ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 13, No. 7, pp. 1620-1631, July 2023
DOI: https://doi.org/10.17507/tpls.1307.03

Uh and *Um* in the Native and Non-Native Speech of Guests on a Saudi English-Language Podcast

Sahar Alkhelaiwi
Department of English and Translation, College of Sciences and Arts in Ar Rass, Qassim University, Saudi Arabia

***Abstract*—Filled pauses (FPs), *uh* and *um*, are an inherent characteristic of impromptu spoken English. However, despite the ubiquity of FP studies across languages and their impacts on speech production and comprehension, they have not been thoroughly examined in the context of second language learners of English whose mother tongue is Arabic. Hence, this study analyzed FPs in the speech of female and male non-native speakers of English and those of an American speaker who were guests on a popular English-language podcast. Combining Praat (speech analysis software), and manual coding of FPs and fillers based on previous studies, native and non-native speech was overall peppered with FPs. Although *uh* was more frequent than *um*, their frequencies among speakers and characteristic positions varied greatly. Whereas the majority comprised standalone FPs, the remaining FPs co-occurred with fillers (*and, but, so, well,* and *you know*) or were aspirated. The average length of the FPs was slightly longer for the native speaker. There were more FPs in the samples taken from early in the podcast episodes than around the middles and sometimes the endings. Regarding gender, male speakers uttered more FPs than the female speaker, whether they are native or non-native speakers.**

***Index Terms*—filled pauses, fillers, podcast, Praat, Saudi context**

I. Introduction

Levelt's (1989) speaking model postulated that people could speak automatically and successfully without deliberate control or attention. Christenfeld and Creager (1996) added that if people thought as much about each word they say as they do about what they wear, "speech would be halting indeed" (p. 459). Various researchers have attempted to define fluent speech and have considered speech "an uninterrupted flow of lexical items" based on perfectly spoken dialogues in plays, films, audio announcements, and scripted lectures (Tottie, 2016, p. 116). However, fluent speech is rare in conversations (Clark & Fox Tree, 2002). When humans speak, they convey messages not only by using words, but also by employing additional components linked to unprepared spoken language, such as fillers, false starts, hesitations, clarifications, and repetitions, which Clark and Fox Tree (2002) referred to as "performance additions" (p. 74). These can occasionally be included in communication, along with other components, such as gestures and changes in tone of voice (Corley & Stewart, 2008). As Kjellmer (2003) stated, this is because speech is characterized by the "frequent indication of hesitation or uncertainty" (p. 170). Hence, in particular, automatic speech production may be littered with filled pauses (FPs)—the focus of the current study—that "provide information about moments when speech is not being produced automatically" (Christenfeld & Creager, 1996, p. 459). Pauses are thus an inherent characteristic of impromptu spoken English (Clark & Fox Tree, 2002; Gilquin, 2008; Tottie, 2017) in both native speech (NS) and non-native speech (NNS; Gilquin, 2008), and they can also be used (primarily intentionally) in writing (Tottie, 2017).

Traditionally, as Tottie (2014) argued, pauses (filled or unfilled), repetitions, false starts, slips of the tongue, and stuttering have, deplorably, been considered "signs of disfluency" and shortcomings, but she refuted this idea, considering them instead "markers of fluency" (p. 26) and stating that we should view them positively (Götz, 2013; Kjellmer, 2003; Tottie, 2011, 2014, 2016). Götz (2013) claimed that what are often called "*dys*fluencies" (i.e., repeats, self-corrections, and FPs) contribute to "performance phenomena," supporting naturalness in the output of a speaker (p. 32). Götz (2013) explained that if performance phenomena are absent from people's speech, speakers may sound "stilted, affected, and unnatural" (p. 32). Götz (2013) argued that spontaneous speech generally imposes "high planning pressure," which tends to increase when speaking a foreign language. Consequently, "this higher planning demand is quite automatically responded to by a higher amount [number] of planning devices or performance phenomena in the learners' output" (p. 34).

Although the use of FPs in NNS is often stigmatized more than in NS (Götz, 2013), Götz (2013) argued that when dysfluencies in NNS are used in a nativelike manner in terms of their presence and distribution, this can reflect non-native speakers' proficient use of speech management strategies (or repair fluency strategies, as Foster & Tavakoli, 2009, called them). FPs, when not overused or used incorrectly, can help non-native speakers sound more natural, nativelike, and fluent (Tottie, 2014). According to Cutting (2006, p. 172; cited in Götz, 2013, p. 34), many foreign language learners are criticized because their speech "sounds like written language".

II. Background of the Study

A. FPs

FPs are usually considered temporal speech variables, among others (e.g., speech rate, mean length of runs, unfilled pauses [UPs], etc.) and have attracted great interest. In particular, regarding NS and NNS *productive fluency* (Götz, 2013), psychologists and psycholinguists have studied FPs using elicitation tests and recordings of actual conversations, and more recently, corpus linguists have studied them using existing corpora (Tottie, 2014). FPs have been examined in several languages (Kosmala & Crible, 2022), although their use and phonological forms vary significantly across languages because they are language-specific (Crible et al., 2017; de Boer & Heeren, 2020; Kosmala & Crible, 2022)—such as *eeto* in Japanese and *pues* in Spanish (Clark & Fox Tree, 2002). However, they generally consist of a central vowel followed by a nasal consonant (Clark & Fox Tree, 2002; Crible et al., 2017). Despite the variation across languages, de Boer et al. (2022) argued that the spectral (sound) properties of FPs are "highly speaker specific"; that is, "speaker specificity means not only that there is variation between different speakers but also that a feature is rather consistent in an individual's speech" (p. 1). Using Praat speech analysis software, de Boer et al. (2022) compared the vowels in FPs of the same undergraduate students in recordings more than 2.5 years apart, and the FPs (especially *uh*) did not change over time in either the students' first language (L1) (Dutch) or second language (L2) (English) speech, confirming acoustic "within-speaker consistency" for FPs across languages (especially L2, despite ongoing English learning) and the robust speaker-specific characteristics of FPs.

In the literature, a distinction is usually made between FPs and UPs (e.g., Götz, 2013) and between *silent pauses* and FPs (e.g., Gilquin, 2008; Kjellmer, 2003). Gut (2009) defined UPs as "silence or the occurrence of non-speech acoustic events such as breathing and noise" (p. 80), whereas FPs are "non-lexical fillers such as '*uh*' and '*erm*' and elongations of sounds (drawls)" (Gut, 2009, p. 80). Fischer (2006) defined FPs as *uh* and *um* that "indicate a current process: 'I am thinking'" (p. 432), and Tottie (2016) favored this definition, pointing out that it is a useful core definition of FPs. Clark and Fox Tree (2002) mentioned that FPs have been so called simply because researchers assume that they are not words—just pauses—although they are filled with sound—not silence. FPs as a concept are in "a state of flux" (Tottie, 2016, p. 99), and researchers use an array of terms: FPs, fillers, hesitators, hesitation disfluencies, hesitations, verbal blunders, dys/disfluencies, fumbles, and even non-lexical perturbations and slips (Tottie, 2014, 2016). Tottie (2014, 2016) referred to them as *planners*.

Tottie (2014, 2016) stated that the use of the vocalizations (ə[:]) or (ə[:]m), often transcribed as *er* and *erm* in British English and *uh* and *um* in American English, and are either nasalized (*um*) or non-nasalized (*uh*; Tottie, 2011). Furthermore, Clark and Fox Tree (2002) claimed that in English, *um* is more frequently used at sentence boundaries than in mid-utterance, which is a more typical place for *uh*. Schneider (2014; cited in Crible et al., 2017) also claimed that FPs occur after the first element in an utterance and at major sentence boundaries. Researchers such as Jessen (2008, p. 690) argued that FPs are used with relatively little conscious cognitive control, and Kjellmer (2003) pointed out that they are often "unintentional" and "perhaps unconscious" (p. 181). However, Clark and Fox Tree (2002) stated that FPs can be "controlled"; that is, when speakers talk, they make high-level language decisions (regarding form/informal speech, adult/baby talk, polite/locker-room language), which are followed by low-level decisions associated with words, phonology, and syntax, including utterances of *uh* and *um*. However, Tottie (2016) did not agree with this view.

(a). The Status of FPs as Words

Many researchers have debated the status of *uh* and *um* as lexical words in speech (Clark & Fox Tree, 2002; Corley & Stewart, 2008; Kosmala & Crible, 2022; O'Connell & Kowal, 2005; Tottie, 2011, 2014, 2016). Corley and Stewart (2008) argued that hesitation disfluencies (their term) are not words because speakers do not have intentional control when they utter them in the same way as they do with words. Tottie (2011), who did not agree with Corley and Stewart (2008), regarded them, when used in speech, as "marginal words," or as "nascent" or "emergent words" (Tottie, 2016, 2017). In this regard, Clark and Fox Tree (2002, pp. 75–76) explained three common types of English FPs: 1) a "*filler-as-symptom*" is an FP that is automatic, involuntary, and a mere consequence of a cognitive demand process during online speech production, but it means nothing; 2) a "*filler-as-nonlinguistic-signal*" refers to an FP used when a speaker simply wants to hold the floor without being interrupted; and 3) a "*filler-as-word*" is an FP treated as an English word, interjected precisely into speech, such as *well*, *oh*, or *say*. Clark and Fox Tree (2002) supported the latter view and also argued that FPs "satisfy the criteria for being English words. They have conventional forms and meanings, conform to the notion of word syntactically and prosodically" (p. 105). However, Kosmala and Crible (2022) indicated that FPs can be situated along a continuum between prosody and discourse markers according to their usage and function, but they are still not words because it is hard to set/confirm their meanings due to their extreme mobility until, maybe, their meanings later become fixed from repeated use, confirming their status as words.

(b). The Functions of FPs

Carney (2022) pointed out that despite researchers generally agreeing that FPs (including fillers such as *like*) do not add propositional content to utterances (and are not real words as discussed in the previous subsection), they are not meaningless and can still serve several functions for both listeners and speakers (Carney, 2022; Clark & Fox Tree, 2002; Götz, 2013; Gilquin, 2008). However, L2 listening research has yielded mixed findings about whether FPs support (e.g.,

Blau, 1991) or hinder L2 listening (e.g., Griffiths, 1991). In a recent study of Japanese undergraduate students listening to unscripted English narrative texts on video, Carney (2022) found that, as revealed by L1 recall, verbal reports, and repetitive data tasks, the FP *um* may cause trouble during L2 listening comprehension regardless of the L2 learners' proficiency (many of their participants mislistened to *um* and heard it as *am* when it occurred between *I* and *had* in an utterance). Carney (2022) attributed this misinterpretation to the position of the FP *um*, especially when it occurred *within* intonation units, making *um* perceptually ambiguous when lexical items surrounded it but not when it occurred at syntactic boundaries of utterances. Carney (2022) held that such misinterpretation could result from mistaken semantic interpretations occurring due to inaccurate phonological decoding, unfamiliarity with English FPs, or the L1 speaker's idiolect.

In speech, Kjellmer (2003) identified five main functions of FPs: hesitation; signposting of speaker turns (turntaking, turnholding, and turnyielding); attraction of listeners' attention; highlighting of semantically important points in a message for listeners; and correction. However, Tottie (2016) argued, based on data from the Santa Barbara Corpus of Spoken American English (SBC), that FPs' major function is "clearly that of helping the speaker to plan what to say next by getting a space for thinking" (p. 110); speakers particularly use them in 1) narratives, longer turns, and thoughtful presentations; 2) answers to questions when one has to think about what to say; 3) word-searches; and 4) as a turn-taking and floor-holding device (although as a minor function). Tottie (2014) also found that FPs can signal an upcoming discourse (prefacing a new paragraph in speech) or be used when rectifying an utterance, searching for a term, seeking precision, clarifying meanings, marking stances, or before talking about a touchy topic that someone is dubious about or unwilling to discuss. Kosmala and Crible (2022) claimed that FPs can have fluent and disfluent functions depending on their use context; for example, they stated that when FPs are used during lexical search or planning or when they occur mid-utterance, often isolated or clustered with a hesitation mark (e.g., a repetition or UP), this can indicate disfluency, but when they occur at initial turns, often clustered with discourse markers, they tend to improve the flow and smoothness of the discourse boundaries in speech.

B. Related Studies

A number of studies have examined the relationships between FPs and several variables that might impact their use in speech. For example, regarding FPs in NS and NNS, language proficiency is a key factor in FP production (Kosmala & Crible, 2022). Taking her data from two corpora compiled at the University of Louvain, Gilquin (2008) examined hesitation markers (FPs, silent pauses, smallwords, and miscellaneous words—her own terms) in NNS (French-speaking advanced learners of English in interviews with non-native and native hosts), comparing them with their use in NS (L1 British speakers). Her study revealed that L2 learners of English "overused" both silent pauses and FPs (except *erm*, which was used less frequently than *eh* and *er* [British transcriptions of FPs she used]) but "underused" smallwords, such as *you know* and *I mean*, when expressing hesitation (especially *like*, which was common in NS).

Similarly, although Kosmala and Crible's (2022) data showed that native French speakers uttered 103 FPs compared to the 120 FPs of American learners of French, both groups of speakers produced *euh* as a phonological form of FP rather than the nasal variant *eum*, and there were no significant differences in language proficiency; however, they attributed this phenomenon to their small corpus. More significantly, they found that FPs in NNS were significantly longer, and standalone positioning of FPs was seen only in NNS, often clustered with fluencemes (a term used to refer to phenomena such as repetitions or false starts), possibly reflecting the fragmented nature of NNS. In contrast, when comparing L1 English speakers with a group of Korean L2 English speakers, Kahng (2014) found that L2 speakers used fewer FPs than L1 speakers (although the difference was not statistically significant), and their FPs occurred in the middle of clauses "for more general planning, such as deciding content of message … and syntactic encoding" (p. 841), which often reflected processing difficulties in speech production (i.e., indicating that the positions of FPs in utterances were important).

Furthermore, in Kosmala and Crible's (2022) study on the presence of FPs in prepared and spontaneous speech, FPs occurred more frequently in long, linguistically complex utterances, when there was no contribution from the listener, and in formal situations (such as speaking in front of audience) than in spontaneous speech (conversations), which contained fewer FPs; anxiety and self-monitoring may have made speakers more self-conscious about their speech, resulting in greater use of FPs. In an experiment with undergraduate students, Christenfeld and Creager (1996) examined the relationship between FPs and anxiety and found significant differences between the low-anxiety group (told that what they said was irrelevant to their task) and the high-anxiety group (told that their speech task was important and would be evaluated), with mean averages of four FPs per minute in the former group and seven in the latter. Christenfeld and Creager (1996) concluded that not all types of anxiety increased the use of FPs, but FP use increased when the speakers paid close attention to their own verbal outputs and became more self-conscious. Christenfeld and Creager (1996) also argued that "people who make an effort to speak well may use *ums* more often than people who do not care how well they are talking" (p. 459).

Kosmala and Crible (2022) also examined the co-occurrence of discourse markers (*and*, *but*, *well*, *I mean*, and *actually*) with FPs and, interestingly, found that the pattern of discourse marker + FP was dominant in NS but used equally on the left or right of an FP in NNS; however, more FPs were followed by discourse markers in NNS, reflecting the online planning and repair processes of non-native speakers. Based on a French–English crosslinguistic study, Crible et al. (2017) examined the combination of FPs and discourse markers and found that FPs occurred 9 times per

1,000 words in English, FPs were rare in formal situations (English political speeches), 71% of FPs occurred in isolation, and most of them occurred "within clauses" than "between clauses," which the researchers argued is a negative speech characteristic and may contribute to disfluency, especially for isolated FPs. The remaining FPs were found to be clustered around discourse markers, particularly the conjunctions (*and*, *but*, and *so*), but not with language-specific interactive expressions (*well* or *I mean*; Crible et al., 2017). Tottie (2016), using SBC data and contrary to her hypothesis, found that most FPs rarely collocated with the four major identified pragmatic markers (*well*, *you know*, *I mean*, and *like*), over 70% did not co-occur with *bona fide* pragmatic markers (e.g., *and*, *but*, *so*, *right*, and *okay*). The remaining FPs were found to co-occurr with *you know* followed by *well*, but more frequently co-occurred with *and* followed by *but.* (She used "pragmatic markers" as an umbrella term for all of these words, including FPs).

Research has also yielded mixed findings about the presence of FPs in relation to formality. Some researchers, such as Swerts (1998), claimed that *uh* and *um* are typically assumed to be features of informal, impromptu conversations, but a study by Stenström and Svartvik (1994) showed that they were used more frequently in court proceedings (formal contexts) than in casual conversations. Tottie (2014), comparing the SBC with the British National Corpus (BNC), detected that the speech situation and text type as an extralinguistic context correlated with the frequency of FP use; *uh* and *um* were more frequent in non-private task-related speech situations where deliberation was essential than in private, casual talks among friends and family members. Acknowledging this as a simplistic explanation, Tottie (2014) attributed the phenomenon to speakers in formal speech situations trying to weigh their words and be more accurate and precise, which required them to plan and give more forethought to their speech; hence, they paused to think and needed more time, which made greater cognitive demands on them and made them resort to using FPs, unlike in informal speech situations where they could use more automatized collations.

Additionally, Tottie (2014) found that nationality can determine the use of FPs. In her study, British speakers produced more *uh* and *um* than American speakers, but she found no significant differences between male and female speakers' use of FPs (although the latter used slightly fewer FPs than the former). In contrast, taking her data from the BNC and the London–Lund Corpus (LLC), while acknowledging issues with the different proportions in her samples and the transcriptions of the corpora, Tottie (2011) has shown that males and females differ in their FP use; men tend to use more FPs than women. She ascertained that "gender is a powerful sociolinguistic variable: men use more fillers than women in impromptu conversation" (p. 182). Nevertheless, according to her corpus data, women used a higher proportion of nasalized *ums* than men, and in general, *uh* was more frequently used than *um*.

Taken together, despite inconsistencies in the findings and methodologies of previous studies, many variables regarding FPs, such as form and frequency, position, length, co-occurrence with discourse/pragmatic or hesitation markers, genre, register, and context, and other sociolinguistic factors, such as nationality, formality, gender, age, and socioeconomic status, were all found to determine the use of *uh* and *um*, among other variables. However, despite the ubiquity of FP studies across languages, the use of FPs has not been thoroughly examined in the context of L2 learners of English who have Arabic as their first language. One reason for this may be the lack of corpus data for this population. To fill this gap, this study sets out to describe the non-lexical status of FPs in native and non-native spoken English using data from an impromptu face-to-face Saudi English-language podcast available online. I examined FPs using different variables: 1) phonological form and frequency, 2) position characteristics, 3) concentration of FPs at various points during a speech encounter, 4) FP length, and within these, gender and nativeness.

III. Methodology

A. Corpus

The present study employed data from *The Mo Show* podcast. As stated on its website, this podcast offers a front-row seat for viewing the life and culture of Saudi Arabia, including personal stories about its citizens' and residents' achievements and experiences, and invites speakers to talk about fashion, wellness, travel, jobs, society, etc. (Islam, 2022). Given the scarcity of an English corpus for Arabic-speaking non-native learners of English, the podcast was chosen for its availability and suitability for the study. Specifically, it is the first English-language (video-recorded) podcast targeting a Saudi audience that includes both native and non-native (Arabic-speaking) speakers of English, together with natural speech involving spontaneous speech features. I selected three recently produced episodes from a larger pool of episodes (as shown in Table 1, correlated to the keys that will be used to refer to them throughout the analyses). I included a female Saudi non-native speaker of English, a male Saudi native speaker of English, and an American native speaker of English in an attempt to reflect gender and nativeness, using data from the American male as a baseline for comparison. I selected Saudi speakers who did not spend their childhood abroad and only undertook higher education in English-speaking countries (based on their online biographies); hence, I excluded non-Saudi guests and a Saudi speaker who spent most of her life abroad, since they were unlikely to represent typical L2 English speakers in Saudi Arabia. Written permission to use *The Mo Show* podcast *was* obtained in this study.

TABLE 1
THE STUDY CORPUS AND CODING FOR EPISODES

Episode No.	Guest	Key	Link
65	A social media influencer and fashion fanatic	Speaker 1	https://www.youtube.com/watch?v=HQ40kn2wN3Y&t=504s
39	An American travel vlogger	Speaker 2	https://www.youtube.com/watch?v=WzAHjrXfQBc&t=3511s
55	Minister of Sports	Speaker 3	https://www.youtube.com/watch?v=4T23VbA6xyU&t=2999s

When creating the small corpus for this study, short audio samples from the three podcast episodes were analyzed to determine the number, type, length and characteristics of FPs. I shall later discuss other factors, such as the gender of the speaker and the concentration of FPs at various points in the podcast content. Three samples of approximately 3 min each were taken from each episode. To avoid any potential effects of scripted openings/closings or awkward introductions/farewells, samples were taken from three unique parts of the episode content, as follows:

1) within the first 10 min of content (1.1, 2.1, 3.1)
2) between 15:00 and 35:00 min of content (1.2, 2.2, 3.2)
3) within the last 10 min of content (1.3, 2.3, 3.3)

This approach was chosen to combat the short length of clips and to provide a holistic view of each episode.

An additional factor in choosing clips was the need for roughly equal amounts of speech from the participants/speakers. Within the limitations listed previously, clips were chosen from conversations that presented a perceived balance between speakers. Speakers with near-equal amounts of speaking time were considered important for ensuring comparable observations of speech patterns; hence, each speaker needed to have a roughly equal opportunity to produce (or not produce) FPs within the selected exchange. Selections were initially chosen by ear, with the understanding that they would be replaced if a significant disparity in speaking time became evident. The number of perceived FPs in a selection did not impact the selection. Generally, selected clips were taken from the beginning of an utterance, although not necessarily at the beginning of a speaker's turn. Similarly, speakers were allowed to finish their utterances before the clip ended, but this did not necessarily indicate an end to the speaker's turn. Placing boundaries around utterances led to some clips being a few seconds longer or shorter than 3 min. Specific clips and durations are detailed in Table 2.

TABLE 2
DURATIONS AND TIME RANGES OF THE AUDIO SAMPLES

	Beginning	Middle	End
Speaker 1	2:22–5:25 (183 s)	21:41–24:43 (182 s)	1:10:00–1:13:04 (184 s)
Speaker 2	1:23–4:24 (181 s)	16:20–19:27 (187 s)	1:03:06–1:06:03 (177 s)
Speaker 3	1:41–4:47 (186 s)	26:14–29:13 (179 s)	48:02–51:03 (181 s)

B. Data Analysis

Clips from Table 2 were annotated in Praat speech analysis software (version 6.2.09; Boersma & Weenink, 2022) by a native English speaker and the researcher until they reached agreement on the appropriateness of the data. Each perceived FP was identified by speaker, sound, and a primary characteristic to describe its position within a larger utterance (see Figure 1). All measurements were taken in seconds and are presented herein as portions of a second to three decimal points; for example, 0.417 s can also be understood as 417 ms.

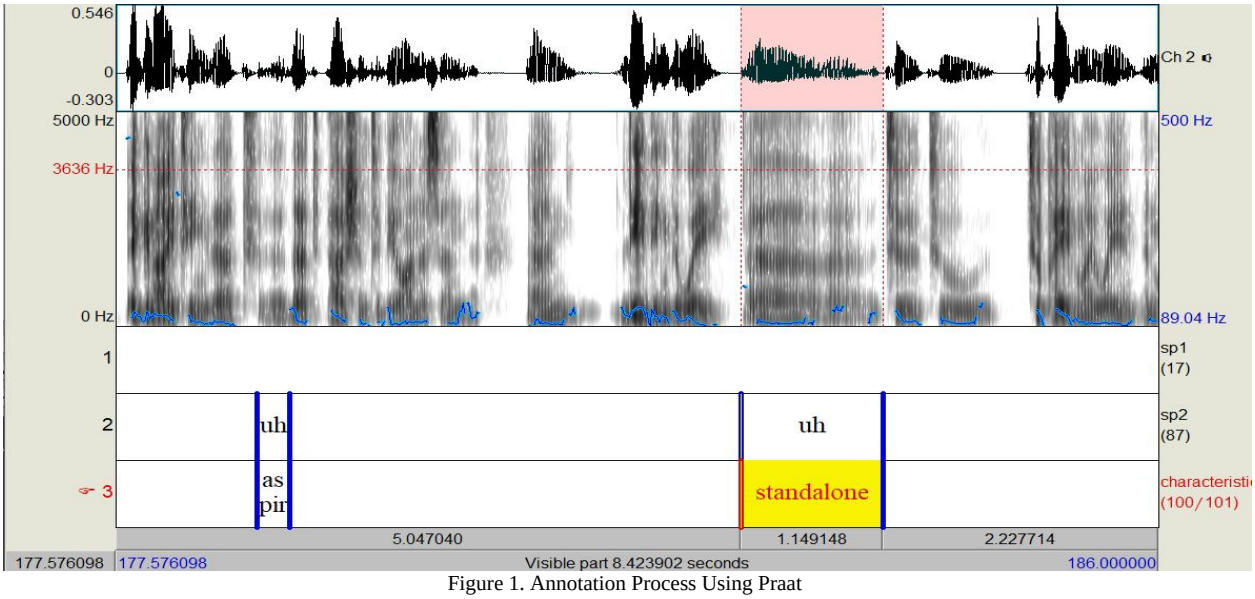

Figure 1. Annotation Process Using Praat

Based on previous research (reviewed in subsection 2.2), position characteristics were assigned based on the frequency of the environments and situations that arose during initial listening. During the review of the patterns and frequencies, the following labels emerged:

1. Standalone: An FP with a "standalone" characteristic occurring during the speaker's typical speech but not co-occurring with another filler and not impacted acoustically by surrounding phonemes.
2. Filler_: An FP occurring immediately after a filler, such as *so*, *you know*, *like*, *but*, *and*, or *well*.
3. _Filler: An FP occurring immediately before a filler, such as *so*, *you know*, *like*, *but*, *and*, or *well*.
4. Aspirated: In some FPs, especially those occurring immediately after /t_ or th_, the FP's vowel is impacted by aspiration from the previous syllable. This was notable for its potential to impact FP length.
5. Doubled: Two distinct FPs occurring one after the other, such as "*um, uh*," or "*um, um*." This only happened when one FP clearly concluded before the next began.

Notably, the term *filler* is used to refer to words such as *well*, *so*, *and*, and, *you know* (following Carney's, 2022, practice), rather than discourse/pragmatic markers or conjunctions, because *filler* seemed to be a more appropriate umbrella term. Additionally, since this study is seeking initial information about existing patterns in the use of FPs, only one characteristic was assigned to each token, even when multiple labels could have been applied, such as in the case of the utterance "*but, uh*." Depending on the situation, this could have represented either an FP immediately following a filler or an FP with environment-based aspiration. Rather than segmenting the data too finely, a hierarchy of characteristics was established to focus on broad patterns.

Essentially, characteristics were prioritized in order of uniqueness. For example, pauses labeled "doubled" could also be labeled "_filler" or "filler_," depending on their position, but the fact that they co-occurred specifically with another FP (and not a filler) indicated that they should be treated as separate and more specific. Likewise, an FP following *but* could be said to be aspirated, but its more specific characteristic was that it occurred with a filler; the designation "aspirated" was therefore reserved for FPs with aspiration carried over from non-fillers, such as in an instance of "I almost, uh," in which the aspiration carried/assimilated forward, making the "uh" sound like "tuh" to the ear.

Doubled > filler_ = _filler > aspirated > standalone

Doubled FPs took the highest priority due to the specific nature of their cooccurrences. Since there were no instances of FPs occurring between the two fillers, the designations filler_ and _filler were treated as equivalent in the hierarchy. Finally, pauses that seemed to be aspirated were documented, with the "standalone" label attached to all remaining tokens.

Using Praat, the length of an utterance was measured by placing boundaries on the corresponding speaker tier, with the pronunciation of the utterance used as a label. These pronunciations were largely determined by ear, annotated simply within Praat, and manually recoded into the international phonetic alphabet (IPA) prior to analysis to allow for a clearer discussion about the actual sound forms. To identify the frequency of each position, identical boundaries were placed on the characteristic tier. Annotations were then saved in a text file and exported into Excel for further analysis. The final raw data in Excel were similar to the sample shown in Figure 2.

	A	B	C	D	E	F	G	H	I	J
1	Clip	Speaker 2 (guest)	M/F	NS/NNS	Characteristic	start (within episode)	end (within episode)	start (within clip)	end (within clip)	length (seconds)
2	3.1	ʌm	M	NNS	_and	223.4007492	223.9263019	122.4007492	122.9263019	0.525552633
3	3.1	ʌm	M	NNS	_and	204.4466032	205.0146294	103.4466032	104.0146294	0.5680262
4	3.1	ʌm	M	NNS	_and	262.3628849	263.0431545	161.3628849	162.0431545	0.68026963
5	3.2	ʌm	M	NNS	_and	1727.127017	1574.404531	152.7224859	153.1270169	0.404530966
6	3.2	æh	M	NNS	_and	1592.697143	1574.427144	18.2699991	18.69714329	0.427144196
7	3.1	ʌh	M	NNS	_but	197.6866606	198.0726914	96.68666056	97.07269135	0.38603079
8	3.1	ʌh	M	NNS	_but	182.5291528	182.9193379	81.52915276	81.91933793	0.390185169
9	3.1	ʌm	M	NNS	_but	198.6424362	199.0495528	97.64243621	98.04955282	0.407116612
10	3.1	ʌm	M	NNS	_but	236.1110838	236.8321772	135.1110838	135.8321772	0.72109344
11	3.2	ʌm	M	NNS	_but	1717.857291	1574.659186	143.198104	143.8572905	0.659186476
12	3.3	ah	M	NNS	_but	2960.679971	2882.383467	78.29650424	78.67997142	0.383467178
13	3.3	ʌm	M	NNS	_but	3049.740991	2882.878687	166.8623039	167.7409907	0.878686829
14	3.1	ʌh	M	NNS	_so	138.8626655	139.1382093	37.86266548	38.13820931	0.27554383
15	3.1	ʌh	M	NNS	_so	231.3369401	231.9538397	130.3369401	130.9538397	0.6168996
16	3.2	ʌm	M	NNS	_so	1695.661539	1574.552251	121.1092879	121.6615392	0.552251327
17	3.2	ʌm	M	NNS	_so	1581.621865	1574.619838	7.002026762	7.621864719	0.619837958
18	3.2	æm	M	NNS	_youknow	1735.985711	1574.424865	161.5608461	161.9857107	0.424864664
19	3.1	ʌh	M	NNS	and_	163.03476	163.3295386	62.03475998	62.32953861	0.294778631
20	3.1	ʌm	M	NNS	and_	186.0716553	186.4526299	85.0716553	85.4526299	0.380974592
21	3.1	ʌh	M	NNS	and_	271.2753205	271.668505	170.2753205	170.668505	0.393184565
22	3.1	ʌh	M	NNS	aspirated	206.0731014	206.3269467	105.0731014	105.3269467	0.253845305
23	3.1	ʌh	M	NNS	aspirated	279.7154228	279.9744736	178.7154228	178.9744736	0.259050844
24	3.1	ʌh	M	NNS	aspirated	221.3733962	221.6674161	120.3733962	120.6674161	0.294019921
25	3.1	ʌh	M	NNS	aspirated	128.0024693	128.3928768	27.00246931	27.39287684	0.390407525
26	3.1	ʌh	M	NNS	aspirated	118.7720494	119.1792468	17.77204939	18.17924684	0.407197448

Figure 2. Sample of Raw Data From Excel

IV. Results and Discussion

A. FP Forms and Frequencies

The forms and frequencies of the FPs generated by the speakers are shown in Table 3.

TABLE 3
FREQUENCIES BY FORM

FP	Frequency
ʌh	84
ʌm	70
æh	9
æm	6
ah	8
am	1

Although the vowels occurring in the FPs varied slightly, the most frequently occurring ones exceeded others by a significant number. These were the sounds typically written as *uh* and *um*, and the other utterances followed a similar pattern of distribution regarding the presence of a nasal consonant. All instances of vowels other than [ʌ] were uttered by non-native speakers. The most frequently used sound was [ʌh], but the difference in frequency between it and [ʌm] was relatively narrow. However, it was worth preserving the distinct sounds of FPs, and when they collapsed into the broader categories identified in the actual analysis, they were treated as broad groups of *-h* and *-m*, reflecting the fact that the ones ending in *h* were all treated as *ʌh*, and the ones ending in *m* were all treated as *ʌm*. Hence, in total, there were 101 instances of *uh*, and 77 of *um.* Speaker 1 produced the least number of FPs (34), Speaker 2 produced 40 FPs, and Speaker 3 produced the highest number of FPs (104), totaling 178 FPs in the present study. (The subsequent quotations show these forms in use.)

1. [ʌh] Speaker 2: We didn't travel too, **uh,** broadly when I was growing up.
2. [æh, ʌm] It was, it was a, **eh**, **um**, it is the highlight of my life, I think.
3. [æm] Speaker 3: Uh, it was a shock, to be honest, but, **em** …
4. [ah] Speaker 1: **Ah**, well, my nanny visa got refused.
5. [am] … manages our competition, **am**, so that's the ladder …

B. FP Position Characteristics

The frequencies of the FP positions are shown in Table 4. It should be noted that both doubled fillers in a pair were marked as "doubled," so it is possible to consider them either as 24 tokens or 12 pairs.

TABLE 4
FREQUENCIES BY POSITION FOR ALL SPEAKERS

Label	Count
Standalone	97
_Filler	24
Doubled	24
Aspirated	20
*Filler_*	13
Total	178

The distribution of FPs by position could be compared to the hierarchy of specificity. The "standalone" category, treated as the most general, was also the most frequent by an enormous margin. However, the "aspirated" category, which was expected to be the next most frequent due to its more general nature, was only the fourth most frequent category. The "_filler" category outstripped the "filler_" category by nearly double; for example, "um, and" was more commonly used than "and, um." Finally, the "doubled" group—the most specific—included 24 tokens, although it is worth noting that they necessarily occurred in pairs, so it would be more accurate to say that 12 instances of "doubled" FPs were recorded. Taking this into account, Figure 3 illustrates the contrasting nature of specificity and frequency.

Specificity: doubled > filler_ = _filler > aspirated > standalone
Frequency: standalone > _filler > aspirated > filler_ > doubled

Figure 3. Specificity vs. Actual Frequency of FP Positions

Rather than a perfect reversal of the specificity dynamic, the three central characteristics shifted, showing that the specificity of the environment did not necessarily predict frequency. Further research with more emphasis on highly specific environments could be fruitful in determining the impact of specificity. (The following quotations show these positions in use.)

1. Standalone: 1. [ʌh] Speaker 2: We didn't travel too, **uh**, broadly when I was growing up.
2. Filler_ 4. [æm] Speaker 3: Uh, it was a shock, to be honest, but, **em**…
3. _Filler 4. [ah] Speaker 1: Ah, well, my nanny visa got refused.
4. Doubled: [æh, ʌm] It was, it was a, **eh**, **um**, it is the highlight of my life, I think.
5. Aspirated: People from different**, uh**, walks of life.

Table 5 shows the data from Table 4, separated by speaker.

TABLE 5
FREQUENCIES BY POSITION FOR EACH SPEAKER

Label	Speaker 1	Speaker 2	Speaker 3
Standalone	19	27	51
_Filler	1	6	17
Doubled	4	--	20
Aspirated	7	3	10
*Filler_*	3	4	6
Total	34	40	104

Speaker 1 was the only speaker to use "filler_" more frequently than "_filler," as shown in Table 5, but her use of both was very infrequent (a different finding from Kosmala & Crible, 2022). She also produced the fewest FPs overall, so it is difficult to say whether the pattern would hold for a larger sample of her speech.

Speaker 2 did not produce any instances of doubled FPs within the selected audio samples. Again, since he produced few fillers overall compared to Speaker 3, a longer sample of his speech might have included some instances. Speaker 2 produced the fewest aspirated FPs. This discrepancy was almost certainly related to the tendency of non-native English speakers to devoice final stops, causing aspiration at the ends of words that are not aspirated via fortis stops in a native English speaker's speech.

Speaker 3, like the others, produced the highest number of "standalone" FPs. He also had a notably high frequency of doubled FPs—ten pairs to Speaker 1's two pairs—and more than double the number of aspirated FPs used by Speaker 2.

Due to the much larger number of FPs produced by Speaker 3, it is likely that his behaviors greatly affected the overall numbers shown in Table 4. Figure 4 illustrates the frequency hierarchy for each speaker. As in Figure 3, it considers "doubled" pauses as pairs rather than individual tokens. Note that since these did not occur in Speaker 2's samples, they were not included in his hierarchy.

Speaker 1: standalone > aspirated > filler_ > doubled > _filler
Speaker 2: standalone > _filler > filler_ > aspirated
Speaker 3: standalone > _filler > aspirated = doubled > filler_

Figure 4. Frequency of FP Positions for Each Speaker

Viewed from this perspective, the only thing clear at first glance was that the "standalone" category was most common across all speakers. Further study may help in identifying other ways to narrow down this category or determine other specific environments, such as *um* appearing before a nasal initial word.

C. Concentration of FPs at Various Points in Podcast Content

Other factors considered when assessing patterns of FPs were speaker characteristics and the distribution of FPs throughout a conversation (a podcast episode). This was the purpose of taking clips from three different locations within each episode: beginning, middle, and end. Table 6 shows both the specific clip totals and the frequencies of FPs by location and speaker.

TABLE 6
FREQUENCIES BY LOCATION AND SPEAKER

	Beginning	Middle	End	Total
Speaker 1	17	5	12	34
Speaker 2	24	9	7	40
Speaker 3	43	23	38	104
Total	84	35	59	178

Speaker 3 (the male non-native English speaker) produced by far the most FPs: a total of 104 across 9.1 min of recorded conversation. Speaker 2 (native speaker of English) produced 40, and Speaker 1 (female non-native English-speaking guest) produced 34. In every case, the section taken at the beginning of the episode yielded the most tokens, with 47.19% occurring at the beginning (n = 84). The middle sections tended to include fewer tokens—19.66% of the total tokens (n = 35)—and the end segments fell somewhere in between, making up the remaining 33.14% (n = 59).

D. FP Length

From this point onward, the focus of the paper is on discussing whether these frequency patterns have any relationship with the lengths of FPs. Lengths varied greatly, even within a single speaker's measurements. Both the shortest FP, at 0.109 s, and the longest, at 1.149 s, were uttered by Speaker 3 (male non-native English speaker). The average length of all FPs generated by podcast guests was 0.418 s. The average for the two non-native speakers was similar at 0.417 s, while the native English speaker's average was slightly higher at 0.424 s. Based on these results, Tables 7 and 8 show more details of length in relation to the factors mentioned earlier in this section: FP form, position, speaker, and location.

TABLE 7
LENGTH CHARACTERISTICS BY FORM IN SECONDS

FP	Mean	Median	Shortest	Longest
ʌh	0.346	0.318	0.109	1.149
ʌm	0.530	0.514	0.169	0.927
æh	0.317	0.305	0.146	0.458
æm	0.516	0.425	0.302	0.789
ah	0.226	0.192	0.129	0.383
am	0.552	0.552	0.552	0.552

The data in Table 7 suggest differences between the relative lengths of FPs ending in nasal sounds and those not so ending. Although the longest individual token was an instance of [ʌh], the average lengths of [ʌm], [æm], and [am] were all above 0.5 s, whereas the average lengths of [ʌh], [æh], and [ah] were all below 0.35 s, partially in line with Clark and Fox Tree (2002).

Regarding length characteristics by position (see Table 8), since the "standalone" category comprised the vast majority of the tokens, a close match with the overall trends was expected; both the shortest and longest tokens produced by Speaker 3 appeared within this group. Neither position nor speaker could therefore be said to solely determine FP length.

TABLE 8
LENGTH CHARACTERISTICS BY POSITION IN SECONDS

Label	Mean	Median	Shortest	Longest
Standalone	0.419	0.400	0.108	1.149
_Filler	0.489	0.435	0.191	0.879
Doubled	0.340	0.424	0.129	0.870
Aspirated	0.374	0.371	0.117	0.927
*Filler_*	0.389	0.393	0.241	0.503

One of the most notable points according to Table 8 is that the average for doubled FPs was the lowest, because they often consisted of two distinct but rapidly produced sounds. The difference between FPs occurring before a filler and those occurring after a filler could be explained by the environment, since some fillers could induce aspiration effects when introducing an FP, such as "but_" or "and_," depending on the dialect. A finer breakdown of the _filler and filler_ categories is shown in Table 9.

TABLE 9
ENVIRONMENTS, INCLUDING FILLERS

Environment	Frequency
_and	*7*
_but	8
_so	6
_well	1
_you know	2
*and_*	6
*but_*	3
*so_*	2
*well_*	1
*you know_*	1

Tables 10, 11, and 12 display the length characteristics of each speaker, broken down according to the three locations in the audio clips.

TABLE 10
LENGTH CHARACTERISTICS BY LOCATION IN SECONDS (SPEAKER 1)

	Mean	Median	Shortest	Longest
Beginning (1.1)	0.323	0.297	0.114	0.927
Middle (1.2)	0.385	0.426	0.238	0.495
End (1.3)	0.281	0.232	0.117	0.529
Total	0.317	0.273	0.114	0.927

TABLE 11
LENGTH CHARACTERISTICS BY LOCATION IN SECONDS (SPEAKER 2)

	Mean	Median	Shortest	Longest
Beginning (2.1)	0.425	0.414	0.163	0.867
Middle (2.2)	0.413	0.468	0.167	0.556
End (2.3)	0.434	0.441	0.263	0.541
Total	0.424	0.435	0.163	0.867

TABLE 12
LENGTH CHARACTERISTICS BY LOCATION IN SECONDS (SPEAKER 3)

	Mean	Median	Shortest	Longest
Beginning (3.1)	0.426	0.403	0.149	1.149
Middle (3.2)	0.418	0.419	0.146	0.752
End (3.3)	0.495	0.527	0.109	0.879
Total	0.450	0.411	0.109	1.149

Tables 10, 11, and 12, on their own, reveal no noteworthy patterns in the speakers' use of FPs. The largest and most significant difference among the speakers was that the number of FPs produced by Speaker 3 surpassed the combined total produced by Speakers 1 and 2 across all clips. Speaker 1 (the only female speaker) generally produced shorter FPs overall. Figure 5 shows a snapshot of her speech, as seen in Praat, which documents a pair of her doubled pauses, representing the seventh and eighth shortest FPs she produced.

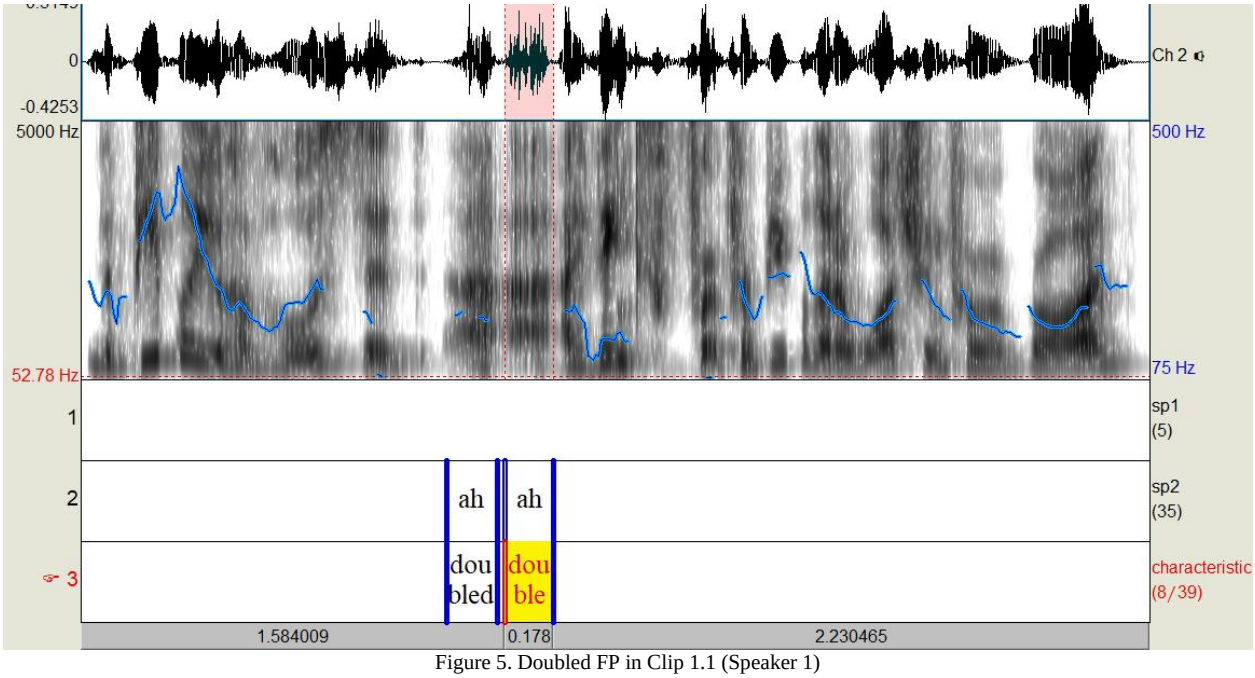

Figure 5. Doubled FP in Clip 1.1 (Speaker 1)

Further analysis of a larger sample would be necessary to draw firm conclusions about the relevance of Speaker 1's gender and non-native speaking status to her use of particular types of FPs. Since she produced fewer and generally shorter FPs than the non-native male speaker (Speaker 3) in the group, but Speaker 2's patterns corresponded more closely—at least in terms of average length—with the patterns of Speaker 3, further exploration of gender's effect on production is warranted. Although Speaker 1's lower production of FPs fell within the expectations set by previous research (Tottie, 2011, 2014), the impact of gender on the length of FPs requires further study.

Speaker 2 (male native speaker) and Speaker 3 (male non-native speaker) produced FPs with similar average lengths but vastly different frequencies. This could be due to many factors, such as words per minute and active speaking time overall in the clips, but the early data suggest the potential for further research to examine differences in native and non-native English speakers' use of FPs. To some extent, this aligns with existing research noting a higher frequency of FPs in the speech of non-native speakers (e.g., Gilquin, 2008; Kosmala & Crible, 2022). However, Speaker 1's data conflicted with this finding because she produced fewer FPs than Speaker 2 (the native speaker), corroborating Kahng (2014). Further investigation with more speakers and longer periods of speech is required to draw any conclusions about whether gender has a greater impact than native speaker status on FP production.

Regarding the position characteristics of FPs, the majority were standalone and did not co-occur with a filler/a discourse/pragmatic marker (terms used interchangeably to refer to the same phenomena)—a result that aligns with Tottie (2016) and Crible et al. (2017). Overall, words such as *well*, *you know*, *and*, *but*, and *so*, which were expected to co-occur with FPs, appeared in the data, and based on the FPs clustered with a (single) filler, they frequently collocated with conjunctions, such as *and*, *but*, and *so*, more than with *well* and *you know*, which aligns with Tottie (2016).

Overall, the bulk of the FPs for all speakers occurred at the beginning of the clips, taken from within the first 5 min of each podcast. This pattern was especially clear in Speaker 2's data, which revealed a markedly significant drop off in his use of FPs in the middle and at the end of clips. This may be related to the establishing nature of the early part of the podcast, which typically includes introductions, the host and speaker familiarizing themselves with each other's conversational patterns, and, sometimes, nervousness or self-consciousness on the part of guest speakers as they adjust to being recorded, in line with Christenfeld and Creager (1996). Based on data from this exploratory study, the middle

and end portions of similar audio sources may provide a clearer view of natural conversations. For the non-native speakers, an increase in FPs occurred again toward the end of the episode, but this was not the case for the native speaker. Like beginnings, endings include a segue into a specific style of conversation, and this may impact non-native speakers to a different degree than native speakers. Moreover, the episode was lengthy, with long speaker turns, and this may have made the endings more cognitively demanding for non-native speakers, leading them to produce more FPs (similar to Kosmala & Crible's, 2022; Tottie's, 2014, arguments). However, the phenomenon may simply have been a quirk of the clips used in this study. Regardless, additional research to specifically compare FP length during natural speech based on sampling from the middle segments of similar audio recordings would be valuable. Greater attention to changes in FP use in different types and segments of conversation would also allow for productive, continued research.

V. Conclusion

The aim of this study was to take FP research into less-researched territory by examining the FPs produced by L2 speakers of English and comparing them with those of an American speaker in a popular podcast. Overall, NNS (produced by Saudi speakers) was peppered with FPs, although *uh* was more frequent than *um*. Notably, according to the data, the Saudi speakers in this study were highly advanced English language users belonging to a high social class, which again aligns with Tottie's (2011) findings that a high level of education and socioeconomic status increase FPs in speech. This was a preliminary study; however, the findings on FPs and the different variables identified to examine the data are useful for underpinning further study. Inevitably, the study has some limitations. As already mentioned, a larger sample and a finer-grained analysis of FPs (e.g., their exact positions in relation to intonational units and their usage-based functions) are necessary to enable more confident inferences to be drawn about FPs and the different components that impact their use. Regarding the study's context, more cross-linguistic (Arabic–English) FP studies are urgently needed.

References

[1] Blau, E., (1991). More on comprehensible input: The effect of pauses and hesitation markers on listening comprehension. *Annual Meeting of the Puerto Rico Teachers of English to Speakers of Other Languages*. Retrieved November 10, 2022, from Education Resources Information Center (ERIC) database. https://eric.ed.gov/?id=ED340234

[2] Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer (version 6.2.09)* [computer program], http://www.praat.org (access 01.01.23)

[3] Carney, N. (2022). L2 comprehension of filled pauses and fillers in unscripted speech. *System*, *105*, (102726), 1–13. https://doi.org/10.1016/j.system.2022.102726

[4] Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, *70*(3), 451–460. https://doi.org/10.1037/0022-3514.70.3.451

[5] Clark, H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

[6] Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, *2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

[7] Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast*, *17*(1), 69–95. https://doi.org/10.1075/lic.17.1.04cri

[8] Cutting, J. (2006). Spoken grammar: Vague language and EAP. In *Spoken English, TESOL and applied linguistics*: *Challenges for theory and practise*, R. Hughes (ed.), (pp. 159–181). Houndmills: Palgrave Macmillan.

[9] de Boer, M. M., & Heeren, W. F. L. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *Journal of the Acoustical Society of America*, *148*(6), 3612-3622.

[10] de Boer, M. M., Quené, H., & Heeren W. F. L. (2022). Long-term within-speaker consistency of filled pauses in native and non-native speech. *JASA Express Letters*, *2*(3), 035201. https://doi.org/10.1121/10.0009598

[11] Fischer, K. (2006). Frames, constructions, and invariant meanings: The functional polysemy of discourse particles. In *Approaches to discourse particles*, K. Fischer (ed.), (pp. 427–447). Amsterdam: Elsevier.

[12] Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, *59*(4), 866–896.

[13] Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 119–149). De Gruyter Mouton.

[14] Götz, S. (2013). *Fluency in native and nonnative English speech*. John Benjamins Publishing.

[15] Griffiths, R. (1991). The paradox of comprehensible input: Hesitation phenomena in L2 teacher talk. *JALT Journal*, *13*(1), 23–38.

[16] Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.

[17] Islam, M. (Host). (2020–present). *The Mo Show Podcast* [audio and video podcast]. Retrieved November 1, 2022, from https://www.themopodcast.com/

[18] Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854. https://doi.org/10.1111/lang.12084

[19] Kjellmer, G. (2003). Hesitation. In defence of er and erm. *English Studies*, *84*(2), 170–198. https://doi.org/10.1076/enst.84.2.170.14903

[20] Kosmala, L., & Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech*, *65*(1), 216–239. https://doi.org/10.1177/00238309211010862
[21] Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671–711. https://doi.org/10.1111/j.1749-818X.2008.00066.x
[22] Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.
[23] O'Connell, D. C., & Kowal, S. (2005). *Uh* and *Um* revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research*, *34*(6), 555–576. https://doi.org/10.1007/s10936-005-9164-3
[24] Schneider, U. (2014). *Frequency, hesitations and chunks. A usage-based study of chunking in English* [Unpublished doctoral dissertation]. Albert-Ludwigs-Universität. Retrieved December 1, 2022, from http://www.freidok.uni-freiburg.de/volltexte/9793/
[25] Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9
[26] Tottie, G. (2011). *Uh* and *um* as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, *16*(2), 173–197. https://doi.org/10.1075/ijcl.16.2.02tot
[27] Tottie, G. (2014). On the use of *uh* and *um* in American English. *Functions of Language*, *21*(1), 6–29. https://doi.org/10.1075/fol.21.1.02tot
[28] Tottie, G. (2016). Planning what to say: *Uh* and *um* among the pragmatic markers. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds.), *Outside the Clause. Form and function of extra-clausal constituents*. (pp. 97–122). John Benjamins.
[29] Tottie, G. (2017). Word-search as word-formation? The case of *uh* and *um.* In *Crossing linguistic boundaries: Systemic, synchronic and diachronic variation in English*, P. Núñez-Pertejo, M. J. López-Couso, B. Méndez-Naya, J. Pérez-Guerra (ed.). Bloomsbury Publishing.

Sahar Alkhelaiwi works as an assistant professor in the Department of English & Translation at Qassim University, Saudi Arabia. She got her PhD in Linguistics from the University of Lancaster, United Kingdom. Her research interests include second language listening comprehension, English for Academic Purposes, needs analysis for curriculum development and cognitive linguistics.