling a within-speaker comparison across the two languages. All participants had received formal instruction in English for at least twelve years and reported no speech, hearing, or neurological disorders.

Each speaker participated in two spontaneous speech sessions, one in Urdu and the other in English. In both sessions, they engaged in dialogues lasting twelve minutes, discussing familiar topics such as daily activities, hobbies, and future plans. Recordings were made under consistent conditions in a sound-attenuated booth, using PRAAT software and a condenser microphone with a sampling rate of 48 kHz. The microphone was positioned approximately thirty centimeters from the speaker. Speakers were unaware that the focus of the study was on disfluency, which helped ensure natural speech production.

The data were manually segmented and annotated using PRAAT. FPs were identified through a combination of auditory judgment and visual inspection of the waveform and spectrogram. Each token was labeled according to its phonetic type—vowel-only, vowel-nasal, or nasal-only—as well as its immediate segmental context and utterance-level position. Context was categorized into four types based on whether the filled pause was flanked by silence (S) or speech (W): S–FP–S, W–FP–S, S–FP–W, and W–FP–W. The threshold for identifying silent pauses was set at 150 milliseconds. Utterance position was determined based on grammatical phrasing and the presence of adjacent pauses, and categorized as start, mid, end, or single, following the criteria outlined by de Boer and Heeren (2020). A second trained annotator independently coded the same data, and inter-annotator agreement was assessed using Cohen's Kappa, yielding a reliability score of 0.89. Discrepancies were resolved through discussion and adjustments to the annotation guidelines.

Acoustic measurements were taken for each filled pause using PRAAT scripts, targeting key phonetic features such as duration, pitch (F0), intensity, and vowel formants (F1 and F2). Formant values were extracted from the midpoint 50 percent of the vocalic portion of the filled pause and normalized using Lobanov's method. The maximum formant ceiling was set at 5500 Hz, appropriate for the all-female speaker sample. Both the total duration of the FP and the individual durations of its vowel and nasal components were measured.

To investigate whether the acoustic realization of FPs varied by language, utterance position, and context, linear mixed-effects models were employed. Each acoustic parameter was treated as a dependent variable in separate models. Fixed effects included language (Urdu vs. English), FP type (vowel, vowel-nasal), utterance position (start, mid, end, single), and segmental context (e.g., speech-silence transitions). Models testing language-by-position interactions were also constructed where relevant. Speaker was included as a random intercept in all models, and random slopes for position and language were added if they improved model fit, as determined by likelihood ratio testing. All statistical analyses were carried out in R, and model comparisons were based on Akaike Information Criterion and significance levels derived from Type III Wald chi-square tests.

By using within-speaker comparisons across equivalent L1 and L2 tasks, this methodology allows for a controlled investigation of whether and how the phonetic realization of FPs differs

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

between Urdu and English for the same speakers, shedding light on both language-specific and speaker-specific influences on disfluency behavior.

4. Results:

De Jong (2016) found that FPs occurring within an utterance are less common in L1 speech, and most common in L2 speech due to perhaps lexical retrieval problems. Contrary to de Jong's (2016) claim, this was the most common position for FPs to occur both in the L1 Dutch (65.3%) and the L2 English (63.8%) in de Boer and Heeren's (2020) study. Zahid (2025) also reported this true in Urdu L1 corpus where FPs occur frequently within an utterance.

4.1.Distribution of FPs:

The distribution of FPs varied considerably across speakers and languages (see fig.1). Out of the 18 speakers, 13 (72%) produced more FPs in English than in Urdu, whereas 3 speakers (17%) produced more in Urdu than in English (Speakers 10, 15, and 16), and 2 speakers (11%) showed nearly balanced use (Speakers 8 and 9). A clear predominance of FPs in English is evident in several cases, such as Speaker 18, who produced 78 FPs in English compared to 24 in Urdu, and Speaker 15, who produced 70 in English compared to 34 in Urdu. In contrast, Speaker 16 showed the strongest tendency toward Urdu, with 36 FPs in Urdu compared to only 14 in English. These results indicate that while FPs were generally more frequent in English, a small proportion of speakers favored Urdu, and a few showed no substantial difference between the two languages.

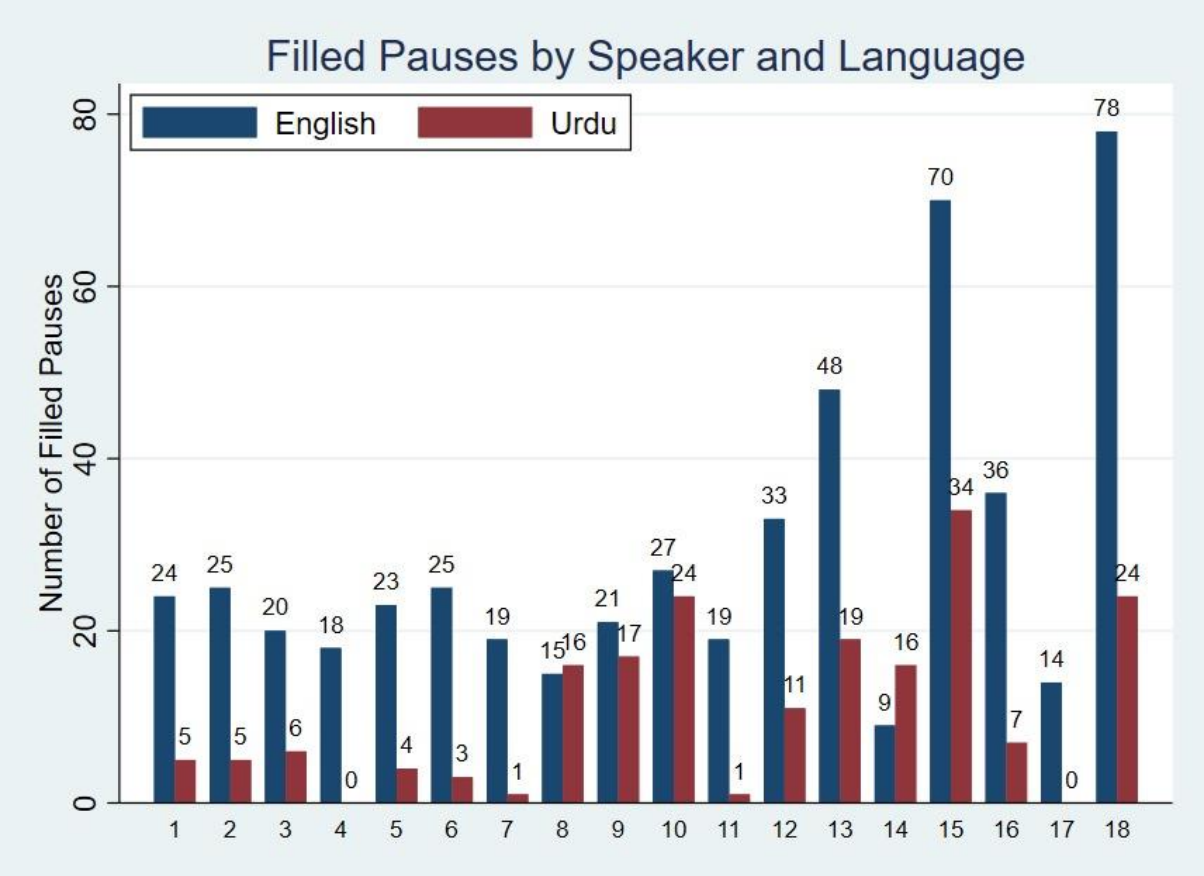

Figure 1: Number of Filled Pauses in both languages per speaker

4.2.Distribution and acoustic characteristics of FPs types:

The distribution of FPs showed a strong dominance of vocalic-only forms in both English and Urdu, though with some differences in proportions (see fig.2). In English, vocalic-only tokens

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

(v) accounted for about 83% of the total, while vocalic-nasal tokens (vn) represented around 12%, and nasal-only tokens (n) made up the remaining 5%. Urdu displayed a similar hierarchy but with slightly higher reliance on vocalic-only forms: vocalic-only tokens comprised about 87%, vocalic-nasal tokens about 10%, and nasal-only tokens just 3%. Across speakers, these patterns remained consistent, with vocalic-only tokens forming the overwhelming majority regardless of individual differences in total counts. Although English yielded higher absolute numbers overall, the proportional breakdown suggests that both languages share a comparable structural preference for vocalic-only FPs, while nasal-only forms remain marginal in both cases.

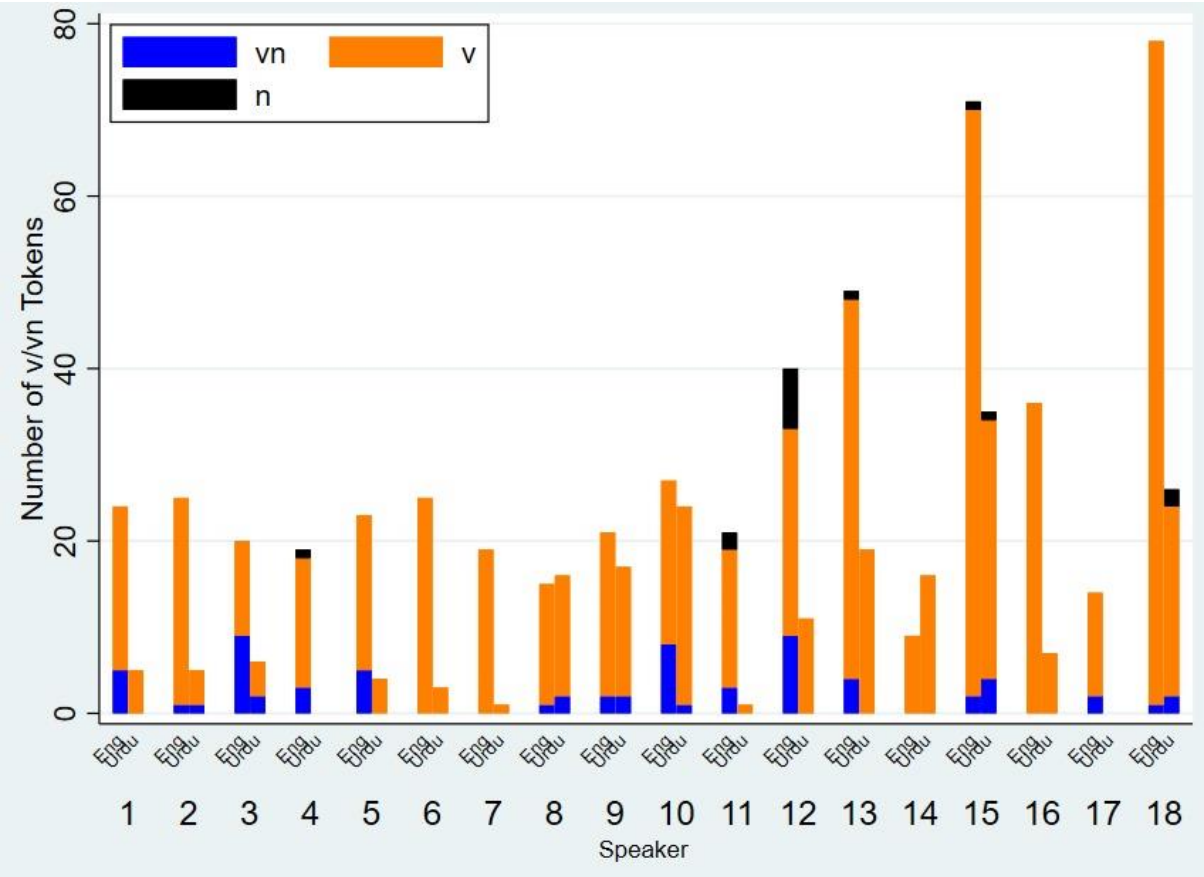

Figure. 2: Distribution of vocalic, nasal and vocalic-nasal FPs in both languages per speaker

The mixed-effects model results examining phonetic differences for vocalic FPs are presented in Table 1, with "vocalic" coded as 1 and other types as 0. For duration (logfp), the effect was non-significant in English [β = 0.039, SE = 0.071] but positive and highly significant in Urdu [β = 0.284, SE = 0.098, p < .001], indicating that speakers were more likely to produce vocalic fillers in their L1 than in their L2. For fundamental frequency (f0sem), vocalic fillers showed a significant negative coefficient in English [β = –1.143, SE = 0.351, p < .001], reflecting lower pitch for these tokens, while in Urdu the coefficient was positive but non-significant [β = 2.324, SE = 2.191], suggesting no consistent effect. Regarding first formant (F1), no significant associations were observed for either language [English: β = –0.220, SE = 0.160; Urdu: β = –0.463, SE = 0.285]. For second formant (F2), vocalic fillers showed a significant positive effect in English [β = 0.370, SE = 0.159, p < .05], indicating a more fronted vowel quality, while no significant effect was observed in Urdu [β = –0.176, SE = 0.287]. Finally, for intensity, the effects were negligible and non-significant in both languages [English: β = 0.037, SE = 0.145; Urdu: β = –0.035, SE = 0.267], showing no systematic difference in loudness levels.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table.1.Phonetic Differences in the Realization of FPs:

VARIABLES	(1) logfp_Eng	(4) logfp_Urdu	(7) f0sem_Eng	(10) f0sem_Urdu	(13) LOB_F1_Eng	(16) LOB_F1_Urdu	(19) LOB_F2_Eng	(22) LOB_F2_Urdu	(25) LOB_intensityI_Eng	(28) LOB_intensityI_Urdu
Type of FP: Vocalic	0.039	0.284***	-1.143***	2.324	-0.220	-0.463	0.370**	-0.176	0.037	-0.035
	(0.071)	(0.098)	(0.351)	(2.191)	(0.160)	(0.285)	(0.159)	(0.287)	(0.145)	(0.267)
Constant	-1.183***	-1.585***	14.057***	12.363***	0.202	0.434	-0.340**	0.164	-0.033	0.033
	(0.072)	(0.106)	(0.443)	(2.178)	(0.153)	(0.276)	(0.153)	(0.278)	(0.138)	(0.257)
Observations	510	193	508	193	498	189	498	189	508	191
Number of groups	17	16	17	16	17	14	17	14	17	14

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3.**Contextual Position of FPs:**

The distribution of FPs across contexts reveals consistent patterns within speakers when using their L1 (Urdu) and L2 (English), though with some notable shifts (see fig.3). The SW (silence–FP–word) context was the most frequent in both languages, accounting for about 42% of FPs in Urdu and 35% in English, showing that speakers strongly favored inserting fillers when initiating a word after silence. The WW (word–FP–word) context followed, contributing around 33% of pauses in Urdu and 29% in English, indicating that speakers also commonly placed FPs between words during ongoing speech. In contrast, the SS (silence–FP–silence) context was used more often in English (24%) than in Urdu (12%), suggesting that speakers produced more stand-alone hesitation markers when speaking in their L2. Finally, the WS (word–FP–silence) context was the least frequent in both languages, with nearly identical proportions (12% in English, 11% in Urdu), reflecting that FPs are rarely used at utterance endings. Taken together, these results indicate that while L1 and L2 speech share a preference for SW and WW contexts, speakers rely more on isolated fillers (SS) in their L2 than in their L1, pointing to greater hesitation or planning difficulty in English.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

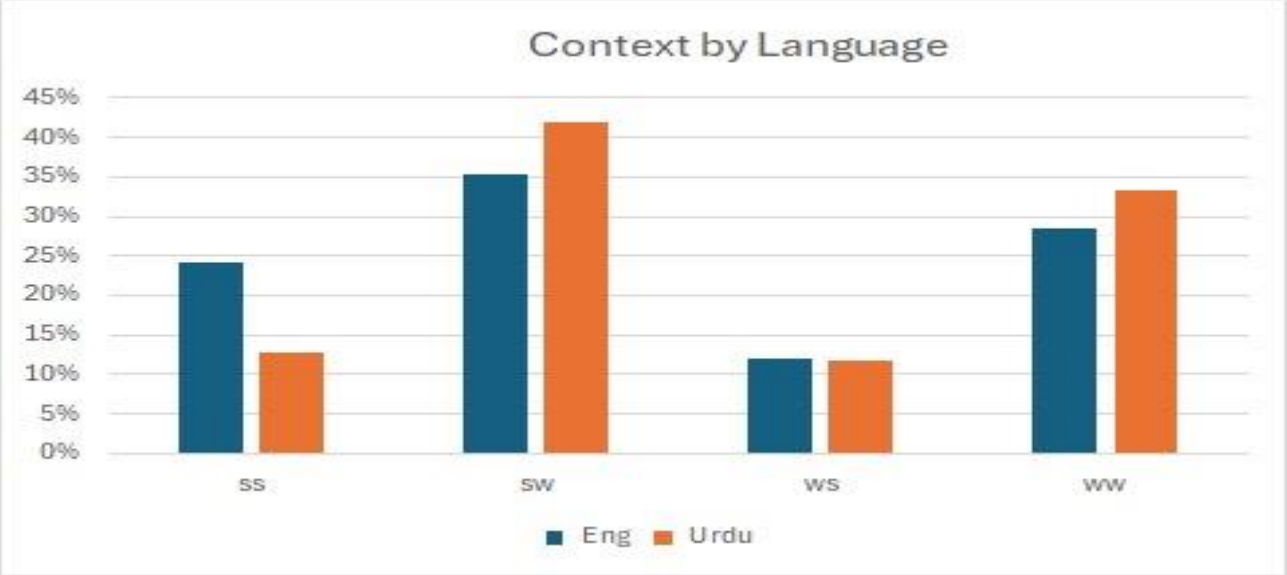

Fig.3 Distribution of FPs in four contexts (SS,WW,WS,SW)in both languages

The mixed-effects model results for contextual effects of FPs are presented in Table X, with SS (silence–FP–silence) serving as the reference category. For duration (logfp), both English and Urdu showed significantly shorter fillers in the SW context [English: β = –0.346, SE = 0.052, p < .001; Urdu: β = –0.153, SE = 0.079, p < .1] and in the WW context [English: β = –0.244, SE = 0.055, p < .001; Urdu: β = –0.212, SE = 0.082, p < .001]. The WS context did not differ significantly from SS in either language.

For fundamental frequency (f0), English showed significantly higher pitch in the SW context [β = 0.609, SE = 0.266, p < .05], whereas no significant effects emerged in Urdu. In contrast, negative but non-significant coefficients for WS and WW in Urdu indicate a tendency toward lower pitch in these contexts.

Regarding formants, English fillers in the WS context showed significantly lower F1 values [β = –0.279, SE = 0.153, p < .1], suggesting higher tongue position, while F2 increased significantly in the WW context [β = 0.509, SE = 0.118, p < .001], indicating more fronted vowels. Urdu coefficients for both F1 and F2 were not significant across contexts.

For intensity, English fillers were significantly stronger in both the SW [β = 0.185, SE = 0.112, p < .1] and WW [β = 0.466, SE = 0.117, p < .001] contexts, while Urdu did not show reliable effects across contexts.

Table.2.Phonetic Characteristics of FPs by Segmental Context in Utterance

VARIABLES	(1) logfp_Eng	(4) logfp_Urdu	(7) f0sem_Eng	(10) f0sem_Urdu	(13) LOB_F1_Eng	(16) LOB_F1_Urdu	(19) LOB_F2_Eng	(22) LOB_F2_Urdu	(25) LOB_intensityI_Eng	(28) LOB_intensityI_Urdu
Sw	-0.346***	-0.153**	0.609**	-0.503	0.006	0.039	0.146	0.017	0.185*	0.182
	(0.052)	(0.079)	(0.266)	(1.790)	(0.115)	(0.224)	(0.114)	(0.223)	(0.112)	(0.226)
Ws	-0.050	0.046	-0.444	-3.599	-0.279*	-0.426	0.199	-0.409	0.203	0.192
	(0.070)	(0.101)	(0.360)	(2.274)	(0.153)	(0.281)	(0.151)	(0.278)	(0.150)	(0.282)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Ww	-0.244***	-0.212***	0.438	-2.148	-0.183	-0.079	0.509***	0.209	0.466***	0.347
	(0.055)	(0.082)	(0.284)	(1.851)	(0.120)	(0.233)	(0.118)	(0.231)	(0.117)	(0.233)
Constant	-0.953***	-1.198***	12.719***	15.949***	0.085	0.060	-0.224**	-0.025	-0.223***	-0.215
	(0.049)	(0.083)	(0.371)	(1.629)	(0.089)	(0.198)	(0.088)	(0.197)	(0.086)	(0.199)
Observations	524	193	522	193	512	189	512	189	522	191
Number of groups	18	16	18	16	18	14	18	14	18	14

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.4.FPs' position in the utterance

The positional distribution of FPs demonstrates a strong preference for medial placement in both Urdu (L1) and English (L2). In the medial (M) position, FPs accounted for the majority of cases, comprising approximately 71% in Urdu and 57% in English, indicating that speakers most frequently inserted fillers between words within ongoing speech. The initial (I) position was the second most frequent, with 24% in English and 15% in Urdu, suggesting that speakers relied more on fillers at the beginning of utterances in their L2. The final (F) position occurred rarely and almost equally in both languages (around 8–9%), while the single (S) position—where a filled pause appeared as a stand-alone token—was least common, observed at about 7% in English and 3% in Urdu. These results highlight that while medial placement dominates across both languages, English speech contains a relatively higher proportion of fillers in initial and single positions, reflecting greater reliance on hesitation markers at utterance boundaries in the L2.

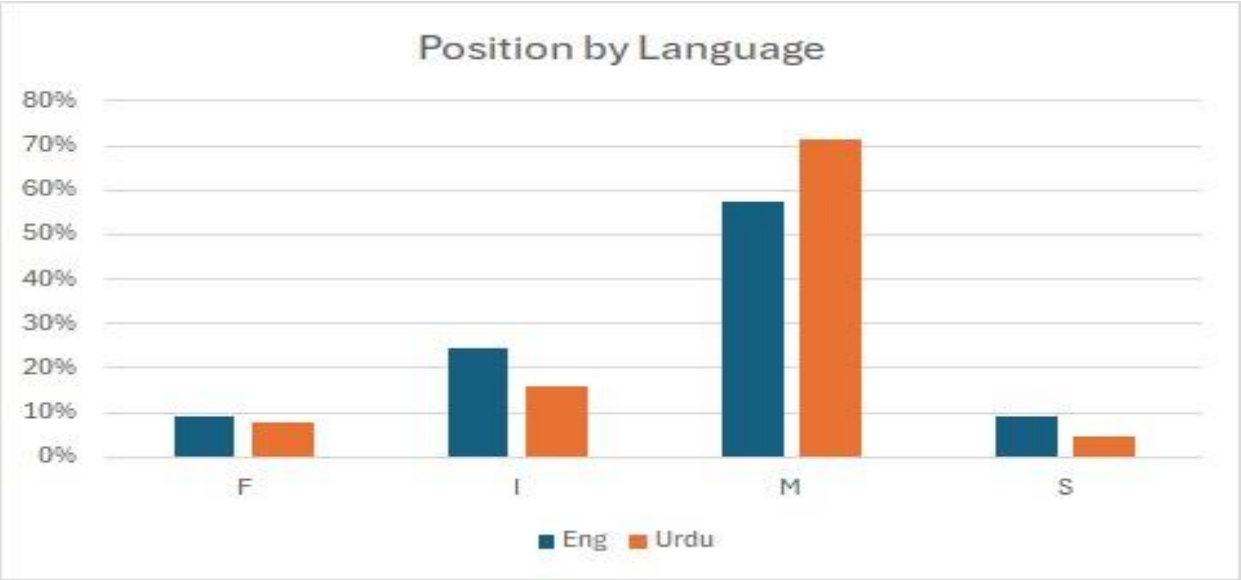

Fig.4 Distribution of FPs in four utterance positions (Final, Initial, Medial, Single) in both languages

The mixed-effects model results examining positional effects of FPs are presented in Table X,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

with the final position treated as the reference category. For duration (logfp), both English and Urdu showed shorter durations in the initial position compared to the final one, with a significant negative coefficient in English [β = –0.136, SE = 0.079, p < .1] and a non-significant effect in Urdu [β = –0.086, SE = 0.112]. A similar pattern was observed in the medial position, where durations were significantly shorter in English [β = –0.122, SE = 0.072, p < .1] but not in Urdu [β = 0.029, SE = 0.096]. In contrast, the single position showed positive but non-significant coefficients for both languages.

For fundamental frequency (f0), initial position fillers in English had a strong positive effect [β = 1.319, SE = 0.390, p < .001], indicating higher pitch relative to the final position, while Urdu showed no significant effect [β = –0.116, SE = 2.487]. Medial and single positions in English also trended toward higher f0 values but did not reach significance, while in Urdu both positions showed negative, non-significant effects.

For formants, the initial position in English was associated with a higher F1 [β = 0.320, SE = 0.166, p < .1], suggesting a lower tongue height, while the effect in Urdu was small and non-significant. Medial and single positions did not produce reliable F1 differences. For F2, a significant negative effect was observed in the single position for English [β = –0.465, SE = 0.206, p < .05], reflecting a tendency toward more back vowel quality, while effects in Urdu were non-significant across positions.

Finally, intensity differences across positions were not significant in either language, though coefficients suggested slightly greater intensity in initial and medial positions in English and lower intensity in Urdu.

Table.3. Phonetic Characteristics of FPs by Position in Utterance

VARIABLES	(1) logfp_Eng	(4) logfp_urdu	(7) f0sem_Eng	(10) f0sem_urdu	(13) LOB_F1_Eng	(16)	(19) LOB_F1_urdu	(22) LOB_F2_Eng	(25) LOB_F2_urdu	(28) LOB_intensityI_Eng	LOB_intensityI_urdu
Initial .position	-0.136**	-0.086	1.319***	-0.116	0.320*	0.116	-0.196	-0.148	0.210	-0.329	
	(0.079)	(0.112)	(0.390)	(2.487)	(0.166)	(0.299)	(0.167)	(0.300)	(0.167)	(0.302)	
Medial .position	-0.122*	0.029	0.381	-1.816	-0.029	-0.222	-0.152	-0.349	0.190	-0.273	
	(0.072)	(0.096)	(0.356)	(2.136)	(0.152)	(0.259)	(0.153)	(0.260)	(0.153)	(0.261)	
Single .position	0.111	0.171	0.559	-2.024	0.198	-0.462	-0.465**	-0.532	-0.108	-0.196	
	(0.096)	(0.150)	(0.477)	(3.323)	(0.205)	(0.416)	(0.206)	(0.418)	(0.201)	(0.420)	
Constant	-1.064***	-1.341***	12.385***	15.938***	-0.077	0.159	0.175	0.296	-0.150	0.257	
	(0.073)	(0.104)	(0.450)	(2.078)	(0.142)	(0.246)	(0.143)	(0.246)	(0.143)	(0.248)	

Observations	524	193	522	193	512	189	512	189	522	191
Number of groups	18	16	18	16	18	14	18	14	18	14

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. Discussion and Conclusion:

The present study examined the use of FPs in sequential bilinguals with Urdu as their L1 and English as their L2, focusing on frequency, distribution, contextual placement, and phonetic realization. The findings reveal both cross-linguistic consistencies and shifts that shed light on bilingual speech planning, as well as broader implications for forensic phonetics.

At the distributional level, most speakers produced more FPs in English than in Urdu, which suggests that L2 speech places higher demands on planning and lexical retrieval. This finding is consistent with accounts of increased disfluency in L2 speech due to greater cognitive load (de Jong, 2016). Still, the presence of speakers who produced more FPs in Urdu or maintained balance across languages highlights that individual speaker differences also play a role in FP production.

When comparing FP types, both languages showed a strong preference for vocalic-only fillers, with *uh*-like forms dominating over *um*-like forms. This pattern diverges from native English norms, where *um* is generally preferred (de Leeuw, 2007; Wieling et al., 2016), and reflects clear L1 transfer into L2 speech. Urdu speakers, even in their L2 English, continued to favor vowel-only fillers, using *um*-type tokens only marginally. This suggests that typological differences in hesitation markers between Urdu and English are not fully neutralized in bilingual speech, but rather speakers project their L1 habits into their L2.

Phonetic analyses highlighted further cross-linguistic differences. In English, vocalic FPs were produced with significantly lower pitch and more fronted vowel qualities (higher F2), while in Urdu they were characterized by longer durations but no consistent pitch or formant effects. This indicates that while the basic structural reliance on vocalic fillers is shared across languages, their acoustic correlates are language-specific. Speakers appear to align with Urdu's longer, central vowel fillers when speaking their L1, but in English they adapt toward more fronted and lowered-pitch realizations. These adjustments show that bilinguals employ a hybrid strategy, maintaining their L1 typological preference but fine-tuning their phonetic realizations to fit L2 articulatory expectations.

The analysis of contextual distribution revealed that both languages favored the SW (silence–FP–word) and WW (word–FP–word) contexts, consistent with the role of FPs as planning aids at utterance onsets and within speech flow. However, English speech contained a higher proportion of SS (silence–FP–silence) fillers compared to Urdu, suggesting that speakers relied more on isolated hesitation markers in their L2. This pattern reinforces the interpretation that L2 speech requires additional planning time, leading speakers to use stand-alone fillers as overt hesitation devices.

Finally, the positional analysis showed a shared preference for medial placement in both languages, but with a stronger weighting in Urdu (71% vs. 57%). English contained relatively more fillers in initial and single positions, accompanied by phonetic adjustments such as higher pitch in initial fillers and backer vowel qualities in single-position fillers. This indicates that

L2 fillers are not only more frequent but also more acoustically marked, functioning as explicit boundary signals that provide planning space in English.

Together, these results suggest that FPs should not be dismissed as mere disfluencies but considered as stable, informative markers in both bilingualism research and forensic applications. Their distributional and acoustic patterns can reveal how bilinguals manage cognitive load in L2 speech, while also offering forensic analysts reliable cues for cross-language speaker comparison.

6. References

Belz, M. (2021). *The phonetics of uh and um: Acoustic variation of filler particles in German* (p. 282). Springer Nature.

Braun, A., & Rosin, A. (2015, August). On the speaker specificity of hesitation markers. In *ICPhS*.

Candea, M., Vasilescu, I., & Adda-Decker, M. (2005). Inter- and intra-language acoustic analysis of autonomous fillers. In *Disfluency in Spontaneous Speech*.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, *59*(2), 367-409.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111.

Cicres, J. (2014). Comparación forense de voces mediante el análisis multidimensional de las pausas llenas. *Revista signos*, *47*(86), 365-384.

Davies, A. (2003). *The native speaker: Myth and reality* (Vol. 38). Multilingual Matters.

De Jong, N. H. (2016). Fluency in second language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 203–218). De Gruyter Mouton.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, Y. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*, 223–243.

De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics, 19*(2), 85–114.

Fehringer, C., & Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory.

Fischer, K., Niebuhr, O., Novák-Tót, E., & Jensen, L. C. (2017). Strahlt die negative Reputation von Häsitationsmarkern auf ihre Sprecher aus? In *Proceedings of the Annual Conference of the German Acoustical Society*.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*(1), 233-277.

García-Amaya, L., & Lang, S. (2020). Filled pauses are susceptible to cross-language phonetic influence: Evidence from Afrikaans-Spanish bilinguals. *Studies in Second Language Acquisition*, *42*(5), 1077-1105.

Giannini, A. (2003). Hesitation phenomena in spontaneous Italian. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2653–2656).

Guz, E. (2015). Establishing the fluency gap between native and non native-speech. *Research Lang. 13,* 230–247.

Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *The International Journal of Speech, Language and the Law, 23*(1), 99–132.

Jabeen, F., & Betz, S. (2022). Hesitations in Urdu/Hindi: Distribution and properties of fillers & silences. In *Interspeech (2022)* (pp. 4491–4495).

Jessen, M. (2008). Forensic phonetics. *Language and linguistics compass*, *2*(4), 671-711.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64*, 809–854.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569-591.

Lo, J. J. (2020). Between äh (m) and euh (m): The distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals. *Language and speech*, *63*(4), 746-768.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1), 19–44.

Pawley, A., & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In *Perspectives on Fluency* (pp. 163–199). University of Michigan Press.

Rose, R. L. (2017). "A comparison of form and temporal characteristics of filled pauses in L1 Japanese and L2 English," *J. Phonetic Soc. Jpn.* 21, 33–40.

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association, 31*(1), 153–169.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies* (Doctoral dissertation, University of California, Berkeley).

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 97–111.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics, 30*(4), 485–496.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*, 71–79.

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication, 52*(3), 201–212.

Vasilescu, I., & Adda-Decker, M. (2007). A cross-language study of acoustic and prosodic characteristics of vocalic hesitations. In A. Esposito, M. Bratanić, E. Keller, & M. Marinaro (Eds.), Fundamentals of verbal and nonverbal communication and the biometric issue (pp. 140–148). Amsterdam, the Netherlands: IOS Press.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50*(2), 81–94.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change, 6*(2), 199–234.

Wiese, R. (1984). Language production in foreign and native languages: Same or different. *Second language productions*, 11-25.

Wong, S. G., & Papp, V. (2018). Transferability of nonlexical hesitation markers across languages: Evidence from te reo Maori-English bilinguals. In *Proceeding of 26th IAFPA* (pp. 35-66).