



Investigating Speech Disfluencies in Spontaneous L2 Speech Production: A Psycholinguistic Study of Indonesian TikTok Users

¹*Shera Yanuar Hardiat, ¹ Udayani Permaludin, ¹ Deni Suswanto

¹ Department of English Literature, Faculty of Adab and Humanities, Universitas Islam Negeri Sunan Gunung Djati. Jl. A.H. Nasution No. 105A, Cibiru, Bandung, West Java, Indonesia

*Corresponding Author e-mail: sheraivanuar@gmail.com

Received: April 2026; Revised: May 2026; Accepted: May 2026; Published: June 2026

Abstract

Speech disfluencies are natural yet meaningful features of spontaneous speech that reflect real-time cognitive processing demands during language production. In second-language (L2) contexts, disfluencies occur more frequently because speakers must simultaneously manage conceptual planning, lexical retrieval, syntactic encoding, and self-monitoring with limited linguistic automaticity. Despite growing research on L2 fluency, disfluency as a psycholinguistic phenomenon remains underexplored in self-initiated digital communication contexts, particularly on platforms such as TikTok, where young adult L2 speakers produce spontaneous, minimally edited English speech as part of self-improvement challenges. This study aims to describe the types of speech disfluencies produced by three Indonesian TikTok users in their spontaneous English speech and to explain the underlying psycholinguistic factors that trigger them. Employing a qualitative descriptive method, nine video samples from the 'Day 1–30 Speaking English Challenge' were analyzed through verbatim transcription and qualitative content analysis, drawing on Levelt's (1989) speech production model and Kormos' (2006) framework of second language speech processing. The analysis identified 227 disfluency tokens distributed across three major categories: pauses, comprising silent and filled pauses (171 instances total); hesitations, encompassing repetitions, false starts, truncated words, prolongations, and self-corrections (84 instances); and lexicalized fillers (9 instances). Five underlying psycholinguistic factors were identified as triggers: conceptual planning difficulty, lexical retrieval effort, syntactic and phonological encoding demands, working memory load, and self-monitoring behavior, each operating at distinct stages of speech production. While the findings are specific to this dataset and are not intended for broad generalization, they demonstrate that disfluencies in spontaneous L2 digital speech are not random errors but systematic indicators of real-time cognitive processing. The study contributes to the extension of Levelt's (1989) and Kormos' (2006) frameworks to digitally mediated speech contexts, with implications for psycholinguistically informed fluency assessment, language pedagogy, and the role of self-initiated digital speaking practice, such as TikTok speaking challenges, in fostering authentic L2 fluency development.

Keywords: Speech disfluency; Second language speech production; Psycholinguistics; TikTok media

How to Cite: Hardiat, S.Y., Permaludin, U., & Suswanto, D. (2026). Investigating Speech Disfluencies in Spontaneous L2 Speech Production: A Psycholinguistic Study of Indonesian TikTok Users. *Journal of Language and Literature Studies*, 6(2), 437-454. doi: <https://doi.org/10.36312/jolls.v6i2.5013>



<https://doi.org/10.36312/jolls.v6i2.5013>

Copyright© 2026, Hardiat et al.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) License.



INTRODUCTION

Speech disfluency is a pervasive yet often overlooked phenomenon in spontaneous spoken language, encompassing pauses, fillers, hesitations — including repetitions, false starts, truncated words, prolongations, and self-corrections — that naturally emerge when speakers produce speech in real time without prior preparation. As a psycholinguistic phenomenon, disfluency reflects the cognitive demands underlying speech production, including conceptual planning, lexical retrieval, syntactic and phonological encoding, working memory load, and self-monitoring (Kormos, 2006; Levelt, 1989). In second language (L2) contexts, these features become even more prominent, as speakers must simultaneously manage multiple competing cognitive processes with limited linguistic

automaticity (Kormos, 2006). While disfluency has long been studied in controlled laboratory and classroom settings, the rise of digital communication has opened new and largely unexplored avenues for observing spontaneous L2 speech in authentic, self-initiated environments. Among contemporary digital platforms, TikTok has emerged as a particularly relevant medium for psycholinguistic investigation of spontaneous L2 speech. Several platform-specific characteristics justify its classification as a context for spontaneous speech production. First, TikTok videos are typically self-recorded in real time without a prepared script, meaning speakers must simultaneously conceptualize, formulate, and articulate their thoughts as they speak — precisely the conditions under which disfluencies are most likely to emerge (Levelt, 1989). Second, the short-form video format imposes time constraints that limit the opportunity for extensive re-recording or editing, preserving natural speech features such as pauses, fillers, hesitations, and self-corrections in their authentic form.

Third, the informal and social nature of TikTok encourages speakers to engage in spontaneous, unmonitored self-expression rather than polished performance, distinguishing it from more structured platforms such as YouTube, which often involves scripted content, or Instagram, which prioritizes visual aesthetics over verbal spontaneity (Kress, 2000). Fourth, the self-initiated nature of speaking challenges such as the ‘Day 1–30 Speaking English Challenge’ means that participants choose their own topics and speak without external prompts or interlocutors, creating a uniquely naturalistic condition for observing L2 speech production under real cognitive load. Indonesian TikTok users represent a particularly meaningful context for this investigation because English functions as a foreign language rather than a second language in Indonesia, meaning speakers have limited opportunities for naturalistic English use outside formal educational settings. Their engagement in self-initiated digital English-speaking challenges, therefore, reflects a form of voluntary, authentic L2 practice that is rarely captured in controlled research settings, making their speech data especially valuable for understanding how cognitive processing demands manifest in spontaneous L2 digital communication (Rahmawati et al., 2023).

Understanding speech disfluency in L2 contexts is both theoretically significant and practically urgent, particularly within the context of self-initiated digital speaking challenges such as those found on TikTok. From a theoretical standpoint, spontaneous speech production involves a cascade of simultaneous cognitive processes: speakers must conceptualize the message they wish to convey, retrieve appropriate lexical items from memory, construct syntactic structures, encode phonological forms, and monitor their own output for accuracy — all within a matter of milliseconds and without the luxury of pre-planning (Levelt, 1989). When any of these processes is delayed or overloaded, disfluencies emerge in the form of silent pauses, filled pauses such as ‘uh’ and ‘um’, lexicalized fillers such as ‘like’ and ‘you know’, repetitions, false starts, prolongations, and self-corrections (Clark, 2002; Lickley, 2015).

In L2 contexts, these disruptions occur more frequently because speakers must distribute limited cognitive resources across multiple competing demands with lower levels of automaticity than native speakers (Kormos, 2006; Segalowitz, 2016). The nature of TikTok speaking challenges amplifies these cognitive demands further. Unlike classroom speaking tasks where learners receive prompts, preparation time, and teacher scaffolding, TikTok speaking challenges require participants to independently select topics, self-record without external support, and publish their speech for public viewing — all of which introduce additional cognitive and affective pressures that intensify the likelihood of disfluency production. From a practical standpoint, this digital context is important for psycholinguistic research precisely because it offers a window into naturalistic L2 speech that is rarely accessible in controlled experimental settings. Traditional fluency assessments often treat disfluencies as indicators of low proficiency, yet psycholinguistic research

consistently demonstrates that they reflect cognitive processing rather than communicative failure (Segalowitz, 2016). If disfluencies in self-initiated digital contexts remain unexamined, our understanding of how fluency develops authentically — outside the classroom and within the digital spaces where many young adult L2 speakers increasingly practice their language — will remain fundamentally incomplete.

Several previous studies have investigated speech disfluency in both first and second language contexts, providing important foundations for the present research. Zulhemindra et al. (2022) analyzed lexicalized and non-lexicalized filled pauses produced by English students and revealed that filled pauses serve communicative, affective, and cognitive purposes. Liao (2023) investigated disfluency and self-repair in L2 presentational and interpersonal speech and found that speech modes significantly affect the distribution and rates of self-repair, highlighting the importance of contextual conditions in shaping disfluency patterns. In a related vein, Wang (2021) examined hesitation markers — including silent pauses, filled pauses, and lexical fillers — in the spontaneous English speech of EFL learners at varying proficiency levels and found that higher-proficiency speakers employed a more diverse range of hesitation strategies than lower-proficiency peers. However, these studies relied on controlled or task-based speech data collected in classroom or experimental settings, which may not fully capture the nature of spontaneous speech production in informal digital environments.

Kirkland et al. (2022) explored how filled pauses influence listeners' perceptions of speakers' confidence and competence, yet their focus remained on listener interpretation rather than on the psycholinguistic mechanisms underlying disfluency production in L2 speech. Jelčić (Jelčić, 2024) investigated pause patterns in spontaneous media speech, but as the participants were professional native speakers, the findings are more aligned with sociolinguistic performance analysis than with psycholinguistic exploration of L2 speech among language learners. Taken together, these studies reveal three critical gaps that the present research seeks to address. First, existing studies have predominantly relied on controlled or task-based data, leaving disfluency as it naturally occurs in self-initiated digital communication largely unexplored — a significant oversight given the growing role of platforms such as TikTok in shaping authentic L2 speaking practice among young adults. Second, the psycholinguistic factors underlying disfluency production — particularly conceptual planning difficulty, lexical retrieval effort, syntactic and phonological encoding demands, working memory load, and self-monitoring behavior — have received limited attention in digitally mediated contexts, where the absence of interlocutors and institutional scaffolding creates unique cognitive conditions for speech production.

Third, no study to date has specifically examined disfluency among Indonesian L2 speakers engaging in spontaneous English speech through social media platforms such as TikTok. Unlike previous studies, therefore, the present research makes three distinctive contributions: it examines naturally occurring, user-generated L2 speech data collected from a digital platform rather than a controlled setting; it applies Levelt's (1989) model of speech production and Kormos' (2006) framework of second language speech processing — frameworks originally developed for laboratory and classroom contexts — to spontaneous digitally mediated speech for the first time in the Indonesian EFL context; and it provides a comprehensive analysis of multiple disfluency types alongside their underlying psycholinguistic triggers, moving beyond descriptive categorization toward theoretical explanation of why disfluencies occur in authentic L2 digital communication.

Building on the gaps identified above, this study aims to investigate speech disfluencies produced by three Indonesian TikTok users in their spontaneous English speech as part of a 30-day speaking challenge. Specifically, it seeks to achieve two objectives: first, to describe how pauses — including silent pauses and filled pauses —

fillers, and hesitations — encompassing repetitions, false starts, truncated words, prolongations, and self-corrections — appear in the participants' spontaneous L2 speech production, following the classification frameworks of Zellner (Zellner, 1994), Clark and Fox Tree (2002), Clark and Wasow (1998), Levelt (1989), Kormos (2006), and Lickley (2015); and second, to explain the underlying psycholinguistic factors — including conceptual planning difficulty, lexical retrieval effort, syntactic and phonological encoding demands, working memory load, and self-monitoring behavior — that trigger these disfluencies during real-time speech production, as theorized by Levelt (1989) and Kormos (2006).

By addressing these objectives and research questions, this study seeks to bridge traditional psycholinguistic frameworks with contemporary digital communication contexts, contributing to a more nuanced and empirically grounded understanding of how L2 fluency develops in authentic, self-directed digital speaking environments (Feng & Guo, 2022; Hanzawa, 2024). To achieve these aims within a clearly defined methodological boundary, the scope of this study is intentionally bounded in several respects. The analysis is limited to three Indonesian TikTok users, which, while sufficient for qualitative depth, does not allow for broad generalization across L2 speaker populations. The data are drawn exclusively from a specific speaking activity — the 'Day 1–30 Speaking English Challenge' — which, although naturalistic, represents only one type of self-initiated digital speaking context. Additionally, as a qualitative study, the findings prioritize descriptive depth and theoretical interpretation over statistical generalizability. Pause durations were categorized auditorily rather than through acoustic measurement tools such as Praat, which may introduce some degree of subjectivity in classification. These limitations notwithstanding, the study contributes a contextually grounded and theoretically informed understanding of disfluency as a psycholinguistic phenomenon in digitally mediated L2 speech and opens productive directions for future research in this emerging area. To achieve these objectives, this study addresses the following research questions: *What types of speech disfluencies — including pauses, fillers, and hesitations — appear in the spontaneous English speech produced by three Indonesian TikTok users participating in the 'Day 1–30 Speaking English Challenge'? What underlying psycholinguistic factors — including conceptual planning difficulty, lexical retrieval effort, syntactic and phonological encoding demands, working memory load, and self-monitoring behavior — trigger the occurrence of these disfluencies during real-time L2 speech production?*

METHOD

Research Design

This study is situated within an interpretive research paradigm, which holds that language use is context-dependent and that meaning must be understood through careful examination of naturally occurring human behavior rather than through experimental manipulation or statistical measurement (Lim, 2025). Within this paradigm, the study employs a qualitative descriptive method to examine speech disfluencies as they naturally occur in spontaneous digital communication. Qualitative descriptive research focuses on representing phenomena as they are directly observable in their natural settings, prioritizing depth and richness of description over generalizability (Colorafi & Evans, 2016). This approach is particularly appropriate for the present study, which aims not to measure disfluency rates across a large population but to describe and interpret how disfluencies emerge and what psycholinguistic processes they reflect in authentic, self-initiated digital speech. The unit of analysis in this study is the individual disfluency token — defined as a discrete instance of a pause, filler, or hesitation marker identifiable within a spoken utterance — embedded within the broader context of each participant's spontaneous speaking episode.

The methodological approach integrates three interconnected layers: qualitative

description, qualitative content analysis, and psycholinguistic interpretation. Qualitative description provides the foundational layer, capturing disfluency phenomena as they naturally appear in the data without imposing predetermined categories. Qualitative content analysis, following Bengtsson (2016) and Lim (2025) provides the analytical layer, enabling systematic coding and categorization of disfluency tokens according to established linguistic frameworks. Psycholinguistic interpretation, drawing on Levelt's (1989) model of speech production and Kormos' (2006) framework of second language speech processing, provides the theoretical layer, connecting each categorized disfluency type to specific cognitive mechanisms operating during real-time speech production. Together, these three layers ensure that the analysis moves beyond surface description toward a theoretically grounded explanation of the cognitive processes underlying spontaneous L2 speech in digital environments. The overall research design of this study is illustrated in Table 1, which outlines the sequential stages of the research process from data source identification through to interpretation and conclusion.

Table 1. Research Design of the Study

Stage	Activity	Output
Data Source Identification	Selecting TikTok videos from 3 participants in the 'Day 1-30 Speaking English Challenge'	9 videos samples
Data Collection	Downloading videos; documenting metadata	Raw video data
Transcription	Verbatim transcription using standardized disfluency symbols	Written transcripts
Data Reduction	Extracting disfluency tokens; excluding irrelevant segments	Focused dataset
Coding & Categorization	Coding tokens by type using establish frameworks	Categorized disfluency data
Interpretation	Analyzing patterns through Levelt (1989) and Kormos (2006) frameworks	Psycholinguistic explanation
Conclusion	Synthesizing findings in relation to research objectives	Research conclusions

Data Source and Participants

The three selected participants are mind of ichi (@mindofichi), Caa (@aiss.yaa), and rahel | jobseeker buddy (@rahelrabani), all of whom were identified through a targeted search on TikTok using the keyword 'Day 1–30 Speaking English Challenge.' From the pool of creators who appeared under this search term, the three participants were selected because they met all inclusion criteria and represented a range of individual disfluency profiles suitable for qualitative psycholinguistic analysis. All three participants share similar demographic backgrounds as young adult Indonesian females and males — P1 @mindofichi and P2 @aiss.yaa are female, while P3 @rahelrabani is male — within the age range of college students and recent graduates. Gender is noted here for descriptive transparency, as it was not a criterion for inclusion or exclusion and is not treated as an analytical variable in this study. In terms of English proficiency, all three participants are assessed as operating at an intermediate level of L2 English proficiency based on auditory evaluation of their spontaneous speech, which revealed characteristics consistent with intermediate-level production: generally intelligible utterances with recognizable syntactic structures, but with frequent disfluencies reflecting ongoing processing demands at the lexical, syntactic, and phonological levels. Among the three, P1 @mindofichi demonstrated the most fluent and varied speech, followed by P2 @aiss.yaa, and P3 @rahelrabani, whose speech was characterized by the highest frequency of disfluency tokens and the exclusive use of prolongation as a hesitation strategy. No formal proficiency test was administered, as the naturalistic focus of this study prioritizes observable speech

behavior over standardized measurement. The detailed characteristics of the participants and their respective data samples are presented in Table 2.

Table 2. Characteristics of Research Participants

Participant	Username	Background	Video Duration	Episode Selected	Total Disfluency Tokens
P1	@mindofichi	Young adult, recent graduate	1-5 minutes	3	84
P2	@aiss.yaa	Young adult, college-age	1-2 minutes	3	41
P3	@rahelrabi	Young adult, recent graduate	1-4 minutes	3	102

Ethical considerations were carefully observed throughout this study. The TikTok videos analyzed in this research were publicly accessible and voluntarily uploaded by the participants on their respective accounts, making them appropriate sources of naturally occurring speech data for qualitative research purposes. The researchers made deliberate efforts to contact all three participants directly through their respective TikTok accounts prior to conducting the analysis. All three participants responded and provided explicit consent to the use of their videos as research data, and expressed interest in receiving the research findings upon publication. No personally sensitive information beyond what was voluntarily shared in the videos themselves was collected or analyzed. The participants' usernames are disclosed in this study in the interest of transparency and verifiability, consistent with standard practice in qualitative research analyzing publicly available digital content. Throughout the research process, the dignity, autonomy, and intellectual contributions of all participants as content creators were respected, and the study was conducted in accordance with the ethical principles governing qualitative research involving human participants in naturalistic digital environments.

Data Collection

Data were collected using documentary and non-participant observational techniques, which are appropriate for qualitative research analyzing naturally occurring linguistic behavior. As Lim (2025) explains, qualitative data collection embraces openness and flexibility, utilizing existing digital records such as online videos and social media content as authentic sources of naturally occurring verbal behavior, while observational approaches enable the examination of verbal behavior as it unfolds naturally without intervention or experimental manipulation.

The data collection process began with a systematic search on TikTok using the keyword 'Day 1–30 Speaking English Challenge,' which yielded a pool of creators producing English-language videos as part of this self-initiated speaking activity. From this pool, creators were evaluated against the inclusion and exclusion criteria outlined in the previous section, and three participants were selected whose videos demonstrated the richest and most analyzable instances of spontaneous L2 speech production. Once the participants were identified, their publicly accessible TikTok accounts were accessed directly, and all available videos associated with the speaking challenge were reviewed. From each participant, three speaking episodes were selected based on two primary criteria: thematic interest and linguistic richness, meaning episodes were prioritized when they contained a high density of observable disfluency phenomena across a range of categories, as well as topically engaging content that reflected genuine spontaneous communication rather than rehearsed or formulaic speech.

The spontaneity of the selected videos was assessed based on several observable indicators. The selected videos were downloaded directly from the participants' public TikTok accounts and archived to prevent data loss. All metadata — including participant

username, video title, date of upload, video duration, and episode number —was systematically documented in a metadata sheet and is listed in full in the Appendix for transparency and verifiability. Each video was then transcribed manually through repeated viewing to ensure accuracy in capturing all verbal details, including disfluency tokens, following McMullin's (2023) recommendation that audio data should be converted into written transcripts to facilitate systematic qualitative analysis. Each transcript was reviewed a minimum of three times against the original video to verify accuracy, and any ambiguous tokens were re-listened to and resolved before finalizing the transcript. Non-relevant segments — including background noise, overlapping music, unrelated commentary, or off-topic remarks — were identified during this review process and excluded from the analytical dataset to ensure focus on speech elements pertinent to psycholinguistic analysis. Pause durations were categorized auditorily following Zellner's (1994) classification. The transcription employed standardized disfluency symbols as presented in Table 3.

Table 3. Transcription Symbols and Their Descriptions

Symbol	Category	Description
(.)	Silent short pause	Less than 0,5 seconds
	Silent medium pause	0,5-1 seconds
(...)	Silent long pause	More than 1 second
uh, um	Filled pause	Vocalized non-lexical hesitation marker
like, you know	Lexicalized filler	Lexical hesitation expression
word ~ word	Repetition	Repeated word or phrase
–	False start	Abandoned utterance before repair
-	Truncated word	Incomplete word
:::	Prolongation	Elongation of a sound or syllable
//	Self-correction	Repair of previous utterance

Transcription symbols were adapted from Jefferson's (1984) conversation transcription system. Pause durations were categorized following Zellner (Zellner, 1994). Filled pauses and lexicalized fillers were identified based on Clark and Fox Tree (2002) while repetitions were analyzed following Clark and Wasow (1998). False starts, truncated words, and self-corrections were identified based on Levelt (1989) and Kormos (2006). Prolongations were identified following Lickley (2015).

Data Analysis

This study employed qualitative content analysis to examine the presence, types, and underlying psycholinguistic factors of disfluencies in the participants' L2 speech. Following Lim (2025) this approach enables the systematic categorization and interpretation of verbal data through open, axial, and selective coding, ensuring that the analysis moves beyond surface description to uncover deeper patterns and meanings embedded in the data. According to Bengtsson (2016), qualitative content analysis involves three interconnected processes: reducing data to manageable units, coding patterns into meaningful categories, and linking findings to established theoretical frameworks. The analysis in this study proceeded through four sequential stages. In the first stage, data reduction was carried out by extracting utterances containing the target phenomena — namely pauses, filled pauses, and hesitation markers — from each transcript. Non-relevant segments such as background noise, unrelated commentary, or off-topic remarks were excluded to ensure that the dataset focused solely on speech elements pertinent to psycholinguistic analysis.

In the second stage, each disfluency token was coded according to established

psycholinguistic frameworks and explicit classification criteria. The coding procedure involved systematic line-by-line examination of each transcript alongside repeated listening to the corresponding video segments to ensure that each token was accurately identified and categorized. A token was classified as a silent pause when a perceptible break in speech occurred without vocalization, categorized by duration following Zellner's (1994) thresholds: short (under 0.5 seconds), medium (0.5–1 second), and long (over 1 second). Filled pauses were identified as non-lexical vocalizations such as 'uh' and 'um' following Clark and Fox Tree (2002).

In the third stage, the coded disfluency tokens were organized into analytical categories corresponding to the two research questions — the types of disfluency produced and the psycholinguistic factors underlying their occurrence — allowing each utterance to be interpreted within a structured theoretical framework. In the fourth stage, the categorized data were interpreted through relevant psycholinguistic theories of speech production, drawing primarily on Levelt's (1989) model of conceptualization, formulation, articulation, and self-monitoring, as well as Kormos' (2006) framework of second-language speech processing.

RESULTS AND DISCUSSION

Results

This section presents the findings of the analysis, organized into two main parts corresponding to the two research questions. The first part describes the types of speech disfluencies identified in the participants' spontaneous L2 speech, while the second part explains the underlying psycholinguistic factors that triggered their occurrence. The analysis was conducted on nine video samples from three Indonesian TikTok users — P1 @mindofichi, P2 @aiss.yaa, and P3 @rahelrabani — each contributing three salient speaking episodes from the 'Day 1–30 Speaking English Challenge.' Table 4 presents the complete distribution of disfluency tokens across participants and categories, providing the quantitative foundation for the analyses that follow.

Table 4. Distribution of Speech Disfluencies Across Participants

Category	P1 @mindofichi	P2 @aiss.yaa	P3 @rahelrabani	Total
Silent short pause (.)	13	8	12	33
Silent medium pause (..)	16	15	1	32
Silent long pause (...)	3	3	0	6
Filled pause (uh,um)	25	8	30	63
Lexicalized filler (like, you know)	4	0	5	9
Repetition (~)	11	2	7	20
False start (–)	3	0	1	4
Truncated word (–)	2	0	0	2
Prolongation (:::)	0	0	43	43
Self-correction (/ /)	7	5	3	15
Total	84	41	102	227

Before proceeding to the detailed analysis of each disfluency category, an overview of the overall findings is presented here. The analysis identified a total of 227 disfluency tokens distributed across three major categories: pauses, hesitations, and lexicalized fillers. In terms of disfluency types, three key findings merit particular attention. First, pauses emerged as the most dominant category, accounting for 171 instances or approximately 75% of all tokens, with filled pauses (63 instances) being more frequent than silent pauses across all duration subcategories, suggesting that participants more frequently relied on vocalized hesitation markers than on silence to manage real-time planning demands. Second, hesitations constituted the second largest category with 84 instances, most notably

dominated by prolongation (43 instances), which was exclusively produced by P3 @rahelrabani and absent in P1 and P2 — making it the most distinctive individual disfluency pattern in this dataset and reflecting a speaker-specific strategy of stretching sounds rather than falling silent to manage cognitive load. Third, lexicalized fillers were the least frequent category with only 9 instances appearing in two of the three participants, suggesting that these speakers relied on other disfluency strategies to manage speech planning demands.

Types of Speech Disfluency

Pauses

Pauses were the most frequently occurring disfluency type identified in this study, with a total of 171 instances across all three participants. Following Zellner's (Zellner, 1994) classification, pauses were categorized into two major types: silent pauses and filled pauses. Silent pauses were further divided into three sub-categories based on duration, namely short pauses (33 instances), medium pauses (32 instances), and long pauses (6 instances), while filled pauses accounted for 63 instances. Among the three participants, P3 @rahelrabani produced the highest number of pauses (80), followed by P1 @mindofichi (57) and P2 @aiss.yaa (34), suggesting considerable variation in pause behavior across participants. Table 5 presents selected examples of pauses identified in the participants' spontaneous L2 speech.

Table 5. Examples of Pauses Found in Participants' Utterances

Code	Participant	Topic	Timestamp	Utterance	Disfluency Type
P1/V1/U1	@mindofichi	Why Overthinking Kills Creativity	00:50-01:00	"and maybe things like that (..) is what (..) makes us (...) overthinking you know?"	Silent medium pause (..), Silent long pause (...)
P1/V2/U1	@mindofichi	A Review, Being Perceived by Others, Going Live	01:28-01:34	"I feel like that would uh make me (..) a better person"	Filled pause (uh), Silent short pause (.)
P2/V1/U1	@aiss.yaa	How I Learn My English	01:32-01:39	"so shadowing method um for example you watch (..) um Hollywood movie"	Filled pause (um), Silent medium pause (..)
P2/V2/U1	@aiss.yaa	Little Advice	01:04-01:14	"I'm not forgetting my problems but(...) it was something like finally (..) released from my mind"	Silent long pause (...), Silent medium pause (..)
P3/V1/U1	@rahelrabani	Just Took the ELT	00:40-00:51	"so I prepare uh this test uh using (..) AI which is ChatGPT from	Filled pause (uh), Silent short pause (.)

Code	Participant	Topic	Timestamp	Utterance	Disfluency Type
				uh exploring about sentences, grammar and uh anything”	
P3/V1/U2	@rahelrabani	Just Took the ELT	01:02-01:06	“and then (.) tonight I just (.) got an email”	Silent medium pause (.), Silent short pause (.)

As presented in Table 5, pauses were identified in various forms across all three participants. In P1/V1/U1, two silent medium pauses (.) and one silent long pause (...) reflect different degrees of planning demand — the medium pauses suggest momentary syntactic interruptions, while the long pause before ‘*overthinking*’ indicates a more substantial conceptual retrieval delay, consistent with Zellner’s (Zellner, 1994) observation that longer pauses cluster before semantically significant words. A similar pattern appears in P2/V2/U1, where the long pause after ‘*but*’ in ‘*I’m not forgetting my problems but (...) it was something like finally (.) released from my mind*’ marks a transition between contrasting ideas, and the subsequent medium pause before ‘*released*’ reflects effortful lexical search.

Fillers

Lexicalized fillers were the second disfluency type identified in this study, with a total of 9 instances found across two participants. Following Clark and Fox Tree’s (2002) framework, lexicalized fillers refer to words or expressions such as ‘like’, ‘you know’, and ‘something like that’ that function as hesitation devices during speech planning. Kosmala and Crible (2022) further demonstrate that filled pauses carry a dual status in spontaneous speech — functioning simultaneously as hesitation markers and as communicative devices that signal upcoming delays to the listener — a distinction that is particularly evident in L2 speakers who rely on these expressions to manage formulation demands while maintaining conversational flow.

P1 @mindofichi produced 4 instances while P3 @rahelrabani produced 5 instances, whereas P2 @aiss.yaa produced no lexicalized fillers throughout the recorded speech samples. Table 6 presents selected examples of lexicalized fillers identified in the participants’ spontaneous L2 speech.

Table 6. Examples of Lexicalized Fillers Found in Participants’ Utterances

Code	Participant	Topic	Timestamp	Utterance	Disfluency Type
P1/V1/U2	@mindofichi	Why Overthinking Kills Creativity	01:17-01:32	“and those things actually kills your creativity you know, and (.) I think we should all just (~) just start doing it you know”	Lexicalized filler (you know)
P3/V3/U2	@rahelrabani	Talk about My Hobby	02:04-02:11	“there is no:::much video that made my content FYP, you know something like that”	Lexicalized filler (you know something like that)

As presented in Table 6, lexicalized fillers were identified in the spontaneous L2 speech of two participants, functioning as hesitation devices that serve both communicative and pragmatic purposes (Clark, 2002). In P1/V1/U2, the lexicalized filler ‘*you know*’ appeared twice within a single utterance produced by P1 @mindofichi: ‘*and those things actually kills your creativity you know, and (..) I think we should all just (~) just start doing it you know.*’ The first occurrence of ‘*you know*’ following ‘*creativity*’ functions as an engagement marker, signaling the speaker’s attempt to ensure listener comprehension and establish shared understanding. The second occurrence at the end of the utterance, however, serves a mitigating function, softening the assertiveness of the preceding statement and inviting the listener’s agreement. This dual functional pattern suggests that lexicalized fillers in L2 speech are not merely hesitation phenomena but also serve interpersonal and pragmatic functions during spontaneous communication.

Hesitation

Hesitations constituted the second largest category of disfluencies identified in this study, with a total of 84 instances across all three participants. Following the frameworks of Clark and Wasow (1998), Levelt (1989), Kormos (2006), and Lickley (2015), hesitations were categorized into five types: repetitions (20 instances), false starts (4 instances), truncated words (2 instances), prolongations (43 instances), and self-corrections (15 instances). Among the three participants, P3 @rahelrabani produced the highest number of hesitations (54), largely driven by the dominance of prolongation, followed by P1 @mindofichi (23) and P2 @aiss.yaa (7). Given that P1 @mindofichi produced the most diverse range of hesitation types, multiple utterances from this participant are presented in Table 5 to represent each distinct hesitation category, while P2 @aiss.yaa and P3 @rahelrabani are represented by their most prominent hesitation types, respectively. Table 7 presents selected examples of hesitations identified in the participants’ spontaneous L2 speech.

Table 7. Example of Hesitations Found in Participants’ Utterances

Code	Participant	Topic	Timestamp	Utterance	Disfluency Type
P1/V2/U2	@mindofichi	A Review, Being Perceived by Others, Going Live	01:51-01:56	“but I’ve been thinking oh why ~ why not should // why not I try that too”	Repetition (~), Self-correction (//)
P1/V3/U1	@mindofichi	A Review on My Confidence Levels	02:05-02:07	“I’ve – I really wanted to do beauty”	False start (-)
P1/V1/U3	@mindofichi	Why Overthinking Kills Creativity	01:35-01:38	“whether it’s perf - perfect enough”	Truncated word (-)
P3/V3/U1	@rahelrabani	Talk about My Hobby	00:25-00:33	“I (:::) could (:::) express what (:::) I feel what (:::) genre or what ~ what taste of music”	Prolongation (:::). Repetition (~)
P2/V1/U2	@aiss.yaa	How I Learn My English	02:15-02:20	“and (..) it help us // it helps us in pronunciation”	Self-correction (//), Silent medium pause (..)

As presented in Table 7, hesitations were identified in various forms across all three participants, reflecting diverse strategies employed by L2 speakers to manage real-time speech planning and production demands. In P1/V2/U2, P1 @mindofichi repeated ‘why’ before repairing the syntactically incomplete ‘why not should’ to ‘why not I try that too.’ The repetition functioned as a planning bridge while the syntactic frame was being reorganized, and the subsequent self-correction reflects active self-monitoring during formulation (Kormos, 2006; Levelt, 1989). In P1/V3/U1, the abandoned present perfect ‘I’ve’ in favor of the simpler past tense ‘I really wanted to do beauty’ suggests that the speaker recognized mid-production that the initially selected syntactic structure was overly complex, demonstrating the simultaneous nature of planning and articulation in spontaneous L2 speech (Levelt, 1989).

Underlying Psycholinguistic Factors

Conceptual Planning Difficulty

Conceptual planning difficulty refers to disruptions arising when speakers struggle to formulate message content before linguistic encoding begins (Levelt, 1989). This factor was evidenced across all three participants, particularly where pauses and prolongations clustered before semantically significant elements. As previously shown in Table 5 in P1/V1/U1, the recurring medium pauses and long pause before ‘overthinking’ suggest that the conceptual relationship between ideas had not yet been fully formed at the preverbal stage. The vagueness of ‘things like that’ further indicates incomplete conceptualization prior to articulation, consistent with Schnadt and Corley’s (2006) finding that conceptual access difficulties trigger prolonged pauses as speakers postpone articulation until the message is sufficiently formed. Similarly, P2/V2/U1’s long pause after ‘but’ and the hedging expression ‘something like’ reflect cognitive effort in transitioning between contrasting ideas while searching for precise conceptual expression. Most strikingly, the sustained prolongations in P3/V3/U2 — distributed across an utterance about emotional expression and musical preference — alongside the subsequent repetition ‘what ~ what’ confirm that even after multiple prolongations, the conceptual representation remained insufficiently formed.

Lexical retrieval Effort

Lexical retrieval effort refers to disruptions occurring when the intended concept is already formed but the corresponding lexical item cannot be accessed efficiently (Levelt, 1989) — distinguishing it from conceptual planning difficulty, which operates at the preverbal level. As previously shown in Table 5 in P1/V2/U1, the filled pause ‘uh’ before ‘make me’ and the subsequent short pause before ‘a better person’ suggest that while the idea of self-improvement was already conceptualized, selecting the most appropriate evaluative expression required additional processing time, consistent with Griffin and Bock’s (2000) observation that lexical competition among semantically related options extends access time. In P2/V1/U1, the recurring ‘um’ in ‘so shadowing method um for example you watch (..) um Hollywood movie’ indicates that the target item ‘Hollywood movie’ could not be immediately retrieved. Most notably, the four clustered ‘uh’ instances in P3/V1/U1 reflect sustained retrieval effort throughout a technically dense utterance — the domain-specific terms ‘AI,’ ‘ChatGPT,’ and ‘exploring about sentences’ likely required more effortful lexical activation, consistent with Kormos’ (2006) observation that low-frequency and domain-specific words are particularly susceptible to retrieval delays in L2 speech production.

Syntactic & Phonological Encoding

Syntactic and phonological encoding demands refer to the cognitive effort required to construct grammatical structures and retrieve sound forms during formulation (Levelt, 1989) — emerging at a later stage than conceptual planning and lexical retrieval. Syntactic encoding demands were evidenced in P1/V2/U2, where the incomplete ‘why not

should’ was repaired mid-utterance to *‘why not I try that too,’* indicating that the syntactic frame had not been fully constructed before articulation began. Similarly, in P1/V3/U1, the abandoned present perfect *‘I’ve’* in favor of the simpler past tense demonstrates that syntactic planning and articulation occur simultaneously, allowing speakers to cancel and restructure frames when the selected structure proves too complex (Bock & Levelt, 1994).

Working Memory Load

Working memory load refers to the cognitive burden experienced when speakers must simultaneously store and manipulate multiple pieces of linguistic information during speech planning and production (Baddeley, 2003). Unlike conceptual planning difficulty or lexical retrieval effort, which involve specific bottlenecks at particular stages, working memory load manifests when the overall demand of managing concurrent cognitive operations exceeds available capacity, resulting in multiple disfluencies distributed across informationally dense utterances. In P1/V1/U2, the medium pause and repetition *‘just ~ just’* within a two-proposition utterance function as tempo-holding strategies, suggesting that maintaining coherence across connected clauses strained the speaker’s working memory capacity (Towell et al., 1996). Similarly, in P3/V1/U1, the four *‘uh’* instances distributed before each new informational unit — the preparation process, the tool used, and the material studied — indicate that the speaker’s working memory was continuously taxed by the need to retain prior content while planning upcoming elements.

Self-Monitoring

Self-monitoring activity refers to the perceptual and cognitive processes through which speakers evaluate and repair their own speech during production (Levelt, 1989). This factor was evidenced across participants through self-corrections that occurred rapidly and without prolonged hesitation, suggesting the operation of an efficient internal monitoring system. As previously shown in Table 7, in P2/V1/U2, the immediate repair of *‘it help us’* to *‘it helps us’* without a significant pause reflects an efficiently operating internal monitor, consistent with Levelt’s (1989) perceptual loop theory, whereby speakers continuously compare intended and produced output and initiate repairs upon detecting discrepancies. A similar pattern was observed in P1/V2/U2, the repair of *‘why not should’* to *‘why not I try that too’* further demonstrates that self-monitoring operates across both grammatical and syntactic levels. Kormos (1999, 2006) notes that L2 speakers rely more heavily on self-monitoring than native speakers due to less automatized grammatical encoding — explaining the relatively high frequency of self-corrections — 15 instances in total — observed across all three participants in this study.

Discussion

The findings of this study offer several interpretive insights that extend beyond the descriptive categorization of disfluency types and psycholinguistic triggers presented in the preceding sections. This discussion situates the findings within the broader context of L2 disfluency research, examines their theoretical and pedagogical implications, and reflects on the contribution of this study to the understanding of spontaneous L2 speech in digitally mediated environments.

The dominance of pauses, particularly filled pauses, across all three participants is consistent with previous research on L2 disfluency, which has consistently identified pauses as the most frequent disruption in spontaneous second language speech (Feng & Guo, 2022; Zellner, 1994). While prior studies such as Liao (2023) and Wang (2021) observed similar patterns in controlled classroom and task-based speech contexts, the present findings demonstrate that this dominance persists in self-initiated digital speaking environments where no external scaffolding, preparation time, or interlocutor support is provided. This parallel suggests that the cognitive demands underlying pause production

are not merely artifacts of experimental or instructional conditions but are fundamental features of L2 speech production that manifest consistently across diverse speaking contexts. More significantly, the present study extends the applicability of Levelt's (1989) and Kormos' (2006) frameworks — originally developed and validated in laboratory and classroom settings — to naturally occurring, user-generated digital speech, demonstrating that these theoretical models retain their explanatory power when applied to contemporary technology-mediated communication environments.

The exclusive presence of prolongation in P3 @rahelrabani's speech is particularly significant from a broader theoretical perspective. Rather than merely confirming that individual speakers differ in their disfluency profiles, this finding raises an important question about the relationship between speaker-specific cognitive strategies and the digital speaking context itself — specifically, whether the absence of interlocutor pressure in self-recorded TikTok videos allows individual stylistic tendencies to emerge more prominently than they might in face-to-face or task-based speech settings (Kormos, 2006; Segalowitz, 2016). The contrast between P3's prolongation-dominant profile and P1's diverse hesitation repertoire further underscores the importance of examining disfluency at the individual level rather than treating L2 speakers as a homogeneous group — a perspective that has significant implications for how L2 speaking fluency is researched, assessed, and developed in both classroom and digital contexts.

The pervasive presence of self-monitoring across all three participants, regardless of individual speaking style, carries direct implications for how L2 fluency is assessed and developed. If self-monitoring is an inherent feature of intermediate-level L2 speech production rather than a marker of low proficiency, then assessment frameworks that penalize self-corrections may systematically underestimate speakers' actual communicative competence (Segalowitz, 2016). This reconceptualization has direct implications for L2 fluency assessment, suggesting that evaluators should distinguish between disfluencies that reflect genuine cognitive effort — such as pauses before semantically complex content or prolongations during abstract conceptualization — and those that may indicate more fundamental processing difficulties. A nuanced, psycholinguistically informed approach to fluency assessment would better capture the cognitive realities of L2 speech production and provide more accurate and fair evaluations of L2 speakers' communicative competence.

From a pedagogical standpoint, the findings suggest that self-initiated digital speaking practice — such as TikTok speaking challenges — represents a valuable and underutilized resource for L2 fluency development. Unlike controlled classroom tasks that provide scaffolding, preparation time, and teacher support, TikTok speaking challenges require participants to independently manage all stages of speech production in real time, creating authentic conditions for the development of cognitive fluency and processing automaticity (Hanzawa, 2024). The public nature of TikTok further introduces an element of communicative accountability that may motivate speakers to sustain their speaking practice over time, as evidenced by the participants' consistent engagement across multiple episodes of the 30-day challenge. Language educators and curriculum designers may therefore consider incorporating self-initiated digital speaking activities into L2 instruction as a complement to traditional classroom-based speaking tasks, particularly for learners who have limited opportunities for naturalistic English use outside formal educational settings — a circumstance that characterizes many Indonesian EFL learners (Rahmawati et al., 2023).

More broadly, TikTok and similar short-form video platforms represent an emerging and largely untapped context for psycholinguistic research on L2 speech. The platform's combination of spontaneity, public visibility, self-initiation, and minimal editing creates a uniquely naturalistic window into the cognitive processes underlying L2 speech production

— one that is difficult to replicate in controlled experimental settings. As digital communication continues to reshape the landscape of language use and language learning, researchers and educators alike stand to benefit from greater engagement with the authentic, user-generated speech data that platforms such as TikTok make available. This study represents a step in that direction, demonstrating that the cognitive dimensions of L2 fluency development are not confined to the classroom but are actively at work in the digital spaces where many young adult L2 speakers choose to practice, challenge, and develop their spoken English.

CONCLUSION

This study set out to describe the types of speech disfluencies produced by three Indonesian TikTok users in their spontaneous English speech and to explain the underlying psycholinguistic factors that triggered their occurrence. Both research questions have been answered within the boundaries of this specific dataset of nine video samples. With respect to disfluency types, pauses emerged as the dominant category, with filled pauses being the most frequent subtype, followed by hesitations — most notably prolongation, which was exclusively produced by P3 @rahelrabani — and lexicalized fillers as the least frequent category. With respect to psycholinguistic triggers, five cognitive factors were identified as operating at different stages of Levelt's (1989) speech production model: conceptual planning difficulty, lexical retrieval effort, syntactic and phonological encoding demands, working memory load, and self-monitoring behavior.

The primary contribution of this study lies in its demonstration that Levelt's (1989) and Kormos' (2006) psycholinguistic frameworks retain their explanatory power when applied to naturally occurring, user-generated L2 speech in a digital context — extending their scope beyond laboratory and classroom settings to technology-mediated communication environments. The findings further contribute to a reconceptualization of disfluency not as a marker of low proficiency but as a meaningful window into the cognitive processes underlying real-time L2 speech production, with implications for how fluency is assessed and developed in both pedagogical and digital communication contexts.

This study acknowledges several limitations. The findings are based on a small, purposively selected sample of three participants and nine video samples, which precludes broad generalization to all Indonesian L2 speakers or TikTok users. Additionally, pause durations were categorized auditorily rather than through acoustic measurement tools such as Praat, which may introduce subjectivity in classification. Future research is encouraged to address these limitations through larger and more diverse participant samples, acoustic measurement tools, and longitudinal designs that track disfluency development over time. As Hanzawa (2024) demonstrates, L2 fluency development is a gradual and context-sensitive process, and self-initiated digital speaking practice — such as TikTok speaking challenges — holds considerable promise as an authentic environment for fostering this development beyond the boundaries of the formal classroom.

RECOMMENDATION

Future research is encouraged to examine larger and more diverse samples of L2 speakers on digital platforms and to employ acoustic measurement tools such as Praat for more precise pause analysis. Longitudinal studies tracking disfluency patterns over time may further illuminate the development of fluency and automaticity in L2 speakers engaged in self-directed digital speaking practice. Additionally, comparative studies involving speakers from different L1 backgrounds or at varying proficiency levels would contribute to a more nuanced understanding of how disfluency patterns vary across learner populations in technology-mediated communication contexts.

ACKNOWLEDGMENT

In preparing this manuscript, the authors made use of several AI-assisted tools for specific purposes. Claude (Anthropic) was used to assist in drafting and revising sections of the manuscript. ChatGPT (OpenAI) was used to assist in identifying potentially relevant references, all of which were subsequently verified independently through academic databases. Grammarly was used for language editing and grammar checking. These tools were used to support the writing process; all intellectual content, analytical decisions, data interpretation, and conclusions remain the sole responsibility of the authors. The authors take full accountability for the accuracy and integrity of the work presented in this manuscript.

REFERENCES

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839. <https://doi.org/10.1038/nrn1201>
- Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, 2, 8–14. <https://doi.org/10.1016/j.npls.2016.01.001>
- Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In *Handbook of psycholinguistics*. Academic Press.
- Chang, H., & Windeatt, S. (2024). Fluency issues in L2 academic presentations: Linguistic, cognitive, and psychological influences on pausing behaviour. *Language Teaching Research*. <https://doi.org/10.1177/13621688241233718>
- Clark, H. H. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1), 73–111. [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–242. <https://doi.org/10.1006/cogp.1998.0693>
- Colorafi, K. J., & Evans, B. (2016). Qualitative descriptive methods in health science research. *HERD: Health Environments Research & Design Journal*, 9(4), 16–25. <https://doi.org/10.1177/1937586715614171>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. <https://doi.org/10.1037/0033-295X.93.3.283>
- Eren, Ö., Kılıç, M., & Bada, E. (2022). Fluency in L2: Read and spontaneous speech pausing patterns of Turkish, Swahili, Hausa, and Arabic speakers of English. *Journal of Psycholinguistic Research*, 51(2), 237–253. <https://doi.org/10.1007/s10936-021-09822-y>
- Feng, R., & Guo, Q. (2022). Second language speech fluency: What is in the picture and what is missing. *Frontiers in Psychology*, 13, Article 859213. <https://doi.org/10.3389/fpsyg.2022.859213>
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10.1111/1467-9280.00255>
- Hanzawa, K. (2024). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, 28(3), 816–838. <https://doi.org/10.1177/13621688211008693>
- Jefferson, G. (1984). Transcription notation. In *Structures of social action*. Cambridge University Press.
- Jelčić, M. (2024). *Pauses in spontaneous speech production in the media*.
- Kirkland, A., Lameris, H., Szekely, E., & Gustafson, J. (2022). Where's the *uh*, hesitation? The interplay between filled pause location, speech rate, and fundamental frequency

- in perception of confidence. In *Interspeech 2022* (pp. 4990–4994). <https://doi.org/10.21437/Interspeech.2022-10973>
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, 49(2), 303–342. <https://doi.org/10.1111/0023-8333.00090>
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.
- 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>
- Kress, G. (2000). Multimodality: Challenges to thinking about language. *TESOL Quarterly*, 34(2), 337–340. <https://doi.org/10.2307/3587959>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Liao, J. (2023). Disfluency and self-repair in presentational and interpersonal speech modalities. *Foreign Language Annals*, 56(2), 401–427. <https://doi.org/10.1111/flan.12658>
- Lickley, R. (2015). Fluency and disfluency. In *The Oxford handbook of speech production* (pp. 445–458). Oxford University Press.
- Lim, W. M. (2025). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal*, 33(2), 199–229. <https://doi.org/10.1177/14413582241264619>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.
- McMullin, C. (2023). Transcription and qualitative methods: Implications for third sector research. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 34(1), 140–153. <https://doi.org/10.1007/s11266-021-00400-3>
- Mizera, G. J. (2006). *Working memory and L2 oral fluency*.
- Rahmawati, A., Syafei, M., & Prasetyanto, M. A. (2023). Improving speaking skills through TikTok application: An endeavour of utilizing social media in higher education. *Journal of Languages and Language Teaching*, 11(1), 137–143. <https://doi.org/10.33394/jollt.v11i1.6633>
- Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual, and planning-based factors on disfluency production. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 750–755).
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54(2), 79–95. <https://doi.org/10.1515/iral-2016-9991>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119. <https://doi.org/10.1093/applin/17.1.84>
- Wang, Y. (2021). A study on the use of hesitation markers in varied-level EFL learners' L2 speaking process. *Open Journal of Modern Linguistics*, 11(5), 823–840. <https://doi.org/10.4236/ojml.2021.115063>
- Zellner, B. (1994). *Pauses and the temporal structure of speech*.
- Zulhemindra, Z., Munir, S., Yulnetri, Y., & Putra, K. P. (2022). Investigating filled pauses found in English students' conversation. *Ahmad Dahlan Journal of English Studies*, 9(1), 14–26. <https://doi.org/10.26555/adjes.v9i1.35>