ISSN: 2339-2940
E-ISSN: 2614-8633

Filler Words and Discourse Functions in NotebookLM-Generated Podcast-Style Audio Overviews

Henry Crisna Susanto[a,1,*], Agustinus Hary Setyawan[b,2]

[a,b]*Universitas Mercu Buana Yogyakarta, Jl. Wates Km. 10 Yogyakarta 55753, Indonesia*
[1]*henrycrisna21@gmail.com**; [2]*agustinus@mercubuana-yogya.ac.id*

ARTICLE INFO

Article history:
Received: 27/4/2026
Revised: 3/7/2026
Accepted:3/7/2026

Keywords:
AI-generated speech;
Audio Overview;
Discourse function;
Filler words;
NotebookLM.

ABSTRACT

This study investigates filler word use in NotebookLM-generated podcast-style Audio Overviews to examine how AI-mediated spoken discourse simulates spontaneous conversation at the discourse level. The study has two objectives: to identify the types of filler words produced in NotebookLM Audio Overviews and to explain their discourse functions. A qualitative descriptive design was employed. Five English-language CNN news articles representing politics, health, entertainment, public relations, and international conflict topics were uploaded to NotebookLM, and the resulting podcast-style Audio Overviews were transcribed manually. The data were coded for filler type, namely lexical and non-lexical fillers, and analyzed functionally using an adapted functional-pragmatic framework based on Ghanmi's discussion of filler word functions. The analysis identified 255 filler tokens. Lexical fillers were dominant, with 233 occurrences, whereas non-lexical fillers occurred 22 times. Filler use appeared in all topics, with the politics transcript showing the highest total. All 255 filler tokens were coded according to one dominant discourse function. Turn-taking signals formed the largest functional category, followed by cognitive planning aids and discourse connectors. The findings show that filler words in NotebookLM-generated podcasts are patterned rather than random and contribute to conversational flow, listener orientation, and discourse-level naturalness. The study concludes that filler words in AI-generated podcasts should be understood as pragmatic discourse resources and not merely as signs of disfluency.

I. Introduction

Artificial intelligence has increasingly transformed language-related activities, including the production of synthetic speech, automated dialogue, and podcast-style spoken content. Earlier forms of speech synthesis began with mechanical attempts to imitate the human vocal tract [1]. Over time, the technology developed into rule-based, concatenative, statistical, and neural forms of text-to-speech systems that are now widely used in media production, accessibility tools, and digital communication [2], [3], [4]. Contemporary text-to-speech systems are increasingly expected not only to convert written text into sound, but also to simulate the rhythm, prosody, and interactional quality of human speech [5], [6].

One important issue in this development concerns naturalness. Speech that is grammatically correct but sounds overly smooth or mechanically polished may still be perceived as artificial. Research on speech synthesis has shown that listeners evaluate synthetic speech partly through its resemblance to spontaneous spoken language, including variation in rhythm, hesitation, and conversational signaling [7], [8], [9]. In real interaction, speech is rarely perfectly fluent. Speakers often use filler words such as um, uh, you know, like, well, and I mean. These forms are commonly associated with hesitation, but they also serve important discourse functions, including signaling planning, maintaining the floor, guiding listener attention, and marking transitions in meaning [10], [11], [12].

ISSN: 2339-2940
E-ISSN: 2614-8633

Filler words are therefore not merely errors or empty elements. Clark and Fox Tree argued that uh and um function as meaningful signals in spontaneous speech [10]. Other studies have also shown that fillers are shaped by age, gender, role, topic, and communicative context [13], [14], [15]. In podcast discourse, Setyowati and Setyawan found that speakers in the DIVE Studios podcast produced both lexicalized and unlexicalized filled pauses, with lexicalized fillers occurring more frequently and functioning as hesitation markers, empathizing devices, mitigating devices, editing terms, and time-creating devices [16]. In second language and classroom contexts, filler words may reflect planning pressure, speech management, and pragmatic control [17], [18], [19]. In discourse-pragmatic terms, fillers can contribute to cohesion, listener engagement, emphasis, and turn management [12].

The emergence of AI-generated podcast-style audio adds a new dimension to this discussion. NotebookLM, a Google application, includes an Audio Overview feature that generates an extended spoken dialogue between two AI hosts based on uploaded sources [20]. Unlike older text-to-speech systems that often sounded robotic, this feature attempts to produce a podcast-like interaction that sounds spontaneous, interpretive, and conversational. Recent work on AI speech and language generation suggests that strategically inserted disfluencies may increase the naturalness of generated utterances [21], [22], [23]. This means that filler words can be examined not only as features of human conversation, but also as indicators of how artificial systems model spoken discourse.

Although previous studies have extensively examined filler words in naturally occurring spoken discourse, including spontaneous conversation, classroom interaction, and human produced podcasts, these studies primarily focus on human speakers and their communicative behavior. Other studies have investigated the insertion of disfluencies into synthetic speech to improve perceived naturalness, but they generally treat fillers as engineering techniques rather than as discourse-pragmatic resources. Consequently, these two lines of research have remained largely separate. Little attention has been paid to how conversational AI systems themselves generate and deploy filler words within extended spoken interaction. In particular, NotebookLM's Audio Overview feature has not yet been systematically examined with regard to the types of filler words it produces and the discourse functions those fillers perform. This omission is important because NotebookLM represents a new generation of AI systems that generates interactive podcast-style conversations rather than simply converting written text into speech. Examining filler word use in this context can therefore contribute not only to discourse-pragmatic research but also to broader discussions of conversational naturalness in AI-generated speech. This study addresses this gap by investigating both the types of filler words produced in NotebookLM Audio Overviews and the discourse functions they perform throughout the generated conversations.

This study focuses on filler words in NotebookLM-generated podcasts. This study addresses two research questions. First, what types of filler words are produced in NotebookLM-generated Audio Overviews? Second, what discourse functions do these filler words perform in the generated podcast-style discourse? The study aims to analyze the types, frequency, and communicative roles of filler words in NotebookLM Audio Overview transcripts. The significance of this study extends beyond describing filler word frequency. By examining how NotebookLM employs filler words as discourse-pragmatic resources, the study contributes to a better understanding of how contemporary conversational AI systems simulate interactional naturalness. The findings are expected to contribute to discourse-pragmatic studies, the evaluation of AI-generated spoken communication, and future research on conversational language generation.

II. Method

A. Research Design

This study employed a qualitative descriptive research design. Qualitative descriptive research is appropriate when a study aims to provide a direct and systematic description of a phenomenon in its natural form [24], [25]. In the present study, the phenomenon under investigation was the use of filler words in AI-generated spoken discourse. The analysis was interpretive, but it remained closely tied to the observable linguistic forms found in the transcripts.

The object of the study was filler word usage in spoken texts generated by NotebookLM's Audio Overview feature. The dataset consisted of five podcast-style transcripts derived from English-language CNN news articles across different thematic domains. The source articles were selected

ISSN: 2339-2940
E-ISSN: 2614-8633

through purposive sampling based on four criteria. First, each article was written in English to ensure linguistic consistency across the dataset. Second, all articles originated from CNN so that differences in journalistic style across news organizations would not influence the generated podcasts. Third, the articles represented different thematic domains, namely politics, health, entertainment, public relations, and international conflict, in order to capture variation in discourse contexts. Finally, each article contained sufficient informational content to enable NotebookLM to generate an extended podcast-style dialogue. These transcripts were analyzed as generated spoken discourse, with attention to filler type and discourse function. The selected articles are as follows:

- Politics: "Trump took the US economy to the brink of a crisis in just 100 days" https://edition.cnn.com/2025/04/28/politics/tariffs-economy-100-days-trump/index.html
- Health: "What you eat in midlife affects how healthy you are at age 70, according to a new study" https://edition.cnn.com/2025/04/06/health/diet-food-aging-nutrition-study-wellness/index.html
- Entertainment: "Screens are the enemy in new 'Toy Story 5' teaser trailer" https://edition.cnn.com/2025/11/11/entertainment/toy-story-5-teaser-watch-now
- Public relations: "JD Vance brushes off speculation about his marriage" https://edition.cnn.com/2025/12/05/politics/jd-vance-usha-vance-marriage
- International conflict: "Could the latest Ukraine talks actually end the war? Here's what to know" https://edition.cnn.com/2025/11/25/europe/ukraine-russia-war-trump-diplomacy-explainer-latam-intl

Although the dataset consisted of only five source articles, this number was considered appropriate for the qualitative descriptive approach adopted in this study. Rather than seeking statistical generalization, qualitative descriptive research emphasizes detailed contextual interpretation of linguistic phenomena. Topic variation was therefore prioritized over sample size to facilitate an in-depth analysis of filler word use across different communicative contexts.

Each article was uploaded into NotebookLM, which then generated an audio conversation between two AI speakers. The resulting audio outputs were transcribed manually and treated as spoken discourse data. The articles were accessed on December 20, 2025. The Audio Overviews were generated once for each source article on December 20, 2025. Because NotebookLM outputs may vary across generation attempts, the analysis was limited to the five Audio Overviews produced during this data collection session.

The selected source articles were chosen to provide topic variation and to make it possible to examine whether filler use changed across different thematic domains. The unit of analysis was the filler word, defined here as a lexical or non-lexical item that does not primarily add propositional meaning but serves a pragmatic or discourse-management function. Lexical fillers included forms such as you know, I mean, like, okay, and well. Non-lexical fillers included um and uh.

B. Data Source and Collection

The data collection procedure followed five steps. First, five English-language news articles were selected. Second, each article was uploaded into NotebookLM to generate an Audio Overview. Third, the audio outputs were exported and transcribed manually. The length of each transcript was recorded to provide contextual information for interpreting filler frequency across topics. Fourth, each transcript was examined line by line to identify potential filler words. A token was identified as a filler only when its primary role was to manage discourse or interaction rather than contribute propositional meaning to the utterance. Lexical items functioning as ordinary content words were excluded from coding. When a lexical item could plausibly function either as a filler or as an ordinary lexical expression, the surrounding utterance was examined to determine its dominant communicative role. Only tokens functioning primarily as discourse management devices were retained for analysis. Fifth, each filler token was coded by type and by discourse function in context. Manual transcription and coding were chosen because they allow close attention to contextual nuance and pragmatic function, which can be difficult to capture through automated extraction alone [26], [27].

ISSN: 2339-2940
E-ISSN: 2614-8633

C. Data Analysis

The analytical framework was adapted from Ghanmi's functional-pragmatic discussion of filler words [12]. Ghanmi treats filler words as meaningful discourse resources that may support continuity, attention, coherence, and communicative intention. Based on this perspective, the present study operationalized five discourse functions for coding: cognitive planning aid, turn-taking signal, attentional cue, emphasis or reassurance, and discourse connector. These functions are defined in **Error! Reference source not found.**

Table 1. Operational Definitions of Filler Word Discourse Functions

Function	Operational Definition
Cognitive planning aid	A filler occurring before elaboration, exemplification, reformulation, or delayed continuation, indicating a simulated planning position in the generated discourse.
Turn-taking signal	A filler marking the beginning, holding, or transition of a speaker's turn
Attentional cue	A filler guiding the listener toward upcoming or important information
Emphasis or reassurance	A filler strengthening, clarifying, softening, or confirming a claim
Discourse connector	A filler linking clauses, ideas, or discourse segments

Each filler token was examined within its immediate utterance context and assigned one dominant discourse function. The coding process also recorded the filler type and its position within the utterance. To enhance coding consistency, the researcher conducted three successive rounds of coding. The first round focused on identifying all potential filler tokens. The second classified each token as lexical or non-lexical. The third assigned each token one dominant discourse function based on its immediate context and the operational definitions presented in Table 1. Ambiguous cases were revisited until a consistent interpretation was reached. Although formal intercoder reliability was not employed because the study was conducted by a single researcher, coding reliability was strengthened through repeated coding, continuous comparison across coding rounds, and systematic reference to the operational definitions throughout the analytical process.

III. Results and Discussion

This section presents the findings according to the two research questions formulated in the Introduction. The first research question examines the types of filler words produced in NotebookLM-generated Audio Overviews, whereas the second investigates the discourse functions performed by those filler words. Across the five generated podcast transcripts, a total of 255 filler tokens were identified. The findings indicate that NotebookLM consistently produced both lexical and non-lexical fillers across all topics, suggesting that filler use was a systematic feature of the generated spoken discourse rather than an isolated phenomenon.

The first result concerns filler type. Table 2 shows that lexical fillers were far more frequent than non-lexical fillers. Of the 255 tokens, 233 were lexical fillers and 22 were non-lexical fillers. This distribution indicates that the NotebookLM outputs contained more discourse-oriented expressions such as you know, I mean, like, okay, and well than hesitation-only markers such as um and uh. This pattern is important because lexical fillers often contribute more directly to the organization of interaction and the framing of meaning than purely non-lexical hesitation sounds [10], [12].

Table 2. Distribution of Filler Word Types

Filler Type	Example Forms	Frequency
Lexical	you know, I mean, like, okay, well	233
Non-lexical	um, uh	22
Total		**255**

To facilitate comparison between lexical and non-lexical fillers, Figure 1 provides a visual representation of their overall distribution in the NotebookLM-generated podcasts.

ISSN: 2339-2940
E-ISSN: 2614-8633

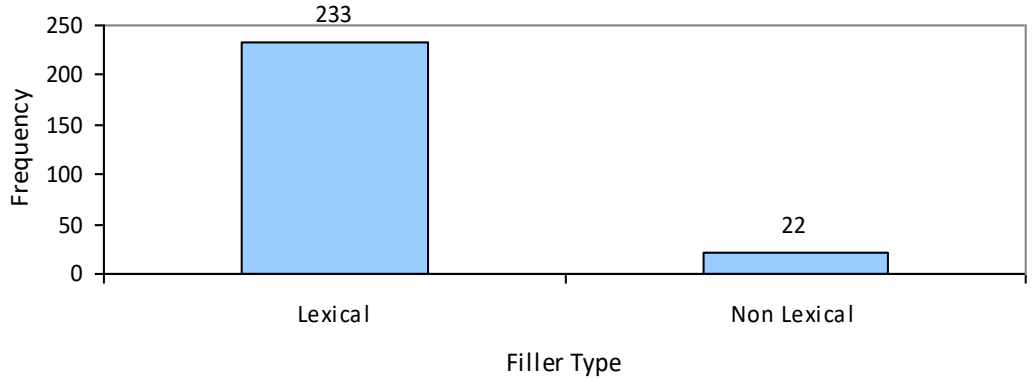

Figure 1. Distribution of Filler Word Types in NotebookLM-Generated Audio Overviews

The predominance of lexical fillers suggests that NotebookLM is not merely simulating speech disfluency but is instead generating discourse markers that organize interaction and maintain conversational flow. In spontaneous discourse, lexical fillers frequently function as turn openers, transition markers, explanation frames, and listener oriented cues [13], [19]. Their high frequency therefore indicates that the generated dialogue resembles conversational interaction rather than a simple oral rendering of written text.

The second result concerns topic distribution. Table 3 shows that filler frequency was not uniform across the five topics. Politics contained the highest number of filler tokens with 63 instances, followed by health with 58, public relations with 49, international conflict with 47, and entertainment with 38. Lexical fillers outnumbered non-lexical fillers in every topic.

Table 3. Distribution of Filler Words Across Topics

Topic	Lexical Fillers	Non-lexical Fillers	Total
Politics	56	7	63
Health	54	4	58
Entertainment	34	4	38
Public relations	44	5	49
International conflict	45	2	47
Total	**233**	**22**	**255**

Figure 2 visually summarizes the distribution of filler words across the five thematic domains. Although lexical fillers predominated in every topic, the overall frequency of filler words varied across the generated podcasts.

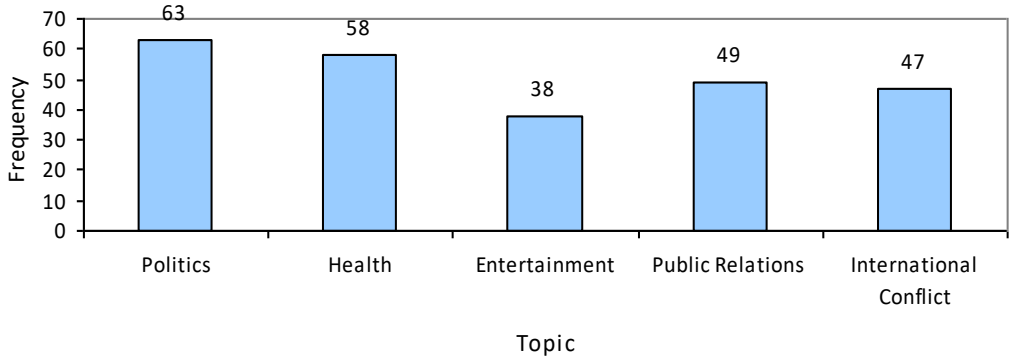

Figure 2. Distribution of Filler Words Across Topic Domains in NotebookLM-Generated Audio Overviews

As illustrated in Figure 2, the politics transcript contained the highest frequency of filler words, whereas the entertainment transcript showed the lowest frequency. This pattern may be interpreted in relation to the complexity of the source text and the explanatory demands of the generated dialogue. Research on spoken production has shown that greater cognitive load can affect temporal organization

ISSN: 2339-2940
E-ISSN: 2614-8633

and increase disfluency patterns [28]. Similarly, Göbel and McCarthy found that increased cognitive load affected bilingual speech production, with speakers showing stronger L1 transfer when attention was divided, suggesting that cognitive demands can shape spoken output patterns [29]. Political discourse often involves abstract claims, competing viewpoints, evaluative language, argumentative framing, and persuasion [30]. In a generated podcast setting, these discourse demands may encourage the use of fillers that simulate deliberation, mark transitions, and guide listeners through complex information. However, this interpretation should be treated cautiously because each topic was represented by only one source article.

Health also showed a relatively high filler count. One possible reason is that health-related material often combines information delivery with explanation and practical interpretation. When the generated dialogue connects evidence, consequences, and recommendations, fillers such as so, well, and you know may help organize the discourse and guide listener attention. By contrast, the entertainment transcript showed a lower frequency of fillers. Entertainment discussion in the present data appeared more narrative and descriptively linear. Familiar topics with concrete referents can be easier to structure, which may reduce the need for hesitation and discourse-management markers. This explanation is consistent with findings that topic familiarity supports conceptualization and can improve spoken production [31]. The public relations and international conflict transcripts also produced substantial numbers of fillers, with 49 and 47 tokens respectively. However, both remained lower than the politics transcript. Entertainment showed the lowest total, with 38 tokens, which may reflect the more linear and narrative structure of the generated discussion in this dataset. However, this should not be interpreted as evidence that these themes are inherently simpler. It is more reasonable to say that filler frequency depends not only on topic label, but also on how the source text frames the topic and how the generated dialogue organizes it. Nevertheless, these differences should be interpreted cautiously because each thematic domain was represented by only one source article. The observed variation therefore reflects patterns within the present dataset rather than definitive evidence that topic alone determines filler word use in AI-generated spoken discourse. Future studies employing larger datasets will be necessary to evaluate whether similar topic-based patterns emerge consistently.

The third result concerns discourse function. Each filler token was coded according to one dominant pragmatic role. Table 4 summarizes the results. Turn-taking signals formed the largest category with 83 tokens, followed by cognitive planning aids with 59, discourse connectors with 57, emphasis or reassurance markers with 34, and attentional cues with 22. These results indicate that fillers in the dataset were used mainly to manage conversational turns, simulate planning-like positions, and connect discourse segments.

Table 4. Distribution of Filler Words by Discourse Function

Discourse Function	Example from Data[a]	Filler Word	Frequency
Cognitive Planning Aid	"the example of, um, playing golf..."	um	59
Turn-Taking Signal	"Okay, let's unpack this new tension..."	Okay	83
Attentional Cue	"You know, the pace of diplomacy..."	You know	22
Emphasis or Reassurance	"I mean, when you look at the core demands..."	I mean	34
Discourse Connector	"Right. So think of this as us doing the heavy lifting..."	So	57
Total			**255**

[a] The examples indicate representative uses found in the dataset, not fixed one-to-one associations between filler forms and discourse functions. Since filler words may be multifunctional, each token was coded according to its dominant function in context.

Figure 3 provides a visual summary of the distribution of filler words across the five discourse functions identified in the analysis. Turn-taking signals constituted the largest functional category, followed by cognitive planning aids and discourse connectors, whereas attentional cues occurred least frequently.

ISSN: 2339-2940
E-ISSN: 2614-8633

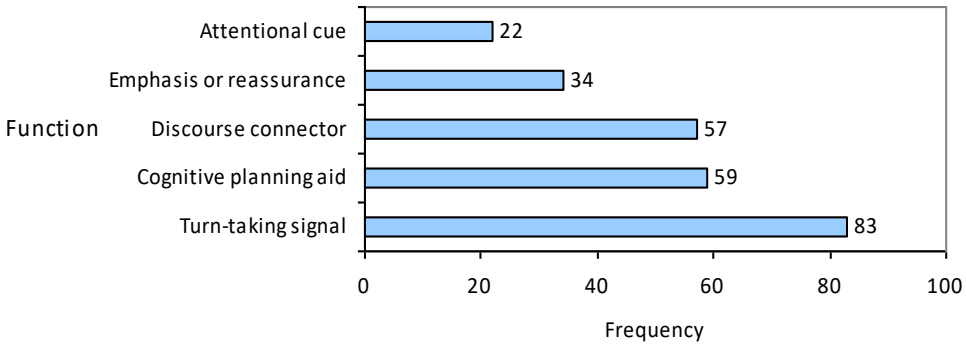

Figure 3. Distribution of Filler Words by Discourse Function in NotebookLM-Generated Audio Overviews

Overall, the functional distribution demonstrates that filler words in NotebookLM-generated podcasts are systematically employed rather than randomly inserted. The predominance of turn-taking signals, cognitive planning aids, and discourse connectors suggests that filler words contribute primarily to organizing interaction and maintaining discourse coherence, thereby enhancing the conversational quality of the generated dialogue.

The prominence of turn-taking signals indicates that NotebookLM frequently uses fillers to organize speaker transitions and maintain the impression of responsive dialogue. Markers such as okay, right, yeah, and well often appear at the beginning of turns or at points of transition, creating a podcast-like interactional rhythm. This pattern is important because the Audio Overview format presents information through two alternating speakers, so turn-management markers help sustain the illusion of conversational exchange.

Cognitive planning aids were the second most frequent category. Although the data are AI-generated and cannot be interpreted as evidence of actual mental hesitation, fillers such as um, uh, well, and like often appear before elaboration, reformulation, or explanatory phrasing. These tokens therefore occupy planning-like positions in the generated discourse and contribute to the simulation of spontaneous spoken production.

Discourse connectors also appeared frequently. Fillers such as so, well, and then helped link one idea, explanation, or topic segment to another. Their frequency suggests that filler words in the dataset were not used merely to imitate hesitation, but also to organize coherence across the generated podcast dialogue.

Emphasis or reassurance was another important functional category. Markers such as I mean and right were often used before clarification or reformulation. This function gives the impression that the speaker is checking alignment with the listener or reinforcing an intended interpretation. In conversational discourse, fillers can be used strategically to soften or strengthen meaning, not simply to fill silence [12], [15]. The presence of this function in the dataset supports the view that NotebookLM-generated speech includes pragmatic features associated with listener orientation.

Attentional cues occurred less frequently than the other functional categories, with 22 tokens. However, they remain analytically important because they show how fillers such as you know and so can orient listeners toward upcoming or salient information. Their lower frequency suggests that listener orientation was present in the dataset, but it was less central than turn management, planning-like positioning, and discourse connection.

Several transcript excerpts illustrate these functions. In the politics transcript, the utterance “the example of, um, playing golf...” shows a non-lexical filler before a concrete explanation, which indicates a simulated cognitive planning position. In the entertainment transcript, “Okay, let’s unpack this new tension...” begins with a turn-taking signal that opens a new analytical segment. In the international conflict transcript, “You know, the pace of diplomacy...” functions as an attentional cue because it orients the listener to the topic under discussion. In another international conflict excerpt, “I mean, when you look at the core demands...” uses a lexical filler to frame clarification and emphasis. Finally, in the politics transcript, “Right. So, think of this as us doing the heavy lifting...” uses so as a discourse connector that links the previous point to the next explanatory move. These examples demonstrate that the fillers were contextually integrated into the generated dialogue rather than

ISSN: 2339-2940
E-ISSN: 2614-8633

inserted in a random way.

Overall, the findings answer both research questions by demonstrating that NotebookLM-generates both lexical and non-lexical fillers and that these fillers consistently perform identifiable discourse-pragmatic functions throughout the generated conversations. Taken together, these results support the argument that filler words in NotebookLM-generated podcasts function as pragmatic resources contributing to discourse-level naturalness. This interpretation is consistent with recent work suggesting that disfluency insertion may enhance the human-like quality of generated utterances [21], [23]. It also resonates with broader work on synthetic speech naturalness, which emphasizes the role of prosody, spontaneity, and conversational texture in user perception [7], [8], [22]. The findings therefore suggest that, in the analyzed dataset, NotebookLM-generated podcast speech does not merely vocalize source content but also simulates selected features of interactional discourse behavior.

At the same time, the findings should be interpreted within the study's limits. The dataset was restricted to five podcasts from one platform and one source domain. Each topic was represented by one source article, so the findings cannot be generalized to all topic types or all AI-generated podcasts. Furthermore, because NotebookLM may produce different outputs across repeated generations, the present findings should be interpreted as representing the specific outputs generated during the documented data collection session rather than all possible outputs produced by the system. The coding process was manual and therefore depended on contextual judgment. In addition, the study focused specifically on filler words and did not examine pause duration, intonation, repair sequences, or other prosodic features that also shape naturalness. Since the study did not involve listener perception testing, claims about naturalness are limited to discourse-level interpretation rather than measured audience perception.

IV. Conclusion

This study examined filler words in NotebookLM-generated podcasts in order to identify the types of fillers produced and explain their linguistic functions. The findings show that NotebookLM-generated both lexical and non-lexical fillers across all five podcast topics. Lexical fillers occurred much more frequently than non-lexical fillers, with 233 lexical fillers and 22 non-lexical fillers out of 255 total tokens. This indicates that the generated speech relies strongly on discourse-oriented expressions rather than on hesitation markers alone.

Functionally, the fillers served as turn-taking signals, cognitive planning aids, discourse connectors, emphasis or reassurance markers, and attentional cues. Among these categories, turn-taking signals appeared most frequently, followed by cognitive planning aids and discourse connectors. This result suggests that filler words in the generated podcasts were used not only to simulate hesitation, but also to manage speaker transitions, organize discourse flow, and create a more interactive listening experience.

Topic variation also mattered. Politics produced the highest number of fillers, followed by health, public relations, international conflict, and entertainment. This suggests that filler frequency may be influenced by topic complexity, source-text framing, and the way the generated dialogue organizes explanation.

Overall, the study concludes that filler words in NotebookLM-generated podcasts should not be viewed merely as signs of disfluency. They function as pragmatic resources that help AI-generated discourse simulate spontaneity, coherence, and human-like interaction at the discourse level. Future research may compare NotebookLM with other AI platforms, expand the dataset, include human podcast data as a comparison, and investigate additional spoken language features such as pause structure, intonation, self-repair, and prosodic variation. Such investigations will contribute to a broader understanding of how conversational AI systems employ discourse-pragmatic resources to approximate naturally occurring spoken interaction and may inform the development of more natural and context-sensitive AI-generated speech.

References

[1] K. OSUKA, "Mechanical Speech Synthesizer and Soft Material," *Nippon GOMU KYOKAISHI*, vol. 78, no. 8, pp. 307–312, 2005, doi: 10.2324/gomu.78.307.

[2] Z. Zhang, "AI-Powered Intelligent Speech Processing: Evolution, Applications and Future Directions," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 2, pp. 918–928, 2025, doi:

ISSN: 2339-2940
E-ISSN: 2614-8633

10.14569/IJACSA.2025.0160291.
[3] S. Latif, A. Shahid, and J. Qadir, "Generative emotional AI for speech emotion recognition: The case for synthetic emotional speech augmentation," *Appl. Acoust.*, vol. 210, p. 109425, Jul. 2023, doi: 10.1016/j.apacoust.2023.109425.
[4] N. S. K. Anuradha, "Text-To-Speech Conversion," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 2, no. 6, pp. 269–278, 2013.
[5] Y. Tabet and M. Boughazi, "Speech Synthesis Techniques . A Survey," *7th Int. Work. Syst. Signal Process. their Appl. SPEECH*, pp. 67–70, 2011, doi: 10.1109/WOSSPA.2011.5931414.
[6] X. Tan, T. Qin, F. Soong, and T.-Y. Liu, "A Survey on Neural Speech Synthesis," 2021, [Online]. Available: http://arxiv.org/abs/2106.15561
[7] R. J. Skerry-Ryan *et al.*, "Towards end-to-end prosody transfer for expressive speech synthesis with tacotron," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 11, pp. 7471–7480, 2018.
[8] L. Bakkouche *et al.*, "Finding the Human Voice in AI : Insights on the Perception of AI-Voice Clones from Naturalness and Similarity Ratings Finding the Human Voice in AI : Insights on the Perception of AI-Voice Clones from Naturalness and Similarity Ratings," no. June, 2025, doi: 10.17605/OSF.IO/WF3HA.
[9] Š. Tomková, "Text-To-Speech Technologies in Online Media," *Media Mark. Identity*, pp. 699–707, 2024, doi: 10.34135/mmidentity-2024-69.
[10] H. H. Clark and J. E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, 2002, doi: 10.1016/S0010-0277(02)00017-3.
[11] M. Corley and O. W. Stewart, "Hesitation disfluencies in spontaneous speech: The meaning of um," *Linguist. Lang. Compass*, vol. 2, no. 4, pp. 589–602, 2008, doi: 10.1111/j.1749-818X.2008.00068.x.
[12] S. Ghanmi, "Filler Words : The Definition and Their Communicative Functions," *Arab. Lang. Lit. Cult.*, vol. 7, no. 2, pp. 6–10, 2022, doi: 10.11648/j.allc.20220702.11.
[13] H. Bortfeld, S. D. Leon, J. E. Bloom, M. F. Schober, and S. E. Brennan, "Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender," *Lang. Speech*, vol. 44, no. 2, pp. 123–147, 2001, doi: 10.1177/00238309010440020101.
[14] S. A. Ruschy, "The Utilization of Filler Words in Relation to Age and Gender," pp. 1–16, 2024.
[15] B. Clancy, "Discourse-pragmatic markers, fillers and filled pauses: Pragmatic, cognitive, multi-modal and sociolinguistic perspectives, edited by Kate Beeching, Grant Howie, Minna Kirjavainen, and Anna Piasecki," *Contrastive Pragmat.*, vol. 0393, pp. 1–6, 2024, doi: 10.1163/26660393-bja10125.
[16] U. Setyowati and A. H. Setyawan, "Fillers Used by Speakers on DIVE Studios Podcast," *Linguist. ELT J.*, vol. 11, no. 2, pp. 131–137, 2023, [Online]. Available: http://journal.ummat.ac.id/index.php/JELTL/article/view/20189
[17] Q. Huang, "A Corpus-based Comparative Analysis on the Use of Non-lexical Filler Words in Spoken English Between Chinese EFL Learners and Native English Speakers," *J. Humanit. Arts Soc. Sci.*, vol. 8, no. 1, pp. 266–272, 2024, doi: 10.26855/jhass.2024.01.046.
[18] S. Khojastehrad, "Hesitation Strategies in an Oral L2 Test among Iranian Students Shifted from EFL Context to EIL," *Int. J. English Linguist.*, vol. 2, no. 3, pp. 10–21, 2012, doi: 10.5539/ijel.v2n3p10.
[19] L. R. Nurjamin, A. Nurjamin, and R. Melati, "What EFL Learners Say in Managing Their Speech During Academic Presentations," in *Proceedings of the Twelfth Conference on Applied Linguistics (CONAPLIN 2019)*, Paris, France: Atlantis Press, 2020, pp. 116–120. doi: 10.2991/assehr.k.200406.023.
[20] Google, "Generate Audio Overview in NotebookLM." Accessed: Jun. 11, 2025. [Online]. Available: https://support.google.com/notebooklm/answer/16212820?ref_topic=16164070&sjid=3352195109469844277-NC
[21] S. Z. Hassan, P. Lison, and P. Halvorsen, "Enhancing Naturalness in LLM-Generated Utterances through Disfluency Insertion," 2024, [Online]. Available: http://arxiv.org/abs/2412.12710
[22] S. Wang, J. Gustafson, and É. Székely, "Evaluating Sampling-based Filler Insertion with Spontaneous TTS," *2022 Lang. Resour. Eval. Conf. Lr. 2022*, pp. 1960–1969, 2022.
[23] R. Sharma, P. V. Shah, and A. M. Joshi, "Naturalization of Text by the Insertion of Pauses and Filler Words," no. August, pp. 1–14, 2020, [Online]. Available: http://arxiv.org/abs/2011.03713
[24] L. Doyle, C. McCabe, B. Keogh, A. Brady, and M. McCann, "An overview of the qualitative descriptive design within nursing research," *J. Res. Nurs.*, vol. 25, no. 5, pp. 443–455, 2020, doi: 10.1177/1744987119880234.
[25] V. A. L. C. E. Lambert, "Editorial: Qualitative Descriptive Research: An Acceptable Design," *Sch. Inq. DNP Capstone*, no. 4, pp. 255–256, 2012.
[26] L. F. de Jesus, T. C. S. Andres, and C. B. Tuazon, "Coding Methods for Enhanced Manual Data Transcripts: Contributing to Sustainable Development Goals (SDG)," *J. Lifestyle SDGs Rev.*, vol. 4, no. 1, p. e01864, 2024, doi: 10.47172/2965-730x.sdgsreview.v4.n00.pe01864.

ISSN: 2339-2940
E-ISSN: 2614-8633

[27] C. Hacking, H. Verbeek, J. P. H. Hamers, and S. Aarts, "Comparing text mining and manual coding methods: Analysing interview data on quality of care in long-term care for older adults," *PLoS One*, vol. 18, no. 11 November, pp. 1–13, 2023, doi: 10.1371/journal.pone.0292578.
[28] J. Bóna and M. Bakti, "The effect of cognitive load on temporal and disfluency patterns of speech," *Target. Int. J. Transl. Stud.*, vol. 32, no. 3, pp. 482–506, 2020, doi: 10.1075/target.19041.bon.
[29] J. Göbel and K. Mccarthy, "The Impact of Cognitive Load on Speech Production in German-English Bilinguals," 2023.
[30] O. Новікова and I. Суїма, "The Strategies of Argumentation in Political Speeches," *J. "Ukrainian sense,"* no. 2, pp. 76–89, Jan. 2024, doi: 10.15421/462320.
[31] X. Qiu, "Functions of oral monologic tasks: Effects of topic familiarity on L2 speaking performance," *Lang. Teach. Res.*, vol. 24, no. 6, pp. 745–764, 2020, doi: 10.1177/1362168819829021.