

FILLED PAUSES IN STUDENT CONVERSATIONS: AN EXPLORATORY MIXED-METHODS SOCIOPHONETIC ANALYSIS

Josan C. Fermano¹ and Charito Ong²

^{1,2}Department of Teaching Languages, University of Science and Technology of Southern Philippines, Cagayan de Oro City, Philippines

Corresponding author: Charito.Ong@ustp.edu.ph

ABSTRACT. Filled pauses are often evaluated as speech defects, although they can also signal planning difficulty, maintain a turn, and organize discourse. We examined the distribution, communicative functions, and acoustic characteristics of filled pauses in the conversations of 250 second-year undergraduate students using an exploratory sequential mixed-methods sociophonetic design. Eight small-group Zoom sessions produced eight hours of speech and 58,990 transcribed words. The transcripts contained 2,480 filled pauses: 1,500 uhm tokens (60.48%) and 980 uh tokens (39.52%). Qualitative analysis indicated that filler use varied with communicative formality, age, gender, sociocultural background, and English proficiency. More proficient speakers used fillers to sustain continuity and hold the floor, whereas some less proficient speakers used silence or shifted to a first language when planning demands increased. In the acoustic subsamples, uhm had a higher mean fundamental frequency and intensity than uh for both male and female speakers. Five measured tokens ranged from 0.28 to 0.50 seconds; mean duration was 0.43 seconds for uhm and 0.33 seconds for uh. Because the acoustic subsamples were small and complete token-level values were unavailable for all measures, the differences are descriptive. The findings show that filled pauses function as context-dependent resources in multilingual student conversation and cannot be interpreted solely as evidence of communicative deficiency.

Key Words: acoustic analysis, filled pauses, multilingual speech, Philippine English, sociophonetics, student conversation

INTRODUCTION

Filled pauses are vocalizations produced during disruptions or anticipated delays in spontaneous speech. English uh and um have traditionally been described as hesitation phenomena, but corpus evidence indicates that speakers use them to announce different degrees of upcoming delay and to manage word search, turn continuation, and formulation [1]. Reviews of speech production and comprehension similarly show that fillers are associated with planning difficulty while remaining informative to listeners [2]. The presence of uh can facilitate recognition of upcoming speech, indicating that listeners do not simply disregard fillers as noise [3].

A sociophonetic perspective is important because the function of a filled pause is partly reflected in its position and acoustic realization. Filled pauses cluster around discourse boundaries, and their duration, segmental form, and prosody may vary according to discourse structure [4]. Their distribution also differs across registers and social groups. Corpus studies have identified variation associated with age, gender, role, topic, relationship, and interactional context [5], [6], while naturalistic language samples show that filler use may index demographic and interpersonal differences without serving as a direct measure of competence [7].

The interpretation of fillers becomes more complex in multilingual and second-language speech. Filled pauses may serve as nonlexical planning devices, co-occur with discourse markers, or alternate with lexical resources from another language. Their frequency and function vary across spontaneous and prepared genres and across proficiency levels [8]. Hesitation markers also have language-specific phonetic forms [9], and combinations of filled pauses and discourse markers contribute differently to fluent and disfluent stretches of speech [10]. Research on learner English further indicates that proficiency operates together with contextual and learning-related factors [11]. Consequently, filler frequency alone cannot establish stronger or weaker communicative ability.

In classroom speaking activities, fillers are nevertheless often treated categorically as undesirable speech habits. Such

evaluation may overlook their interactional value and may disadvantage multilingual students who use a first language, an English filler, or silence as alternative planning resources. At the same time, dense or poorly placed fillers can interfere with comprehensibility. Research has not sufficiently explained how filler frequency, communicative function, acoustic realization, and sociocultural interpretation converge in Philippine student conversation. Much of the established evidence comes from native-speaker corpora or European languages, with limited sociophonetic work integrating acoustic measures and speakers' reported experiences in multilingual Philippine educational settings.

We therefore examined the distribution of uh and uhm in student conversations, the social and contextual patterns associated with their use, their acoustic characteristics in terms of fundamental frequency, intensity, and duration, and the way qualitative and quantitative evidence jointly explained their planning, discourse, and interpersonal functions. The analysis was informed by sociolinguistic variation as socially structured practice [12] and by established principles of acoustic phonetics [13].

MATERIALS AND METHODS

Research Design

The researchers employed an exploratory sequential mixed-methods sociophonetic design [14]. Naturalistic conversations and semi-structured interviews were first examined to identify filler forms, interactional positions, social contexts, and participants' interpretations. The identified filler tokens were then subjected to acoustic analysis. We integrated the qualitative and acoustic strands through narrative comparison by examining whether the observed distributional and acoustic patterns converged with the communicative functions described in the transcripts and interviews.

Participants and Speech Corpus

The participants were 250 second-year undergraduate students selected through purposive sampling. The sample comprised 95 males (38%) and 155 females (62%). Ten participants were 19 years old, 239 were 20 years old, and

one was 21 years old, resulting in a mean age of 19.96 years ($SD = 0.21$). Their first languages included Cebuano, Maranao, and Tagalog, and they used English as a second language. English proficiency was described using a department-designed teacher-made examination. Scores ranged from 88 to 95, with a mean of 92.00; these scores were interpreted only within the study sample and were not treated as equivalent to a standardized international proficiency measure.

Participants were included when they were full-time second-year students, spoke a local first language and used English as a second language, had completed the teacher-made examination, and consented to audio recording. Students who reported a speech or hearing condition affecting acoustic measurement, identified English as their native language, or produced recordings with severe background noise were excluded from detailed acoustic analysis.

Table 1. Participant and corpus profile

Item	Study information
Participants	250 second-year students
Sex	95 male; 155 female
Age	19–21 years; $M = 19.96$, $SD = 0.21$
First languages	Cebuano, Maranao, Tagalog
English proficiency	88–95; $M = 92.00$
Recording sessions	8 sessions; about 60 min each
Corpus duration	480 min / 8 h
Corpus size	58,990 transcribed words
Filled pauses	2,480 tokens

Speech Elicitation and Recording

The students participated in small-group discussions about familiar and relatively spontaneous topics, including academic experiences, daily routines, and communicative practices. Eight Zoom sessions were conducted, each lasting approximately 60 minutes. The resulting corpus contained approximately 480 minutes, or eight hours, of recorded speech and 58,990 transcribed words. Recordings were captured through Zoom using built-in digital microphones at an approximate desktop distance of 30–50 cm. The audio files were imported into Praat for acoustic analysis [16]. Because the recordings were collected through conferencing software, microphone quality, speaker-to-microphone distance, internet transmission, and automatic audio processing may have affected the acoustic measurements.

Transcription and Qualitative Coding

The recordings were transcribed orthographically. Filled pauses, repetitions, false starts, and relevant instances of code-switching were retained verbatim, while unclear portions were marked as [inaudible]. The filler analysis focused on the nonlexical forms *uh* and *uhm*. The transcripts and corresponding audio were reviewed to identify these tokens. Tokens affected by overlapping speech, severe noise, clipping, or insufficient acoustic clarity were excluded from detailed phonetic measurement.

Conversation and interview transcripts were analyzed using a codebook thematic approach informed by the research questions and refined inductively [15]. Initial codes

represented planning difficulty, lexical search, turn-holding, stance, politeness, social context, proficiency, and perceived communicative effectiveness. Related codes were organized into higher-order themes and checked against the audio to preserve prosodic and sequential context.

Acoustic Measurement and Data Analysis

A purposive acoustic subsample was selected from the complete filler corpus to represent both filler types and male and female speakers, subject to the availability of clear audio. The final subsample consisted of 35 tokens for fundamental-frequency analysis, 32 tokens for intensity analysis, and five tokens for duration analysis. Fundamental frequency (f_0) was measured in hertz, intensity in decibels, and duration in seconds. Initial pitch ranges were 95–400 Hz for male speakers and 250–650 Hz for female speakers. Pitch contours were visually inspected, and speaker-specific adjustments were made when necessary to reduce octave-tracking errors.

We computed frequencies, percentages, means, ranges, standard deviations, and confidence intervals when the required individual measurements were available. Complete token-level values were not available for all pitch and intensity groups; consequently, inferential comparisons were not conducted. The acoustic findings were interpreted descriptively and were integrated with the qualitative patterns by examining convergence, complementarity, and dissonance [17].

Ethical Considerations

Participation was voluntary. Informed consent covered audio recording, transcription, anonymized reporting, and acoustic analysis. The study treated recordings and transcripts as confidential research data and followed accepted principles for research involving human participants [18].

RESULTS

Distribution of Filled Pauses

The corpus contained 2,480 filled pauses. *Uhm* was the more frequent form, accounting for 1,500 tokens (60.48%), while *uh* accounted for 980 tokens (39.52%). Thus, almost three of every five identified fillers were *uhm*.

Table 2. Frequency and percentage of filled pauses

Filler	Frequency	Percentage
Uh	980	39.52%
Uhm	1,500	60.48%
Total	2,480	100%

Sociolinguistic and Contextual Patterns

Qualitative analysis indicated that filler use varied across speakers and communicative situations. Age-related variation was reported, with younger participants described as using fewer fillers, although the narrow age range and absence of subgroup rates limited this interpretation. Gender-related differences were also observed, but the available results did not establish a specific or statistically reliable direction of difference.

Communicative formality influenced the visibility and placement of fillers. Less formal exchanges contained more

noticeable filler use, whereas formal settings appeared to encourage greater self-monitoring. Sociocultural background was associated with variation in filler forms and judgments of appropriateness, indicating that filler use was situated within multilingual interaction rather than determined by a fixed cultural characteristic.

English proficiency was associated with the planning resource selected by speakers. More proficient speakers used fillers to preserve continuity, retain the floor, and organize an upcoming explanation. In some episodes, less proficient speakers paused silently or shifted to a first language when formulation demands increased. The pattern showed that fillers, silence, and code-switching served as alternative resources for managing real-time speech production.

Fundamental Frequency

The fundamental-frequency subsample comprised 35 tokens. Mean f_0 was higher for uhm than for uh in both speaker groups. Among male speakers, the mean was 326.67 Hz for uh and 459.88 Hz for uhm. Among female speakers, the mean was 490.89 Hz for uh and 545.90 Hz for uhm. Female speakers also displayed higher mean f_0 values than male speakers for both filler forms. These results describe the selected clear tokens and do not represent inferential group differences.

Table 3. Mean fundamental frequency by speaker group and filler

Sex	Filler	n	Mean f_0 (Hz)
Male	Uh	12	326.67
Male	Uhm	8	459.88
Female	Uh	10	490.89
Female	Uhm	5	545.90

Intensity

The intensity analysis included 32 tokens. Mean intensity was 88.90 dB for male uh, 96.70 dB for male uhm, 92.30 dB for female uh, and 99.98 dB for female uhm. The weighted mean across the four groups was 93.54 dB. In both male and female subsamples, uhm had higher mean intensity than uh. Because the individual measurements were not available for all groups, dispersion statistics and inferential comparisons were not computed.

Table 4. Mean intensity by speaker group and filler

Sex	Filler	n	Mean intensity (dB)
Male	Uh	10	88.90
Male	Uhm	8	96.70
Female	Uh	9	92.30
Female	Uhm	5	99.98

Duration

Five filler tokens were available for duration analysis. Durations ranged from 0.28 to 0.50 seconds, with an overall mean of 0.37 seconds ($SD = 0.08$), 95% CI [0.26, 0.47]. The three uh tokens had a mean duration of 0.33 seconds, whereas the two uhm tokens had a mean duration of 0.43 seconds. The small subsample made these values preliminary descriptive observations.

Table 5. Duration of the measured filler tokens

Token	Filler	Duration (s)
1	Uh	0.32

Token	Filler	Duration (s)
2	Uhm	0.35
3	Uh	0.28
4	Uhm	0.50
5	Uh	0.38

Integration of Qualitative and Acoustic Findings

The qualitative and acoustic evidence converged around the interpretation of filled pauses as planning and interactional resources. Fillers occurred during formulation difficulty, yet they also enabled speakers to sustain continuity and hold a conversational turn. Their placement changed with formality and audience, supporting the view that filler behavior was style-sensitive rather than a stable personal defect. The higher mean f_0 and intensity of uhm within both speaker groups, together with its greater frequency and slightly longer duration in the small duration subsample, indicated that the two forms were not acoustically interchangeable in the selected tokens. In multilingual episodes, code-switching and silence complemented filled pauses as alternative strategies for managing speech planning.

DISCUSSION

Filled Pauses as Context-Dependent Resources

The principal finding is that filled pauses cannot be interpreted uniformly as errors. The planning, turn-holding, and continuity functions identified in the conversations are consistent with the proposal that uh and um signal an upcoming delay [1] and with evidence that listeners use disfluency information during comprehension [2], [3]. A filler may reflect production difficulty while simultaneously providing useful information to an interlocutor. Its communicative value therefore depends on its sequential position, frequency, and prosodic realization.

Social Variation and Multilingual Proficiency

The observed variation across age, gender, formality, and sociocultural background broadly corresponds with corpus studies showing that filler rates change with speaker characteristics, topic, relationship, role, and register [5]–[7]. The present results, however, were descriptive. The age distribution was highly concentrated at 20 years, gender groups were unequal, and token-normalized subgroup rates were unavailable. The patterns are therefore more appropriately interpreted as contextual tendencies within the corpus than as stable characteristics of Philippine student groups.

The association between proficiency and filler function also requires a functional interpretation. More proficient speakers used fillers strategically to maintain continuity and organize speech, consistent with research showing that fillers vary by proficiency, genre, and discourse function [8, 11]. A high filler count does not necessarily indicate fluency, and a low count does not necessarily indicate proficiency. Placement, clustering, repair success, speech rate, and comprehensibility remain important. Similarly, a shift to Cebuano, Tagalog, Maranao, or another familiar language may represent purposeful multilingual resource use rather than failed English retrieval.

Acoustic Characteristics

The selected uhm tokens had higher mean f0 and intensity than uh for both male and female speakers, and the small duration subsample also showed a longer mean for uhm. These directions are compatible with evidence that fillers near discourse boundaries may differ segmentally and prosodically from phrase-internal fillers [4]. Higher f0 or intensity may reflect local prominence, turn competition, phrase position, or the onset of repair, while longer duration may reflect greater planning demand or the additional nasal segment in uhm. The purposive subsamples and variable Zoom recording conditions, however, prevent these explanations from being treated as confirmed causal mechanisms.

Implications for Oral-Language Assessment

The findings support a functional approach to oral-language assessment. Teachers' pedagogical beliefs and social expectations influence assessment practices [19]. Automatic penalties for every filler may privilege rehearsed delivery and overlook interactional competence, planning under real-time pressure, and effective turn management. Assessment is more valid when it distinguishes occasional functional fillers from dense, disruptive disfluency and considers comprehensibility, organization, repair, responsiveness, and audience awareness. The results also indicate that multilingual planning strategies, including code-switching and silence, require interpretation within the communicative context.

Limitations

The study used a purposive sample with a narrow age distribution and a teacher-made proficiency examination. The Zoom recordings were not produced under controlled laboratory conditions, and variation in microphones, distance, transmission, and automatic audio processing may have influenced f0 and intensity. Acoustic analysis relied on small purposive subsamples, particularly the five duration tokens. Complete token-level values were unavailable for all pitch and intensity groups, which restricted the analysis to descriptive comparisons. The qualitative results were not accompanied by inter-coder reliability statistics, and the cross-sectional design could identify recurrent practices but could not demonstrate that filler use developed into fixed speech habits.

CONCLUSION

This exploratory mixed-methods sociophonetic analysis linked filler frequency, social context, multilingual proficiency, acoustic realization, and reported communicative function in student conversations. Uhm occurred more frequently than uh and displayed higher mean f0 and intensity in both male and female acoustic subsamples. The limited duration data also indicated a longer mean for uhm. Qualitative evidence showed that filled pauses supported planning, turn maintenance, and discourse organization, while their use varied with formality, speaker characteristics, and multilingual planning strategies. These results challenge the categorical treatment of fillers as speech deficiencies. Because the social and acoustic comparisons were descriptive, the study does not establish causal effects or statistically reliable group differences, but it provides context-specific evidence that filled pauses are meaningful resources in Philippine student English conversation.

REFERENCES

- [1] Clark, H. H. and Fox Tree, J. E., "Using uh and um in spontaneous speaking," *Cognition*, **84**(1): 73–111 (2002). [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- [2] Corley, M. and Stewart, O. W., "Hesitation disfluencies in spontaneous speech: The meaning of um," *Language and Linguistics Compass*, **2**(4): 589–602 (2008). <https://doi.org/10.1111/j.1749-818X.2008.00068.x>
- [3] Fox Tree, J. E., "Listeners' uses of um and uh in speech comprehension," *Memory & Cognition*, **29**(2): 320–326 (2001). <https://doi.org/10.3758/BF03194926>
- [4] Swerts, M., "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, **30**(4): 485–496 (1998). [https://doi.org/10.1016/S0378-2166\(98\)00014-9](https://doi.org/10.1016/S0378-2166(98)00014-9)
- [5] Tottie, G., "Uh and um as sociolinguistic markers in British English," *International Journal of Corpus Linguistics*, **16**(2): 173–197 (2011). <https://doi.org/10.1075/ijcl.16.2.02tot>
- [6] Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. and Brennan, S. E., "Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender," *Language and Speech*, **44**(2): 123–147 (2001). <https://doi.org/10.1177/00238309010440020101>
- [7] Laserna, C. M., Seih, Y.-T. and Pennebaker, J. W., "Um... who like says you know: Filler word use as a function of age, gender, and personality," *Journal of Language and Social Psychology*, **33**(3): 328–338 (2014). <https://doi.org/10.1177/0261927X14526993>
- [8] Kosmala, L. and Crible, L., "The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence," *Language and Speech*, **65**(1): 216–239 (2022). <https://doi.org/10.1177/00238309211010862>
- [9] De Leeuw, E., "Hesitation markers in English, German, and Dutch," *Journal of Germanic Linguistics*, **19**(2): 85–114 (2007). <https://doi.org/10.1017/S1470542707000049>
- [10] Crible, L., Degand, L. and Gilquin, G., "The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency," *Languages in Contrast*, **17**(1): 69–95 (2017). <https://doi.org/10.1075/lic.17.1.04cri>
- [11] Götz, S., *Fluency in Native and Nonnative English Speech*. John Benjamins Publishing Company, Amsterdam (2013).
- [12] Labov, W., *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia (1972).
- [13] Ladefoged, P. and Johnson, K., *A Course in Phonetics*, 7th ed. Cengage Learning, Stamford, CT (2015).
- [14] Creswell, J. W. and Plano Clark, V. L., *Designing and Conducting Mixed Methods Research*, 3rd ed. SAGE Publications, Thousand Oaks, CA (2018).
- [15] Braun, V. and Clarke, V., "Using thematic analysis in psychology," *Qualitative Research in Psychology*, **3**(2): 77–101 (2006). <https://doi.org/10.1191/1478088706qp063oa>
- [16] Boersma, P., "Praat, a system for doing phonetics by computer," *Glott International*, **5**(9/10): 341–345 (2001).
- [17] Fetters, M. D., Curry, L. A. and Creswell, J. W., "Achieving integration in mixed methods designs: Principles and practices," *Health Services Research*, **48**(6 Pt 2): 2134–2156 (2013). <https://doi.org/10.1111/1475-6773.12117>
- [18] Council for International Organizations of Medical Sciences, *International Ethical Guidelines for Health-related Research Involving Humans*. CIOMS, Geneva (2016).