Disfluencies in Spoken Language: Analyzing Fillers and Repetitions in Relation to Speech Rate

Linnéa Rydén

Department of Linguistics
Bachelor's Thesis 15 ECTS credits
Linguistics: Experimental Linguistics – Bachelor's Course, LIN633
Bachelor's Programme in Linguistics 180 ECTS credits
Spring semester 2025
Supervisor: Julia Uddén
Swedish title: Talets dysfluens: En studie av utfyllnadsord och upprepningar i relation till talhastighet

Stockholm University

Disfluencies in Spoken Language: Analyzing Fillers and Repetitions in Relation to Speech Rate

Linnéa Rydén

Abstract

This study investigates the relationship between disfluencies and hesitation by examining how these are reflected in speech rate. Disfluencies such as fillers ("uh", "uhm") and repetitions are common in spontaneous speech, but their connection to cognitive processes such as hesitation and uncertainty remains the subject of ongoing research. Using an existing dataset, the study included annotation of various types of disfluencies and analysis of their distribution in relation to speech rate. The study centers around the following questions: Are fillers associated with speech rate, and could this pattern reflect a link to hesitation or uncertainty? Furthermore, do fillers differ from repetitions in how they relate to speech rate — and thus in their connection to hesitation or uncertainty? The results show that speakers use fillers more often during slower speech than they use repetitions. This suggests that underlying processes such as hesitation and uncertainty are more likely to lead to the use of fillers. These can reduce speech rate and more strongly signal hesitation in the face of uncertainty than repetitions do.

Keywords

Disfluency, hesitation, speech rate, fillers, repetitions, uncertainty, spontaneous speech

Dysfluenser i talat språk: En analys av fyllnadsord och upprepningar i relation till talhastighet

Linnéa Rydén

Sammanfattning

Denna studie undersöker sambandet mellan dysfluenser och tvekan genom att studera hur dessa yttrar sig i talhastighet. Dysfluenser såsom fyllnadsord (’eh’, ’ehm’) och upprepningar är vanliga i spontant tal, men deras koppling till kognitiva processer som tvekan och osäkerhet är fortfarande föremål för pågående forskning. Med hjälp av ett befintligt dataset omfattade studien annotering av olika typer av dysfluenser samt analys av deras fördelning i relation till talhastighet. Studien kretsar kring frågorna: Är fyllnadsord associerade med talhastighet och kan detta mönster spegla en koppling till tvekan eller osäkerhet? Vidare, skiljer sig fyllnadsord från upprepningar i hur de hänger ihop med talhastighet – och vad kan det i så fall säga om deras koppling till tvekan eller osäkerhet? Resultaten visar att talare använder fyllnadsord oftare i långsammare tal än de använder upprepningar. Detta tyder på att underliggande processer såsom tvekan och osäkerhet snarare tycks leda till användning av fyllnadsord. Dessa kan sänka talhastigheten och i högre grad signalera tvekan vid osäkerhet än vad upprepningar gör.

Nyckelord

Dysfluenser, tvekan, talhastighet, fyllnadsord, upprepningar, osäkerhet, spontant tal

Contents

1 Introduction 1

2 Background 2
- 2.1 Conversation . . . 2
- 2.2 Hesitation and disfluency . . . 2
  - 2.2.1 Fillers . . . 3
  - 2.2.2 Repetitions . . . 3
  - 2.2.3 Speech rate . . . 3
- 2.3 Purpose and research questions . . . 4

3 Method 5
- 3.1 Data . . . 5
  - 3.1.1 Participants . . . 5
  - 3.1.2 Procedure . . . 5
- 3.2 Annotation . . . 5
- 3.3 Analysis . . . 6

4 Results 7

5 Discussion 8
- 5.1 Results, hypotheses and research questions . . . 8
- 5.2 Method discussion . . . 8
  - 5.2.1 Data . . . 8
  - 5.2.2 Procedure . . . 8
  - 5.2.3 Annotation . . . 8
  - 5.2.4 Analysis . . . 9
- 5.3 Future research . . . 9

6 Conclusion 11

References 12

Appendix 14

1 Introduction

Disfluency in speech impact the overall conversation in different ways. We, as human beings, use disfluency every day in our natural language. However, there is not much litterature on the patterns of how we display disfluencies. Speech disfluency in general have been studied for some time, at least, since the 1930s (Eklund and Ingvar 2016, p.1). A study by De Oliveira, C. M. C., et al. (2013) showed a tendency to a higher frequency for common disfluencies with increased speech rate. In a study by Oomen and Postma (2001), the authors induced different speech rate conditions together with the disfluency types fillers (eg. filled pauses) and repetitions. They found that participants, in the faster speech rate condition, seemed to use more repetitions than fillers. However, the rate at which fillers were used remained consistent throughout the experiment. (Oomen and Postma 2001b, Corley and Stewart 2008). This gives rise to the question of whether there is a correlation between speech rate and disfluency — particularly regarding the relationship between speech rate, fillers, and repetitions. This study investigates two central research questions: *Are fillers associated with speech rate, and could this pattern reflect a link to hesitation or uncertainty?* And *do fillers differ from repetitions in how they relate to speech rate — and thus in their connection to hesitation or uncertainty?*

2 Background

2.1 Conversation

In conversation, language processing is the dominant explanandum in psycho- and neurolinguistics. To be in conversation is to entail the roles of speaker and listener while considering contextual factors such as linguistic and social (Arvidsson et al. 2024, p.1). The process of conversation contains a lot of different steps. The participants has to do certain tasks to engage successfully in the conversation. These tasks include opening up for conversation and being engaged in it, create meaning to the conversation and evolve it, converge on agreement and finally taking action or do a transaction (Dubberly and Pangaro 2009, p.23-24). In conversation, speech flow can be an important factor in keeping the conversation going.

2.2 Hesitation and disfluency

Hesitation often involves a temporary break in speech flow. This can be manifested by momentary silence, syllable elongation, by using a filled pause or a lexical filler. It can also be achieved by repeating the onset of the current phrase or openly expressing uncertainty (Lickley 2015, p.456). Speakers commonly use language marked with hesitations, false starts and repetitions; and making sounds like *uh* and *uhm*, or choose to elongate forms of words (Arnold et al. 2004, Brennan and Schober 2001). These disfluencies play a role not only in speech production but also in how listeners interpret speech. During disfluent speech, the listener must edit out the disfluencies. This means understanding that there is a problem with the utterance, determining what the problem is, and also getting to terms with how to repair the utterance. All this requires the listener to identify different intervals, where the *reparandum* is one. The reparandum contains fluent speech up until the *interruption site*, which is were the speaker stops speaking fluently (Brennan and Schober 2001, p.1).

Fox Tree argues that different types of disfluencies lead to different effects in conversations. Hence, one type of disfluency may cause more or less trouble for comprehension depending on its place in the utterance (Tree 1995, p.730). This relates to the findings of Oomen and Postma (2001) which were mentioned in the *introduction*, section 1. They suggested that fillers and repetitions might be controlled by different processes (p.180). In addition to their communicative effects, disfluency rates themselves can vary depending on a number of factors, such as task complexity and the amount of preplanning or rehearsal involved (E. E. Shriberg 1994, p.17). Fraundorf and Watson (2014, p.1094) argue that fillers are typically used when speakers are planning upcoming speech and have not recently articulated repeatable material. This implies that fillers mark moments of increased cognitive load or uncertainty, which may slow down speech. This claim is directly relevant to the current research questions, which examine whether fillers are associated with slower speech rate — and whether this differs from the use of repetitions — as potential indicators of hesitation or uncertainty.

In addition to verbal cues, non-verbal cues such as gaze can play a certain role in manifesting hesitation. Beattie (1978) proposed that gaze aversion (i.e. looking away while speaking) may be linked to hesitation or uncertainty. Furthermore, Schultz et al. (2008, p.3010) showed that different neural signals are associated with different levels of uncertainty. Among the different ways people show hesitation, verbal cues, such as fillers, are some of the most frequently used.

2.2.1 Fillers

Fillers are perhaps the most common type of disfluency, examples being *uh* and *uhm* in the English language-sphere (Corley and Stewart 2008, p.1–2). Fillers, or *filled pauses*, are mostly defined by frequency, duration, fundamental (pitch) and location. They are often produced in conjunction with silent pauses and prolongation. The form of a filler can vary between different languages and it seems to be common that filled pauses have at least two distinct forms in a language (Lickley 2015, p.458; Williams 2022, p.75). Another common form of disfluency is repetition, which, like fillers, can take several forms and serve different functions in speech.

2.2.2 Repetitions

There are different types of repetition. First, there are what are called sublexical repetitions; these are sound or part-word repetitions. Then there are lexical repetitions, which are repeated words. There are also supralexical repetitions, which are phrase or multiple word repetitions (Oomen and Postma 2001a, p.1001). This study uses lexical repetitions; see *Method* in section 3.

What makes a repetition disfluent is the context in which it occurs. Furthermore, repetitions tend to appear in similar contexts as other disfluent pauses (MacGregor et al. 2009, p.1–2). However, prosodic differences, such as when a speaker repeats a word to emphasize, persuade, or for rhetorical effect, can distinguish fluent repetitions from disfluent ones. When a speaker disfluently repeats part of an utterance, that repetition will typically be prosodically similar to the original word, for instance, in pitch level. A disfluent repetition is also often accompanied by other signs of disfluency, such as silent pauses or sound prolongation (Lickley 2015, p.459–460). Fraundorf and Watson (2014, p.1094) showed that disfluent repetitions are typically used while a speaker is already articulating a segment of speech that can be easily repeated. This suggests that repetitions may function as more local disfluencies — they are closely tied to the immediate speech stream and rely on material that has just been produced. Because repetitions involve surface-level articulation rather than the formulation of new content, they may reflect a relatively automatic strategy for maintaining fluency with minimal cognitive disruption. In contrast, fillers such as uh or um often occur at points of greater planning difficulty, indicating potential hesitation or uncertainty. To explore these functional differences further, this study investigates the role of speech rate in the production of fillers and repetitions, asking whether speech rate can help differentiate these two types of disfluencies in terms of their connection to hesitation processes.

2.2.3 Speech rate

Oomen and Postma (2001) argue that a faster speech rate can be described as speech with shorter pauses and a higher articulation rate. According to Oomen and Postma, this is because speech rate, in fact, consists of variations in articulation rate and pause duration (Oomen and Postma 2001b, p.168). This relationship between articulation rate and pause duration is often calculated using the formula ‘number of syllables divided by total time, including silences’ (Bosker et al. 2013, p.161), which is the same measure this study uses to calculate speech rate, but applied per utterance. See *Method*, section 3.

In their experimental study, Oomen and Postma manipulated time pressure by asking participants to describe visual networks at normal and fast rates. They found that increased speech rate led to a rise in speech errors and a significant increase in repetitions, while filled pauses remained stable. Their analysis suggested that different disfluency types may stem from dis-

tinct underlying mechanisms: repetitions were linked to automatic timing problems in speech planning under high articulation pressure, while filled pauses appeared less sensitive to such constraints (Oomen and Postma 2001b). Furthermore, a study by Huttunen et al.(2011, p.1588) suggests that an increase in cognitive load leads to a slower speech rate. This highlights the role of cognitive factors in shaping how we speak, especially when hesitation or uncertainty is involved. In this study, hesitation is interpreted as a potential marker of uncertainty, a sign that the speaker may still be engaged in planning or decision-making. This link between hesitation and uncertainty is central to understanding how disfluencies reflect cognitive processes in real time during speech.

2.3 Purpose and research questions

The purpose of this study is to investigate two types of speech disfluencies and their potential relationship to speech rate, as well as to examine whether differences in speech rate can be observed. More broadly, it aims to explore how hesitation relates to disfluency, using speech rate as an indicator.

As already stated, the main research questions are as follows: Are fillers associated with speech rate, and could this pattern reflect a link to hesitation or uncertainty? Furthermore, do fillers differ from repetitions in how they relate to speech rate — and thus in their connection to hesitation or uncertainty? The hypothesis of this study is that fillers and repetitions differ in their association with speech rate, and therefore hesitation or uncertainty, with fillers being more closely related to these cognitive processes. The research questions are grounded in the assumption that different types of disfluencies, such as fillers and repetitions, may reflect different underlying cognitive processes during speech production. By examining how these disfluencies relate to speech rate, this study seeks to determine whether the type of disfluency the speaker uses depends on indications of hesitation already in the reparandum. Speech rate is used here as an indicator of hesitation or cognitive load, allowing for a comparison of the speech patterns that occur when speakers produce fillers versus repetitions. This study seeks to contribute to our understanding of how speakers manage planning difficulties, or hesitation, in natural speech.

3 Method

3.1 Data

3.1.1 Participants

This study utilized data from a dataset of human-human interaction by Torubarova et al. (2025). The dataset originally consisted of 33 participants who engaged in three runs of 10-minute free conversations. Although, two participants were excluded from the dataset because of technical issues during the data collection. The participants were all right-handed adults of mixed ages and genders. They were all healthy and their first language were Swedish (Torubarova et al. 2025). One additional participant were excluded from the dataset because they were not a native speaker of swedish. This means that the dataset consisted of 30 participants from the start. However, for five of the participants, transcriptions did not exist in the database due to different technical issues. This means that, for this study, 25 participants were available. During the stage of the analysis for this thesis, six participants got excluded because they did not meet the requirements of using both disfluency types. Therefore, this study used only 19 participants. These were 8 female and 11 male with an age-span of 21-39 years old (Torubarova et al. 2024).

3.1.2 Procedure

The participants were conversing with a confederate who were the key to manipulate the conversation across three different levels of engagement: Engaged Communicator (EC), Active Listener (AL), and Passive Listener (PL). In the EC condition, the confederate was actively engaged in the conversation by furthering the topic, asking questions, frequently using backchannels, and employing paraphrasing and summarization. In the AL condition, the confederate demonstrated engagement through backchannels and summarization but did not proactively drive the conversation forward. In the PL condition, the confederate remained largely unresponsive, providing minimal feedback and asking no questions — although they still responded when directly addressed by the participant. The conversations where all held in swedish and were about opinions on different ethical dilemmas (Torubarova et al. 2025, p. 2-3). This study focused exclusively on the Active Listener level as this was thought to most ressemble natural conversation, see *Discussion*, section 5.

Throughout the experiment, participants were conversing from inside an fMRI scanner - while the confederate were conversing from a different room. The participants had been encouraged to maintain a natural conversational flow, speaking as they would with a friend. They had also been instructed to first read the dilemma presented on a screen and then indicate how much they agreed with the dilemma statement using a five-point Likert scale, where 1 corresponded to "completely disagree" and 5 to "completely agree." One example of a dilemma is *Would you take a DNA test before a first date with a potential partner?*. After the conversations, they were asked to state their opinion again (Torubarova et al. 2025).

3.2 Annotation

This study utilized pre-existing transcriptions and speech units from the dataset, which was automatically segmented with a 200 ms minimum silence duration threshold (Torubarova et al. 2025, p.4). The conversations had been transcribed orthographically, preserving the original spoken content without phonetic modifications.

The data were annotated in ELAN, a specialized annotation tool for language analysis (Tacchetti 2017). Annotations were based on participants' utterances and categorized into two types: the fillers *uh* and *uhm* (corresponding to 'eh' and 'ehm' in Swedish), and repetitions, thus referring to lexical repetitions (see *Repetitions*, 2.2.2), where applicable. Utterances were annotated up to the point of the disfluency, while those consisting solely of a disfluency were excluded. Disfluent repetitions (see *Repetitions*, Section 2.2.2) were distinguished from fluent ones based on auditory discrimination of prosodic features.

It was from the onset of the utterance and up until the disfluency where the hypothesized differences in terms of speech rate would be observable. For utterances containing both repetitions and fillers, a few procedures were taken depending on the contexts:

- If a sentence started with a filler, this filler got ignored and the counting of syllables started afterwards.
- If a disfluency came after a repetition, with a few words in between, the later of the repeated word(s) counted as a normal word.
- If there were no words between two disfluencies, only the first one got counted.

Utterances which started with a disfluency and did not contain anymore disfluencies were ignored. This approach was chosen to focus on the fluent speech preceding the disfluency rather than the disfluency itself, allowing for an analysis of speech rate before the interruption site.

The annotations, and corresponding transcriptions, were subsequently used to calculate speech rate in utterances containing disfluencies, forming the basis for the statistical analysis. For examples of annotations with two disfluencies in the same utterance, see the *Appendix* in section 6.

3.3 Analysis

The data analysis involved the development of custom scripts in Visual Studio Code to align transcription files with annotation files, utilizing timestamps from the dataset and from the annotations. Furthermore, for each utterance containing disfluencies, syllables per second (syllables/s) were calculated. Utterances with fewer than three syllables were excluded, as they were deemed irrelevant to the study objectives.

A linear mixed model was implemented using the JASP program (JASP Team 2025) to examine the relationship between disfluency type and speech rate. Given that each participant contributed multiple utterances, a mixed-effects modeling approach was chosen to account for both within- and between-participant variability. In this model, disfluency type was treated as a fixed effect, as it remained consistent across participants, while participants were treated as a random variable to capture individual differences in speech rate. The dependent variable, or *response* variable, was speech rate, which constituted the primary outcome of interest in the analysis (Bolker 2015, p. 312–315). Since the linear mixed model required data in which both types of disfluencies were present for each participant, individuals exhibiting only one type of disfluency were excluded from the study.

The statistical analysis focused on testing the significance of the relationship between speech rate and types of disfluencies, with the expectation that speech rate would differ in terms of being faster with the usage of repetition rather than fillers. The alpha-level was set to 0.05. Finally, a plot was generated in JASP to visually represent the distribution of speech rates between the two disfluency types, illustrating potential differences in speech rate associated with each type. This approach provided a rigorous evaluation of whether disfluency type distribution varied systematically with speech rate while accounting for individual variability.

4 Results

The analysis of the data showed $F_{1,12.91} = 13.679, p = 0.003$. Table 1 shows statistics for speech rate and the distribution of the two types of disfluencies per participant. All rates are reported in syllables per second (syllables/s).

For fillers, there were 359 valid instances. Participants produced an average of 15.8 filler instances each (range: 2–45). The mean speech rate during utterances containing fillers was 5.7 syllables per second, with a minimum of 2.3 and a maximum of 12.5 syllables per second.

For repetitions, 116 valid instances were identified. Participants produced an average of 14.7 repetition instances (range: 2–17). The mean speech rate for utterances with repetitions was 6.8 syllables per second, with values ranging from 1.4 to 16.2 syllables per second.

Table 1: Speech rate and labels per participant

	Syllables/s	
	Fillers	Repetitions
Valid (n)	359	116
Mean valid n (per participant, unweighted)	15, 8	14, 7
Min valid n	2	2
Max valid n	45	17
Mean (rate)	5.7	6.8
Min	2.3	1.4
Max	12.5	16.2

The plot in figure 1 is visually representing the difference between the two disfluency types in regards to speech rate. The grey dots shows the instances of the disfluencies and its speech rate, while the black area shows the average speech rate for the two groups. The lines and asterisks above the figure indicate statistically significant results, based on the alpha level defined in the *Analysis*, section 3.3.

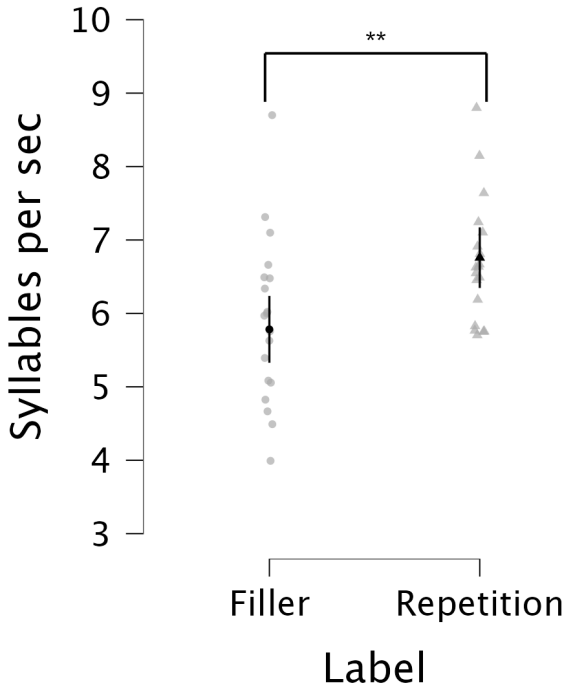

Figure 1: Disfluency types corresponding to speech rate; calculated using syllables/s

5 Discussion

5.1 Results, hypotheses and research questions

This study posed the questions: Are fillers associated with speech rate, and could this pattern reflect a link to hesitation or uncertainty? Furthermore, do fillers differ from repetitions in how they relate to speech rate — and thus in their connection to hesitation or uncertainty?

The results show that a slower speech rate is associated with a higher use of fillers rather than repetitions. This is indicated both in figure 1 and table 1, as seen in the *Results*, section 4. Table 1 presents the valid number of disfluencies. Though the greater number of fillers, repetitions showed a stronger association with higher speech rates. This is also reflected in figure 1, where the average speech rate is higher for repetitions than for fillers.

One hypothesis for this thesis was, as indicated by the second research question and in *Purpose and research questions*, section 2.3, that fillers would differ from repetitions in matter of hesitation or uncertainty. If theoretically, a slower speech rate equals more hesitation, this study points to fillers being used more in a hesitation context as opposed to repetitions. This insight gives light to various different questions, see *Future research*, section 5.3.

5.2 Method discussion

5.2.1 Data

The number of participants in this study were quite high. Though as previously mentioned, there were a lot of technical issues that led to a lot of participants having to be excluded from the dataset and the study. More participants or more data overall could maybe had given more data on repetitions, as it would possibly have given an estimate on how much they relate to speech rate.

5.2.2 Procedure

By choosing to focus on the *active listener* level of engagement, this study opened up for more questions and thoughts, see *Procedure*, section 3.1.2. One thought is that the active listener level may enhance the number of disfluencies because the participant becomes required to speak more, as the confederate becomes less of an active speaker in the conversation. This could potentially have lead to more disfluencies than it would have for the *engaged communicator* level, where the participant hypothetically could speak less because the confederate would take up more space in the conversation. In contrast, the *passive listener* condition may have elicited even more disfluencies, as participants might have felt the need to keep their turn and sustain the conversation without much input from the confederate. However, exploring this further would necessitate a broader investigation into turn taking dynamics and whether different types of disfluencies serve distinct functions, such as signaling hesitation versus maintaining the speaker's turn.

Because the confederate, in the passive listener stage, would not give that much feedback, this engagement-level could be subject to a lot of hesitation regarding whether the participant should keep talking or not, see *Future research*, section 5.3. There is a lot to consider regarding this study's procedure, from the used dataset to the annotation procedures.

5.2.3 Annotation

The annotations had some procedures that may not have been the best approaches to take. The annotational procedures were thought up as the annotation went along. Therefore, one could

think that using existing annotation guidelines beforehand could have made it easier for others to reciprocate the study (see *Annotation*, section 3.2 - for example of annotation, see *Appendix*, section 6). However, the procedures taken were in the interest of contributing more data for the analysis, which it ultimately did (see, *Analysis*, 3.3). This is something that may be thought about regarding the results, and something that might need to be revisited in future research.

5.2.4 Analysis

Another potentially important aspect of the study would have been to analyze the utterances that did not contain disfluencies and to calculate an average speech rate for each participant. Including such measures could have provided a broader and more comprehensive analysis, potentially offering stronger support for the study's conclusions. Additionally, it might have helped account more effectively for individual variability in speech patterns. Unfortunately, this could not be done within the time constraints of the study's procedure.

Regarding the transcriptions, which were used for the analysis together with the annotations, there are some important considerations. Since the transcriptions were made orthographically, some words may have been written in their full form even if they were not fully articulated in speech. This means that slightly more syllables could have been counted in the analysis than were actually spoken by the participants. As a result, the calculated speech rate may have been skewed in favor of faster speech. In other words, if a speaker reduced or skipped parts of words in real-time speech - but the transcription reflected the full word — this would lead to a higher syllable count per second than what was actually produced.

5.3 Future research

For future research on the topic, there is considerable potential for expansion. On the one hand, more levels of engagement could be interesting to analyze and compare to one another by using the dataset by Torubarova et al. 2025 (see *Participants*, section 3.1.1). This would give more insight into how engagement influences disfluency. Questions that could be asked in this case would be if the rates of disfluencies notably changes with engagement and which of the disfluency types changes the most.

On the other hand, more participants could be included, as explained in the *Data* discussion (section 5.2.1). In addition, other types of disfluencies and contributing factors could be analyzed. This study focuses on only two disfluency types, but others—such as silent pauses or syllable prolongation—may also be worth investigating.

Furthermore, additional indicators of hesitation or uncertainty, briefly mentioned in the *Background*, section 2, include measures like eye-tracking or brain imaging, both methods of which are part of the dataset used in this study. Based on the earlier findings, these factors could help provide more insight into the subject and are definitely worth exploring further. Eye-tracking could be a key to see how gaze relates to disfluencies. As suggested by Beattie (1978), gaze aversion might be linked to these underlying processes of hesitation and uncertainty. To study gaze aversion with material already existing could answer the questions: Do people look away more when they use fillers versus repetitions and is there a difference to where they look directon-wise (up, left, down, right)? Furthermore, a study by Pistono et al. (2021) found that lexical access difficulties in the initial stage resulted in specific disfluencies and eye-movements. To study eye-movements with disfluencies could further investigate if certain eye-movements relates to certain disfluencies.

Brain imaging could maybe give information about brain activity and if that is related to the choice of disfluency type. The results posed by Schultz et al. (2008) is pointing towards

different neural signals being associated with different levels of uncertainty. Relating to this studys results, could the reason why utterances using filler-disfluencies generelly have a slower speech rate, than that for repetitions, be a result of more brain activity? Do people experience higher levels of cognitive load when they are using fillers rather than when they are using repetitions? And, finally, does different neural signals relate to different disfluency types? A study by Theys et al. (2020) showed that producing disfluencies in non-habitual speech led to more brain activity than normal habitual speech. The findings by Theys et al. support the idea that some disfluencies might signal more active planning, aligning with the thought that speakers may indeed be experiencing higher levels of cognitive load when using some types of disfluencies. Future research incorporating brain imaging could provide more precise insights into which types of disfluencies are associated with increased neural activity.

As a final point, this study contributes to a better understanding of disfluencies, which may be valuable in the development of language models for both human and machine processing (E. Shriberg 2001, p.167). Disfluencies also play a role in social and cognitive contexts. For instance, suspects often exhibit more pauses when lying (Vrij et al. 2006), suggesting that disfluencies may serve as markers for hesitation or uncertainty. Further exploration of this connection could aid in detecting deception.

6 Conclusion

This study aimed to investigate two types of speech disfluencies —fillers and repetitions— and their potential relationship to speech rate. Specifically, it examined whether differences in speech rate between disfluency types might reflect underlying cognitive processes such as hesitation or uncertainty. The findings shows that speakers use fillers more in slower speech than they use repetitions. This in turn could indicate that different cognitive processes, for instance hesitation, may underlie the productions of disfluencies.

The results support the hypothesis that fillers are more closely associated with hesitation or uncertainty, while repetitions appear to reflect more localized or surface-level planning issues. This distinction contributes to a deeper understanding of how speakers manage cognitive load in spontaneous speech. This conclusion is drawn despite different limitations such as technical issues, data imbalance between disfluency types and limited demographic diversity - which may have influenced the generalizability of the findings.

Understanding the roles of disfluencies has different types of practical implications. Not only can theories of speech production benefit from research on disfluencies, but also language models and forensic linguistics can have use for this kind of research. Future research could expand on the subject by including other disfluency types, incorporating eye-tracking or brain imaging data and examining the impact of conversational engagement levels.

References

Arnold, Jennifer E, Michael K Tanenhaus, Rebecca J Altmann and Maria Fagnano (2004). The old and thee, uh, new: Disfluency and reference resolution. In: *Psychological science* 15.9, pp. 578–582.

Arvidsson, Caroline, Ekaterina Torubarova, André Pereira and Julia Uddén (2024). Conversational production and comprehension: fMRI-evidence reminiscent of but deviant from the classical Broca–Wernicke model. In: *Cerebral Cortex* 34.3, bhae073.

Beattie, Geoffrey W (1978). Sequential temporal patterns of speech and gaze in dialogue. In.

Bolker, Benjamin M (2015). Linear and generalized linear mixed models. In: *Ecological statistics: contemporary theory and application* 2015, pp. 309–333.

Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders and Nivja H De Jong (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. In: *Language Testing* 30.2, pp. 159–175.

Brennan, Susan E and Michael F Schober (2001). How listeners compensate for disfluencies in spontaneous speech. In: *Journal of memory and language* 44.2, pp. 274–296.

Corley, Martin and Oliver W Stewart (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. In: *Language and Linguistics Compass* 2.4, pp. 589–602.

Dubberly, Hugh and Paul Pangaro (2009). What is conversation? How can we design for effective conversation. In: *Interactions Magazine* 16.4, pp. 22–28.

Eklund, Robert and Martin Ingvar (2016). Supplementary motor area activation in disfluency perception: An fmri study of listener neural responses to spontaneously produced unfilled and filled pauses. In: *Understanding Speech Processing in Humans and Machines, September 8-12, 2016, The Hyatt Regency, San Francisco, California, USA*. ISCA-INT SPEECH COMMUNICATION ASSOC, pp. 1378–1381.

Fraundorf, Scott H and Duane G Watson (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. In: *Language, Cognition and Neuroscience* 29.9, pp. 1083–1096.

Huttunen, Kerttu H, Heikki I Keränen, Rauno J Pääkkönen, R Päivikki Eskelinen-Rönkä and Tuomo K Leino (2011). Effect of cognitive load on articulation rate and formant frequencies during simulator flights. In: *The Journal of the Acoustical Society of America* 129.3, pp. 1580–1593.

JASP Team (2025). *JASP (Version 0.19.3)[Computer software]*. URL: https://jasp-stats.org/.

Lickley, Robin J (2015). Fluency and disfluency. In: *The handbook of speech production*, pp. 445–474.

MacGregor, Lucy J, Martin Corley and David I Donaldson (2009). Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. In: *Brain and language* 111.1, pp. 36–45.

Oomen, Claudy CE and Albert Postma (2001a). Effects of divided attention on the production of filled pauses and repetitions. In: *Journal of speech, language, and hearing research* 44.5, pp. 997–1004.

— (2001b). Effects of time pressure on mechanisms of speech production and self-monitoring. In: *Journal of Psycholinguistic Research* 30, pp. 163–184.

Pistono, Aurélie and Robert J Hartsuiker (2021). Eye-movements can help disentangle mechanisms underlying disfluency. In: *Language, Cognition and Neuroscience* 36.8, pp. 1038–1055.

Schultz, Wolfram, Kerstin Preuschoff, Colin Camerer, Ming Hsu, Christopher D Fiorillo, Philippe N Tobler and Peter Bossaerts (2008). Explicit neural signals reflecting reward uncertainty. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1511, pp. 3801–3811.

Shriberg, Elizabeth (2001). To 'errrr'is human: ecology and acoustics of speech disfluencies. In: *Journal of the international phonetic association* 31.1, pp. 153–169.

Shriberg, Elizabeth Ellen (1994). Preliminaries to a theory of speech disfluencies. PhD thesis. Citeseer.

Tacchetti, Maddalena (2017). User's Guide for ELAN Linguistic Annotator. In: *The Language Archive, MPI for Psycholinguistics, Nijmegen, The Netherlands.[Google Scholar]*.

Theys, Catherine, Silvia Kovacs, Ronald Peeters, Tracy R Melzer, Astrid van Wieringen and Luc F De Nil (2020). Brain activation during non-habitual speech production: Revisiting the effects of simulated disfluencies in fluent speakers. In: *Plos one* 15.1, e0228452.

Torubarova, Caroline Arvidsson, Jonathan Berrebi, Julia Uddén and André Pereira (2024). *'NeuroEngage'*. OpenNeuro. DOI: doi:10.18112/openneuro.ds004996.v1.0.1.

— (2025). NeuroEngage: A Multimodal Dataset Integrating fMRI for Analyzing Conversational Engagement in Human-Human and Human-Robot Interactions. In: *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '25. Melbourne, Australia: IEEE Press, pp. 849–858.

Tree, Jean E Fox (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. In: *Journal of memory and language* 34.6, pp. 709–738.

Vrij, Aldert, Ronald Fisher, Samantha Mann and Sharon Leal (2006). Detecting deception by manipulating cognitive load. In: *Trends in cognitive sciences* 10.4, pp. 141–142.

Williams, Simon (2022). *Disfluency and proficiency in second language speech production*. Springer.

Appendix

This appendix shows an example of how disfluencies were annotated using ELAN. The figures below illustrates the four tiers used in the annotation process, with accompanying description of the tiers.

The example in figure 2 shows how a line with two relevant disfluencies used was annotated for the study.

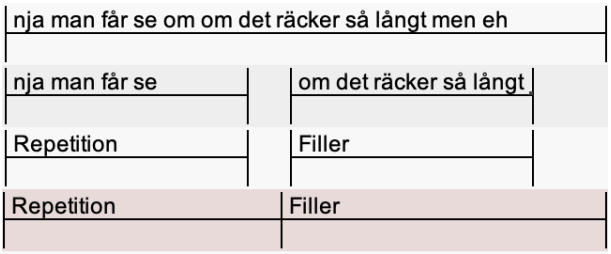

Figure 2: Annotation example 1 – from ELAN. Tier 1: the line said as it is. Tier 2: the line split for the respective disfluencies said. Tier 3: the annotations for each part of the line. Tier 4: a test annotation-line for if the whole line would be used unsplit (not used in this analysis).

The example in figure 3 shows how a line with two disfluencies, whereas one is non-relevant, was annotated for the study.

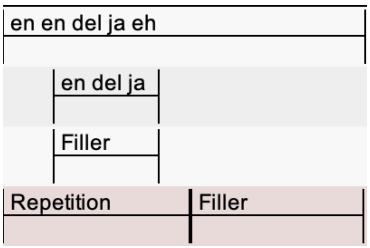

Figure 3: Annotation example 2 – from ELAN. Tier 1: the line said as it is. Tier 2: the line split for the relevant disfluency said. Tier 3: the annotations for the split part of the line. Tier 4: a test annotation-line for if the whole line would be used unsplit (not used in this analysis).

Stockholm University
SE-106 91 Stockholm, Sweden
Telephone +46 (0)8 16 20 00
https://www.su.se/