Studies in Second Language Acquisition (2025), **47**, 560–591
doi:10.1017/S0272263125000245

RESEARCH ARTICLE

Disfluency doesn't happen in isolation

Exploring how individual disfluency features co-occur in L2 speaking performances

Xun Yan[1,2], Ping-Lin Chuang[3], Yulin Pan[1], Huiying Cai[1], Shelley Staples[4] and Mariana Centanin Bertho[5]

[1]Department of Linguistics, School of Literatures, Cultures and Linguistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA; [2]Beckman Institute for Advanced Science and Technology, Urbana, IL, USA; [3]Duolingo Inc, Pittsburgh, PA, USA; [4]Department of English, University of Arizona, Tucson, AZ, USA and [5]Department of Spanish and Portuguese, Yale University, New Haven, CT, USA
Corresponding author: Xun Yan; Email: xunyan@illinois.edu.

(Received 28 November 2023; Revised 15 February 2025; Accepted 04 March 2025)

Abstract

The construct of second language (L2) utterance fluency is typically operationalized through various individual temporal features. However, in natural speech, fluency (or disfluency) is often characterized by the clustering of multiple temporal features, collectively revealing the speaker's effort in speech production or disfluency recovery. In this study, we explore the co-occurrence patterns of disfluency features in L2 speech and their associations with speakers' L2 oral proficiency. We initially segmented all speech samples into analysis of speech (AS)-units. Within each AS-unit, six individual fluency features were manually coded, standardized, and subsequently subjected to a hierarchical-based *k*-means cluster analysis to examine their co-occurrence patterns. The results revealed four distinct disfluency clusters. A subsequent qualitative analysis of disfluencies in each cluster revealed distinct distributional patterns, disfluency makeup, and communicative functions. Additionally, the proportions of different disfluency clusters were significantly influenced by speakers' proficiency level, first language background, and their interaction. These findings carry implications for L2 speaking research in general, shedding light on the intricate nature of speech fluency and presenting an alternative approach to the operationalization of this multidimensional construct.

Keywords: disfluency cluster; L2 oral proficiency; L2 speech fluency

Introduction

Speech fluency is a multifaceted and dynamic construct that plays a pivotal role in effective communication and language acquisition. On the one hand, human speech

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

universally features disfluencies (Clark & Fox Tree, 2002). On the other hand, within the realm of second language (L2) research, fluency stands as a central element of language proficiency, reflecting a speaker's capacity to effortlessly, smoothly, and appropriately pace their speech (Segalowitz, 2010). Abundant empirical evidence has revealed that utterance fluency features can distinguish between first language (L1) and L2 speakers (Götz, 2013) or predict proficiency levels of L2 speakers (de Jong, 2018; Ginther, Dimova & Yang, 2010; Koizumi, Jeon & In'nami, 2022). Higher utterance fluency typically correlates with higher proficiency, and an excessive amount of disfluencies can indicate developing L2 oral proficiency (de Jong, 2018; Ginther et al., 2010; Koizumi et al., 2022; Lennon, 1990; Yan, Kim & Kim, 2021). Skehan (2003) further categorized L2 utterance fluency into speed, breakdown, and repair features, providing a valuable framework for operationalizing speech fluency across various dimensions. However, despite the practical utility of this framework, it does not provide a full picture of the nature of (dis)fluencies in L2 speech.

Examining the nature of disfluencies can provide insights into the cognitive processes involved in speech production. In L2 speech, Segalowitz (2010) identified seven vulnerability points in fluency that occur at various stages of speech production: microplanning, grammatical encoding, lemma retrieval, morphophonological encoding, phonetic encoding, articulation, and self-perception. His specifications suggest that disfluencies predominantly stem from difficulties in lexico-grammar and content processing, and this can be a sign of developing L2 oral proficiency. Influenced by this conceptual model, recent L2 fluency research in both natural and assessment contexts has incorporated more fine-grained features related to pausing (e.g., frequency, type, and location), leading to the development of more nuanced breakdown features (Kahng, 2014, 2018; Park, 2016; Tavakoli, Nakatsuhara & Hunter, 2020; Yan et al., 2021). These fine-grained, micro-fluency features appear to offer more nuanced information and implications for the cognitive processes of speech production and oral proficiency for L2 speakers.

Despite advancements in the refinement of individual fluency features, in natural speech, breakdowns and repairs tend to co-occur during instances of disfluencies, offering insights into the effort required for disfluency recovery (Shriberg, 1994). Thus, there is a need for further research to explore the co-occurrence or clustering of disfluency features in L2 speech and examine whether and the extent to which the clustering effects of disfluency features are associated with L2 oral proficiency. This study explores the co-occurrence of disfluency features (e.g., number of reformulations, number of repetitions) in an International English Language Testing System (IELTS) speaking performance corpus consisting of speakers from different L1 backgrounds and various proficiency levels, to explore the relationship between speech (dis)fluency and oral proficiency from a different perspective.

Construct of L2 speech fluency

With L1 speakers in mind, Levelt (1993) conceptualizes speech production as a cognitive process involving three stages: conceptualization, formulation, and articulation. In the conceptualization stage, a preverbal message is formed using world knowledge. This message is then processed for lexical, grammatical, and phonetic encoding during formulation and articulated into speech while being monitored by the speaker. Segalowitz (2010) adapted this model for L2 speakers, identifying seven vulnerability points that may trigger disfluencies: (a) microplanning, (b) grammatical encoding, (c) lemma retrieving, (d) morphophonological encoding, (e) phonetic

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

encoding, (f) articulation, and (g) self-perception. Findings from studies that compared L2 speakers across proficiency levels or with L1 speakers appear to support this view, revealing that pause type and location are associated with lexico-grammar or content processing and can explain the differences between L1 and L2 speech or in L2 speech across proficiency levels (e.g., Kahng, 2014, 2018; Skehan & Foster, 2008; Tavakoli et al., 2020; Yan et al., 2021). Segalowitz also outlined three fluency types: cognitive, utterance, and perceived fluency. Cognitive fluency reflects mental efficiency in speech production, utterance fluency captures measurable speech flow features, and perceived fluency relates to the listener's impression of smoothness and naturalness. These dimensions are interconnected, with cognitive fluency influencing utterance fluency, which shapes perceived fluency.

In terms of operationalization, Skehan (2003) divided utterance fluency into three subdimensions: speed, breakdown, and repair fluency. Speed fluency measures speech rapidity through speech rate, mean length of utterance, and articulation rate. Breakdown fluency assesses pausing features, including frequency, duration, and location of filled (e.g., "uh", "um", "you know") and silent pauses. Repair fluency examines repair behaviors, such as repetitions, reformulations, and their outcomes (e.g., success rate; Yan et al., 2021). Recent research has also focused on further differentiating individual utterance fluency features to clarify their specific roles within the broader construct of fluency. Tavakoli et al. (2020) highlighted the importance of composite features, which integrate multiple dimensions (e.g., speech rate, mean length of run) to capture fluency more comprehensively. Yan et al. (2021) distinguished between macro-and micro-fluency features based on operational complexity. They considered macro-fluency features as features that can be easily automated through duration or frequency measures (e.g., articulation rate, number of pauses); in contrast, micro-fluency features were those whose computation requires fine-grained differentiations in breakdown and repair type and location (e.g., features related to pause location) or combines more than one subdimension (e.g., mean length of run). Although this might be a crude distinction, existing research on L2 disfluency seems to suggest that while macro-fluency features predict L2 oral proficiency well, micro-disfluency features can offer additional, nuanced insights into L2 speech production process and oral proficiency in both assessment and natural settings (Kahng, 2014, 2018; Yan et al., 2021).

Nature of disfluencies in L1 speech

Disfluencies in L1 speech, defined as unintended interruptions that do not contribute to the propositional content of an utterance, are a natural aspect of speech production (Fox Tree, 1995). In L1 speech, disfluencies occur approximately 6 to 10 times per 100 words (Fox Tree, 1995; Shriberg, 1994) and are typically classified into four main types: filled pauses, silent pauses, repetitions, and repairs (Bergmann, Sprenger & Schmid, 2015; Götz, 2013; Maclay & Osgood, 1959; Shriberg, 1994). Filled pauses include non-lexical fillers (e.g., "uh" and "um"; de Jong, 2016) as well as lexical fillers (e.g, "like" and "you know"; Carney, 2022; Clark & Fox Tree, 2002). Silent pauses, on the other hand, represent unfilled hesitations longer than typical pauses in fluent speech. Repetitions involve the restatement of words or phrases, while repairs involve revisions or corrections to prior speech (Maclay & Osgood, 1959; Shriberg, 1994).

In L1 speech, disfluencies often reflect cognitive challenges during speech production, such as retrieving or organizing linguistic or conceptual information (Clark, 1996; Fraundorf & Watson, 2013). For example, filled pauses are linked to discourse-level planning, particularly at the beginning of an utterance, while silent pauses tend to indicate

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

difficulties with lexical, syntactic, or phonological processing (Bortfeld, Leon, Bloom, Schober & Brennan, 2001; Clark, 1996; Clark & Fox Tree, 2002). Repetitions and repairs typically occur after problematic material has been articulated, serving to resolve errors or clarify intent (Clark & Wasow, 1998). That said, sociocultural and individual factors also influence the occurrence and nature of disfluencies. Filled pauses, for instance, can signal imminent delays to the listener, convey feelings like anxiety or confidence, or function as a mechanism to hold the conversational floor (Brennan & Williams, 1995; Clark & Fox Tree, 2002; Maclay & Osgood, 1959). Additionally, variables such as gender, age, conversational context, and topic can impact the frequency and type of disfluencies (Bortfeld et al., 2001; Choo, Smith & Seitz, 2024; Schneider, 2014).

L2 versus L1 speech disfluencies: Higher frequency and distinct distributional patterns

Compared to L1 speech disfluencies, L2 speech disfluencies often occur with greater frequency and intensity due to the lack of oral proficiency and additional cognitive demands of managing two linguistic systems (Bergmann et al., 2015; Choo et al., 2024; Lennon, 1990). L1 speech generally exhibits fewer disfluencies due to high automaticity in lexical retrieval and grammatical processing. Disfluencies tend to occur in complex or unfamiliar contexts, such as when discussing abstract topics or integrating new information (Levelt, 1993). In contrast, L2 speech typically features a higher frequency of disfluencies due to less automaticity, lower oral proficiency, and greater cognitive effort required for lexical retrieval, syntactic structuring, and pronunciation (Derwing, Munro, Thomson & Rossiter, 2009). In addition, disfluencies in L2 speech are subjected to similar cognitive and sociocultural factors as those in L1 speech, although the degree of L2 speech disfluency can also be affected by factors such as the speaker's L1 background (Eren, Kılıç & Bada, 2022), age of acquisition (Götz, 2019), task type (Liao, 2023), and language proficiency (Saito, Ilkan, Magne, Tran & Suzuki, 2018).

Frequency of occurrence aside, L2 speech disfluencies display distinct distributional patterns and functions than L1 speech disfluencies. In L1 speech, disfluencies typically arise from natural cognitive demands during speech production, such as planning, monitoring, or repairing utterances (Levelt, 1989). In terms of silent and filled pauses, disfluencies often occur at clause boundaries where speakers are planning discourse or integrating complex or unfamiliar information (Goldman-Eisler, 1968). In contrast, L2 speakers frequently pause mid-clause, reflecting the additional cognitive effort required for lexical retrieval, grammatical structuring, or pronunciation (Derwing et al., 2009; Kahng, 2014). L2 speakers need to allocate additional cognitive resources for language processing, leading to longer pauses and more interruptions within clauses (Segalowitz, 2010). The difference in mid-clause pause rate between L1 and L2 speakers also prompted scholars to speculate that disfluencies at syntactic boundaries are more reflective of discourse planning in the conceptualization stage, whereas disfluencies within syntactic boundaries are more indicative of processing difficulties at the formulation stage (e.g., Huensch, 2023). Similarly, repetitions and reformulations, while common in both L1 and L2 speech, might differ in their underlying causes. In L1 speech, repetitions often serve rhetorical purposes or aid in clarification (Clark & Wasow, 1998), whereas in L2 speech, they frequently signal hesitation or linguistic difficulty (Kormos, 1999). Additionally, repair behaviors in L2 speech are often more explicit and laborious, reflecting limited linguistic resources compared to more efficient and effective repairs by L1 speakers (Fox Tree, 1995).

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

In terms of listener perceptions of disfluencies, L1 disfluencies are often interpreted as natural markers of thought or emphasis (Brennan & Schober, 2001); however, L2 disfluencies are often perceived as indicators of lower proficiency, especially when they disrupt comprehension or fluency (Rossiter, 2009). In L2 speaking assessment, higher oral proficiency is typically associated with longer and more complex runs of speech, faster speaking rates, and fewer pauses or repairs (Ginther et al., 2010; Saito et al., 2018; Tavakoli et al., 2020). In particular, the frequency and duration of silent pauses are considered the most reliable indicators of proficiency (Koizumi et al., 2022; Suzuki, Kormos & Uchihara, 2021), while filled pauses and repairs are less consistent in their predictive power, although notable exceptions exist (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Révész, Ekiert & Torgersen, 2016). However, the prominence of silent pauses in measuring proficiency seems to contrast with how listeners perceive fluency. From an assessment perspective, listeners (i.e., test raters) may be sensitive to not only breakdowns but also repairs when judging proficiency. This is reflected in descriptors from major language proficiency frameworks like the Common European Framework of Reference for Languages (CEFR) and rating scales from tests, such as the IELTS and the Pearson Test of English (PTE) Academic. For instance, the CEFR B2 descriptor highlights speakers being "hesitant as they search for patterns and expressions" with "few noticeably long pauses," while at the B1 level, more frequent pauses for lexical and grammatical planning are evident (Council of Europe, 2001). Similarly, IELTS band 8 describes fluent speech with "only occasional repetition or self-correction" and rare hesitation due to language searches, while band 5 reflects slower speech that relies on repetition, self-correction, and pauses (IELTS, 2023). Interestingly, on the PTE Academic, at the highest level (level 5), the descriptor suggests "native-like" fluency with "no repetitions, hesitations, false starts, or non-native phonological simplifications" (PTE, 2024). These frameworks suggest that listener perception of fluency is shaped by a constellation of disfluency features, rather than solely the presence of silent pauses.

Disfluency clusters in L1 and L2 speech

One caveat of earlier research in L1 and L2 disfluency is that disfluency features tend to be examined individually. However, in natural speech, disfluency is never an isolated phenomenon and is often marked by the co-occurrence of multiple temporal features. This co-occurrence can be observed at two levels. First, at the local level, when pauses occur, they are frequently followed by repairs, an attempt to return to the original utterance (i.e., a cluster of disfluency features; Shriberg, 1994, 2001). Second, at the global level, within a speech utterance, a disfluency cluster might deplete unplanned cognitive resources, disrupt the planning and execution of the utterance stream, and consequently lead to more disfluency clusters within the remainder of the utterance (i.e., a spillover effect; Riggenback, 1991; Shriberg, 1994). In general, longer disfluency clusters tend to occur more frequently in nonfluent speakers than fluent speakers and might entail greater cognitive effort in disfluency recovery (Riggenback, 1991).

Research in L1 corpus linguistics has documented disfluency clustering effects at the local level. While corpus-based studies tend to examine specific combinations of disfluencies, most of these combinations center around non-lexical filled pauses (e.g., "uh" and "um"). The findings show that non-lexical filled pauses often co-occur with

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

silent pauses, lexical fillers, or repetitions (Clark & Wasow, 1998; Crible, Degand & Gilquin, 2017; Degand & Gilquin, 2013; Schneider, 2014), suggesting that these disfluency clusters may serve complementary cognitive or interactional functions, such as holding the floor or facilitating planning and transition in real-time speech (Crible et al., 2017). In terms of distributional patterns, combinations of fillers and silent pauses tend to signal hesitation or processing difficulty and occur at syntactic boundaries. In contrast, combinations of fillers and repetition (hesitation disfluencies; e.g., *I uh I want to*) are commonly found at the start of utterances, associated with discourse planning (e.g., to plan more complex utterances the speaker has initiated) (Clark & Wasow, 1998; Crible et al., 2017). Especially when speech repair is involved, speakers tend to utilize the insertion of filled pauses as a strategy to signal the delay in the intended speech content while dealing with errors or challenges in content and linguistic processing (Clark & Wasow, 1998). These clustering and distributional patterns suggest that the occurrence of disfluencies in L1 speech does not necessarily indicate the lack of proficiency; instead, they are part of the typical discourse planning and speech production process and can even serve facilitative communicative functions.

In contrast, L2 disfluency clusters have received less attention, particularly in relation to oral proficiency. As an exception, Götz (2013) cluster analyzed fluency and disfluency features among L1 and L2 speakers and found distinct patterns in disfluency clusters and fluency enhancement strategies both between L1 and L2 speakers and among L2 speakers. Despite the few studies of L2 disfluency clusters, studies that employed factor analytic approaches to examine the dimensionality of utterance fluency have revealed meaningful covariances among individual speed, breakdown, and repair features (Suzuki & Kormos, 2023; Yan et al., 2021), suggesting that the subdimensions of fluency are closely associated and can collectively explain variances in L2 oral proficiency. These covariance patterns suggest that disfluencies in L2 speech likely co-occur frequently; however, more studies are needed to closely examine the nature of disfluency co-occurrence, particularly in relation to L2 oral proficiency.

Uncovering the co-occurrence patterns of disfluency clusters in L2 speech and their relationships with oral proficiency can have meaningful implications on L2 speech perception and assessment. Listeners may judge proficiency based on the frequency and distribution of multiple disfluency clusters at scale throughout the speech (as reflected in descriptors of the CEFR and IELTS). Therefore, taking a global approach to understanding disfluency clustering could yield more meaningful insights into L2 proficiency. To analyze co-occurrence at both local and global levels, disfluency features can be examined within AS-units (de Jong, 2016; Foster, Tonkyn & Wigglesworth, 2000; Huensch & Tracy–Ventura, 2017). Analyzing clusters within these meaning-bound units can reflect the cognitive effort required by speakers to convey meaning and thereby their language proficiency. We also speculate that across AS-units, there might be different types of disfluency clusters (i.e., patterns of co-occurrence among disfluency features), and the proportion of different types of disfluency clusters a speaker employs in speech production might provide insights into their L2 oral proficiency.

That said, the relationships between disfluency clusters and proficiency can vary based on speakers' L1 backgrounds. This is a reasonable speculation as previous research has revealed distinct distributional patterns of disfluencies as well as their relationships with oral proficiency across speakers' L1 backgrounds (e.g., Eren et al., 2022; Ginther et al., 2010; Götz, 2019; Park, 2016). The cross-L1 effect may suggest a link between L1 and L2 fluency, with L2 speech possibly inheriting characteristics from L1 speech. Existing evidence from L2 fluency research supports this idea (de Jong, Groenhout, Schoonen & Hulstijn, 2015), though the influence of L1 on L2 fluency

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

varies across different fluency subdimensions (Gao & Sun, 2023a, b) and individual fluency features (Peltonen, 2018).

In summary, although disfluency has gained significant interest in L1 corpus-based or psycholinguistics research, similar investigations are relatively few in L2 research; in particular, the examination of disfluency clusters and their relationship with language proficiency is rare. To address this gap, our study investigates the nature of disfluency in L2 speech in assessment contexts, focusing on IELTS speaking performances. Specifically, we ask

1. What types of disfluency clusters are commonly present in IELTS speaking performances?
2. Are there meaningful associations between different types of disfluency clusters and L2 oral proficiency or L1 background in the assessment context?
3. How do different disfluency clusters manifest across L2 oral proficiency levels?

Methods

The IELTS speech corpus

The spoken corpus for this study consists of 272 benchmark speech samples from responses to the long run (Part II) task of the IELTS speaking test. In Part II, test takers give a 2-min monologue on a pre-selected topic after 1 min of preparation. This task was chosen for two reasons: (a) It allows for the extraction of a long, uninterrupted speech, and (b) The topics, though varied, are controlled for difficulty and are comparable across test administrations. While there is variability in topics across proficiency levels and language backgrounds in the corpus due to test security and task nature, all tasks involved independent narrative descriptions (e.g., describing a favorite place, an admired person, or a time of difficulty). Table 1 provides a breakdown

Table 1. Corpus used for the study.

Level	L1	*n*	Total number of words	Mean length of interaction (words)	Total number of AS-units
Band 5	Chinese	23	16,774	236.25	1,287
	Punjabi	22			
	Arabic	25			
	Subtotal	70			
Band 6	Chinese	21	20,329	282.35	1,312
	Punjabi	23			
	Arabic	23			
	Subtotal	67			
Band 7	Chinese	24	23,118	312.41	1,319
	Punjabi	24			
	Arabic	25			
	Subtotal	73			
Band 8	Chinese	18	21,118	324.89	1,331
	Punjabi	20			
	Arabic	24			
	Subtotal	62			
Total	Chinese	86	81,339	288.25	5,249
	Punjabi	89			
	Arabic	97			
	Subtotal	272			
Gender	Female	109			
	Male	68			
	Unknown	95			

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

of speakers across band levels, showing a relatively even distribution across the four bands. Each speech sample was edited to include only the test taker's initial response, excluding follow-up questions from examiners to maintain consistency in analysis.

Data analysis

Speech transcription

All speech files were first converted from .mp3 to .wav and edited to include only the first part of the monologic turn by the test taker using Audacity (Audacity Team, 2022). Five undergraduate and graduate research assistants were trained to edit the files and transcribe the data. The research assistants transcribed the speech samples according to the TOEFL 2000 Spoken and Written Academic Language corpus transcription guidelines (Biber, 2006) using ELAN software (Sloetjes & Wittenburg, 2008).

Pause segmentation

Before transcription, files were segmented by silent pauses that were 250 ms or greater, using the automatic segmentation script in Praat (Boersma & Weenink, 2023). Then, adjustments were made to the pause boundaries in ELAN, alongside the transcription. An additional segmenter/transcriber cross-checked the files for the accuracy of the transcription and pause boundaries. Pauses and speech were separated by tiers in ELAN.

AS-unit segmentation

Using the speech tiers (not the pause tiers; marked as RUN in Figure 1), we segmented each speech sample into individual analysis of speech (AS)-units. We followed the definition of AS-unit by Foster et al. (2000): "An AS-unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, along with any subordinate clause(s) associated with either" (p. 365). In addition, we allow fragmentary AS-units that are bounded by intonation and pauses. Three research assistants conducted segmentation with an inter-coder reliability of .91 and resolved discrepancies through iterative discussion. The process resulted in 5,249 AS-units.

Fluency and disfluency analysis

Six fluency features were extracted in the dimensions of speech duration, breakdown, and repair features, namely, (a) AS-unit duration, (b) silent pause, (c) filled pause (e.g., *"uh", "um", "hmm", "well", "you know", "like"*), (d) repetition (e.g., "*I want to share share my experience in…*"), (e) reformulation ("*The school I go I went to…*"), and (f)

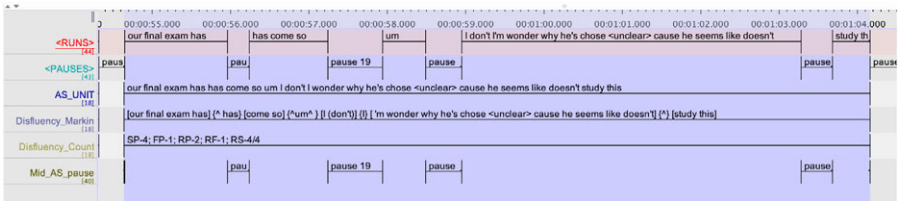


Figure 1. Screenshot of disfluency coding on ELAN.
Note: In the Disfluency_Markin tier, [] = original or intended utterance; () = reparandum; {} = editing phase; and ^ = silent pause.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Table 2. Structure of disfluency.

(Prior context)	Reparandum	Editing phase	Repair	(Continuation)	Disfluency cluster	Repair success
Parking is				*free on Sunday*	#	+
Parking		*<silent pause>*		*is free on Sunday*	SP	+
Parking is		*um/uh*		*free on Sunday*	FP	+
Parking	*is*	*is*		*free on Sunday*	RP	+
Parking	*was*		*is*	*free on Sunday*	RF	+
Parking	*on*	*<silent pause>*	*is*	*free on Sunday*	SP + RF	+
Parking	*is*	*you know um is*		*free on Sunday*	FP + RP	+
Parking	*was*	*um <silent pause> uh*	*is*	*free on Sunday*	FP + SP + RF	+
Parking is	*fr-*	*f- fr- <silent pause> uh*	*free*	*on Sunday*	SP + FP + RP + RF	+
Parking	*is fr-*	*fr- <silent pause> is fr- uh*		*(abandoned)*	SP + FP + RP	–

Source: Adapted from Shriberg (1994).
Note: FP = filled pauses; RF = reformulations; RP = repetitions; SP = silent pauses.

repair success (a binary choice of whether the speaker failed to return to the original utterance after disfluencies; see Yan et al., 2021). As mentioned above, silent pauses were automatically extracted; filled pauses, repetitions, reformulations, and repair success were manually coded directly on ELAN, with inter-coder reliability of .93 for filled pause, .94 for repetition, .96 for reformulation, and .89 for repair success. After all the disfluencies were coded, we marked the boundaries of each disfluency following Shriberg's (1994) definition of disfluency structure. According to her definition, a disfluency instance comprises three components: reparandum, editing phase, and repair. The reparandum signifies the portion of the speech that the speaker intends to revise or replace. The editing phase, immediately following reparandum, can involve silent pauses, filled pauses, repetitions, any combinations of them, or nothing (i.e., empty editing phase). This component reflects the effort made to restore fluency and is the main region of analysis in this study (See Table 2 for examples of the disfluency structure).

Statistical analysis

All statistical analyses in this study were performed using RStudio, version 1.1.383 (RStudio Team, 2016). We first computed the frequencies of pauses and repairs for each AS-unit and examined the Spearman correlation between each variable and the IELTS score band. Upon checking the descriptive statistics, we noticed a skewed distribution in some disfluency features and thus performed log transformations[1] (Warner, 2012), after which variables fell within acceptable skewness and kurtosis values (–3, +3). In addition, the average rate of successful repair per AS-unit (*M*) was 0.939, and the standard deviation (*SD*) of repair success was rather small. When we examined the entire speech corpus, 94.81% of the disfluencies were successfully repaired. Thus, we decided to remove this feature as it is unlikely to generate much variance among speakers across proficiency levels.

[1]Among the different transformation techniques, we chose log transformation because it maintains the original order in the values of each variable, allowing readers to still compare the individual observations on each variable.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Next, to answer research question (RQ) 1, we performed a cluster analysis on the disfluency features across AS-units, not across learners, to identify common types of AS-units based on the disfluency features they might contain (i.e., our operational definition of disfluency clusters). Our assumption is that AS-units that are classified within the same cluster tend to share similar patterns across different disfluency features and differ in those features from AS-units in other clusters (Kaufman & Rousseeuw, 2009). In this step, we first standardized all transformed disfluency variables into *z* scores and then employed a hierarchical-based *k*-means clustering approach (Ginther & Yan, 2018; Pourahmad, Basirat, Rahimi & Doostfatemeh, 2020). Hierarchical-based *k*-means clustering uses hierarchical clustering and *k*-means clustering successively and has been shown to be an effective approach to combine the advantages and meanwhile ameliorate the shortcomings of both approaches (Pourahmad et al., 2020). First, AS-units were grouped into clusters using distance-based, agglomerative hierarchical clustering. The similarity between AS-units was determined by Euclidean distance and the decisions of successive merges of clusters were made using Ward's method. The result of hierarchical clustering was represented by a dendrogram, from which the optimal number of clusters ($k = 4$) was determined. Then, the centers of the four clusters derived from hierarchical clustering (i.e., cluster centroids) were used as the initial cluster centers for *k*-means clustering, and through iterations *k*-means clustering will gradually update the cluster centroids and assign a cluster membership for each AS-unit. In cluster analysis, cluster centroids are computed as the mean of the data points in a cluster, calculated separately for each dimension (or feature) in the dataset. Thus, they can be interpreted as the "average point" of each disfluency feature for all AS-units that belong to the same cluster. To answer RQ2 and RQ3, we computed the proportions of different disfluency clusters of each speaker (i.e., speech sample) and then performed multivariate analysis of variance (MANOVA) and Spearman correlations to examine the impact of proficiency and L1 backgrounds on these proportions. Specifically, proficiency was operationalized as IELTS score bands (i.e., 5, 6, 7, 8), a categorical variable with four levels; L1 backgrounds was operationalized as a categorical variable with three levels (i.e., Chinese, Punjabi, Arabic). We also created boxplots to visualize the associations between score bands and disfluency clusters across L1 backgrounds.

Additionally, we performed a qualitative analysis of the disfluencies within AS-units, to examine (a) disfluencies at various locations within AS-units and (b) the makeup of disfluencies (i.e., the combination of different types of disfluencies). Specifically, we treated each editing phase, following Shriberg's (1994) definition of disfluency structure, in the speech transcripts as an occurrence of disfluency chain. We coded all AS-units in the speech corpus in terms of disfluency location and disfluency makeup. For disfluency location, we coded whether the disfluency chain occurs in the initial, medial, or final position of the AS-unit. In terms of makeup, a disfluency chain can consist of zero or multiple disfluency elements, namely, (a) filled pause, (b) silent pause, (c) word repetition, and (d) word reformulation (zero elements would mean an empty editing phase). We did not code the order of disfluency elements; thus, the possible combinations of disfluency elements equal $2^5 = 32$. We underwent the following coding process: First, three researchers in our team first read through five speech samples (including 57 AS-units) and discussed the common combinations of disfluency elements in the speech samples; next, all three researchers coded 17 speech files (313 AS-units). The inter-coder agreement (in terms of percent agreement) was 99.12% for disfluency location and 98.33% for disfluency

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

makeup, respectively; since the two codes were not high inference, after a discussion to resolve the disagreements, they divided the remaining 255 speech files and each coded a third of them. This analysis allowed us to explore the association between disfluency cluster type and the location and makeup of disfluencies. We triangulated the findings of this analysis with the results of quantitative analyses to provide a nuanced interpretation of the nature of disfluencies and their relationships with language proficiency.

Results

Descriptive statistics of individual disfluency features

Table 3 summarizes the descriptive statistics for all raw, untransformed micro-disfluency features examined in this study. Among all the AS-units (k = 5,249), the mean duration was 5.46 s. On average, one AS-unit contains 1.47 filled pauses and 1.28 silent pauses but with only 0.47 repetitions and 0.40 reformulations. To produce a correlation matrix for the micro-disfluency features and IELTS scores across individual speakers, we derived the total frequencies of disfluency features normalized by AS-unit duration for each speech (e.g., number of silent pauses/second) and correlated them with one another and IELTS scores (see Table 3). The correlations between individual disfluency features and IELTS scores were either weak or not statistically significant. In contrast, the macro-fluency features (mostly speed fluency features not included in this study, e.g., $r_{\text{speech rate}} = .67$, $r_{\text{articulation rate}} = .67$, $r_{\text{mean length of run}} = .40$, and $r_{\text{mean length of silent pauses}} = -.44$), tended to show much stronger correlations (see Yan & Staples, 2023). These findings align with previous meta-analyses demonstrating that macro-fluency variables are more effective predictors of language proficiency or perceived fluency (Koizumi et al., 2022; Suzuki et al., 2021; Yan & Staples, 2023).

RQ1: What types of disfluency clusters are commonly present in IELTS speaking performances?

Using hierarchical k-means cluster analysis, we assessed the optimal number of clusters based on the dendrogram and scree plot (Figure 2). Both 3- and 4-cluster solutions were viable; however, the 4-cluster solution provided a clearer interpretability with adequate observation counts per cluster. Table 4 and Figure 3 present the cluster centroids and the descriptive statistics in terms of raw disfluency features, interpreted as follows:

- Cluster 1: medium runs with more silent pauses but fewer repairs

Table 3. Descriptive statistics and Spearman correlations of individual disfluency features.

	M	*SD*	IELTS	SP	FP	RP	RF	RS
AS	5.457	4.094	–.201					
SP	1.277	1.693	–.136	—				
FP	1.469	1.819	–.164	.016	—			
RP	0.466	0.910	–.116	.142	.349	—		
RF	0.402	0.715	–.109	.120	.382	.576	—	
RS	0.939	0.130	.007	.044	.019	–.097	–.106	—

Note: AS = AS-unit duration; FP = number of filled pauses, RF = number of reformulations; RP = number of repetitions; RS = repair success rate; SP = number of silent pauses.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

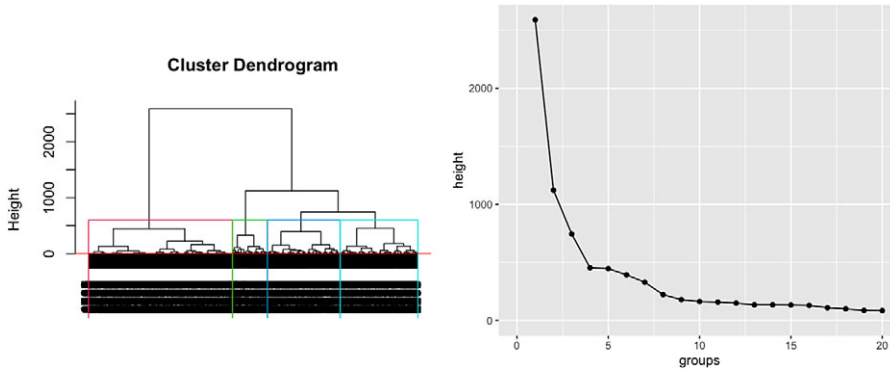


Figure 2. Dendrogram and scree plot for hierarchical cluster analysis.

Table 4. Cluster centroids of micro-disfluency features.

	k	%	AS	SP	FP	RP	RF
Cluster 1	1,586	30.22	0.373	0.837	0.062	0.013	–0.446
Cluster 2	1,349	25.71	0.936	0.902	0.809	0.756	1.513
Cluster 3	882	16.81	–0.436	–1.153	0.352	–0.106	0.017
Cluster 4	1,431	27.27	–1.130	–1.153	–1.317	–0.648	–0.665

Descriptive statistics of micro-disfluency features by cluster

		AS		SP		FP		RP		RF	
	k	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cluster 1	1,586	5.962	3.479	1.625	1.001	1.257	0.231	0.463	0.834	0	0
Cluster 2	1,349	8.663	4.659	2.891	2.002	2.856	1.375	1.117	1.238	1.361	0.738
Cluster 3	882	4.809	2.579	0	0	2.112	0.871	0.125	0.416	0.145	0.383
Cluster 4	1,431	2.273	1.196	0.156	0.391	0	0	0.068	0.305	0.104	0.317

Note: AS = AS-unit duration; FP = number of filled pauses; RF = number of reformulations; RP = number of repetitions; RS = repair success rate; SP = number of silent pauses.

- Cluster 2: long runs with the most pauses and repairs
- Cluster 3: medium runs with more filled pauses and some repairs
- Cluster 4: short runs with the least pauses and repairs

Clusters 2 and 4 displayed distinct contrasts: Cluster 2 included the longest AS-units with the most silent pauses, filled pauses, repetitions, and reformulations, while cluster 4 comprised the shortest AS-units with minimal disfluencies. Clusters 1 and 3 were similar in AS-unit length (i.e., medium length) but differed in disfluency types, with cluster 1 exhibiting more silent pauses and fewer repairs and cluster 3 more filled pauses and fewer repairs. Cluster sizes were as follows: 1,586 (cluster 1); 1,349 (cluster 2); 882 (cluster 3); and 1,431 (cluster 4).

RQ2: Are different types of disfluency clusters associated with oral proficiency?

To examine the relationships between disfluency clusters and L2 oral proficiency, we first computed the proportions of different disfluency clusters for each speaker. This is

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

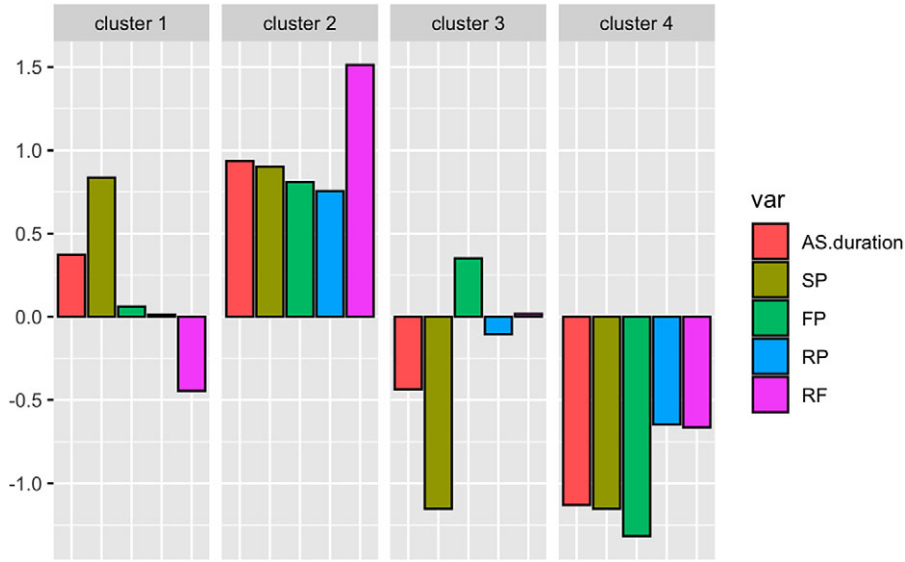


Figure 3. Cluster centroids of micro-disfluency features.
Note: FP = filled pause; RF = reformulation; RP = repetition; SP = silent pause.

because disfluency cluster is defined by each AS-unit and each speaker produces multiple AS-units. Next, we conducted a MANOVA analysis to examine the effects of proficiency level, speaker's L1 background, and their interaction on the proportions of different disfluency clusters. The results (see Table 5) showed significant effects for proficiency, $V = .137$, $F(4, 263) = 4.851$, $p < .001$; for L1, $V = .181$, $F(8, 528) = 9.493$, $p < .001$; and for the interaction between speakers' L1 background and L2 proficiency level, $V = .107$, $F(8, 528) = 1.928$, $p = .045$. Further univariate analyses of variance (ANOVAs) indicated significant main effects of proficiency on cluster 2: $F(1, 266) = 21.085$, $p < .001$, $\eta 2 = .09$ and cluster 4: $F(1, 266) = 31.586$, $p < .001$, $\eta 2 = .11$. To further illustrate the main effects, we plotted the boxplots by cluster, proficiency level, and L1 background and computed Spearman correlations between proficiency score and cluster proportions for all speakers (see Figure 4).

Based on the results, disfluency clusters 2 and 4 had overall correlations of –.26 and .32 with proficiency, respectively. Although these correlations are not strong effects per se (Plonsky & Oswald, 2014), the boxplots showed a steady increasing or decreasing trend on these clusters with proficiency. That is, as L2 oral proficiency increases, speakers tend to produce a higher proportion of short speech runs without many disfluencies as well as fewer long runs with lots of pauses and repairs (i.e., laborious speech runs). That said, we note that even speakers at band 5 were able to produce short and fluent speech runs, suggesting that short and fluent runs are not exclusive to high proficiency L2 speakers and that not all speech runs need to be long and complex; similarly, even high proficiency L2 speakers (at band 8) still produce a substantial proportion of AS-units classified as cluster 1, 2, and 3, suggesting that disfluencies are common across all levels. Alternatively, it is also possible that some low proficiency L2 speakers might use a strategy to avoid producing long utterances so that they do not exhibit disfluent features excessively.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Table 5. MANOVA and ANOVA analysis of proficiency level and L1 on disfluency cluster proportions.

Variables	*df*	Pillai's trace (V)	df_{NUM}	df_{DEM}	F	p
L1	2	0.181	8	528	9.493	< .001
Score	1	0.137	4	263	4.851	< .001
L1 × score	2	0.107	8	528	1.928	.045

Variables	*df*	*SS*	*MS*	*F*	*p*	Eta squared
Cluster 1						
L1	2	0.518	0.259	13.302	< .001	.091
Score	1	0.047	0.047	2.391	.123	.009
L1 × score	2	0.002	0.001	0.044	.957	< .001
Residuals	266	5.181	0.019			
Cluster 2						
L1	2	0.412	0.206	8.723	< .001	.062
Score	1	0.498	0.498	21.085	< .001	.073
L1 × score	2	0.160	0.080	3.388	.035	.025
Residuals	266	6.279	0.024			
Cluster 3						
L1	2	0.276	0.138	9.353	< .001	.066
Score	1	0.001	0.009	0.672	.413	< .001
L1 × score	2	0.006	0.003	0.197	.821	.002
Residuals	266	3.927	0.015			
Cluster 4						
L1	2	0.226	0.113	5.286	.006	.038
Score	1	0.675	0.675	31.586	< .001	.106
L1 × score	2	0.129	0.064	3.023	.051	.022
Residuals	266	5.686	0.021			

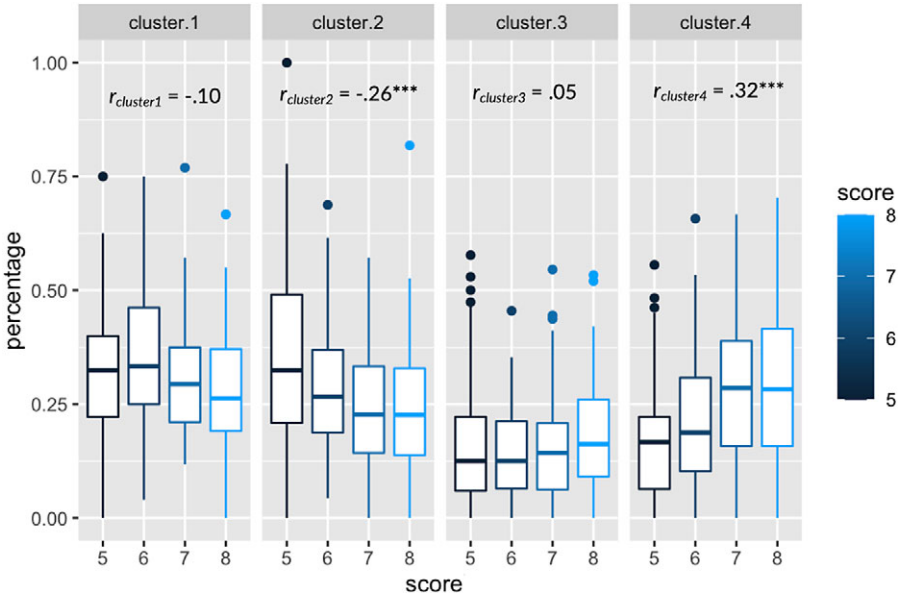


Figure 4. Correlations between proportion of disfluency clusters and IELTS scores.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Interestingly, clusters 1 and 3 did not correlate meaningfully with IELTS scores, suggesting that the presence of silent or filled pauses alone does not indicate higher proficiency. Instead, it is the co-occurrence of pauses and repairs that collectively mark lower L2 oral proficiency. In addition, the lack of meaningful associations for clusters 1 and 3 might be because when silent or filled pauses occur without much explicit repair behavior, the speakers are either attempting to formulate content or search for lexico-grammatical items as part of the normal disfluency phenomenon observable in natural speech.

The main effects of speaker's L1 background suggest that the speakers from Chinese, Punjabi, and Arabic L1 backgrounds displayed group-level contrasts in the disfluency cluster proportions in their L2 speech. L1 Chinese speakers produced a higher proportion of AS-units in cluster 1, while L1 Punjabi speakers had the smallest proportion. L1 Punjabi speakers exhibited a higher proportion of long, disfluent utterances in cluster 2 compared to L1 Arabic speakers, who produced the fewest utterances in this cluster. Conversely, L1 Arabic speakers produced the highest proportion of AS-units in cluster 4.

The significant interaction effect between speaker's L1 background and proficiency level indicates that the impact of L2 oral proficiency on disfluency cluster proportions is different across speakers' L1 backgrounds. Further ANOVAs showed that the interaction effect was likely due to the differential impact of proficiency level on cluster 2, $F(2, 266) = 3.388$, $p = .035$, $\eta^2 = .025$. Specifically, when broken down by L1 backgrounds, the correlations for cluster 2 were stronger for L1 Chinese ($r_{cluster2} = -.34$) and Pubjani ($r_{cluster2} = -.37$) than for L1 Arabic (see the boxplots and correlations in Figure 5 and the descriptive statistics in Appendix A). These findings suggest that the significant multivariate interaction between speakers' proficiency level and L1 background is mostly because cluster 2 failed to discriminate Arabic speakers across proficiency levels.

RQ3: How do different disfluency clusters manifest across L2 oral proficiency levels?

Following the cluster analysis, we examined the qualitative nature of disfluencies within AS-units for each identified cluster type. Cluster 1 predominantly featured individual silent pauses and occasional filled pauses. These disfluencies tended to occur at both phrasal and clausal boundaries, perhaps indicating natural speech formulation processes rather than overt planning difficulties. As the examples in Table 6 show, example 1 (band 7) shows that the filled pause "uh" might be occurring as a result of lexico-grammar search for "healthy oxygen". In example 2 (band 5), the silent pauses seem to occur as the speaker is formulating each subsequent clause or phrase. We did not notice a clear difference in this type of AS-units across proficiency levels, although the speech at lower band scores tend to show more grammatical errors and the lexico-grammar features at higher band levels tend to be more sophisticated. This suggests that disfluencies in cluster 1 are perhaps more indicative of normal content formulation or lexico-grammar search at the local level.

Cluster 2 contained AS-units characterized by significantly longer utterances. Within this cluster, we identified two notable types of disfluencies. The first type resembled those in cluster 1, featuring a high concentration of individual silent and filled pauses that occur at various points in the utterance. However, this type of AS-unit does not seem to show noticeable differences in the nature of disfluencies from AS-units in cluster 1. For instance, example 3 (band 6) consists of a high number of filled and silent pauses, and these disfluencies also tend to occur at either phrasal or clausal

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

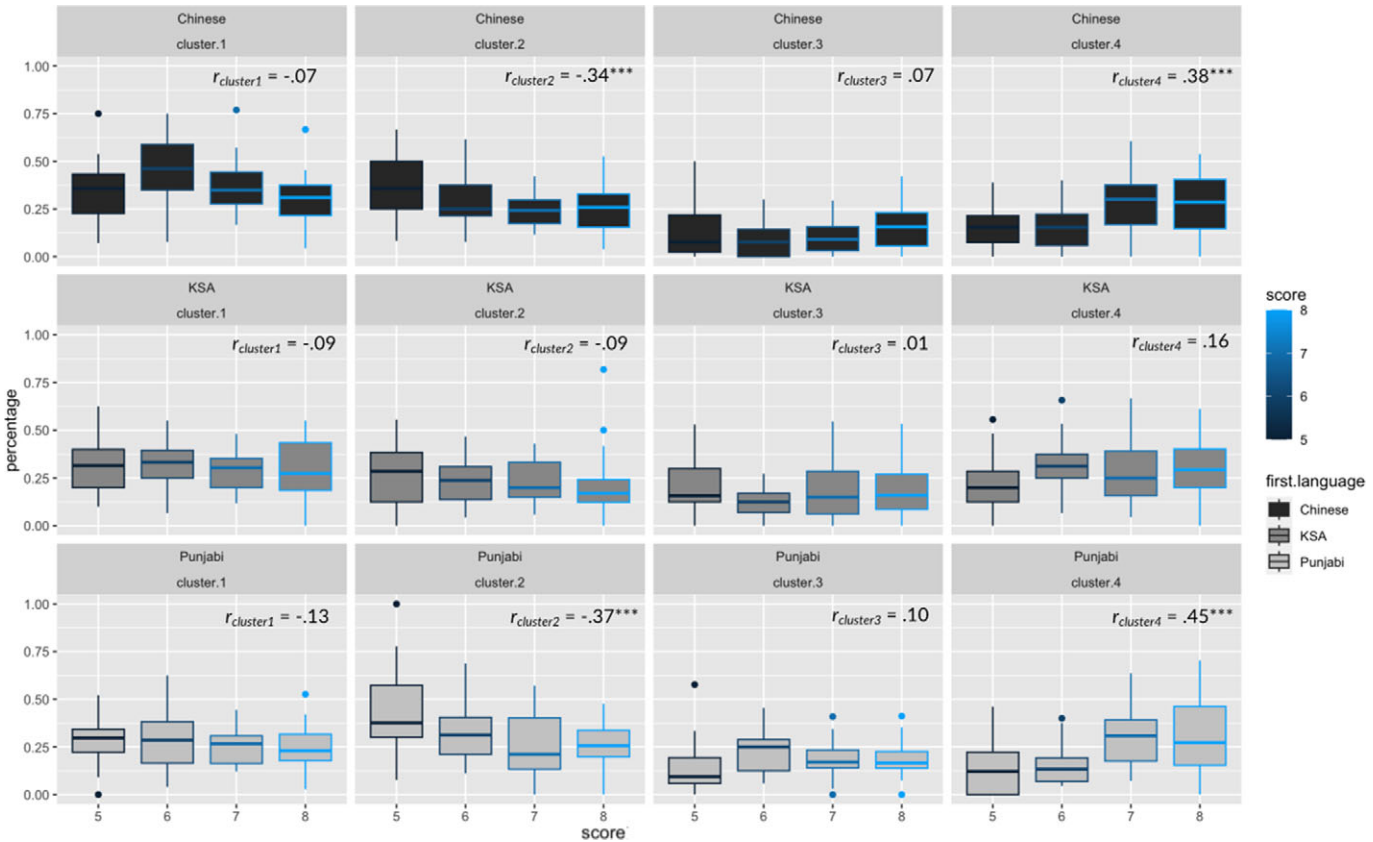


Figure 5. Correlations between proportion of disfluency clusters and IELTS scores by speaker L1.
Note: * $p < .05$; ** $p < .01$; *** $p < .001$. KSA = Kingdom of Saudi Arabia.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Table 6. Example AS-units within each cluster.

Example	Begin time	End time	AS-unit	Band level	L1
Cluster 1					
1	00:38.5	00:43.5	[and] {^} [we wouldn't be able to get] {uh} [health oxygen]	7	Chinese
2	00:59.3	01:03.9	[you can say] {^} [it's warm] {^} [and sometimes rainy] {^} [so no cold or hot weather]	5	Arabic
Cluster 2					
3	00:30.5	00:59.9	{uh} [the reason why] {uh} [our university held this party is because] {^ um ^} [(the male)] {} [the number of male students in my university] {uh} [is much more than the female students] {^} [university wanted to] {uh ^} [use this party as an opportunity] {^} [to increase] {uh} [the friendship among all the students]	7	Chinese
4	01:48.9	01:57.0	[and] {uh^uh} [then I] {^and uh then I} [(fe-)] {um then I} [felt the] {uh} [(fees)] {} [exam fees] {} []	5	Chinese
5	00:54.4	00:59.4	[you can] {uh} [(borrow)] {you can uh^} [use it] {uh} [(inside)] {} [outside the center]	5	Punjabi
6	01:25.8	01:38.1	[she wasn't educated (in)] {she wasn't educated} [to speak English] {^} [so] {^} [I understand that] {^} [she needs me to do all the things she can't do]	6	Arabic
7	00:38.9	00:55.4	[and] {^} [I] {^I} [have lots of] {I- I have lots of} [things to carry on so] {^} [I also called my friend to help me to move (to)] {move} [the] {^} [boxes] {^} [to the taxi]	6	Arabic
8	01:49.7	01:56.5	[we mainly (mo-)] {} [motivate] {^ motivate} [each other] {^} [so (h-)] {} [that's how we get good marks]	5	Arabic
9	01:24.1	01:32.1	[I have already (t-)] {^} [told that this is the time period of nineteen sixties and seventies] {^} [and (w-)] {^uh} [why I've become interested]	8	Punjabi
Cluster 3					
10	00:33.8	00:39.3	[then] {uh} [after that I also searched on the internet to know about] {uh} [how to prepare this type of assignment]	8	Punjabi
11	01:36.2	01:40.7	[and he had] {uh uh he had} [a lot of experience about different kinds of things]	5	Punjabi
12	01:38.5	01:46.2	{so um} [I consulted about different types of modules that are going to be there in IELTS exam] {^} [on the internet]	7	Arabic
13	00:43.8	00:52.0	[(that's perhaps)] {uh that's uh that's uh} [that happens almost every day] {uh} [because every day I have several things to do]	6	Chinese
Cluster 4					
14	00:47.5	00:49.4	[it's full of sugar and fat]	8	Arabic
15	00:50.9	00:53.7	[it] {it} [had screen sensitivity issues]	6	Punjabi
16	00:38.7	00:41.4	[but that photo I have sent to (my)] {} [one of my friend]	7	Chinese

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

boundaries. Disfluencies in this type of AS-unit, while prevalent, still seem to show expected content formulation and lexico-grammar search during speech production. Although these disfluencies might make the speech more laborious, it is perhaps because the utterances are longer with more elaborate content.

The second type of AS-units from cluster 2 tend to feature sequences of combinations of different types of disfluencies. These sequences seem to indicate clear composing effort on the part of the speaker, either searching for or correcting a particular word or phrase. As shown in example 4, the speaker (band 5) shows two sequences of disfluencies with the same AS-unit. The disfluency chains seem to show the speaker's effort to produce the word *"felt"*. The speaker first produces filled pauses *"uh"* followed by a repetition of the original utterance *"then I"*; after this accumulation of these disfluencies, the speaker utters a partial word *"fe-"* (the recording of the speech sounds like *"fi:"*). However, the speaker seems to quickly notice the incorrect tense, so he produces another filled pause + repetition disfluency chain (i.e., *um then I*) before eventually producing the correct form "felt". Right after the disfluency chain, the speaker also made a quick repair of "fees" into "exam fees". It is possible that overcoming these two lexico-grammatical difficulties within a sentence has taken up a great amount of cognitive resources, so the speaker ends up not completing the sentence and abandoning the original utterance. In contrast, example 9 is produced by a speaker at band 8. In this AS-unit, the speaker also produces two disfluency chains, namely, a sequence of silent pause + reformulation (i.e., *[…already (t-)] {^} [told that …]*) and another sequence of silent pause + filled pause + reformulation (i.e., *[and (w-)] {^uh} [why I've become interested]*). However, these disfluency chains seem to be brief as the speaker is formulating a longer and complex dependent clause. Thus, taken together, it seems that the second type of AS-unit in disfluency cluster 2 indeed indicates composing effort from the speaker. However, a close examination of the utterances seems to suggest a proficiency difference in that the higher proficiency speakers produce disfluencies as a result of formulating complex lexico-grammatical structures, whereas for the lower proficiency speakers, the disfluencies seem to involve more repetition than reformulation, and the composing effort reflected in these disfluencies are not necessarily associated with sophisticated lexico-grammar.

AS-units in cluster 3 tend to feature disfluencies that are centered around filled pauses. These disfluencies largely take four forms, namely, (a) individual non-lexical fillers (i.e., *uh* or *um*; e.g., example 10); (b) non-lexical fillers + repetition (e.g., example 11); (c) non-lexical fillers + lexical fillers (functioning as discourse markers, e.g., *so uh, uh you know*; example 12); and (d) non-lexical fillers + repetition + repair (e.g., example 13). The majority of disfluencies in this cluster occurred at the beginning of AS-units, suggesting that speakers were perhaps engaging in discourse planning and content formulation. Importantly, the distribution of these forms did not show clear proficiency-related differences, indicating that filled pauses may be an inherent aspect of spoken discourse across proficiency levels.

In contrast to the other clusters, disfluencies in cluster 4 were more clearcut. AS-units in this type are short utterances with either no disfluency or very few disfluencies. When disfluencies occur, they tend to be individual occurrences of silent pause, filled pauses, or a quick repair. Aside from the proportion of this type of AS-units, we also did not observe proficiency differences in the makeup of disfluencies. We also included some examples of AS-units in cluster 4 in Table 6 (examples 14 and 15).

To illustrate the differences in disfluency clusters across L2 oral proficiency levels and speaker L1 backgrounds, we include excerpt transcripts of speakers in Appendix B, two in each L1 background, to showcase the different proportions of disfluency

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

clusters. In each L1 background, one speaker was rated at band 5, whereas the other was rated at band 8. In each excerpt, each line represents an AS-unit; the editing phase —where most disfluency features are located—is marked by curly brackets {}, whereas silent pauses are marked by ^, filled pauses are directly transcribed, repetitions are enclosed in the curly brackets, and reformulations are indicated by the combination of a reparendum (indicating something needs to be edited) and a square bracket immediately following the curly bracket (editing phase), respectively. As shown in the Appendix B, speakers 1 and 2 are both L1 Chinese speakers. Speaker 1's (band 8) speech showed a higher proportion of disfluency cluster 4 (i.e, short but fluency runs) (e.g., "*[and she works very hard]*"). Although the speaker also produced disfluency cluster 2 (i.e, long but laborious, disfluency runs) (e.g., "*[I give it to her (because I)] {^uh I^} [as a] {^} [memento] {^} [because I] {^um^because I^} [wanted to remind her of our time]*"), the proportion of this type of disfluency cluster (13.33%) was much smaller than that of short fluency runs (40%). In contrast, speaker 2 (band 5) produced a much higher proportion of long and laborious runs (i.e., cluster 2, 42.86%; e.g., "*[and] {uh^ um^} [from (her)] {^um from} [his] {^um} [songs I felt the pure] {^} [and] {uh} [the great passion for the music]*") and a lower proportion of short and fluency runs (7.14%; e.g., "*[and he's a singer]*"). Similar contrasts can be found in the pair of Punjabi speakers (speakers 3 and 4). However, for L1 speakers of Arabic, the contrast in the proportions of disfluency clusters is less prominent. Despite being of lower proficiency, speaker 5 (band 5) produced similar proportions of disfluency clusters 2 and 4, 31.58% and 36.84%, respectively. Similarly, for speaker 6 (band 8), the proportions were 38.46% and 30.77%, respectively. Although the proportions of disfluency cluster types did not show noticeable differences, interestingly, we noted that speaker 6's utterances were less syntactically and lexically complex. In addition, all speakers across L1 backgrounds have fair proportions of disfluency cluster 1 (i.e., medium runs with more silent pauses and fewer repairs) (e.g., "*[and as they also] {^} [improve their reading skills]* ") and disfluency cluster 3 (i.e., medium runs with more filled pauses and fewer repairs) (e.g., "*{um} [and I know] {you know} [the elementary or rudimentary] {^} [concept of its development]*"). These proportions do not seem to differ noticeably between the two proficiency levels.

Discussion and implications

This study explores the co-occurrence of disfluencies in L2 speech to understand their nature and association with L2 oral proficiency. Analyzing open-ended responses from the long run speaking task on the IELTS, we annotated various disfluency features in responses from L1 Chinese, Punjabi, and Arabic speakers. Using a mixed-methods approach, we identified four distinct disfluency clusters, each exhibiting unique characteristics and relationships with L2 oral proficiency. The findings indicate that disfluency is complex, marked by multiple temporal features. In this section, we discuss our findings in relation to previous research on L1 and L2 disfluency and provide practical implications and recommendations for L2 pedagogy and assessment.

Our findings support the clustering nature of disfluency features in L2 speech. We identified four disfluency clusters: (a) short spurts with few disfluencies (cluster 4); (b) medium spurts with silent pauses but few repairs (cluster 1); (c) medium spurts with filled pauses and some repairs (cluster 3); and (d) long spurts with many disfluencies (cluster 2). The emergence of different types of disfluency clusters suggests that disfluency is a multifaceted phenomenon and can manifest in various

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Table 7. Summary of findings.

Disfluency cluster	Cluster description	Example AS-units	Relationship with L2 oral proficiency
1	Medium spurts with silent pauses but few repairs	*[and as they also] {^} [improve their reading skills]*	n.s.
2	Long spurts with many disfluencies	*[and] {uh^ um^} [from (her)] {^um from} [his] {^um} [songs I felt the pure] {^} [and] {uh} [the great passion for the music]*	–
3	Medium spurts with filled pauses and some repairs	*{um} [and I know] {you know} [the elementary or rudimentary] {^} [concept of its development]*	n.s.
4	Short spurts with few disfluencies	*[and he's a singer]*	+

Note: n.s. = not significant.

forms (i.e., combinations of disfluency features). Approximately 25.71% of speech runs were lengthy and laborious (cluster 2), while 27.27% were fluent but short (cluster 4). The remaining half of the runs fell into clusters 1 and 3, showing that disfluencies are common in L2 speech (Clark & Fox Tree, 2002), but excessive disfluency features are relatively rare. These clusters further revealed distinct distributional patterns, makeup, and relationships with L2 oral proficiency. The findings of our analyses are summarized in Table 7.

Disfluency clusters indicative of typical discourse planning and formulation processes

Clusters 1 and 3—comprising 50% of AS-units—seem to indicate typical speech composing processes. In terms of cluster makeup, in cluster 1, silent and filled pauses can occur independently within AS-units, whereas in cluster 3, filled pauses frequently co-occur with silent pauses, lexical fillers, or repetitions. The qualitative findings further suggested that these clusters display distinct patterns: In cluster 1, silent pauses tend to occur at both phrasal and clausal boundaries, while in cluster 3, the disfluencies around filled pauses typically appear at the start of utterances. These findings align with L1 psycholinguistic and corpus-linguistic research on the distributional patterns and makeup of disfluency clusters (Clark & Wasow, 1998; Crible et al., 2017; Degand & Gilquin, 2013; Schneider, 2014), suggesting that similar kinds of disfluencies also occur in L2 speech to signal discourse planning and formulation processes.

The absence of significant correlations for clusters 1 and 3 further strengthens the argument that silent or filled pauses may not necessarily signal low proficiency but rather can reflect typical discourse planning processes. That said, there are interesting nuances. In previous research, disfluency clusters, especially similar to those observed in cluster 3, were viewed as a compensatory strategy that has the potential to facilitate communication (e.g., Crible et al., 2017). However, we did not observe a positive relationship between L2 oral proficiency and cluster 3, either. There might be several plausible interpretations and implications. First, when viewed as a composing or recovery strategy, disfluency clusters are an integral part of the fundamental communication resources in both L1 and L2 speech that do not require the acquisition of a new language. Thus, they do not contribute to differentiating levels of L2 oral

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

proficiency. Second, listener judgments of L2 fluency or oral language proficiency might be more attuned to difficulties or unexpectedness in the speaker's speech. Thus, compensatory strategies that can facilitate speech comprehension or communication might be less noticeable and subconsciously underweighted. Although it is not possible to find direct evidence for this interpretation, previous research on juncture and non-juncture pauses might offer relevant insights. While previous research found converging evidence that non-juncture pauses are negatively associated with L2 oral proficiency (e.g., Huensch, 2023; Kahng, 2014, 2018), the relationship between juncture pauses and L2 oral proficiency is not necessarily positive or even significant (e.g., Yan, Lei & Pan, 2025; Koizumi et al., 2022). That said, more research is needed to further explore the associations of disfluency clusters as compensatory strategies and L2 oral proficiency.

Disfluency clusters indicative of L2 oral proficiency

In contrast, the prominence of clusters 2 and 4 in relation to L2 oral proficiency indicates that these types of disfluency clusters might be more characteristic of L2 speech. Higher oral proficiency was linked to a greater proportion of short, fluent runs and fewer lengthy, laborious ones. The noteworthy relationships observed for clusters 2 and 4 carry theoretical implications regarding the cognitive processes underlying L2 speech production and the connection between utterance fluency and L2 oral proficiency. First, the positive association between cluster 4 and L2 oral proficiency should be discussed. In this study, cluster 4 features short speech runs without much disfluencies. Although fluency has been frequently associated with automaticity in producing long and complex runs with accuracy and fewer disfluencies (e.g., Ginther et al., 2010; Tavakoli et al., 2020; also supported by meta-analyses conducted by Koizumi et al., 2022, and Suzuki et al., 2021), the findings of this study suggest that from a micro-perspective L2 oral proficiency may be more closely associated with the ability to produce short bursts of runs with minimal disfluencies (cluster 4). We argue that this finding reflects the nature of L2 speech production and comprehension. Unlike written registers that utilize longer syntactic structures and sophisticated lexical items to convey nuanced, compressed meanings, speech is characterized by brief runs to facilitate efficient communication. The timing factor in oral communication can make the production of long and complex runs more challenging and counterproductive. Similarly, from the listener's perspective, longer runs demand more cognitive effort in comprehension as they often incorporate more intricate syntactic structures and contain a greater amount of information.

Conversely, cluster 2 features long spurts, in which silent pauses often coincided with various forms of repair phenomena. These disfluency clusters suggest challenges in content formulation or lexico-grammar processing. When breakdowns are excessive or highly concentrated alongside repeated repair attempts within lengthy speech runs, it may make the speech more challenging to comprehend, ultimately resulting in the perception of lower L2 oral proficiency. This finding further suggests that laboriousness in speech production or fluency recovery plays an important role in perceived fluency or L2 oral proficiency. When evaluating speakers' L2 oral proficiency, listeners are sensitive to longer sequences of disfluencies and consider the clustering of disfluencies as a sign of developing L2 proficiency (i.e., knowledge and automaticity of lexico-grammar) and strategies to rectify disfluencies (Tavakoli et al., 2020; Yan et al., 2021).

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

That said, while we found meaningful associations for clusters 2 and 4, these correlations are weaker than those for commonly used macro-fluency features (e.g., speech rate, mean length of run, length of silent pauses). Consequently, if the aim is to predict L2 oral proficiency (e.g., for the development of an automated scoring system), macro-fluency features may still surpass disfluency clusters in terms of predictive power. However, the meaningful relationships between disfluency clusters and L2 oral proficiency can, to some extent, help explain cognitive and performance-related processes underlying L2 speech disfluencies in ways the macro-fluency features cannot. That said, it is worth noting that the associations between disfluency clusters and L2 oral proficiency differed across L1 backgrounds, which merits further investigation. L1 Chinese and Punjabi speakers exhibited similar trends, while L1 Arabic speakers showed a wider spread in cluster proportions, leading to non-significant results. One plausible explanation of these cross-L1 differences might be that variability in correlational patterns emerges from the linguistic properties of the three L1s (e.g., their respective linguistic distance to English), and these cross-linguistic differences might have an impact on the L2 fluency-proficiency relationship (e.g., Eren et al., 2022; Götz, 2019). In addition, individual differences in L1 fluency might also account for some of the variability in L2 fluency features across speakers (de Jong et al., 2015; Gao & Sun, 2023a, b). To further decompose these variabilities, more follow-up studies are needed to closely examine disfluency features in each L1 background.

Implications for L2 fluency research and practice

Our findings contribute to the discussion on the dimensionality of L2 utterance fluency. Although we did not perform factor analysis on the disfluency features in this study, our findings are in broad alignment with findings in previous research (e.g., Suzuki & Kormos, 2023; Yan et al., 2021). That is, breakdown and repair fluency might be operationally distinguishable; these subdimensions of utterance fluency also showed meaningful association with each other, evidencing disfluency co-occurrence at the global level (Riggenback, 1991). That said, in the current study, the emergence of different disfluency clusters (e.g., cluster 1 versus 3) suggests some independence or distinction between breakdown and repair features. Pauses can both occur independently and co-occur with repair features. In regard to breakdown-repair associations, our interpretation resonates with Tavakoli et al. (2020), who speculated that different clusters of repair and breakdown might be a strategy employed by speakers at different proficiency levels to compensate for the linguistic and cognitive resources consumed while controlling for complexity and accuracy.

Under this background, we view our study as a call for future L2 fluency research to incorporate features that tap into the interplay between breakdown and repair rather than examining them in isolation. Different disfluency clusters can reflect the speakers' automaticity in retrieval of lexico-grammatical resources during the speech production process and different strategies to maintain the flow of speech (Tavakoli et al., 2020). Focusing on disfluency clusters allows for a more balanced operationalization of L2 speech fluency and places features of different subdimensions on a comparable level of importance.

Finally, this study has practical recommendations and implications for language pedagogy and assessment. Our findings challenge the prevailing assumption that fluency primarily hinges on only a few prominent utterance fluency features, equating fluency to the ability of producing long and complex utterances with a high speed and

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

few disfluencies. We argue that although laborious disfluency clusters (e.g., *[and] {uh^ um^} [from (her)] {^um from} [his] {^um} [songs I felt the pure] {^} [and] {uh} [the great passion for the music]*) can indicate developing proficiency, other clusters may reflect typical discourse planning and formulation processes (e.g., *[and as they also] {^} [improve their reading skills]*). Thus, it is perhaps unrealistic to expect or target disfluency-free speech at the advanced level of oral proficiency in language pedagogy and assessment (e.g., see the PTE scale descriptors). Disfluencies are natural in speech, marked by the co-occurrence of multiple features. If the goal of teaching and assessment is the production of disfluency-free speech, it might even prompt learners and test takers to adopt avoidance strategies, to sacrifice complexity for fluency (the trade-off hypothesis; Skehan, 1998).

Limitations and future research

This study has several limitations that suggest areas for future research. First, while AS-units provide valuable insights into disfluency, the presence of subtypes within disfluency clusters 2 and 3 indicates that a more fine-grained analysis might yield deeper understanding of speech disfluency. Future studies could adopt a more detailed approach, such as examining disfluency structures in Shriberg (1994). Second, the speech corpus in this study represented a range of narrative task topics; thus, topic or content variation may impact the nature of disfluency especially if certain topics might be more familiar to the test takers given their previous experience or prior world knowledge. Third, given the observed cross-L1 differences, future research should consider recording both L1 and L2 speech from each speaker. This would allow for comparisons of disfluency clusters across both languages, improving the understanding of L2 disfluencies relative to L1 backgrounds. Fourth, this study focused on IELTS band 5 to band 8 performances (corresponding to CEFR B1 to C2 levels, https://ielts.org/organisations/ielts-for-organisations/compare-ielts/ielts-and-the-cefr); examining lower proficiency levels could reveal how disfluency clusters manifest in less proficient L2 speech. Fifth, the current research primarily addresses micro-disfluency features. Future studies could explore the co-occurrence of disfluencies with non-temporal features, including prosodic cues and gestural signals, to better understand underlying cognitive processes. Sixth, data used in this study were collected from speaking performances in the assessment context, where fluency is an explicit scoring criterion. Future research should verify the findings of the study in natural conversational settings. Finally, while this study explores the relationship between disfluency clusters and L2 oral proficiency, future research should examine their connections with constructs such as comprehensibility and perceived fluency to build a causal framework among these related factors.

Conclusions

Limitations notwithstanding, this exploratory study suggests that disfluencies in L2 speech do not occur in isolation. Although we did not directly address the dimensionality of speech fluency, the co-occurrence of various disfluency types and their associations with language proficiency imply that while the three dimensions of utterance fluency are operationally distinct, breakdown and repair features are meaningfully interconnected. Disfluency features, even those with individually weak proficiency correlations, should not be dismissed. When considered collectively, they provide a contextualized explanation of L2 proficiency and reflect the cognitive challenges

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

speakers face in speech production. While disfluencies are prevalent in spontaneous speech, not all indicate developing language proficiency. A high concentration of laborious disfluencies can signal developing proficiency, while other clusters may reflect normal discourse planning. Moreover, an advanced level of L2 proficiency is not necessarily marked by a high proportion of long, complex, and disfluency-free utterances. In contrast, short and fluent speech runs are perhaps more realistic, efficient, and effective communication strategies in the exchange of information and the negotiation of meaning. These findings prompt a reassessment of expectations regarding fluency-proficiency relationships in language pedagogy and assessment, particularly at advanced levels. Finally, while associations between disfluency clusters and proficiency are observed, these relationships vary across L1 backgrounds, highlighting the need for further research into the linguistic, educational, and cultural factors influencing these dynamics.

Acknowledgement. We appreciate the comments from the reviewers and the guidance from Kazuya Saito throughout the peer review process. If there is any disclosure of grants, our research is supported by the IELTS joint-research grant program.

Competing interest. The authors do not have competing interests in the study presented in this manuscript.

References

Audacity Team. (2022). *Audacity* (Version 3.2.0) [Computer application]. https://audacityteam.org/

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, *161*, 25–35.

Boersma, Paul & Weenink, David (2023). *Praat: doing phonetics by computer* (Version 6.3.18) [Computer program]. http://www.praat.org/

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123–147.

Brennan, S.E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of memory and language*, *44*(2), 274–296.

Carney, N. (2022). L2 comprehension of filled pauses and fillers in unscripted speech. *System*, *105*, 102726.

Choo, A. L., Smith, S. A., & Seitz, S. (2024). The relationship between executive function, age of L2 acquisition and speech disfluencies in bilinguals. *Journal of Monolingual & Bilingual Speech*, *6*, 21–43.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

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

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*, 201–242.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis) fluency. *Languages in Contrast*, *17*, 69–95.

Degand, L., & Gilquin, G. (2013). *The clustering of 'fluencemes' in French and English* [Paper presentation]. 7th International Contrastive Linguistics Conference (ICLC 7): 3rd Conference on Using Corpora in Contrastive and Translation Studies (UCCTS 3), Ghent.

de Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*, 113–132.

de Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*, 237–254.

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

de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*, 893–916.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533–557.

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*, 237–253.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709–738.

Fraundorf, S. H., & Watson, D. G. (2013). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*, 1083–1096.

Gao, J., & Sun, P. P. (2023a). To correct or not: The role of L1 fluency in understanding and measuring L2 fluency. *TESOL Quarterly*, *57*, 643–655.

Gao, J., & Sun, P. P. (2023b). How does learners' L2 utterance fluency relate to their L1? A meta-analysis. *International Journal of Applied Linguistics*, *34*, 276–291.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*, 379–399.

Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, *35*, 271–295.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press, New York.

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

Götz, S. (2019). Filled pauses across proficiency levels, L1s and learning context variables: A multivariate exploration of the Trinity lancaster corpus sample. *International Journal of Learner Corpus Research*, *5*, 159–180.

Huensch, A (2023). Effects of speaking task and proficiency on the midclause pausing characteristics of L1 and L2 speech from the same speakers. *Studies in Second Language Acquisition*, *45*, 1031–1055.

Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, *101*, 275–293.

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

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

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

Koizumi, R., Jeon, E., & In'nami, Y. (2022). L2 speaking and its internal correlates: A meta-analysis. in E. Jeon & Y. In'nami (Eds), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 307–338). John Benjamins.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning*, *49*(2), 303–342.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387–417.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT press.

Liao, J. (2023). Disfluency and self-repair in presentational and interpersonal speech modalities. *Foreign Language Annals*, *56*, 401–427.

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

Park, S. (2016). *Measuring fluency: Temporal variables and pausing patterns in L2 English speech* [Unpublished doctoral dissertation]. Purdue University.

Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *The Modern Language Journal*, *102*, 676–692.

Pourahmad, S., Basirat, A., Rahimi, A., & Doostfatemeh, M. (2020). Does determination of initial cluster centroids improve the performance of-means clustering algorithm? Comparison of three hybrid methods by genetic algorithm, minimum spanning tree, and hierarchical clustering in an applied study. *Computational and Mathematical Methods in Medicine*, *2020*, 1–11.

PTE Academic. (2024). PTE Academic: Score Guide. [https://www.pearsonpte.com/pte-academic/scoring]

Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, *37*, 828–848.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441.

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*(3), 395–412.
RStudio Team. (2016). *RStudio: Integrated Development for R* (Version 1.1. 383)[Computer Software]. Boston, MA: RStudio, Inc.
Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid-and high-level second language fluency. *Applied Psycholinguistics*, *39*, 593–617.
Schneider, U. (2014). *Frequency, chunks and hesitations: A usage-based analysis of chunking in English* [Doctoral dissertation]. Albert-Ludwigs-Universität Freiburg.
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies* [Unpublished doctoral dissertation]. University of California, Berkeley.
Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*, 153–169.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*, 1–14.
Skehan, P., & Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, and I. Vedder (Eds.), *Complexity, accuracy, and fluency in second language use, learning, & teaching* (pp. 207–226). KVAB.
Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 816–820). European Language Resources Association.
Suzuki, S., & Kormos, J. (2023). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, *45*, 38–64.
Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, *105*, 435–463.
Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, *104*, 169–191.
Warner, R. M. (2012). *Applied statistics: From bivariate through multivariate techniques*. Sage Publications.
Yan, X., Kim, H. R., & Kim, J. Y. (2021). Dimensionality of speech fluency: Examining the relationships among complexity, accuracy, and fluency (CAF) features of speaking performances on the Aptis test. *Language Testing*, *38*, 485–510.
Yan, X., Lei, Y., & Pan, Y. (2025). Diving deep into the relationship between speech fluency and second language proficiency: A meta-analysis. *Language Learning*.
Yan, X., & Staples, S. (2023). *Investigating the cognitive and social aspects of IELTS speaking performances across proficiency levels: Comparing the CAF-based and register-linguistic analyses* (IELTS Research Reports Online Series, No. 3/23). British Council, Cambridge Assessment English, and IDP: IELTS Australia

Appendix A. Descriptive statistics for disfluency cluster proportions by proficiency level and L1 background.

			Cluster 1		Cluster 2		Cluster 3		Cluster 4	
L1	IELTS score	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Chinese	5	23	0.34	0.16	0.38	0.18	0.13	0.14	0.15	0.10
	6	21	0.46	0.18	0.31	0.16	0.08	0.08	0.15	0.11
	7	24	0.37	0.14	0.25	0.09	0.10	0.09	0.28	0.16
	8	18	0.32	0.16	0.26	0.14	0.15	0.12	0.27	0.16
Arabic	5	25	0.32	0.12	0.25	0.16	0.21	0.15	0.22	0.15
	6	23	0.33	0.11	0.23	0.12	0.13	0.08	0.31	0.13
	7	25	0.30	0.11	0.23	0.12	0.20	0.16	0.28	0.15
	8	24	0.29	0.15	0.22	0.17	0.19	0.14	0.30	0.16
Punjabi	5	22	0.29	0.13	0.43	0.22	0.14	0.14	0.13	0.14
	6	23	0.28	0.16	0.33	0.16	0.22	0.11	0.17	0.11
	7	24	0.25	0.09	0.27	0.17	0.18	0.09	0.30	0.14
	8	20	0.25	0.12	0.25	0.13	0.19	0.10	0.31	0.20

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. Excerpts from speakers at bands 5 and 8 across L1 backgrounds.

Speaker 1: Chinese band 8
Cluster 1: 33.33%, cluster 2: 13.33%, cluster 3: 13.33%, cluster 4: 40%

Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:01.3	00:07.2	{um^} [I'd like to share a party with you] {^uh} [which happened several days ago]	SP–2; FP–2; RP–0; RF–0	1
00:07.8	00:11.2	{uh} [it was for (my)] {^} [one of my besties]	SP–1; FP–1; RP–0; RF–1	4
00:11.2	00:14.0	[she was (my)] {} [one of my best friends]	SP–0; FP–0; RP–0; RF–1	4
00:14.5	00:19.1	[actually she was striving to pass her IELTS exam with high band]	SP–0; FP–0; RP–0; RF–0	4
00:19.4	00:20.8	[and she works very hard]	SP–0; FP–0; RP–0; RF–0	4
00:21.7	00:23.9	{uh} [we (hold)] {we} [held this party]	SP–0; FP–1; RP–1; RF–1	3
00:24.5	00:29.4	{uh^} [and I] {^ I} [designed the whole party]	SP–2; FP–1; RP–1; RF–0	1
00:29.7	00:34.5	[it was held in an] {^} [ornate and] {uh} [exquisite cafe]	SP–1; FP–1; RP–0; RF–0	1
00:35.6	00:39.0	{uh} [I give her a present which was bought in Bali]	SP–0; FP–1; RP–0; RF–0	3
00:39.0	00:45.7	[it was a] {^} [intricate and impressive and unique] {^uh} [handmade stitching]	SP–2; FP–1; RP–0; RF–0	1
00:46.3	01:00.8	[I give it to her (because I)] {^uh I^} [as a] {^} [memento] {^} [because I] {^um^because I^} [wanted to remind her of our time]	SP–7; FP–2; RP–2; RF–1	2
01:01.5	01:14.2	[and the party's] {^} [ambience was] {uh^ uh was} [very (happy)] {^was very} [exciting] {^um} [but a little bit sentimental because she was leaving]	SP–4; FP–3; RP–1; RF–1	2
01:14.8	01:20.5	{uh} [and (she)] {^ she} [her destination it was Britain]	SP–1; FP–1; RP–1; RF–1	1
01:20.5	01:23.2	[but my destination was Australia]	SP–0; FP–0; RP–0; RF–0	4
01:23.7	01:25.7	[so (we)] {^} [actually we]	SP–1; FP–0; RP–0; RF–1	4

Speaker 2: Chinese band 5
Cluster 1: 28.57%, cluster 2: 42.86%, cluster 3: 21.43%, cluster 7.14%

Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:02.3	00:07.6	[as far as this topic I want to] {uh} [talk about Ma Buyi]	SP–0; FP–1; RP–0; RF–0	1
00:07.7	00:13.8	{uh} [he is a] {^um} [person I] {^uh I} [would like to meet]	SP–2; FP–3; RP–1; RF–0	2
00:16.3	00:18.2	{uh} [he's] {uh} [Chinese]	SP–0; FP–2; RP–0; RF–0	3
00:18.6	00:20.0	[and he's a singer]	SP–0; FP–0; RP–0; RF–0	4
00:22.1	00:26.6	[and he's a very famous <segerror: famerous>] {^uh} [in China]	SP–1; FP–1; RP–0; RF–0	1
00:26.7	00:40.4	[and] {uh^} [I know about this person] {uh} [when I] {^uh} [listen] {uh} [(beautiful)] {^um} [(songs)] {^} [a beautiful song]	SP–4; FP–5; RP–0; RF–2	2
00:40.9	00:45.0	[and] {uh} [I think the melody is very beautiful]	SP–0; FP–1; RP–0; RF–0	3

(*Continued*)

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. (*Continued*)

Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
Speaker 2: Chinese band 5 *Cluster 1: 28.57%, cluster 2: 42.86%, cluster 3: 21.43%, cluster 7.14%*				
00:45.1	00:52.5	[and] {uh} [I] {uh} [ask other people] {uh} [who] {uh^ who} [is this person]	SP–1; FP–4; RP–1; RF–0	1
00:52.5	00:56.7	[and] {uh} [(he)] {and uh} [others told me this is Mao Buyi]	SP–0; FP–2; RP–0; RF–1	3
00:56.7	01:08.5	[I think] {uh} [I] {I} [want to meet this person because] {um^} [her songs] {^uh} [are very] {^um} [beautiful]	SP–3; FP–4; RP–1; RF–0	2
01:08.5	01:17.9	[and] {uh uh} [I think] {uh} [he's] {uh^ uh} [very] {^um^} [work hard person]	SP–3; FP–6; RP–0; RF–0	2
01:18.1	01:34.3	[and] {uh^ um^} [from (her)] {^um from} [his] {^um} [songs I felt the pure] {^} [and] {uh} [the great passion for the music]	SP–5; FP–5; RP–1; RF–1	2
01:34.4	01:49.5	[I think] {uh^ h- uh} [his] {uh his} [songs can give me] {uh^ uh} [power to] {^um^} [solve problem]	SP–4; FP–6; RP–1; RF–0	2
01:49.5	01:51.6	[and] {uh} [when I]	SP–0; FP–1; RP–0; RF–0	3
Speaker 3: Punjabi band 8 *Cluster 1: 27.78%, cluster 2: 16.67%, cluster 3: 5.56%, cluster 4: 50.00%*				
00:00.0	00:02.3	{well} [actually history is not my cup of tea]	SP–0; FP–1; RP–0; RF–0	4
00:02.7	00:08.2	[so I would say that I would love to learn more about nineteen sixties or seventies]	SP–0; FP–0; RP–0; RF–0	4
00:08.6	00:18.1	[during that time there] {there} [were many incidents that happened in the past for example green revolution in India] {^} [and more importantly] {^} [the evolution of the internet]	SP–2; FP–0; RP–1; RF–0	1
00:18.3	00:19.7	[the reason is quite simple]	SP–0; FP–0; RP–0; RF–0	4
00:19.7	00:26.5	[the internet is now] {i- now internet is} [(m-)] {} [making our lives more convenient and easy]	SP–0; FP–0; RP–1; RF–1	4
00:26.8	00:33.4	[so (I)] {^} [when I studied] {^} [my childhood <unclear: accountancy> we used to have a subject (le-)] {} [of computers]	SP–2; FP–0; RP–0; RF–2	1
00:33.7	00:36.2	[and the first chapter was its history]	SP–0; FP–0; RP–0; RF–0	4
00:36.5	00:45.2	[when I studied it it was firstly designed (by)] {^you know} [for a millitary purpose like an intranet not actually the internet but for the intranet]	SP–1; FP–1; RP–0; RF–1	2
00:45.6	00:53.6	[so I] {^} [became fascinated by (what)] {^} [the person who (have invented the internet)] {^who} [had invented the internet]	SP–3; FP–0; RP–1; RF–2	2
00:53.6	00:57.3	[(whe-)] {} [what was the logic behind his ideology]	SP–0; FP–0; RP–0; RF–1	4

(*Continued*)

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. (*Continued*)

Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
Speaker 3: Punjabi band 8 *Cluster 1: 27.78%, cluster 2: 16.67%, cluster 3: 5.56%, cluster 4: 50.00%*				
00:57.3	00:58.8	[what was his methodology]	SP–0; FP–0; RP–0; RF–0	4
00:59.3	01:10.4	[although we are not] {^} [usually taught about history in our academic courses] {^you know} [in] {^} [detail unless we are] {^} [choosing that particular trade to learn]	SP–4; FP–1; RP–0; RF–0	1
01:10.7	01:23.0	[so when I was reading I become fascinated that whenever I will get a free time I definitely would learn the] {^} [concept or the ideology of that particular person]	SP–1; FP–0; RP–0; RF–0	1
01:24.1	01:32.1	[I have already (t-)] {} [told that this is the time period of nineteen sixties and seventies] {^} [and] {uh} [(w-)] {} [why I've become interested]	SP–1; FP–1; RP–0; RF–2	2
01:32.4	01:33.1	[but the] {} []	SP–0; FP–0; RP–0; RF–0	4
01:33.6	01:39.3	{um} [and I know] {you know} [the elementary or rudimentary] {^} [concept of its development]	SP–1; FP–2; RP–0; RF–0	3
01:39.4	01:55.8	[but I want to go in detail what actually the steps were there in the history so that I can also] {^} [plan and think about that strategy for that enterpreneur had in their mind] {^} [while developing] {^} [the internet]	SP–3; FP–0; RP–0; RF–0	1
01:56.1	02:00.3	{well} [I would love to read more history so that's why I followed a page on]	SP–0; FP–1; RP–0; RF–0	4
Speaker 4: Punjabi band 5 *Cluster 1: 33.33%, cluster 2: 38.89%, cluster 3: 11.11%, cluster 4: 16.67%*				
Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:00.9	00:10.1	[books are very important for every person because] {uh} [when people (book read)] {^} [read a book that then that time they come to know about different type of things]	SP–1; FP–1; RP–0; RF–1	2
00:10.2	00:13.6	[and as they also] {^} [improve their reading skills]	SP–1; FP–0; RP–0; RF–0	1
00:13.6	00:21.1	[but here I would like to talk about a book which I] {^uh which I} [mostly] {uh} [read in online] {^uh} [last year]	SP–2; FP–3; RP–1; RF–0	2
00:21.8	00:24.8	[it (i-)] {uh it} [was a book regarding history]	SP–0; FP–1; RP–1; RF–0	3
00:25.1	00:28.0	[the name of book was history of India]	SP–0; FP–0; RP–0; RF–0	4
00:28.7	00:33.6	[(rec-)] {uh} [in last year I visit a museum with my friends] {uh} [in my summer vacation]	SP–0; FP–2; RP–0; RF–1	3
00:33.9	00:36.3	[for that time I read this book]	SP–0; FP–0; RP–0; RF–0	4

(*Continued*)

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. (*Continued*)

Speaker 4: Punjabi band 5 *Cluster 1: 33.33%, cluster 2: 38.89%, cluster 3: 11.11%, cluster 4: 16.67%*				
Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:38.2	00:47.3	[this book (is regarding the)] {^uh this book reg-} [describe the all] {^} [history and culture of] {^} [India]	SP–3; FP–1; RP–1; RF–1	2
00:47.8	00:52.8	[and] {uh^} [from this book I] {^} [come to know about] {^} [different kind of things]	SP–3; FP–1; RP–0; RF–0	1
00:52.8	01:06.7	[like I] {^} [come to know about my (p-)] {} [past culture and tradition] {^} [which is very important for] {^} [me] {^} [because nowadays mostly youngster] {^uh^} [forget their] {^} [culture and tradition]	SP–7; FP–1; RP–0; RF–1	2
01:07.0	01:20.5	[because] {uh} [they mostly prefer to] {^uh^} [they use technological gadgets and] {uh^} [like to] {^uh^ like to} [improve (their se-)] {} [themself] {^} [in this society]	SP–6; FP–4; RP–1; RF–1	2
01:20.9	01:26.4	[that's why I come to know about different kind of things from this] {^} [book]	SP–1; FP–0; RP–0; RF–0	1
01:26.9	01:31.5	{uh} [I also] {^} [come to know about my] {^uh} [cultural things]	SP–2; FP–2; RP–0; RF–0	1
01:31.5	01:36.3	{like} [I belong to a sikh family] {^} [so I come to know about my] {^} [sikhism]	SP–2; FP–1; RP–0; RF–0	1
01:36.6	01:44.8	{like} [I come to know about] {uh} [many things regarding my sikhism] {^} [which is very important (to v-)] {important} [for me]	SP–1; FP–2; RP–1; RF–1	2
01:45.1	01:51.1	[and] {uh} [my grandfather also tell about] {^uh^} [(me a-)] {about} [my culture and traditions]	SP–2; FP–2; RP–1; RF–1	2
01:51.1	01:54.0	[that's why it is very] {^} [important for me]	SP–1; FP–0; RP–0; RF–0	4
01:55.1	01:59.6	[moreover] {^} [when I read this book at that time I felt ve-]	SP–1; FP–0; RP–0; RF–0	1
Speaker 5: Arabic band 5 *Cluster 1: 21.05%, cluster 2: 31.58%, cluster 3: 10.53%, cluster 4: 36.84%*				
Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:17.9	00:21.2	[so I've enjoyed working with my close friend]	SP–0; FP–0; RP–0; RF–0	4
00:21.5	00:22.4	[her name is Mariam]	SP–0; FP–0; RP–0; RF–0	4
00:23.1	00:26.0	[okay I've been with her since Elementary school]	SP–0; FP–1; RP–0; RF–0	4
00:26.0	00:37.0	[and what's (make)] {^ what} [makes it fun and enjoyable to (w-)] {} [work with her and] {^ and} [study together is because she's] {^} [trustworthy and reliable]	SP–3; FP–0; RP–2; RF–2	2
00:39.1	00:41.6	[and] {uh} [I can tell her everything]	SP–0; FP–1; RP–0; RF–0	3

(*Continued*)

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. (*Continued*)

Speaker 5: Arabic band 5 *Cluster 1: 21.05%, cluster 2: 31.58%, cluster 3: 10.53%, cluster 4: 36.84%*				
Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:41.8	00:44.6	[we] {^} [work] {work} [in a projects]	SP–1; FP–0; RP–1; RF–0	1
00:44.6	00:46.3	[and we never fight]	SP–0; FP–0; RP–0; RF–0	4
00:46.8	00:49.2	[that's the] {uh that's the} [main reason]	SP–0; FP–1; RP–1; RF–0	3
00:50.9	00:53.7	[and also I can talk to her whenever I want]	SP–0; FP–0; RP–0; RF–0	4
00:54.4	00:58.0	[we also (g-)] {} [go off together to study or have fun]	SP–0; FP–0; RP–0; RF–1	4
01:01.3	01:11.8	[also she (is close to me so)] {^ she} [lives close to me so I can] {^ I can} [go in five minutes to her house] {^} [and talk to her about anything]	SP–3; FP–0; RP–2; RF–0	2
01:15.8	01:21.8	[we] {^ we} [study mainly science] {^} [subjects and]	SP–2; FP–0; RP–1; RF–0	1
01:23.1	01:29.1	[and she (he-)] {she} [helps (me)] {^ she helps} [to explain to me some details that I don't understand]	SP–1; FP–0; RP–2; RF–2	2
01:31.4	01:36.5	[and yeah also] {^} [Can I describe someone else?]	SP–1; FP–0; RP–0; RF–0	1
01:36.9	01:45.0	[yeah also my other friend Zaina] {uh^} [she's another friend go so we all go (each oth-)] {^uh} [to each other houses]	SP–2; FP–2; RP–0; RF–1	2
01:45.0	01:49.0	[and we] {^} [study] {^} [to get good (gr-)] {} [grades]	SP–2; FP–0; RP–0; RF–1	1
01:49.7	01:56.5	[we mainly (mo-)] {} [motivate] {^ motivate} [each other] {^} [so (h-)] {} [that's how we get good marks]	SP–2; FP–0; RP–1; RF–2	2
02:00.7	02:07.8	[we also] {like^} [watch TV shows together] {^} [and talk about everything and] {everything}	SP–2; FP–1; RP–1; RF–0	2
02:09.4	02:10.7	[there is no racism]	SP–0; FP–0; RP–0; RF–0	4
Speaker 6: Arabic band 8 *Cluster 1: 23.08%, cluster 2: 38.46%, cluster 3: 7.69%, cluster 4: 30.77%*				
Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:00.0	00:05.9	[so I bought a car back in when I used to live in Saudi Arabia before I joined the American University of Sharjah]	SP–0; FP–0; RP–0; RF–0	4
00:06.6	00:16.6	{um uh} [the reason I bought it in Saudi Arabia is that] {^uh} [to get from one] {^} [place to the other it's a very long distance and the public transportation over there it's not very good]	SP–2; FP–3; RP–0; RF–0	2
00:16.6	00:21.9	[so] {^} [you actually need a car or you're gonna pay a lot of money on taxis so this is why I bought it]	SP–1; FP–0; RP–0; RF–0	1
00:22.6	00:33.8	{uh} [the reason I'm not using it very much] {^} [in the last four years when I (was in)] {when} [i'm in the American University of Sharjah is that] {^ uh} [I only use it] {^} [to get to Dubai]	SP–3; FP–2; RP–1; RF–1	2

(*Continued*)

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press

Appendix B. (*Continued*)

Speaker 6: Arabic band 8
Cluster 1: 23.08%, cluster 2: 38.46%, cluster 3: 7.69%, cluster 4: 30.77%

Begin time	End time	AS-unit transcript	Disfluencies	Cluster type
00:33.8	00:42.8	[but once I'm in Dubai] {^} [(cause)] {^} [it's cause there is a lot of traffic and all that I] {^} [generally use the public transportation as it can get you nearly everywhere]	SP–3; FP–1; RP–0; RF–1	1
00:43.4	00:45.7	[and plus it costs less than fuel]	SP–0; FP–0; RP–0; RF–0	4
00:46.0	00:51.0	[and] {uh and} [you would save a lot of time that you would waste on traffic I] {uh} [like the Dubai metro]	SP–0; FP–2; RP–1; RF–0	3
00:53.4	01:01.3	{uh} [and another reason is I] {uh} [for] {for} [the past three years I've been on the soccer team of my] {uh ^} [university]	SP–2; FP–3; RP–1; RF–0	2
01:01.3	01:03.5	[and I like to get active and all that]	SP–0; FP–0; RP–0; RF–0	4
01:03.5	01:07.4	[I would rather walk if the weather is good then] {^} [take a car and all that]	SP–1; FP–0; RP–0; RF–0	1
01:08.2	01:09.5	[so yeah that's basically it]	SP–0; FP–0; RP–0; RF–0	4
01:12.2	01:23.6	{uh} [another reason is that] {^} [the increased amount of] {uh} [accidents that happen due to the traffic might] {uh ^ might} [cause lots of deaths and injuries and all that]	SP–2; FP–3; RP–1; RF–0	2
01:23.6	01:34.5	[so I'd rather stay safe use a metro or (a)] {^} [(bu-)] {or} [public bus where it's much safer than driving a car] {^} [on the highway where speeds can reach up to hundred thirty kilometer per hour so]	SP–2; FP–0; RP–1; RF–2	2

Note: [] = original or intended utterance; () = reparandum; {} = editing phase; ^ = silent pause; FP = filled pause; RF = reformulation; RP = repetition; SP = silent pause. Each row represents a separate AS-unit, and the duration of AS-unit is computed as the difference between the end time and the begin times. This applies to all the tables in Appendix B.

Cite this article: Yan, X., Chuang, P.-L., Pan, Y., Cai, H., Staples, S., & Bertho, M. C. (2025). Disfluency doesn't happen in isolation: Exploring how individual disfluency features co-occur in L2 speaking performances. *Studies in Second Language Acquisition*, 47: 560–591. https://doi.org/10.1017/S0272263125000245

https://doi.org/10.1017/S0272263125000245 Published online by Cambridge University Press